STUDY TO SUPPORT AN IMPACT ASSESSMENT FOR THE REVIEW OF THE DATABASE DIRECTIVE

Final Report

AUTHORS :
DR NORBERT MAIER
DR FEDERICO DE MICHIEL

AUTHORS :
DR VIOLA PETER
MARIA DEL CARMEN CALATRAVA MORENO

AUTHOR :
CHIARA PANCOTTI
FRANCESCA MONACO

AUTHOR :
MAURICE SCHELLEKENS
DR INGE GRAEF
DR NADYA PURTOVA

Individual Expert:

Prof. Andreas Wiebe (G.A. Univ.Götttingen)

Internal identification

Contract number: LC-01634139
VIGIE number: 2021-0465

EUROPEAN COMMISSION

Directorate-General for Communications Networks, Content and Technology
Directorate I — Media Policy
Unit I. 2 — Copyright

Contact: CNECT-I2@ec.europa.eu (functional e-mail)

European Commission
B-1049 Brussels

STUDY TO SUPPORT AN IMPACT ASSESSMENT FOR THE REVIEW OF THE DATABASE DIRECTIVE

Final Report

Directorate-General for Communications Networks, Content and Technology
2022 EN

EUROPE DIRECT is a service to help you find answers to your questions about the European Union

Freephone number (*):
00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you)

LEGAL NOTICE

PDF ISBN 978-92-76-46550-8 doi: 10.2759/647387 KK-09-22-006-EN-N

Manuscript completed in January 2022
1st edition

Luxembourg: Publications Office of the European Union, 2022

ABSTRACT

Access to data and the ability to use it have become an EU policy priority as they are considered essential for innovation and economic growth in an increasingly data-driven society. Particular attention is given to broadening access to and use of data generated/collected by sensors and machines in the Internet of Things (IoT) environment. As part of the actions to achieve this objective, the European Commission is reviewing the Directive 96/9/EC on the legal protection of databases ("Database Directive" or "Directive"), within the broader Data Act initiative.

The Database Directive aims to stimulate the creation of databases in the EU by providing legal protection (through a *sui generis* right) to database makers that made substantial investments in obtaining, verifying or presenting the contents of a database. Currently, there is legal uncertainty around the scope of the Directive, in particular in relation to machine-generated data. The study supports the impact assessment for the review of the Database Directive in the context of the Data Act initiative to ensure that the Directive is fit for the new data economy and does not risk becoming an obstacle to sharing and usage of data within the EU and across sectors.

Based on information gathered from stakeholders and legal experts through an online survey, in-depth interviews and a closed workshop, and further complemented with desk research, this study offers a set of policy options for the Commission's consideration to review the Directive and provides an assessment for their expected effectiveness, efficiency, and coherence.

EXECUTIVE SUMMARY

Purpose and scope of the study

This study, conducted on behalf of the European Commission, Directorate-General for Communications Networks, Content and Technology (DG CONNECT), supports the Impact Assessment of the review of the Directive 96/9/EC of 11 March 1996 on the legal protection of databases ("Database Directive" or "Directive") as part of the broader Data Act policy initiative[1].The study provides evidence for defining and assessing the likely impact of different policy options for the review of the Database Directive with a particular focus on the *sui generis* right and its relationship with machine generated data ("MGD"), especially with databases containing data generated/collected by sensors in the Internet of Things (IoT) environment.

The study presents evidence and inputs collected from key experts and stakeholders via an online survey, in-depth interviews with business stakeholders, and a workshop organised with legal experts. Furthermore, it also relies on desk research, review of existing literature within the field, findings of the previous extensive evaluation of the Database Directive conducted in 2018[2] and on contributions to the Open Public Consultation (OPC) for the Data Act[3].

Background and problem assessment

The Database Directive adopted 26 years ago in 1996, protects the intellectual creativity embodied in the design of databases through copyright (Art. 3) and the investment in the collection, verification, and presentation of content through the *sui generis* right (Art. 7). Databases are defined in Art 1(2) of the Database Directive as "a collection of independent works, data or other materials which are arranged in a systematic or methodical way and are individually accessible by electronic or other means". The *sui generis* right applies regardless of whether the content qualifies for copyright protection or if the database is innovative or original.

The considerable changes in the technological landscape and economic context over the past 26 years have tested the applicability of the Directive. As also pointed out in the 2018 Evaluation, the Database Directive may not be fit for purpose in the new data economy primarily due to the unclarity of the *sui generis* right in relation to MGD. The applicability of certain provisions and legal requirements such as "substantial investments", "insubstantial extraction", and the distinction made by previous decisions of the Court of Justice of the European Union ("CJEU") between "creation" and "obtaining/collection" of data remain unclear and may not be aligned with new technological developments. The economic relevance of data has increased as well as the different activities linked to data and databases such as their creation, collection, structuring, maintenance and analysis. At the same time, progress in automated data processing may drive down costs for those activities. Furthermore, the automated collection of data via sensors and its continuous processing pose technical and

---

1 See: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Data-Act-&-amended-rules-on-the-legal-protection-of-databases_en

2 See: https://digital-strategy.ec.europa.eu/en/library/staff-working-document-and-executive-summary-evaluation-directive-969ec-legal-protection-databases

3 See: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Data-Act-including-the-review-of-the-Directive-96-9-EC-on-the-legal-protection-of-databases-/public-consultation_en

legal challenges for clearly identifying instances where a database that contains MGD ("MGD databases") could be granted protection.

The unclear scope of the *sui generis* right in relationship to MGD risks to become an obstacle for the implementation of the European strategy for data[4] that has a key objective in stimulating data access and sharing within the EU Single market and across sectors. The possible consequences of the legal uncertainty around the scope of the Directive amplifies as the data economy continues to grow, reaching €355 bn in 2020[5], and as IoT and MGD solutions are increasingly adopted across different sectors, with the market for sensors expected to more than triple in 2025[6].

In particular, the legal uncertainty around the scope of the Directive may lead to underutilisation of data due to the undue Intellectual Property (IP) protection that it may provide over data. It may be exploited to further protect data monopolies and generate data lock-ins. This may increase transaction costs for data users seeking access to data and hamper data-driven innovation and competition in primary and secondary markets.

Policy options

Following the analysis of the problems, the general objective of the review of the Database Directive is to increase legal certainty and facilitate data access and use in the EU Single market. This must be achieved by balancing the protection required to incentivise the creation of databases and the societal benefits derived from access to data. The key specific objective is to clarify the status of the *sui generis* right in relation to databases containing MGD to ensure that the Directive does not pose an obstacle to the sharing of and trading of MGD.

Several policy options (PO) have been defined in order to address the identified problems and achieve the objectives formulated above. Some of them relate specifically to MGD as the main immediate challenge, while the rest are possible supplementary options offered for the Commission to consider that would apply to all protected databases. After an initial scoping phase, some options have been discarded and/or not explored further while the remaining ones were further elaborated and assessed against the baseline option (PO0) of no action. The options further assessed are defined and explained below.

PO2: Options to exclude MGD from the scope of the *sui generis* right. Two general approaches could be taken to implement the option: either a) directly by explicit exclusion of MGD or b) indirectly by adapting legal concepts already existing in the Directive. Alternatives to applying the second approach would be, for example, to exclude investments in the generation of data or to establish a minimum standard of substantial investments that will be hard to overcome by MGD. All approaches present their technical and legal challenges and uncertainties. Exclusion could apply also to mixed databases (i.e. containing both MGD and non-MGD), which, according to some stakeholders, is common in some industries.

4 See https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066

5 See European Commission (2020), The European Data Market Study Update

6 Cision PR Newswire (2019): Global IoT Sensors Market Analysis, Trends, and Forecasts, 2019-2025 - Global Market is Projected to Grow by US$35.2 Billion, at a CAGR of 34%

This option is further split into three retained sub-options.

PO2a: Exclusion of MGD from the scope of the *sui generis* right. This sub-option would bring legal clarity and ensure that the *sui generis* right does not become an additional layer of indirect protection for MGD that could generate opportunistic litigations and stand in the way of possible obligations to allow access to such data or otherwise impede use where access to such data is technically feasible. This option considers the fact that i) other types of protections are already used by database makers, and ii) firms may not need further incentives to produce MGD.

PO2b: Exclusion of MGD from the scope of the *sui generis* right combined with substitution of the current infringement test with a more flexible test related to economic detriment. This sub-option would expand PO2a by creating a more flexible test drawing from the recent *CV v Melons* case[7] that would apply to the other databases still under the scope of *sui generis* right. This test could include a weighing of interests with an emphasis on the ability to redeem the risk of investment by the database maker as well as the purpose of the taking by the user (e.g. follow-on innovation). The new infringement test may also have the effect of overcoming the unpractical distinction between insubstantial and substantial taking.

PO2c: Exclusion of MGD from the scope of the *sui generis* right combined with the introduction instead of an alternative control mechanism applied only to MGD databases. This sub-option would substitute the *sui generis* right with an alternative, lighter protection mechanism over MGD databases not based on exclusive rights. The protection would be limited and targeted against unauthorised and unfair competitive uses by third parties and would provide a basis for MGD database makers to share and trade their databases. This alternative control mechanism could draw from how unlawful uses are defined in the context of trade secret protection.

PO3: Options that include MGD in the scope of the *sui generis* right. The most direct approach to implement the inclusion of MGD would be to count in the requirement of substantial investment also investments in the generation of data. However, uncertainty remains in identifying relevant investments, especially in the context of MGD. On one side, data may be collected as a by-product of another economic activity carried out by the database maker. On the other side, the creation and processing and refurbishing of raw data in different ways may be inextricably linked.

One sub-option was explored further after the scoping phase:

PO3a: Inclusion of MGD in the scope of the *sui generis* right. This option would bring legal clarity by avoiding the need for a distinction between “generation” and ”collection” of data. Moreover, similar to 2c, MGD database makers might have more incentives to make their databases public, as the exclusive right would protect them against misuse of the pertinent data by third parties. New problems would include drawing a line for relevant investment into the generation of data.

Supplementary policy options (SPO) for a potential wider review of the Database Directive. This set of options is treated separately as their scope is different to the main options that deal with the most immediate challenge of the interaction between MGD and the *sui generis* right. These policy options that apply to all databases protected by the *sui generis*

[7] Case C-762/19, SIA ‘CV-Online Latvia’ v SIA ‘Melons’

right, deal with a different array of topics and should not be seen as mutually exclusive. The possible supplementary options were defined jointly with the Commission upon a scoping phase and they take the following forms:

SPO S1: would expand exceptions to *sui generis* right in line with broader general copyright exceptions. Under this supplementary option, limitations newly introduced into copyright would be extended to the *sui generis* right. As to existing copyright limitations, SPO S1 would extend all or some of them to the *sui generis* right avoiding new uncertainties and fragmentation. The most promising existing exceptions would be private use, quotation and reporting.

SPO S2: would strengthen the exception for research. Under this supplementary option, a newly expanded research exception could have the following characteristics: a) it could be made mandatory, b) it could encompass the re-utilisation and the reproduction of a database as a whole for scientific purposes, c) it could apply also to infrastructure operators of non-commercial scientific databases and d) it could be extended for commercial research (under certain conditions). The new exception should not contradict or change the existing Text and Data Mining (TDM) exceptions in Art. 3 and Art. 4 of Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market (the DSM Directive).

SPO S3: would exclude public bodies from the *sui generis* right. Under this supplementary option, it would be made clear that public bodies do not hold *sui generis* rights.

Assessment of the impacts

The options are assessed and compared against three criteria of effectiveness, efficiency and coherence. Furthermore, the study explored the impact of each policy on the fundamental rights to protection of personal data as well as to property and freedom to conduct business without finding major conflicts in any policy.

Assessing the options presents challenges due to i) technical aspects that are not defined yet and may affect the judgment on the overall option, such as the approach to exclude MGD; and ii) there being little tangible evidence available. The assessment is qualitative and mainly based on the anticipated possible effects reported in selected studies and academic work and on the opinion of the survey's respondents and interviewees. Despite these challenges, the following findings can be considered to have an indication of the impact of the options.

Policy options related to MGD

Effectiveness and Efficiency

These options are assessed against the baseline option. At the current stage, the Directive does not seem to pose a serious obstacle to the sharing and usage of MGD. However, this might change in the future as databases containing MGD become more important as an input for innovation and competition, e.g., in the context of Artificial Intelligence (AI) applications, allowing database makers to use the legal uncertainty for purposes not intended by the initial objectives of the Directive. The efficiency of the Directive in the broader data economy context remains unknown as its practical relevance is still limited and contractual agreements usually overcome its provisions.

- **PO2a** shows **effectiveness** in bringing legal clarity and facilitating data access. PO2a would reduce the risk that the *sui generis* right might extensively be claimed by MGD database makers opportunistically seeking extra protection beyond factual control. While unlikely to have a major, immediate effect on the sharing and usage of MGD, this option has a forward-looking effect, especially in combination with other possible

actions within the Data Act to introduce access and usage rights to data. In terms of **efficiency**, the survey conducted for this study shows that respondents expect this option to bring high benefits and no additional costs compared to the baseline, although some compliance and enforcement costs may materialise due to difficulties in its implementation. The majority of survey respondents maintain that excluding MGD will have a positive effect on obtaining legal certainty. Direct benefits are also expected from reduced information and transaction costs for database users due to less room for opportunistic litigation on third-party data use. In terms of indirect costs, while few interviewees worried that the exclusion of MGD would trigger protective behaviour on the side of the database makers, most survey respondents do not think exclusion will translate in lock-in situations or less access to data for users. By contrast, as indirect benefits, survey respondents see positive effects for innovation and research activities and for revenues generated from the production and/or exploitation of databases as a consequence of excluding databases containing MGD.

- **PO2b** is expected to possess similar **effectiveness** to PO2a with regard to MGD. Moreover, the infringement analysis inspired by the recent *CV v Melons* case aims to promote innovation and competition beyond MGD. Nevertheless, the criteria for the test must be further defined to avoid further legal uncertainty and fragmentation at the national level to arise. In addition, uncertainty on the side of the database maker, due to the fact that the *sui generis* protection would depend on other contextual factors, may disincentivize the database makers to make the data public, thus restricting *de facto* ability of users to access them. In terms of **efficiency**, PO2b would bring positive impacts in terms of flexibility and competition compared to the baseline and PO2a. However, as it is currently envisaged, it is likely to present important compliance and enforcement costs. Further analysis is needed as the CJEU ruling on the *CV v Melons* case is very recent and the effects are still uncertain. Finally, this CJEU judgment itself is expected to have some effect on the *acquis*, notably on the infringement test without further codification.

- **PO2c** could, in theory, bring similar or higher **effectiveness** than PO2a and PO2b by negating exclusive rights on MGD databases while still offering database makers a sort of control mechanism that would protect them from unlawful uses of their databases, thus encouraging data sharing. However, there is no evidence that the effect would materialise and some authors have warned about the risk that the defensive right, by creating a layer of protection to databases containing MGD, will discourage their use. As for its **efficiency,** PO2c might be counterproductive concerning the objective of a REFIT initiative[8]: bringing simplification and improved efficiency. As underlined by consulted legal experts**,** creating a defensive right applicable only to databases containing MGD, which would run in parallel to the *sui generis,* would create additional complexity and uncertainty, e.g. by creating overlapping schemes for mixed databases. Further enforcement and compliance costs may arise if the requirements for protection are too vaguely defined and difficult to prove in practice. Especially the criterion of “wrongfulness” and the associated weighing of interests will be difficult to evaluate by the parties in advance. Hence, the risk of a negative impact on the overall clarity and functioning of the Directive seems to overtake the possible benefits. Nevertheless, most survey respondents would agree with the possibility of replacing the *sui generis* right with an alternative protective mechanism against certain unauthorized uses.

[8] European Commission's regulatory fitness and performance programme (REFIT)

- **PO3a** has similar **effectiveness** as PO2a and PO2b in increasing legal clarity on the scope of the *sui generis* right in relation to MGD. However, it would add an additional layer of exclusive rights for database makers and could row against the main vision of the European strategy for data. As in most cases, no new incentives are needed to produce MGD and it might overly restrict access and use of MGD to the detriment of general competitive interests and innovation. This risk is aggravated as many databases containing MGD are sole-source databases (e.g. MGD related to the performance of a device). By contrast, some interviewees pointed out that the protection may encourage, similarly to PO2c, database makers to make their databases public and contract with users licencing terms for the use of their data, thus also promoting in a way data sharing. As for the **efficiency** of the option, survey results suggest that costs would slightly exceed the benefits compared to the baseline option. Moreover, while survey respondents report that the inclusion of MGD (with an access right) would affect innovation and competition more positively than negatively overall, the view is less strong than for PO2a. Yet, the positive effect on data sharing pointed out by one interviewee may bring indirect and direct benefits with clearer rules and protection structure and more choice for database seekers.

Coherence

PO2a and PO2b would not create any conflicts with other legal instruments for the protection of data as they limit the scope of the *sui generis* right. PO2c could possibly interfere with trade secrets protection. However, the relevance of trade secrets may be reduced in an increasingly networked data economy where data is shared and exchanged. PO3a is also coherent with other existing legal frameworks as it would only clarify the application of the already existing *sui generis* database right to MGD. With respect to PO3a, possible conflicts may arise with the upcoming Data Act which may consider the introduction of an access right. The *sui generis* protection of MGD may pose an impediment to data access as right holders may appeal to the exclusive right. As for PO2c, the new right bears the risk of working as an additional layer of protection. By contrast, the PO2a and PO2b, appear more aligned with the possible intentions of the Data Act than PO2c and PO3a, to avoid the Database Directive becoming an obstacle for sharing and trading of MGD.

Supplementary policy options

This group of options dealing with different topics are not directly comparable nor are strictly related to the ongoing legislative initiative (i.e. the Data Act) but are rather offered for the Commission's consideration. Therefore, they are only assessed with respect to the baseline option.

Effectiveness and Efficiency

SPO S1 could be effective in decreasing legal uncertainties and increasing the use of data beyond MGD databases. This would be particularly important in the context of big data use cases, where data users may need complete databases (not just insubstantial takings as now in the exceptions of the *sui generis* right). As for the efficiency, SPO S1 can be positively assessed against the status quo, considering that the additional costs would be minimal, whereas it would bring benefits in terms of increased clarity and uniformity of rules between exceptions to copyright and *sui generis* right. According to the survey, the benefit of the option would exceed its costs. Legal experts consulted also support the option. With **SPO S2**, the strengthened exception on research would improve harmonisation across Member States, support and have a positive societal impact through research and innovation in the EU. However, the exception must take into account (and be compatible with) the TDM exception for research that is currently being transposed by Member States. **SPO S3** eliminates the

uncertainty regarding the permissibility of using the contents of the databases held by public bodies**.** For a large number of occurrences, the Open Data Directive (ODD)[9] already resolved the tension between access and database protection for public sector databases. Therefore, the problem that this option addresses is significantly narrower than what existed prior to the ODD. SPO S3 entails low enforcement costs. Also, further efficiency gains are expected compared to the baseline by reducing the disputes on possible inconsistencies with the ODD and the related transaction costs. The activity of self-founded public bodies relying on the *sui generis* right may risk stopping being economically viable if the option does not explicitly exclude private organisations active in competitive environments.

Coherence

SPO S1 will be coherent with the intention of European institutions to increase access and use of data within EU and will harmonise the legal framework for copyright in the Digital Single Market. However, uncertainty remains on the alignment with the Art. 5 ("Exceptions and limitations") Directive 2001/29/EC on Copyright in the Information Society depending on whether the exceptions would be made mandatory or left to Member States' discretion. **SPO S2** does not appear to interfere with other existing legal frameworks once compatibility with TDM exception is ensured. **SPO S3** is in line with the objectives of the ODD and the Data Governance Act. Uncertainties could arise with regard to the different scope of the exclusion proposed in the ODD. However, both approaches have the same goal.

Comparison of policy options related to MGD

Following the assessment, the option related to MGD that appears to be the best choice based on the evidence gathered is PO2a: the exclusion of MGD. This will bring legal clarity to the relationship between the Database Directive and MGD – increasingly relevant for the data economy in the EU Single Market – while at the same time ensuring that the Directive will not become an obstacle for the sharing of data across sectors. Both PO2a and PO2b are the most effective and coherent ones compared to the baseline. However, PO2a appears the most efficient of the two as PO2b carries potential costs related to the uncertainties around the introduction of the infringement test and taking into account that the judgement (*CV Online v Melons*) underlying PO2b already forms part of *acquis communautaire* and should be observed by the national courts. PO2c and PO3a are seen as only slightly effective or neutral while being less efficient than the status quo as well as less coherent with the intention of the Data Act.

[9] See the Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.

RÉSUMÉ

Objet et champ de l'étude

La présente étude, menée pour le compte de la Commission européenne, Direction générale des Réseaux de communication, contenu et technologies (DG CONNECT), vient appuyer l'analyse d'impact de l'évaluation de la directive 96/9/CE du 11 mars 1996 sur la protection juridique des bases de données (« directive sur les bases de données » ou « directive ») dans le cadre plus vaste de l'initiative stratégique d'acte législatif sur les données[10].Cette étude apporte des preuves permettant de définir et d'évaluer l'impact probable des différentes options pour l'évaluation de la directive sur les bases de données, en mettant l'accent sur le droit *sui generis* et sa relation avec les données générées par des machines (« DGM »), en particulier avec les bases de données contenant des données générées/collectées par des capteurs dans l'environnement de l'Internet des objets (IdO).

L'étude présente des preuves et des contributions provenant d'experts et d'acteurs reconnus par l'intermédiaire d'une enquête en ligne, d'entretiens approfondis avec des acteurs économiques, et d'un atelier organisé avec des experts juridiques. Elle s'appuie également sur une recherche documentaire, sur l'examen de la littérature dans le domaine, sur les résultats d'une évaluation approfondie de la directive sur les bases de données menée en 2018[11], et sur les contributions de la consultation publique pour l'acte législatif sur les données.[12]

Contexte et évaluation des problèmes

La directive sur les bases de données, adoptée il y a 26 ans en 1996, protège la créativité intellectuelle faisant partie intégrante de la conception de bases de données par le droit d'auteur (art. 3) et l'investissement dans la collecte, la vérification et la présentation de contenu via le droit *sui generis* (art. 7). L'article 1, paragraphe 2, de la directive sur les bases de données définit les bases de données comme « un recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique et individuellement accessibles par des moyens électroniques ou d'une autres manière ». Le droit *sui generis* s'applique, peu importe si le contenu est éligible à la protection du droit d'auteur ou si la base de données est innovante ou originale.

Les changements considérables du paysage technologique et du contexte économique au cours des 26 dernières années ont mis à l'épreuve l'applicabilité de la directive. Comme le soulignait l'évaluation de 2018, la directive sur les bases de données n'est sans doute plus adaptée à la nouvelle économie des données, essentiellement en raison du manque de clarté du droit *sui generis* en ce qui concerne les DGM. L'applicabilité de certaines dispositions et obligations juridiques telles que l'« investissement substantiel », « l'extraction non substantielle », et la distinction établie par des décisions antérieures de la Cour de justice de l'Union européenne (« CJUE ») entre « création » et « obtention/collecte » de données, reste

[10] Voir : https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Acte-legislatif-sur-les-donnees-et-modification-des-regles-relatives-a-la-protection-juridique-des-bases-de-donnees_fr

[11] Voir : https://digital-strategy.ec.europa.eu/en/library/staff-working-document-and-executive-summary-evaluation-directive-969ec-legal-protection-databases

[12] Voir : https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Data-Act-including-the-review-of-the-Directive-96-9-EC-on-the-legal-protection-of-databases-/public-consultation_fr

incertaine et peut s'avérer inadaptée aux dernières évolutions technologiques. L'importance économique des données a augmenté, ainsi que les différentes activités liées aux données et aux bases de données, comme leur création, leur collecte, leur structuration, leur maintenance et leur analyse. Dans le même temps, les progrès dans le domaine de l'automatisation du traitement des données peuvent entraîner une baisse des coûts de ces activités. De plus, la collecte automatisée de données par l'intermédiaire de capteurs et leur traitement en continu pose des difficultés techniques et juridiques pour identifier sans ambiguïté les cas où une base de données contenant des DGM (« bases de données DGM ») peut se voir accorder une protection.

Le champ mal défini du droit *sui generis* par rapport aux DGM risque de devenir un obstacle à la mise en œuvre de la Stratégie européenne pour les données[13], dont l'un des objectifs principaux est d'encourager l'accès aux données et leur partage au sein du marché unique européen et entre les différents secteurs. Les conséquences potentielles de l'incertitude juridique autour du champ de la directive augmentent au rythme de la croissance de l'économie des données, qui représentait 355 milliards d'euros en 2020[14], et de l'adoption des solutions d'IdO et de DGM dans différents secteurs, le marché des capteurs devant plus que tripler d'ici 2025[15].

Plus particulièrement, l'incertitude juridique autour du champ de la directive peut conduire à une sous-utilisation des données en raison d'une protection excessive de la propriété intellectuelle qu'elle pourrait accorder aux données fournies. Cette situation pourrait être exploitée pour protéger encore davantage les monopoles sur les données et créer des blocages. Ceci pourrait augmenter le coût des transactions pour les utilisateurs cherchant à accéder aux données, et nuire à l'innovation basée sur les données, ainsi qu'à la concurrence sur les marchés primaires et secondaires.

Options stratégiques

Après analyse des problèmes, l'objectif général de l'évaluation de la directive sur les bases de données consiste à augmenter la sécurité juridique et à faciliter l'accès aux données et leur utilisation sur le marché unique européen. Pour y parvenir, il faut impérativement équilibrer la protection nécessaire pour encourager la création de bases de données et les avantages sociétaux découlant de l'accès aux données. Le principal objectif spécifique consiste à clarifier la position du droit *sui generis* par rapport aux bases de données contenant des DGM afin de faire en sorte que la directive ne s'oppose pas au partage et aux échanges de DGM.

Plusieurs options stratégiques (OS) ont été élaborées pour faire face aux problèmes recensés et pour atteindre les objectifs susmentionnés. Certaines d'entre elles concernent plus spécialement les DGM, considérées comme la principale difficulté immédiate, tandis que les autres représentent des possibilités supplémentaires applicables à toutes les bases de données protégées si la Commission le jugeait nécessaire. Après un premier examen, certains scénarios ont été rejetés et/ou ne seront pas étudiés plus avant, tandis que les scénarios restants seront plus approfondis et évalués par rapport à l'option de référence sans action

[13] Voir : https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066

[14] Voir : Commission européenne (2020), The European Data Market Study Update

[15] Cision PR Newswire (2019) : Global IoT Sensors Market Analysis, Trends, and Forecasts, 2019-2025 - Le marché mondial devrait croître de 35,2 milliards USD, avec un taux de croissance annuel moyen de 34 %

particulière (OS0). Les options retenues pour un examen complémentaire sont décrites ci-dessous.

OS2 : Options visant à exclure les DGM du champ du droit *sui generis*. Deux approches générales différentes sont possibles pour mettre en pratique cette option: soit a) directement en excluant explicitement les DGM, soit b) indirectement en adaptant les concepts juridiques déjà présents dans la directive. Les alternatives à l'application de la seconde approche consisteraient, par exemple, à exclure les investissements dans la génération de données, ou à mettre en place un niveau minimum d'investissements substantiels qui sera difficile à surmonter pour les DGM. Toutes les approches comportent des difficultés et des incertitudes techniques et juridiques. Cette exclusion pourrait également s'appliquer aux bases de données mixtes (c'est-à-dire contenant à la fois des DGM et des données non générées par des machines), ce qui, d'après certains acteurs, est une pratique courante dans certains secteurs économiques.

Ce scénario se subdivise en trois sous-options retenues.

OS2a : Exclusion des DGM du champ du droit *sui generis*. Cette sous-option apporterait de la sécurité juridique et ferait en sorte que le droit *sui generis* ne devienne pas une couche supplémentaire de protection indirecte des DGM, susceptible de générer des litiges opportunistes et de faire obstacle aux obligations potentielles d'autoriser l'accès à ces mêmes données, voire d'empêcher l'usage de ces données alors qu'elles sont tout à fait accessibles sur le plan technique. Cette option tient compte du fait que i) d'autres types de protection sont déjà employés par les créateurs de bases de données, et ii) qu'il n'est pas nécessaire d'inciter davantage les entreprises à produire des DGM.

OS2b : Exclusion des DGM du champ du droit *sui generis*, associée au remplacement du test de violation par un test plus souple portant sur le préjudice économique. Cette sous-option permettrait d'élargir le champ de OS2a en créant un test plus souple à partir de l'arrêt récent *CV contre Melons*[16], qui s'appliquerait aux autres bases de données toujours couvertes par le champ du droit *sui generis*. Ce test pourrait inclure une pondération des intérêts, en mettant l'accent sur la capacité du créateur de la base de données à compenser le risque d'investissement, ainsi que le but de l'adoption par l'utilisateur (innovation subséquente, par exemple). Ce nouveau test de violation pourrait également permettre de surmonter la distinction peu pratique entre adoption substantielle et non substantielle.

OS2c : Exclusion des DGM du champ du droit *sui generis*, associée à la mise en place d'un autre mécanisme de contrôle appliqué uniquement aux bases de données de DGM. Cette sous-option entraînerait le remplacement du droit *sui generis* par un autre mécanisme de protection, plus léger, pour les bases de données de DGM non basées sur des droits exclusifs. Cette protection se limiterait à et ciblerait les utilisations déloyales et non autorisées par des tiers, et offrirait aux créateurs de bases de données une base pour partager et commercialiser leurs bases de données. Ce mécanisme de contrôle pourrait s'appuyer sur la définition des utilisations illégales dans le cadre de la protection du secret des affaires.

OS3 : Options consistant à inclure les DGM dans le champ du droit *sui generis*. L'approche la plus directe pour mettre en œuvre l'inclusion des DGM serait de s'appuyer sur l'obligation d'investissement substantiel, à savoir d'investissements dans la génération de

[16] Arrêt C-762/19, SIA « CV-Online Latvia » contre SIA « Melons »

DGM. Une certaine incertitude subsiste néanmoins pour identifier les investissements pertinents, en particulier dans le contexte des DGM. D'une part, les données pourraient être collectées en tant que sous-produit d'une autre activité économique menée par le créateur de la base de données. D'autre part, la création, le traitement et la restauration de données brutes peuvent être indissociables.

Une sous-option a été étudiée de manière plus approfondie après la première selection :

OS3a : Inclusion des DGM dans le champ du droit *sui generis*. Cette option permettrait de lever l'incertitude juridique en évitant le besoin d'établir une distinction entre création et collecte. De plus, comme pour l'option 2c, les créateurs de bases de données pourraient être incités à rendre leurs bases de données publiques, car le droit exclusif les protégerait contre les utilisations abusives des données pertinentes par des tiers. De nouveaux problèmes feraient leur apparition pour la définition des investissements pertinents dans la génération de données.

Options stratégiques supplémentaires (OSS) pour une éventuelle révision de plus grande ampleur de la directive sur les bases de données. Cet ensemble d'options fait l'objet d'un traitement séparé, car leur champ est différent de celui des options principales qui traitent de la difficulté la plus immédiate, c'est-à-dire l'interaction entre DGM et droit des bases de données. Ces options stratégiques, qui s'appliquent à toutes les bases de données protégées par le droit *sui generis,* traitent d'un ensemble de sujets totalement différents et ne devraient pas être considérées comme incompatibles. Les possibles options supplémentaires ont été définies conjointement avec la Commission au cours d'une première phase de sélection. Elles prennent les formes suivantes :

OSS S1 : élargirait les exceptions au droit *sui generis* en les alignant sur les exceptions plus larges du droit d'auteur. Avec cette option supplémentaire, de nouvelles limitations ajoutées au droit d'auteur seraient étendues au droit *sui generis*. Pour ce qui est des limitations existantes au droit d'auteur, OSS S1 les étendrait toutes, ou certaines d'entre elles, au droit *sui generis* pour éviter la fragmentation et de nouvelles incertitudes. Les exceptions les plus prometteuses pourraient être l'usage privé, le droit de citation, et le signalement.

OSS S2 : renforcerait l'exception à des fins de recherche. Avec cette option supplémentaire, une exception élargie à des fins de recherche pourrait présenter les caractéristiques suivantes : a) elle pourrait devenir obligatoire, b) elle pourrait englober la réutilisation et la reproduction d'une base de données complète à des fins scientifiques, c) elle pourrait s'appliquer aux exploitants d'infrastructures de bases de données scientifiques non commerciales, et d) elle pourrait être étendue à la recherche commerciale (sous certaines conditions). Cette nouvelle exception ne devrait pas entrer en conflit ou modifier les exceptions en faveur de la fouille de textes et de données (« text and data mining » ou TDM en anglais) dans l'Art. 3 et l'Art. 4 de la Directive (EU) 2019/790 sur le droit d'auteur et les droits voisins dans le marché unique numérique (ou « Digital Single Market »).

OSS S3 : exclurait les organismes publics du droit *sui generis*. Avec cette option supplémentaire, le droit *sui generis* préciserait expressément que les organismes publics ne disposent pas de droits *sui generis*.

Analyse des impacts

Ces options sont analysées et comparées selon trois critères d'efficacité, d'efficience et de cohérence. En outre, l'étude a examiné l'impact de chaque scénario sur les droits fondamentaux à la protection des données à caractère personnel, ainsi qu'à la propriété et à

la liberté d'entreprendre sans rencontrer de conflit majeur dans une quelconque mesure réglementaire.

L'analyse de ces options présente un certain nombre de difficultés en raison i) des aspects techniques qui ne sont pas encore définis et peuvent affecter le jugement sur l'option dans son ensemble, comme l'approche consistant à exclure les DGM, ii) du peu de preuves tangibles disponibles. L'évaluation est qualitative, elle repose essentiellement sur les effets prévisibles mentionnés dans diverses études et travaux universitaires, ainsi que sur l'avis des personnes ayant répondu à l'enquête et participé aux entretiens. Malgré ces difficultés, les résultats suivants peuvent être considérés comme des indicateurs de l'impact des différentes options.

Options stratégiques liées aux DGM

Efficacité et efficience

Ces options sont évaluées par rapport à l'option de référence. En l'état, la directive ne semble pas représenter un obstacle majeur au partage et à l'utilisation des DGM. Cela pourrait toutefois changer à l'avenir, les bases de données contenant des DGM prenant de plus en plus d'importance comme support d'innovation et de concurrence, par ex. dans le contexte des applications d'intelligence artificielle, permettant aux créateurs de bases de données d'exploiter cette incertitude juridique pour des usages non prévus par les objectifs initiaux de la directive. L'efficience de la directive dans le contexte économique plus vaste reste largement inconnue, car sa pertinence pratique reste limitée et les accords contractuels vont généralement au-delà des dispositions de la directive.

- **OS2a** fait preuve d'**efficacité** pour lever l'incertitude juridique et faciliter l'accès aux données. OS2a permettrait de réduire le risque d'un recours excessif au droit *sui generis* de la part de créateurs de bases de données DGM cherchant une protection supplémentaire allant au-delà du contrôle de fait. Bien qu'un impact majeur et immédiat sur le partage et l'utilisation des DGM soit peu probable, cette option présente une approche prospective, en particulier en association avec les autres actions possibles dans le cadre de l'acte législatif sur les données, pour mettre en place les droits d'accès et d'utilisation des données. Pour ce qui est de l'**efficience**, l'enquête menée pour cette étude montre que les répondants s'attendent à ce que cette option offre des avantages importants sans coût supplémentaire par rapport à l'option de référence, malgré la probable apparition de coûts de mise en conformité et de mise en œuvre liés aux difficultés de mise en œuvre. La majorité des personnes ayant répondu à l'enquête soutiennent que l'exclusion des DGM aurait un impact positif sur la sécurité juridique. La baisse des coûts d'information et de transaction pour les utilisateurs de bases de données occasionnée par la réduction des possibilités de litiges opportunistes portant sur l'utilisation des données par des tiers devrait également donner lieu à des avantages directs. En matière de coûts indirects, malgré l'inquiétude de quelques-unes des personnes interrogées quant à l'adoption d'une posture protectionniste par les créateurs de bases de données, la plupart des répondants ne pensent pas que l'exclusion se traduira par des situations de blocage ou de restriction d'accès aux données pour les utilisateurs. En revanche, les répondants s'attendent à des avantages indirects, notamment des effets positifs pour les activités d'innovation et de recherche et pour les revenus tirés de la production et/ou de l'exploitation des bases de données à la suite de l'exclusion des bases de données contenant des DGM.
- **OS2b** devrait présenter la même **efficacité** que OS2a en ce qui concerne les DGM. De plus, l'analyse de violation inspirée du récent arrêt *CV contre Melons* vise à favoriser l'innovation et la concurrence au-delà des DGM. Quoi qu'il en soit, il est

impératif d'affiner les critères de test pour éviter que l'incertitude juridique et la fragmentation au niveau national ne s'aggravent. De plus, l'incertitude du côté des créateurs de bases de données, due au fait que la protection *sui generis* dépendrait d'autres facteurs contextuels, pourrait dissuader ces créateurs de rendre les données publiques, réduisant de fait la possibilité pour les utilisateurs d'y accéder. Sur le plan de l'**efficience**, OS2b aurait des répercussions positives en matière de flexibilité et de concurrence par rapport à l'option de référence et à OS2a. Mais dans sa forme actuelle, cette option devrait présenter des coûts élevés d'exécution et de mise en conformité. Une analyse plus poussée est nécessaire car l'arrêt de la CJUE dans l'affaire *CV contre Melons* est très récent, et ses effets ne sont pas encore très clairs. Enfin, l'arrêt de la CJUE proprement dit devrait avoir des répercussions sur l'acquis, notamment sur le test de violation sans codification supplémentaire.

- **OS2c** pourrait théoriquement offrir une **efficacité** comparable ou supérieure à celle de OS2a et de OS2b en réduisant les droits exclusifs sur les bases de données DGM, tout en offrant les créateurs de bases de données d'un mécanisme de contrôle qui les protégerait contre les utilisations illégales de leurs bases de données, encourageant ainsi le partage de données. Néanmoins, la concrétisation de cet effet est loin d'être garantie, et certains auteurs attirent l'attention sur le risque que le droit protecteur pourrait décourager le recours aux bases de données contenant des DGM en créant une couche de protection supplémentaire les concernant. Sur le plan de l'**efficience**, OS2c pourrait être contre-productive par rapport à l'objectif de l'une des initiatives REFIT[17] : apporter de la simplification et améliorer l'efficience. Comme le soulignent les experts juridiques consultés**,** la création d'un droit défensif portant uniquement sur les bases de données contenant des DGM, en parallèle du droit *sui generis*, pourrait ajouter de la complexité et de l'incertitude, par exemple en créant des dispositifs faisant doublon pour les bases de données mixtes. Des exigences de protection trop vagues et difficiles à démontrer dans la pratique pourraient entraîner des coûts d'exécution et de mise en conformité supplémentaires. Le critère de « caractère illicite » et la pondération des intérêts qui y est associée seront particulièrement difficiles à évaluer à l'avance pour les parties concernées. Le risque d'impact négatif sur la clarté globale et le fonctionnement de la directive semble donc supérieur aux avantages potentiels. La plupart des répondants à l'enquête sont toutefois favorables à la possibilité de remplacer le droit *sui generis* par un autre mécanisme de protection contre certaines utilisations illégales.
- L'**efficacité** de **OS3a** est comparable à celle de OS2a et de OS2b en levant l'incertitude juridique sur le champ du droit *sui generis* concernant les DGM. Mais cela ne ferait qu'ajouter une couche supplémentaire de droits exclusifs pour les créateurs de bases de données, à contre-courant de la vision principale de la Stratégie européenne pour les données. Car, le plus souvent, il n'est pas nécessaire de mettre en place de nouvelles incitations à produire des DGM, cela ne ferait que restreindre à l'excès l'accès aux et l'utilisation des DGM au détriment de l'intérêt général en matière de concurrence et d'innovation. Ce risque est d'autant plus important que de nombreuses bases de données contenant des DGM sont des bases de données à source unique (par ex. des DGM liés aux performances d'un appareil). En revanche, certaines des personnes interrogées ont fait remarquer que, comme pour OS2c, la protection pourrait encourager les créateurs de bases de données à rendre leurs bases de données publiques et à signer des contrats de licence avec les utilisateurs, favorisant ainsi une

[17] Programme pour une réglementation affûtée et performante (REFIT) de la Commission européenne

certaine forme de partage des données. Pour ce qui est de l'**efficience** de cette option, les résultats de l'enquête semblent indiquer que les coûts seraient légèrement supérieurs aux avantages par rapport à l'option de référence. De plus, alors que les répondants à l'enquête indiquent que l'inclusion des DGM (avec un droit d'accès) aurait un effet global plutôt positif sur l'innovation et la concurrence, cet avis semble moins tranché que pour OS2a. Néanmoins l'effet positif sur le partage des données mentionné par l'une des personnes interrogées pourrait apporter des avantages directs et indirects, avec des règles et une structure de protection plus claires pour les acteurs à la recherche de bases de données.

Cohérence

Les options OS2a et OS2b ne donneraient lieu à aucun conflit avec d'autres instruments juridiques de protection des données car elles limitent le champ du droit *sui generis*. OS2c pourrait éventuellement interférer avec la protection du secret des affaires. Le secret des affaires perdrait toutefois de sa pertinence dans une économie des données de plus en plus interconnectée, dans laquelle les données se partagent et s'échangent. OS3a est également compatible avec les autres cadres juridiques existants, car elle se contente de clarifier l'application du droit *sui generis* existant en matière de DGM. Dans le cas de OS3a, des conflits potentiels pourraient se faire jour, car le futur acte législatif sur les données pourrait mettre en place un droit d'accès. La protection *sui generis* des DGM pourrait entraver l'accès aux données car les ayants droit peuvent avoir recours au droit exclusif. Pour OS2c, le nouveau droit présente le risque de fonctionner comme une couche de protection supplémentaire. OS2a et OS2b, en revanche, semblent plus en phase avec les intentions possibles de l'acte législatif sur les données que OS2c et OS3a, à savoir éviter que la directive sur les bases de données ne devienne un obstacle au partage et à l'échange de DGM.

Options stratégiques supplémentaires

Ce groupe d'options traite de différents sujets qui ne sont pas directement comparables et ne concernent pas exclusivement l'initiative législative en cours (c'est-à-dire l'acte législatif sur les données), il s'agit de simples propositions à la Commission. C'est pourquoi elles n'ont été évaluées que par rapport à l'option de référence.

Efficacité et efficience

OSS S1 pourrait faire preuve d'efficience pour lever certaines incertitudes juridiques et augmenter l'utilisation des données au-delà des bases de données DGM. Ce point serait particulièrement important dans le contexte d'utilisation de gros volumes de données, où les utilisateurs peuvent avoir besoin de bases de données complètes (et pas seulement des parties non substantielles, comme dans les exceptions actuelles au droit *sui generis*). Sur le plan de l'efficience, OSS S1 peut bénéficier d'une évaluation positive par rapport au status quo, dans la mesure où les coûts supplémentaires seraient faibles, largement contrebalancés par le renforcement de la sécurité juridique et l'harmonisation des règles entre les exceptions du droit d'auteur et du droit *sui generis*. D'après l'enquête, les bénéfices de cette option seraient supérieurs à ses coûts. Les experts juridiques consultés sont également partisans de cette option. Avec **OSS S2**, le renforcement de l'exception au titre de la recherche pourrait améliorer l'harmonisation entre États membres, soutenir la recherche et l'innovation au sein de l'UE, ce qui aurait un impact positif sur la société. L'exception doit cependant tenir compte de, et être compatible avec, l'exception pour le TDM au titre de la recherche en cours de transposition dans les États membres. **OSS S3** élimine l'incertitude relative à la licéité de l'utilisation des contenus de bases de données appartenant à des organismes publics. Dans un grand nombre de cas, la directive concernant les données ouvertes et la réutilisation des

informations du secteur public (« Open Data Directive » ou ODD en anglais)[18] a déjà résolu les tensions entre l'accès et la protection des bases de données pour les bases de données du secteur public. Le problème auquel s'attaque cette option est donc considérablement moins étendu qu'avant l'existence de la directive ODD. Les coûts de mise en œuvre de OSS S3 sont faibles. D'autres gains en efficacité pourraient être obtenus par rapport à l'option de référence en réduisant les différends sur les incohérences potentielles avec la directive ODD et les coûts de transaction associés. La viabilité économique de l'activité des organismes publics auto-financés dépendant du droit *sui generis* pourrait être compromise si l'option n'exclut pas explicitement les structures privées actives dans les environnements concurrentiels.

Cohérence

OSS S1 sera en phase avec la volonté des institutions européennes de développer l'accès à et l'utilisation des données au sein de l'UE, et harmonisera le cadre juridique du droit d'auteur dans le marché unique numérique. Mais l'incertitude persiste en ce qui concerne l'alignement sur l'article 5 (« Exceptions et limitations ») de la directive 2001/29/CE sur le droit d'auteur dans la société de l'information, selon que les exceptions seront rendues obligatoires ou laissées à la discrétion des États membres. **OSS S2** ne semble pas interférer avec d'autres cadres juridiques existants une fois la compatibilité avec l'exception TDM garantie. **OSS S3** est en phase avec les objectifs de la directive ODD et avec le règlement sur la gouvernance européenne des données. Des incertitudes pourraient naître de la différence avec le champ de l'exclusion proposée par la directive ODD bien que les deux approches aient le même objectif.

Comparaison avec les options stratégiques relatives aux DGM

À la lumière de cette évaluation, l'option concernant les DGM qui paraît être le meilleur choix au vu des éléments de preuve à disposition est OS2a : l'exclusion des DGM. Cela permettrait de lever l'incertitude juridique dans la relation entre la directive sur les bases de données et les DGM, de plus en plus importante pour l'économie des données dans le marché unique européen, tout en veillant à ce que la directive ne fasse pas obstacle au partage des données entre différents secteurs de l'économie. Les options OS2a et OS2B sont les plus efficaces et les plus cohérentes par rapport à l'option de référence. OS2a semble être la plus efficiente des deux, car OS 2b comporte des coûts potentiels liés aux incertitudes entourant la mise en place du test de violation et la prise en compte du fait que l'arrêt (*CV Online contre Melons*) sur lequel s'appuie OS 2b fait déjà partie de l'acquis communautaire et devrait être respecté par les justices nationales. L'efficacité de OS2c et OS3a est jugée faible, voire nulle, tandis que leur efficience est inférieure à celle du status quo, avec une cohérence moindre par rapport à l'intention de l'acte législatif sur les données.

[18] Voir la directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public

TABLE OF CONTENTS

1 Introduction

This chapter illustrates the scope of the study and describes the methodology used to develop the analysis undertaken that led to conclusions on the viability of policy options targeting an improved sharing of machine-generated data ("MGD"), for which the relevance of the *sui generis* right defined in the Database Directive has been unclear.

1.1 Scope of this study

The present study contains the findings of the work undertaken on behalf of the European Commission (also "Commission") to support an Impact Assessment for the review of the Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases ("Database Directive" or "Directive") under the Framework Contract SMART 2019/0024. The various policy options assessed in the study have been defined and further elaborated in consultation with the Commission throughout the period of the study. The structure of this final report follows closely the Commission's Better Regulation Guidelines. The report aims at providing support to the Commission in its preparation of the Impact Assessment supporting the "Data Act" initiative.

1.2 Methodology for the assignment

This section provides a high-level presentation of the methodological choices taken for the following tasks:

- Stakeholder mapping;
- Semi-targeted survey;
- Interviews;
- Desk research and literature review;
- Workshops.

As part of the data-gathering exercise the team launched an online survey to collect information on the applicability of the *sui generis* right for databases containing MGD ("MGD databases") and also on views regarding the applicability, costs and benefits of various related policy options that could improve the sharing of MGD for the benefit of the society. The survey was initially conceived as targeted to specific industries relying on the Internet of Things ("IoT"), however, as applications of the IoT can be found across numerous sectors, the target was very broad, spanning several sectors. The geographic scope covered Denmark, Italy, Germany, France, Netherlands, Poland, Romania, Spain, and Sweden. To obtain a view of research organisations, the survey was further sent out by the EC to stakeholder organisations. Therefore, the geographic coverage extends beyond the initial focus. More details on the implementation of the survey can be found in the Annex.

The study team also carried out individual interviews with business stakeholders, companies, or business associations, in key sectors relying on MGD to discuss and gather evidence on the support of different policy options and the costs and benefits entailed. Furthermore, an online group discussion was organized with academic legal experts to receive inputs on the elaboration of the policy options.

Moreover, legal experts from the study team undertook an extensive legal analysis, both from a legal and Intellectual Property (IP) angle, based on desk research and a literature review of recent publications related to the Database Directive and more generally the data economy

and IoT environment. Desk research was also used to gather qualitative evidence on the expected impacts of policy options to compensate for the limited available evidence from the survey and interviews.

Finally, the study also relies in part on contributions to the Open Public Consultation (OPC) for the Data Act. The consultation activity of the OPC was not an integral part of this supporting study. In fact, the OPC ran in parallel with the study and remained open until the 9th of September when the study was already in its final phase. Nevertheless, the best effort was made to consider, to the extent possible, also this source for the present analysis.

Limitations of the analysis

The analysis has some limitations. In particular, the response rate in the stakeholder consultations was low, suggesting little awareness of the *sui generis* right. From the previous evaluation conducted in 2018, it was known that the Database Directive was used by sectors that traditionally rely on databases, i.e. publishers, while broader industries that are starting to collect and use MGD are less aware of the instrument and its potential use in the data economy.

Given that there has been little clarity and use of the Directive in the context of MGD, the analysis is based on limited empirical evidence. It mainly relies on the views of legal experts in industry, research, and academia as well as legal practitioners. Due to the thus far low application level of the *sui generis* right and a quite complex review and range of policy options, it was not possible to obtain reliable estimates on costs and benefits expected for each policy option.

2 Background and problem assessment

2.1 The background

Among the EU actions put forward in the communication on "a European strategy for data" adopted by the European Commission in February 2020[19], the Data Act has the objective to promote a European Single Market for data, fostering innovation and improving the competitiveness of European industries.[20] The Data Act will foster business-to-government (B2G) data sharing for the public interest, support business-to-business (B2B) data sharing, and evaluate the Intellectual Property Right ("IPR") framework. It will do this with a view to further enhancing data access and use, by also addressing some of the issues brought up by current related regulations and directives, including, among others, the Trade Secrets Directive, the GDPR, the Regulation on the Free Flow of Non-personal Data, and indeed the Database Directive. The review of the Database Directive was presented alongside the Data Act in the European Commission Work Programme 2021 as the "Data Package".[21]

The Data Strategy's review of the EU IPR framework includes the Database Directive and the Trade Secrets Directive. In the Commission's IP Strategy (November 2020), the Commission indicated a particular interest in the role of both directives in relation to MGD and other IoT data.[22] The review is equally linked to a related discussion, namely on the legal IPR framework in the area of artificial intelligence (AI). In the European Commission's "Trends and Developments in AI – Challenges to the IPR framework" (December 2020), the report indicates that "the creation/obtaining distinction in the *sui generis* right is a cause of legal uncertainty regarding the status of MGD that could justify review or clarification of the EU Database Directive."[23]

The Data Act should also be seen in light of the Digital Single Market Strategy, which identified several relevant legal areas for the construction of a data economy including "the right to use data". Possible rights to use data are to be seen as part of a comprehensive legal framework for a data economy.[24]

---

19 European Commission (2020b): Communication from the Commission, A European strategy for data, COM (2020) 66 final, 2.

20 This can be traced back to European Commission (2015): European Commission's Communication, A Digital Single Market Strategy for Europe, COM(2015) 192 final, 6 May 2015.

21 See European Commission Work Programme 2021: https://ec.europa.eu/info/publications/2021-commission-work-programme-key-document

22 See European Commission (2020c): Communication from the Commission, Making the most of the EU's innovative potential, COM(2020) 760 final

23 See JIIP, IViR – University of Amsterdam (2020): Trends and Developments in AI – Challenges to the IPR framework, Report for the European Commission.

24 Zech H (2016), "A Legal Framework for A Data Economy In The European Digital Single Market: Rights To Use Data", Journal Of Intellectual Property Law & Practice, Vol. 11, 460-470

One part of the existing legal framework for the data economy is the Database Directive. Following its evaluation in 2018 ("Evaluation 2018")[25], as part of the forthcoming Data Act, the Commission included the review of the Database Directive.

Key features of the Database Directive

The Database Directive[26], was adopted 25 years ago in 1996. At that time, the rapid expansion of the Internet led the European Commission to provide a harmonising legislation on databases within the European Union. The Database Directive thus addressed expected transformational failures linked to the exponential growth in the amount of information generated and processed with the rise of the Internet. An expected important role of databases "in the development of an information market within the community" (Recital (10), (9)), and economic concerns on unequal investments in the database sector led to the adoption of the Directive.

Its primary objectives are:

- harmonisation of a legal protection regime for the rights of database makers across Member States,
- promotion of investment in the production of databases – aiming to increase the competitiveness of the European digital industry,
- safeguard of the balance of interests between database users and database makers

Databases are defined in Art 1(2) of the Database Directive as "a collection of independent works, data or other materials which are arranged in a systematic or methodical way and are individually accessible by electronic or other means". Originally, databases store information in the form of lists, forms or tables related to one another. They can include personal or non-personal data. Non-personal data can be anything such as weather forecasts, data on the performance or maintenance requirements of commercial planes or industrial machinery.

The Database Directive protects the intellectual creativity embodied in the design of databases through copyright (Art. 3), and the investment in the collection, verification, and presentation of content through the *sui generis* right (Art. 7). This *sui generis* right, the main focus of the present study, applies whether or not the content qualifies for copyright protection or if the database is innovative or original.

The key features of the *sui generis* right are:

a) Protection of "substantial investment"
b) Protection against acts of extraction and re-utilization
c) Exclusion of "insubstantial parts" from the protection
d) Exceptions for certain uses
e) A protection period of 15 years
f) Available only to EU nationals or habitual residents; to non-EU nationals only on the basis of reciprocity

[25] JIIP and Technopolis (2018): Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases, Report for the European Commission

[26] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

2.1.1 Emerging challenges

Today, organisations can entirely transform their businesses by gaining access to valuable sets of data. This leads to a rush to either build or buy data [27] and contributes to the rise in data analytics. “Traditional” or non-tech firms need to get familiar with complex legal questions on acquiring rights or managing restrictions on the use of data and data analytics.

The European Commission has identified a number of challenges in its European Strategy for data [28], some also relevant in context of the Database Directive:

- Growing data volumes but insufficient availability of data for use and re-use,
- Fragmentation and still limited availability of public data (government to business (G2B) data sharing),
- Sensitive data in databases of the public sector are not made available for research,
- Insufficient data sharing among companies due to
    - lack of economic incentives,
    - lack of trust and imbalances in negotiating power,
    - fear of misappropriation of the data by third parties,
    - lack of legal clarity on who can do what with the data,
- Imbalances of market power in terms of data infrastructures and cloud services leading to
    - data analytics advantages,
    - unilateral setting of conditions for access and use of data,
    - unilateral advantages for expanding into new markets.

The review of legislative actions that affect the relations between actors in the “data-agile economy”, such as the review of the Database Directive, are part of the Data Act. The key focus of the Data Act, i.e. how to increase access and further use of data – including MGD – for the wider benefit of private and public users, is thus also considered in the review of the Database Directive.

The Inception Impact Assessment of the Data Act [29] highlighted problems in relation to data access and use in business-to-business situations (in-line with the above-cited Communication), among others, the imbalances in negotiating power between data holders and access seekers. This may lead to imposing unfair data licence terms on the data recipient,

---

[27] For example, Microsoft acquiring LinkedIn, Google acquiring FITBIT, Facebook/WhatsApp, Clarivate acquiring DartsIP, TrademarkVision, ProQuest – among other IPR specialised data providers.

[28] European Commission (2020b)

[29] Inception impact assessment Data Act (including the review of the Directive 96/9/EC on the legal protection of databases) (2021): Inception impact assessment -Ares(2021)3527151

or a refusal to provide access to data.[30] If access and use of MGD are curbed they generate a further potential problem with a societally sub-optimal re-use of data and non-realised innovations. In this context the Database Directive with its legal uncertainly on its application to MGD may create additional complexity and uncertainty, thus becoming an obstacle for the use of that data.

The identified economic problems are not all based on are derived from observable technical trends and case studies.[31] However, in terms of the planned review of the Database Directive, its 2018 evaluation and the consultation activities in the context of the Data Act provide a basis for potential changes.

2.1.2 Evaluations of the Database Directive

The Database Directive has been evaluated twice since entering into force, first in 2005 and more recently in 2018.[32] Both evaluations focussed on the "database sector", i.e., publishers, suppliers of data and information, database producers, distributors etc.

The 2005 evaluation was highly critical of the application and effects of the Database Directive and concluded that the economic impact of the *sui generis* right on database production was unproven and invited further analysis. Despite its limited effectiveness, the evaluation did not recommend the repeal of the Database Directive since it did "*not impose significant administrative or other regulatory burdens on the database industry*"; a repeal would create a risk that Member States fall back on national and fragmented legal options; in particular, they could limit protection only to "original" databases. The database industry on the other hand appreciates the *sui generis* protection.

The Evaluation 2018, aimed to assess the effectiveness, efficiency, relevance, coherence, and EU-added value of the Database Directive against its following three aims: (1) to harmonise database protections; (2) stimulate investment in databases; and (3) safeguard the balance between database makers and users.

An open public consultation carried out in 2017 indicated the heterogeneity in perceptions about the achievements of the Database Directive:

- While about two-thirds of the respondents agreed that the Database Directive sufficiently protects investments in creating databases, the views on its effects on the creation of the databases remained mixed.
- There was a rather clear opinion on the effects on data access and re-use of data: about one-third found the Database Directive had positive effects on data access and re-use, while almost two-thirds certified negative or no effects.

---

30 See also Bird and Bird (2021): Commission Inception Impact Assessment on a proposed Data Act: What does it mean for IP owners? Lexology

31 See for example OECD (2015), European Commission (2020a) and Datalandscape (2018): The European Data Market Monitoring Tool

32 See European Commission (2005): First evaluation of Directive 96/9/EC on the legal protection of databases. DG Internal Market and Services Working Paper, and JIIP, Technopolis (2018) Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases.

- There were also strong concerns regarding the expected balance between right-holders and users – two-thirds of the respondents found this unbalanced.
- A small majority did not find the Database Directive fit for purpose in a data-driven economy.
- A very clear majority was against the proposal to apply the *sui generis* right to MGD.[33]
- There were mixed views on the scope of the *sui generis* right. Almost half of the respondents found it satisfactory or too narrow, while slightly more than half of the respondents found it too broad or unclear.

The diversity of opinions about the usefulness of the Database Directive came with the differences of stakeholders: while the publishing sector was, by and large, satisfied with the Database Directive and its scope, more criticism came from other sectors such as transport or IT, as well as research. The latter was inter alia concerned about legal certainty for users in research and academia concerning the access to and re-use of data.

Results of the Evaluation 2018 remained critical concerning the goal attainment: based on survey outcomes, positive effects of the *sui generis* right to **stimulate investment on databases remained unproven.** Yet in workshops and interviews, several database makers and user-makers considered the *sui generis* as essential for the growth of their business. This group did not confirm positive effects on further investments in databases triggered by the *sui generis* right. Positive effects were further contested, and the *sui generis* right was seen as a barrier for scientific databases.[34]

Only database makers found the *sui generis* right an effective mean to protect databases together with the range of other protection means (such as contracts).

- In terms of costs and benefits, the effects of the *sui generis* right on database users and user-makers were found to be small. The benefits were incurred mainly by database makers through improved legal certainty and better protection of their databases against unlawful access and use by third parties. For the users, the improved certainty as to what constitutes lawful use was seen as a potential benefit through the reduction of legal costs.
- Overall, there was a very low awareness about the *sui generis* right on the side of database users, while database makers did not consider it to be a reason for further investments in the generation and collection of data.
- The Evaluation 2018 noted that the scope of the database rights was restricted to the collection of data and not more broadly to the creation of data, so that most websites and MGD databases (such as the ones containing data automatically produced by connected devices in the IoT environment) would fall outside its scope.

The legal analysis of the supporting study for the Evaluation 2018 and the interaction with stakeholders in 2018 suggested that the Database Directive provided an "**outdated legal**

[33] See https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-evaluation-directive-969ec-legal-protection-databases

[34] See p.26 of Annex 2 of Evaluation 2018.

framework" since it was not in line with the recent technical developments of the data economy and contained a number of unclear provisions.

2.1.3 Clarification through case law

One of the main caveats in interpreting the extent of protection of the *sui generis* right has been the lack of clear definitions contained in the Database Directive. Since 2004, judgements of the Court of Justice of the European Union ("CJEU") have brought further clarification. The first cases[35] that focused on the *sui generis* right (Art. 7) in the context of sports data fixture lists – and their unauthorised use by betting companies – have been particularly important.

The *British Horseracing Board v William Hill case ("BHB v Hill"),* provided clarification about the qualification for protection for database makers in terms of the **substantial investment requirement.** According to Art. 7(1) of the Database Directive, in order to qualify for protection the database maker must show that there was "a qualitative and/or quantitative substantial investment in either the obtaining, verification or presentation of the contents" of the database. The investment in:

- "obtaining" requires resources to be spent in seeking out independent material and collecting them into the database;
- "verification" relates to resources to be spent in ensuring the reliability of the information contained in the database and monitoring its accuracy;
- "presentation" relates to the investments in the arrangement of the data in the database.

The CJEU thus held that only investments into "obtaining" the contents of a database will be relevant for the substantiality threshold, whereas investments into the "creation" of material are irrelevant. Several authors have interpreted this ruling that the investments of "producers" of data - be they sport event organisers or firms that collect MGD or sensor-generated data ("SGD") - would be excluded from the *sui generis* right since in most cases, such investments would be considered in the "creation" of data.[36] The narrow interpretation of the Court in this ruling, would exclude a lot of applications relevant in big data scenarios. The "BHB v Hill doctrine" was confirmed in subsequent judgements. In 2015, the Court provided in Case C-490/14, *Freistaat Bayern v Verlag Esterbauer* that investments into the establishment of an infrastructure for obtaining, verifying of presenting data in order to obtain certain pre-existing use, sales or geographical data, will be relevant for assessing substantiality under Art. 7 (1). [37]

The most recent ruling of the CJEU on C762/19 *CV-Online Latvia v Melons* ("*CV v Melons*") provides an additional interpretation of "extraction" and "reutilisation" of substantial parts of a

---

[35] C-203/02, The British Horseracing Board Ltd & Ots v. William Hill Organization Ltd; C-46/02, Fixtures Marketing Ltd v. Oy Veikkaus AB; C-338/02, Fixtures Marketing Ltd v. Svenska Spel AB; C-444/02, Fixtures Marketing Ltd v. Organismos Prognostikon Agonon Podosfairou

[36] See European Commission (2017a): Staff Working Document on the free flow of data and emerging issues of the European data economy, SWD (2017) 2 final, 20 It is thus this "narrow' interpretation.

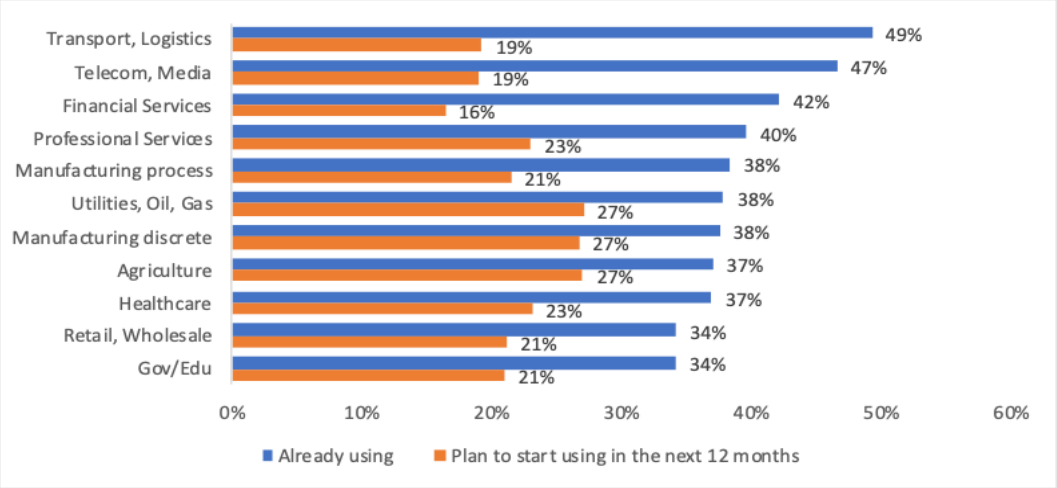
[37] See Leistner, M. (2018): Big Data and the EU Database Directive 96/9/EC: "Current Law And Potential For Reform", SSRN and Noto La Diega, G. (2019): Artificial Intelligence and Databases in the Age of Big Machine Data

database. According to Derclaye and Husovec, this ruling "has a different flavour" from previous ones in the sense that it requires a balance between the database maker's investment on the one hand, and the legitimate interest – even of competitors - on the other hand[38] (see also legal assessment of policy option 2b in Section 3.3 for a more detailed analysis).

2.1.4 Shift of scope of application of the Directive

The current review thus assesses whether the Database Directive can and will have a broader potential scope of application. While the current Directive mainly interests the more "traditional" database industry (i.e. commercial database providers and publishers), provisions concerning MGD and SGD could potentially open up the scope to a much wider community of players in sectors that generate and use MGD.[39] Currently, the uptake of sensors and IoT applications can be seen in sectors such as transport, energy, agriculture, and many manufacturing sectors which already are core IoT users. In the near future, however, the expected growth in the use of sensors may expand into all manufacturing and other industries (see Figure 1) with increasingly more MGD data collected in the private sector. Policy makers fear this data is not widely shared, nor its access granted, jeopardising the positive effects assumed from aggregation and combination of data.

Figure 1: Uptake of IoT, selected industries



Source: Data: IDC 2020, compilation: Technopolis group

[38] See, Husovec, M., Derclaye, E. (2021): Access to information and competition concerns enter the sui generis right's infringement test – The CJEU redefines the database right, Kluwer Copyright Blog. Available at: http://copyrightblog.kluweriplaw.com/2021/06/17/access-to-information-and-competition-concerns-enter-the-sui-generis-rights-infringement-test-the-cjeu-redefines-the-database-right/

[39] See, for example, https://ati.ec.europa.eu/.

2.1.5 Data and MGD

Data is an input of increasing importance for a growing number of economic activities that exhibit a number of unique characteristics.

The often-heard terms “data is the new oil” or “data is the oil of the digital era” indicates that carries great value. However, contrary to oil (which is a depletable resource):

- data can be used and reused multiple times by different individuals. In economic terms data is characterised by non-rivalry, which means that someone’s use of the data does not constrain someone else from using the same data (unless access to data is technologically or administratively blocked). Social welfare will in principle be maximised by guaranteeing full access to data. Introduction of exclusive rights needs a special justification.
- obtaining, verifying, and presenting data within databases may require a substantial investment, but data gathering, processing, and sharing is not necessarily costly,
- raw data can be any information or characteristic capable of being processed by computers (e.g., text, characters, symbols, photos, time),
- data may have a “use by date” but it can be endlessly reused and repurposed,
- the value of data varies depending on its attributes, the amount it is used, and the creativity employed when using it. The value may be difficult to ascertain until it is combined with other data.[40]

Among the key issues raised in relation to the *sui generis* right is that it does not directly protect access to data as such, but protects only indirectly against taking and re-using of data stemming from the protected database. Thus, the *sui generis* right protects data only in an indirect manner.

Due to some of its special characteristics, data is a difficult subject matter for ownership rights. Special problems include specification to avoid a “Super-IP” that would result in an unqualified protection of data and information without any balance of countervailing interests of other stakeholders and the public. Moreover, it would be very difficult to allocate initial ownership because of the many stakeholders involved in the value chain (co-producers, data holders, intermediaries etc.).[41]

Recently, the notion of MGD has attracted considerable attention both legally and economically. To approach this notion, in a technical sense, it starts with everyday business activities that generate data. In the value chain of IoT information, there are business activities such as manufacturing goods but also healthcare systems, smart transport systems, smart buildings etc. The devices and machines used in the production and operations of these services and goods include sensors that produce data by detecting changes in the environment and functions/performance of the machine. Commonly used sensors can measure, among other things, temperature, pressure, humidity, radio frequencies, etc. Krishnamurthi et al.

---

[40] McKenna and Olswang (2021): Embracing open data is now more important than ever (open data note 2 of 2), Drexl (2018): Data Access amd Control in the Era of Connected Devices.

[41] See Hugenholtz, P.B. (2017): Data Property in the System of Intellectual Property Law.

(2020),[42] in a technical article that contains a high-level overview of the basic architecture behind IoT sensors networks, explain that the majority of IoT sensor data "incorporates real-time processing". Often this includes data outlier detection, missing data imputation and data aggregation as raw data composed by sensors signals is often voluminous and unclean requiring too much computing power and storage to be analysed as the network scales up. This already (pre) processed data is then transmitted usually wirelessly to a computer processor or cloud storage. Therefore, in the data value chain after the acquisition data is often formatted and filtered to be subsequently further processed.[43] On any step of the value chain data analytics may be applied that can lead to descriptive, predictive, or cognitive analyses and insights.[44]

Databases will store data and their use in practice cannot be reduced to certain phases of the data life cycle. Sensor data are heterogeneous, comprising different data sources from structured datasets to real-time data-generation networks, social network media data streams, and other participatory sensor networks[45]. The format of data will be determined by the nature of the connected stage of the value chain and it assures the persistence and management of IoT data to satisfy the need of the application. E.g., if data is stored for data analytics, the preceding formatting or other processing may be related to the purpose of the database and, hence be said to constitute an investment relating to the database. This may be the case at any stage of the data life cycle. The processes may be inextricably linked with the consequence that investment in the different phases and processes could not be separately identified. E.g., processing may be done by the sensors themselves.[46] More detailed technical discussion on sensor data, its transmission and data management, as well as use cases are presented in the Annexes.

Box 1 below provides the working definition of MGD used in the present study.

Box 1: Working definition of MGD used in the present study.

One of the challenges that arise when thinking about possible legal measures addressing databases including MGD, lies in the difficulty in providing a legally stable, definable notion of MGD.

In its 2017 Communication "Building A European Data Economy", the European Commission defined MGD as data "*created without the direct intervention of a human by*

[42] Krishnamurthi, R., Kumar, A., Gopinathan, D., Nayyar, A., & Qureshi, B. (2020). An overview of IoT sensor data processing, fusion, and analysis techniques. Sensors, 20(21), 6076.

[43] CREATE IoT Project (2017): CROSS FERTILISATION THROUGH ALIGNMENT, SYNCHRONISATION AND EXCHANGES FOR IoT, D05.03 IoT Data Value Chain Model. p. 24-25.

[44] Deloitte insights see Deloitte (2018): The internet of things: A technical primer or McGrath et al (2013): Sensor Technologies: Healthcare, Wellness and Environmental Applications.

[45] Krishnamurthi et al. (2020).

[46] See example p. 110, Final Report, Database Evaluation 2018.

computer processes, applications or services, or by sensors processing information received from equipment, software or machinery, whether virtual or real."[47]

However, this definition is contentious as various authors and legal experts[48] have remarked that humans have an inevitable role in the creation of data, even that generated by artificial Intelligence or machine-learning.[49] Starting from the same observation, an interviewee active in the automotive sector expressed his concerns about the uncertainty of defining MGD as "independent" from human action. As a way to overcome this, Drexl (2018) uses a broader definition and considers MGD those generated by "connected devices", being the latter connected to other devices or persons. In Drexl's definition "devices used by humans for the purpose of communication, such as PCs, tablets and smartphones, are equally covered, because it is not relevant to which extent data is stored or processed by the device with or without being influenced by the decisions of a natural person. In fact, most data collection through smart devices is influenced to some extent by decisions of the user as a natural person, such as in the case of connected cars or household devices".

For the purpose of this study and during the stakeholder consultations the following working definition of MGD has been used:

MGD is defined as data recorded, collected, or generated independent of direct and economically significant human intervention by:

- *sensors processing information received from equipment, software or machinery, whether virtual or real*
- *computer processes, applications or services.*

The present definition may include data that are the result of observation of (acts of) humans (e.g. which pages within a website a person visits). However, data that are the result of humans consciously providing information/choices are excluded (persons typing in their name, address etc. to create an account). MGD include SGD-IoT as well as data generated and used by some online intermediary application.

The definition "data recorded, collected, or generated by sensors" refers primarily to sensor-generated data in IoT environment. In line with the approach taken by the Deloitte-led supporting study for the Data Act this type of data includes:

- data generated by sensors about the sensor and machine itself, e.g. data on machine performance;

---

47 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "Building a European Data Economy", COM(2017) 9 final.

48 This has been expressed during the legal expert workshop conducted for the present study

49 A recent study on the Trends and Developments in Artificial Intelligence (2020) commissioned by the European Commis-sion maintains that "fully autonomous creation or invention by AI does not exist" and that AI systems will continue to be tools in the hands of human operators. See JIIP and IViR (2020): Trends and Developments in Artificial Intelligence

- data generated/observed by sensors observing the environment in which sensors and machines operate, e.g. information on the soil recorded by sensors in smart tractors;
- the data resulted from the aggregations and processing of the two types of data above.

It must be noted that the above definition of MGD used for the purposes of the study may not be the same as what a revised Database Directive would use as a new legal category. Moreover, a definition encompassing (to some extent) this type of data may also be provided in the Data Act.

2.1.6 Economic Context

The European Data Market study (2017) expected that the value of the EU data economy would more than double from almost €300 bn in 2016 to €739 bn in 2020 (including the UK). [50] This estimate was revised down in 2020, to €355 bn in 2020 excluding the UK, (or €443 bn including the UK). [51] Further growth in the EU data economy is predicted to continue in a high growth scenario to reach €827 bn for the EU-27 in 2025. The data economy will need to be based on accessible data, often also referred to as "open data". The expected growth in "open data" is correlated to economic growth and technological change: it is expected that there will be accelerated public and private investments in key enabling technologies such as artificial intelligence ("AI"), robotics and automation. "Open data" has the potential to lead to more innovation, low data power concentration, a high level of data sharing and a wider distribution of the benefits derived from data innovation.

In addition to the clear economic benefits, there are social and environmental impacts emerging from access to public and private sector data. According to IDC/Lisbon Council (2017) "*it is a safe assumption that the vast majority of data globally are held by private companies*" and this was considered to "*offer radical new insight on aspects of life that were never measured before and could drive an unprecedented understanding of human behaviour as well as natural phenomena"*.[52] Impact cases of the use of open data are clearly seen in health, economic development, but also environment (climate change), crisis response, agriculture etc. A recent example of how open data helps society can be seen with the open data principle used during Covid 19. Many database providers and publishers have issued over 80,000 pre-prints and peer-reviewed articles and research data has been shared helping to share knowledge about the virus in an unprecedented way. It allowed scientists to scrutinise scientific results and transfer the scientific findings to drug development and inform policy makers in developing better prevention schemes. This case demonstrates that open data can have massive societal benefits.

---

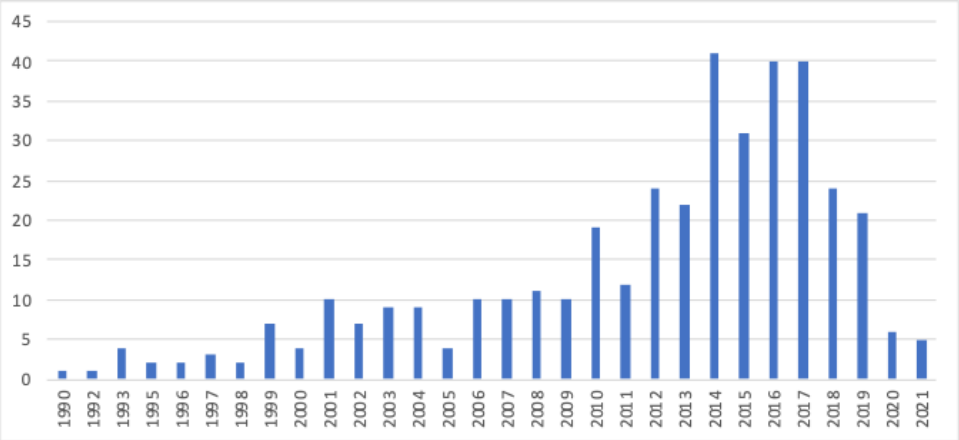
50 See European Commission (2017b), The European Data Market Study Update.

51 See European Commission (2020d), The European Data Market Study Update

52 IDC (2021): Global Datasphere And Storagesphere Forecasts IDC/Lisbon Council.

The increasing demand for value-added activities linked to aggregation and/or combination of data can be seen in the increase of data analytics services as evidenced by the number of start-ups active in this area in the EU- 27 (see Figure 2).

Figure 2: Number of start-ups in "Data analytics", EU-27



Source: Crunchbase, treatment: Technopolis Group

The current and predicted numbers in the area of big data and MGD suggest a dramatic surge in this type of data collected in the private sector:

- The size of database management systems (DBMS) market worldwide increased from US$ 39 bn in 2017 to US$ 65 bn in 2020. Most of the growth accounted for the migration of database systems to cloud platforms.[53]
- The total amount of data created, captured, copied, and consumed globally was forecasted to increase rapidly, reaching 64.2 ZBs in 2020 and growing to more than 180 ZBs by 2025 [54]
- The data volume of IoT connections is expected to grow by a factor of 10 to almost 80 ZBs between 2019 and 2025.
- The number of publicly known IoT platforms was more than 620 by the end of 2019, more than double the number in 2015 (260).[55]
- The growth of MGD can also be proxied by the growth in the market of sensors: worldwide, revenues from IoT sensors amounted to US$ 12 bn in 2019. It is expected that this will more than triple and reach US$ 43 bn in 2025. At the same time, the price of sensors is expected to fall, contributing further to mass deployment of the IoT endpoints in industry. The rise of 5G and cellular IoT modules are key to the rise of connected devices. In 2020, there were 12.4 bn connections. By 2026, this will rise to

[53] See Statista (2021): Size of the database management system (DBMS) market worldwide from 2017 to 2020

[54] See IDC (2021): Data Creation and Replication Will Grow at a Faster Rate Than Installed Storage Capacity, According to the IDC Global Datasphere And StorageSphere Forecasts

[55] IoT Analytics

26.4 bn.[56] Key sectors which are already today relying on cellular IoT is the automotive sector, smart home devices and smart agriculture.[57]

The European data market value, defined as the place of services and products bought following the elaboration of raw data, has equally seen impressive increases from €54 bn in 2015 to €80 bn in 2020.[58] In the same period, also the number of data suppliers and data users has grown. While the former grew from 250,000 to 290,000 (average annual growth of 3.6%), the number of data users rose from 650,000 to 726.000 (2.2%).

While the above focused on global trends, the survey conducted for this study ("Survey for the Database Directive Review"), as part of the stakeholder consultation, attempted to understand this subject in relation to the present situation. Questions asked, for example, whether MGD was relevant for the business, what type of databases were used to store MGD, as well as issues related to data access and sharing.[59]

The 114 respondents were mainly a mix of data holders, data re-users and data re-users/data intermediaries. They were asked about general statements on data, which indicate the **relevance of data** as part of their businesses. The "generation of data only for own purposes" and "using data to design innovative solutions and products" were considered by most as important. By contrast, "selling/licensing data to third parties" was rather unimportant, while "sharing data for free on a mutual basis with business partners and platforms" was very important. "Providing free services to third parties based on data", "obtaining data from third parties through contracts/licensing" or "optimisation of production processes/devices", and "using data for the training of AI" were all very important aspects too.

The participants see the value of data beyond their internal usage in the organization. While respondents did not know whether selling or licensing provided extra profits to the company, they strongly disagreed that **selling or licensing to third parties** would put exclusive data analytics insights in danger. Furthermore, they agree that the selling or licensing of data to third parties - and thus enabling the re-use of data - enables additional user and consumer **benefits**.

The responding organisations also **invested substantially in the data infrastructure** which was described as the IT system, database structure, protection and training aspects. In terms of the investments in the infrastructures, a rather diverse picture emerged: average investment costs ranged from below €5,000 to more than €100,000, while the annual average operating costs was reported as mainly a function of the license model, backup and restore capabilities, administrator (employment) costs among other factors. Investments in MGD databases were undertaken by respondents' own companies but also by device operators.

This suggests that organisations know and use data – including MGD – for their own purposes but at the same time, they see the value of the re-use of data beyond their own organisation.

---

[56] See Erikson Mobility Report (2021): 5G on the road to mass market

[57] Cision PR Newswire (2019): Global IoT Sensors Market Analysis, Trends, and Forecasts, 2019-2025 - Global Market is Projected to Grow by US$35.2 Billion, at a CAGR of of 34%

[58] See Datalandscape (2018): The European Data Market Monitoring Tool.

[59] A full analysis is provided in the Annex.

This position may be explained with the view that the organisations benefit from buying third-party data themselves. In fact, as mentioned by several respondents, **access to third party data** is often fundamental for the business model of companies (e.g., aftermarket sales).

As for **data sharing**, respondents strongly agree that data sharing for free on a mutual basis with business partners or on a platform provides them with benefits. However, when it comes to the actual sharing, two-thirds of the respondents had encountered problems when requesting access to other companies' data. As an example, a transparency platform (an open platform in the energy sector) reported that it was "caught up in the sui generis right but the Directive does not inform about public licensing*"*. Another example mentioned by participants concerned vehicle manufacturers that either refuse to grant access to real-time data or put burdensome restrictions in place when data access is granted. The sharing of databases, however, does not occur often. In fact, only one in three IT experts reported that they share databases "on a daily basis", "several times" and/or "only a few times".

The responses obtained on the **infrastructures** used to handle databases suggest that, in the context of this survey, the **use of cloud services is rather limited**: only 15% of IT experts indicated that their organisation uses cloud services. Cooperation with other companies in establishing and running databases happens but the frequency varies from a few cases to a wide extent. Cooperation is more often seen with companies within the same industry but suppliers and – to a lesser extent – customers can be relevant partners too. Cooperation with companies from other industries, research or the public sector were more limited. The collected data – including MGD – is most often stored in the in-house databases and more seldomly in the cloud.

It has been stated earlier that value from data is generated through **data analytics**. Respondents that conducted supplementary data analytics were asked where the newly obtained data would be stored. While almost half of the respondent IT experts did not know the answer to the question, the remaining half was evenly split between the options to include it in the original database or a new database.

The **importance of MGD** for the business and the good functioning of operations was confirmed by more than half of the IT experts participating in the survey while a couple of them indicated that it is becoming increasingly relevant.

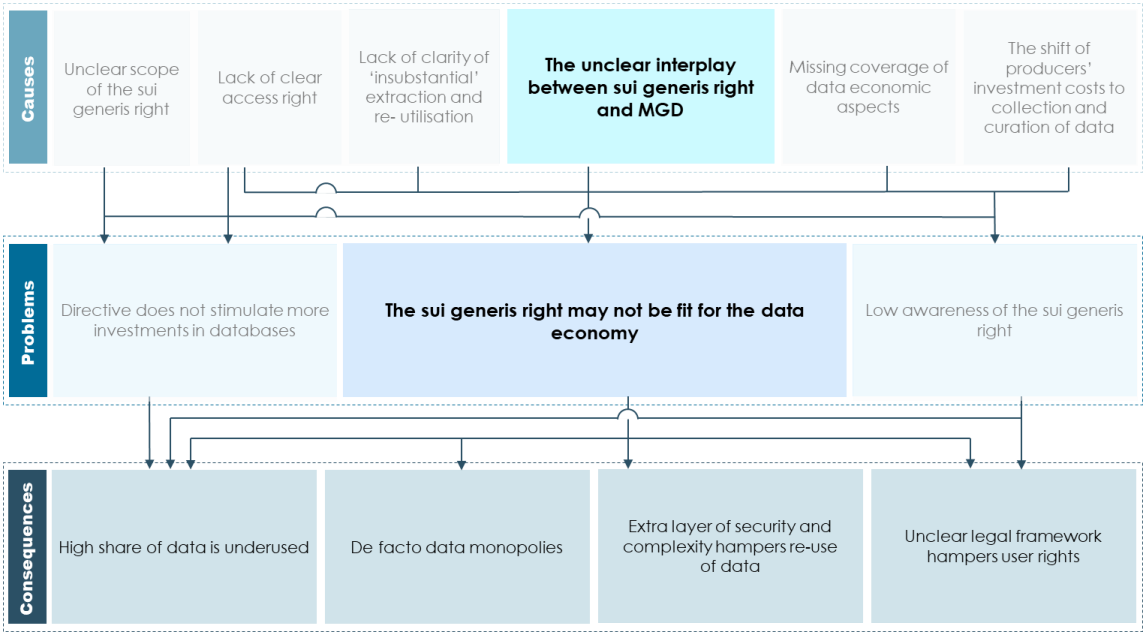
However, almost half of the IT experts acknowledge that companies may, in some cases, become the only providers of certain data (i.e. **sole-source databases**), thus establishing *a de facto* data monopoly in relation to that particular data. The issue was reported in relation to both MGD and non-MGD data. Again, more than half indicated problems when seeking to obtain access to databases. Hampering factors were (in decreasing order of magnitude) the fact that the database was legally protected, the fact that no licensing was available, the lack of interoperability, the fact that data was kept secret, and too high costs.

2.2 Problem assessment

This section presents the problem assessment in relation to the emerging challenges presented above, drawing on evidence from the previous Evaluation 2018 as well as from desk research and the findings of the stakeholder consultations undertaken for this study. These will serve as the basis for the formulation of the objectives of the review which will focus on the *sui generis* right part of the Directive and its interplay with MGD, the possible actions to achieve the objectives and the understanding of the likely impacts.

A problem tree summarising the problems, their drivers, their effects and their logical links is presented in Figure 3.

Figure 3: Problem tree for the review of the Database Directive



2.2.1 The Problem

The objective of the *sui generis* right in the Database Directive has always been to promote the production of databases by protecting database makers that undertook a substantial investment in either the obtaining, verification, or presentation of the contents of the database from extraction and/or re-utilization of the whole or of a substantial part of such database content.

However, the considerable changes in the technological landscape and economic context over the past 26 years have tested the applicability of the Directive. The Evaluation 2018 found that **the Database Directive may not be fit for purpose in the new data economy. A key element of this general problem is the lack of clarity of *sui generis* right with respect to MGD and the IoT context**. Therefore, the most helpful action – and focus of this review - will be to address the related cause, namely the unclear interaction between the *sui generis* right's protection and MGD.

The Evaluation 2018 also concluded that the Directive had not been effective in incentivising database creation in the EU and that there was a low awareness of the *sui generis* right as a protection tool for databases. These could be viewed as two subproblems of the key problem that the Directive - particularly the *sui generis* right - may not be fit for the data economy. We note that the low awareness could be also seen working as a driver to the fact that the Database Directive is not stimulating more investments in the production of databases in the EU.

2.2.2 Causes of the problem

The Database Directive adopted in 1996 is often criticised, in particular by legal academic experts, as **being too general and vague.** This concerns several aspects of the *sui generis* right, starting with a definition of a database that results in uncertainty as to the extent it covers databases created in an IoT context. Moreover, it is not clear which type of investment is to be considered in the assessment of substantiality and how the exact definition of the thresholds of protection on the side of the subject matter as well as the scope of protection should be taken. The Court of Justice has provided some clarification over time that need to be taken into account when interpreting the relevant provisions of the Directive, yet several aspects remain vague.

Below we provide a structured discussion on the lack of precise definitions in the Database Directive which tend to be associated with the problems identified in the previous section.

2.2.2.1 Unclear scope of the sui generis right

Legal academic experts identified problems arising from the scope of protection of the *sui generis* database right. For example, Art. 7(1) of the Database Directive recognises a too broad right of the database maker "to prevent extraction and/or reutilisation of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of the database". [60] In Art. 7 (2), the two sub-rights to prevent extraction and to prevent reutilisation are defined more concretely. However, in the current data economy, the most pertinent question is how to construct the requirement for the use of a "substantial part of the contents of the database". If the provision protected against any extraction or (re)utilization of data that is contained in a protected database, the *sui generis* right could amount to a far-reaching right to control the use of information. [61] When comparing exceptions to the *sui generis* right, legal experts find the Database Directive to be too narrowly designed. Leistner (2020) warns the issue represented by the narrow design of the exceptions is even more critical with the new challenges of the data economy, where big data applications often necessitate the use of (multiple) complete databases. [62]

2.2.2.2 Lack of clear access rights

Much of the legal and economic discussion about data centres around access to data and access rights. Our analysis of relevant documents (included in the reference list) was no exception (see Figure 4).

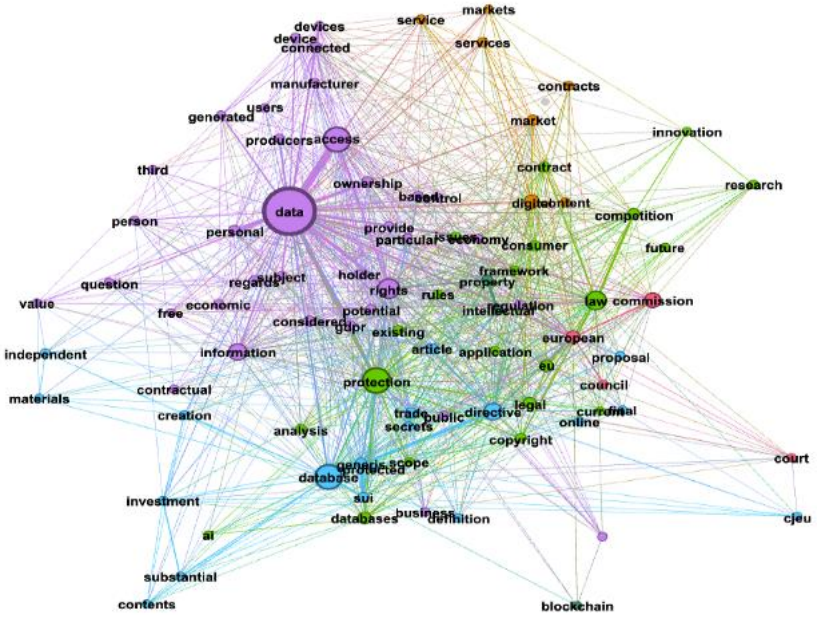
---

[60] See Noto La Diega, G. (2019)

[61] See Myška, M., & Harašta, J. (2016): Less is More? Protecting Databases in the EU after Ryanair

[62] See Leistner, M. (2020): The existing European IP rights system and the data economy – An overview with particular focus on data access and portability.

Figure 4: Core semantic network within the body of legal discourse documents



Source: Technopolis

In the context of MGD, manufacturers of connected devices are technically able to control access to data. While they do not necessarily "own" the data, as there is no ownership right in the legal sense defined for data, they are "data holders". They can thus commercialise the data by licensing it to third parties and they can equally exclude others by refusing such licensing. A second area of potential risks of data lock-ins concerns services, for example aftermarket services such as repair, predictive maintenance, or data analytics, where the access to the MGD may be considered instrumental. The data lock-ins problem in aftermarkets has also been mentioned by associations of auto parts makers that participated in our survey.

Adding to existing technical control, data holders might use the *sui generis* right on databases as a legal instrument to grant, even if indirectly, legal "ownership" to data. This may risk to further decrease the accessibility and usage of MGD databases.

2.2.2.3 Lack of clarity of "insubstantial" extraction and re-utilisation

The Database Directive does not clearly define the terms "insubstantial part", "substantial part", "quantitative" or "qualitative". The resulting uncertainty is aggravated by the fact that the CJEU has so far decided that the terms "extraction" and "re-utilisation" were to be given a wide meaning.[63]

The substantiality of the contents taken are to be evaluated in terms of quantity or quality (Art. 8(1) Database Directive) A quantitative evaluation is based on the volume of the extracted or re-utilised data from the database while a "qualitatively substantial part" refers to the scale of investment in the obtaining, verification or presentation of the database contents, regardless of whether that part is a quantitatively substantial part of the contents. This implies that

[63] See Evaluation 2018. This was also further discussed in a group interview with legal experts.

obtaining, verifying or presenting a quantitatively small part of the content, may still require a substantial investment.

The CJEU did not consider the intrinsic value of the extracted or re-utilised content to be a relevant criterion for the assessment of whether a part is substantial or not. In fact, the CJEU also ruled that a part which does not fulfil the requirement of a substantial part is automatically an insubstantial part[64]. Since the CJEU has so far not given more specific guidance on what a quantitatively substantial or insubstantial part is, e.g., in terms of percentages of a database, the use and interpretation of these terms continue to be not clear.

The recent *CV v Melons* case seems to introduce an additional infringement test that is added to the finding of taking of substantial parts of the contents and includes a balancing of interests. The taking of substantial parts would only be prohibited where those acts adversely affect its investment in the obtaining, verification or presentation of that content, namely that they constitute a risk to the possibility of redeeming that investment through the normal operation of the database in question as the main criterion, which it is for the referring court to verify. While this does not affect the distinction of substantial and insubstantial taking, in fact it applies criteria similar to those that have been applied as to insubstantial taking. It could be a fair analysis of the decision that it de facto abolishes this distinction in practice.

2.2.2.4 The unclear interplay between sui generis right and MGD

The growth of data gathered and created through sensors and machines raises the issue whether databases containing this kind of data can fall under the *sui generis* right. As discussed above, and found in the Evaluation 2018, it remains unclear whether the Directive **does or does not include or exclude databases including MGD.**

The reason why it is often argued in the academic literature that the *sui generis* database right will not apply to MGD relates to the distinction between “creating” and “obtaining*”* data as introduced by the CJEU in its case law.[65] The Evaluation 2018 summarized that in the current context, it seems that the Database Directive does not apply to the databases generated with the means of machines, sensors and other new technologies (such as the IoT or AI). In fact, the generation of these databases is closely interlinked with the creation of their content (i.e. data) and “case law indisputably excludes investments in data creation from the scope of the *sui generis* right”.[66]

By limiting investments in the generation of data relevant for the substantiality assessment to those that could be separately demonstrated, the *Fixtures* and *BHB v Hill* cases 2004 could be interpreted as indirectly resulting in excluding MGD from the *sui generis* scope since some argue that most investments of MGD producers go into the “creation” of this data[67]. However, as discussed above and further below, without a clear legal clarification uncertainty remains as the distinction between generation and collection is difficult to draw in the context of MGD. Moreover, efforts undertaken by an MGD producer to transform, clean, arrange and aggregate the raw and unstructured volume of data generated by a network of sensors in order to utilise

---

64 See for example the CJEU decisions on C444-02 and C46/02, Fixtures Marketing

65 See case C-444/02, Fixtures Marketing and C-203/02, The British Horseracing Board.

66 Evaluation 2018, p.ii.

67 Evaluation 2018, p.20.

it in the decision making and analysis could be interpreted as investment in "verification" and "presentation" of the pertinent database where the data is stored simultaneously or right after these acts, thus relevant for the protection of the *sui generis*[68]. While the *BHB v Hill* decision provides for a narrow interpretation of the CJEU, its repercussions on the protection of MGD are far from clear and caused legal academics to point out that upcoming issues may be clarified in future judgements (Leistner, 2018). However, this may also lead to significant (and unpredicted) shifts in case law, as interpreted by Derclaye and Husovec (2021) in the most recent *CV v Melons* case.[69] The legal analysis within the 2018 evaluation warned that relying on case law may not be a sufficient solution given the rise of MGD and SGD (sensor-genereted data), and the economic importance of data analytics and the risk that the current exclusion of MGD from the *sui generis* right as provided through earlier decisions may be revised. In such a scenario, clarification on the database maker, owner, and substantial investments would be needed.

In the literature, considerable efforts with different approaches were made to apply this distinction to MGD. While acknowledging legal uncertainty as to the current state of the law, Leistner and others suggested a possible delineation to be implemented by case law or statutory clarification[70]. Relating to the erection of barriers to market entry, Leistner proposes that sensored data could be considered as collected while data produced by a machine would be regarded as generated.[71] The delineation as to observed or measured data would be specifically drawn by considering whether the data could also be collected independently by a third party or only by the data producer. On one side, data sensored from the environment would be regarded as collected, since there would be multiple sources available to obtain the data. On the other side, data internally produced by the machine itself or the internal (real-time) operation of a product or service would be regarded as generated. This would roughly correspond to the distinction between data sensored by the machine itself and those sensored from the environment. Hence, the application of SGD to MGD could be linked or combined with a definition of MGD reflecting the distinction put forward by Leistner.

While in theory this distinction between "sensor-data" and machine-generated data to delineate between obtaining and creating data seems comprehensible, the achievement of the goal of improving legal certainty through such distinction becomes questionable when taking a closer look. A first approach of using sensors as such to be a criterion for delineation may not be

---

68 This data curation process is further explained in Box: Example of a curation needs. In a recent study about challenges in the AI legal framework (European Commission (2020), "Trends and Developments in Artificial Intelligence: Challenges to the Intellectual Property Rights Framework"), the authors argue that substantial investments in developing and implementing AI techniques should qualify for substantial investment. Given the current vagueness concerning the creation/obtaining distinction, the study leaves it open if MGD with the help of AI techniques could be factored in the equation. The study suggests that indeed the data obtained from sensors (such as weather data) could be classified as 'obtained' but that the data analytics of AI could be interpreted as 'creation'. (European Commission (2020), "Trends and Developments in Artificial Intelligence: Challenges to the Intellectual Property Rights Framework", pp.93f).

69 See Derclaye and Husovec (2021): Access to information and competition concerns enter the sui generis right's infringement test – The CJEU redefines the database right.

70 Leistner (2020), Wiebe, A (2017): Schutz von Maschinendaten durch das sui-generis Schutzrecht für Datenbanken, GRUR, 338.

71 See Leistner (2018), footnote 10 where he points out to "real-time operational data" a machine 'produces' (i.e., in technical language this is called status data).

viable since sensors are used for both scenarios – the gathering and creation of data – which shows that in the data economy the distinction between data gathered or generated by machines and sensors would be artificial. Therefore, such distinction alone seems not to be sufficient and without a legal clarification the sensor "obtained" data may still incidentally be considered as covered by the *sui generis* right. A second approach could look at whether the source of the sensors is under exclusive control, as with sensoring of machine performance, or whether the sensoring can be done by third parties. In practice, however, new problems would also emerge with this second approach. One problem would be the role played by possible restrictions on access to the sensored environment. For example, the collection of data about the soil of farmland sensored by a tractor could be seen as a "collection" of data in a legal sense since anybody could theoretically gather the data. On the other hand, it could be also regarded as a generation, if only the farmer as landowner has exclusive control over the gathering of data. A similar case could be made as to the gathering of information in a production line by sensors. The example shows that the distinction proposed above may not guarantee a clear delineation between generation and collection of MGD in the end, which is only proving the close interlinkage between "sensor-data" and other types MGD. Moreover, this distinction could be easily circumvented by the database maker claiming activities of verification or maintenance on the data that would trigger inclusion of the pertinent investments in the scope of protection.

Sattler (2020)[72] stipulated additional arguments for the difficult case of distinguishing generation and collection in the context of MGD. If data that is machine generated in a production line are first categorised and then automatically and methodologically arranged by the production robots, a distinction between generation and collection would hardly be feasible according to Sattler. The process described above would include and make practically inseparable the data generated by sensors about the machine itself and the data resulting from aggregations and processing of this data with data collected by sensors about the environment. The processes performed on raw data after their generation may be inextricably linked with the creation process with the consequence that investment in the different phases and processes could not be identified separately.[73] Even if the functions of measuring and curating[74]/aggregating the data could be clearly and economically separated, e.g. by offering different price levels or an additional service contract for the latter by the machine producer, it would depend on the contractual arrangements whether the investment for curating or aggregating would constitute investments into the database of the machine or robot user or constitute data usage for a separate service by the producer without relevance to the database of the user.

Moreover, with increasing distributed data networks it becomes more and more difficult to allocate relevant investment to certain parties, e.g., separate investments of producer and operator of machines collecting data. This not only has repercussions on the verification of sufficient investment to establish protection but also on the question of whom is to be considered the database maker. Most respondents to the survey conducted for this study reported that they cooperate with other companies in establishing and running databases. If different parties contribute to the relevant investment the concept of co-ownership would have to be invoked that is well-known from copyright but not explicitly included in the database right.

---

72 Sattler (2020) Rechtshandbuch Industrie 4.0 und Internet of Things, Sassenberg/Faber (Hrsg.)

73 See example p. 110, Final Report, Database Evaluation 2018.

74 The process of data curation is further explained below in the section.

Further clarification would be needed as to what threshold of investment would trigger co-ownership and how the internal and external relations of the co-owners should be designed. Analogies to copyright are problematic since the database right is based on investment and not creativity as is copyright.

These examples show that the status of MGD as to the application and scope of SGD is in urgent need of clarification. Whether MGD should be included in (possibly with specific obligations) or excluded from the Database Directive, needs clarification of a full range of other factors such as a definition of what MGD is.

The above discussion highlights the view of several legal academic authors worried that the *sui generis* right may raise serious information and transaction cost problems due to the legal uncertainty of its potential to protect volunteered or observed data. [75] This particularly concerns the use scenario in which firms need access to complete, aggregated data sets to access the primary market or certain entirely new, complementary or aftermarkets.

2.2.2.5 Missing coverage of data economy aspects

The Evaluation 2018 broadly concluded that the Directive does not take into account the data economy and is in that respect outdated. Indeed, the Directive does not take into account the economic value of data – thus of the content – and does not distinguish between various database-related tasks such as data creation, collection, arrangement, update, maintenance, verification, management, etc. It equally does not provide for the situation where data aggregation and combination of databases offers new business models for makers, or new methods for analysis, information and decision making for users.

Another important factor concerns “ownership” of the database. In the context of sensor and MGD, where data is increasingly generated in network structures, it becomes particularly difficult to determine whom to consider as a database maker. Multiple individuals or entities (from sensor and device makers to the operator of the device) can claim having taken the initiative and risk of investing, thus “owning” jointly the database and benefitting jointly from the *sui generis* right. The Database Directive does however not define and include the terms “ownership” or “joint ownership” for the part related to the *sui generis* right. In fact, the discussion in the literature about “joint ownership” reflects on the lack of such a provision in the Database Directive[76] adding thus another layer of uncertainty.[77]The uncertainty on who is the right holder may create hold-up problems.

2.2.2.6 The shift of producers’ investment costs to collection and curation of data

The Evaluation 2018 pointed out the judicial limit on the scope of “substantial investment”, and that it prevents organisations benefitting from database rights where their investment targets the creation of content of the database. By limiting “substantial investment” to cover the actual data storage and processing systems and not the creation of the collected content, the CJEU

---

[75] See Evaluation 2018, section 5 and Leistner (2020), See Cremer, J. de Montjoye Y-A, Schweitzer, H. (2019): Competition policy for the digital age. European Commission.

[76] See Hugenholtz (2016)

[77] See Noto La Diega (2019)

made a clear decision that the purpose of the Database Directive is to promote storage systems and not the underlying data.

The mentioning of a shift in investments towards the "curation" of data focuses the attention on an area in the data value chain, which goes beyond the creation of data. Data curation encompasses activities such as content creation, selection, classification, transformation, validation, and preservation.[78] Investment in this type of activities could be interpreted as investment in "verification" and "presentation" under the Database Directive.

In fact, following Freitas and Curry (2016), data curation is emerging as a key data management activity. The activities depend on the data variety, which is in particular common in big data environments. While data management activities have dealt with data quality and data heterogeneity for a long time, variety requires a lot of resources to gain insights out of the data. The authors also expect that data curation will evolve from a niche activity to "become more present within the average data management environment".[79]

When it comes to MGD, data curation seems to be a key cost-driving factor. In fact, in an IoT environment, different types of devices send different data structures, which prevents efficient networking.

Box 2: Example of a curation needs

By way of illustration, a company uses various liquids in production. The corresponding containers detect the fill levels via the installed sensors and send them to a production planning system. However, the sensors collect different units as one filling quantity is reported in litres, the other in millilitres, and a third one is reported as the relative fill quantity in a percentage. A comparison of the fill levels is thus not straightforward and requires the introduction of a data interpretation logic in the respective production planning system.

The logic puts all the collected data into a uniform form. In this example, it converts the filling quantity into millilitres and calculates the absolute quantity from the percentage information - and thus enables the use of the data within the application. However, this logic must be kept in every application that receives and processes the data from the networked devices. The effort quickly becomes unmanageable when the data structures become more complex and the number of devices increases. If the data structure of a single device changes, all instances of an application must be updated. In theory, this is feasible, but in practice it usually proves to be too inflexible and expensive. Therefore, comprehensive updates are costly, complex and sometimes highly error-prone.

Source: https://www.industry-of-things.de

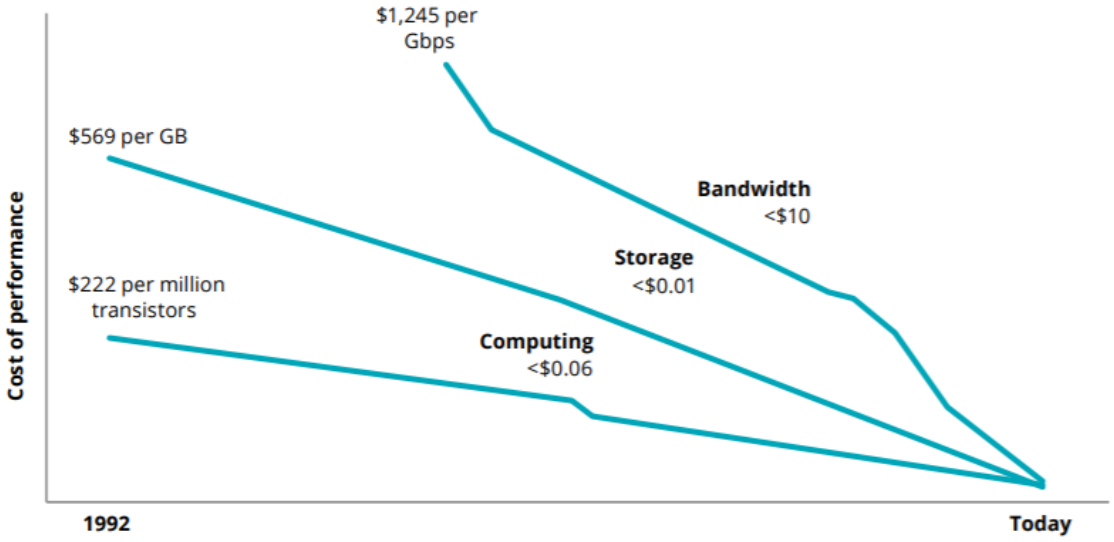
The increase in the number of devices and the increase in the *variety* of data requires companies to transform the data into a useable structure before they can use it. Quick solutions where the data from a smart device is readable by a particular app only leads to fragmentation. If companies want to bring together data from different devices, these data need to be

[78] See Freitas A., Curry E. (2016): Big Data Curation. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy

[79] See Freitas A., Curry E. (2016)

interpreted and harmonised. In an IoT environment, the increase in the number of devices may increase the complexity of the data structure and the effort to handle it. A uniform data structure of the devices through manufacturer firmware updates is thus an option and in case of over the air (OTA) updates, also a less costly one. For companies with available devices worldwide, this is by far too costly and resource-intensive and thus not a viable option. Flexible solutions distinguish between the device and the user software. Adjustments happen in the middleware. Depending on the architecture of the IoT solution, a very large amount of data needs to be sent. Less costly alternatives make use of devices for data transfer between a local network and the cloud as this requires less computing power and storage capacity. For further sustainable data models, data curation is essential. Companies need to think ahead whether to design an isolated solution or a more forward-looking data model.[80] In this respect, data curation starts with the design and implementation of a data model that allows MGD not only to be generated and transmitted by a sensor, but also to be processed and structured already in preparation for data usage. This was also confirmed by a large MGD database maker interviewed for the study which stressed the relevance of the investments undertaken in the design and set up phase of the system ahead of the data generation/collection. In a MGD environment, it is the investment in the data model that provides the value added. However, as pointed out in the open public consultation of the previous Evaluation 2018 – selection and structuring of data are becoming increasingly automated and thus the investment part in these activities decreases. In general, the technological progress in recent years has shown a general decline in the cost (per performance) of activities related with the IoT data value chain such as bandwidth, storage and computing power as shown in Figure 5.

Figure 5: Declining costs in bandwidth, storage, and computing



Source: The rise of the digital supply network, Deloitte University Press[81]

[80] See: Heger, S (2021): Einheitliche Datenstruktur: Ohne nahtlose Zusammenarbeit kein IoT

[81] See https://www2.deloitte.com/global/en/insights/focus/internet-of-things/technical-primer.html

2.2.3 The effects of the problem

The problem that the Directive may not be fit for the data economy primarily due to the legal lack of clarity in relation to MGD, implies that it may not help to effectively reach and maintain a state of optimal level of access to data and databases, i.e. where the benefits of data are fully explored and the incentives of those contributing to the creation, collection and organizing data are also preserved. In particular, it may cause underutilisation of data due to the extra layer of protection that it provides.[82]

The consequences of this key problem can be best understood by recalling the following main economic characteristics of data:

- Non-rivalry: someone's use of the data does not constrain someone else to use the same data (unless access to data is technologically or administratively blocked)[83]
- Economies of scope: combining data from multiple sources may create more value than the sum of the values created by using data from different sources in isolation[84]
- Co-creation of data: data is created through the contribution of multiple parties and efficient side-contracting between any of these two parties is not always possible[85]

The first two characteristics suggest that stimulating access to data and databases could provide more value and business opportunities than what would be the case for traditional products or services. The fact that the Database Directive may not be fit for the data economy primarily due to its unclear relation to MDG, because its uncertainty may create excessive protection for database makers, would lead to economic harm for consumers and society. The third characteristic, which also lies at the basis of no ownership being assigned to data, further exacerbates the problem.

Access to data (and databases), or the lack of it, is especially problematic in cases when a firm has exclusive control over certain data for which there are no close substitutes – the case of (*de facto*) data monopolies. If the firm does not grant access to the data (in the absence of close substitutes) it has control over:[86]

- There may be little or no competition in the market of products and services to which that data is an input.

---

82 See Duch-Brown et al. (2017).

83 See Lambrecht, A., & Tucker, C. E. (2015). Can big data protect a firm from competition? Available at SSRN 2705530. and Jones, C. I., & Tonetti, C. (2020). "Nonrivalry and the Economics of Data". American Economic Review, 110(9), 2819-58.

84 See Duch-Brown et al. (2017).

85 For example, a piece of data denoting the creation of a 'like' on Facebook in the Facebook mobile app is created through the interaction of the handset manufacturer, the mobile operating system supplier, the internet service provider, Facebook and the user. It is almost prohibitively difficult to have contracts that would efficiently assign the benefits realised from the creation from this 'like' among the five (or possibly more) stakeholders involved in its creation.

86 See Abrahamson, Z. (2014). "Essential data". Yale LJ, 124, 867.

- There will be too little innovation – in any market – that would use that data as an input.[87]
- There would be no extra value materializing by combining the data monopolist's data with data from other sources, thus exploiting economics of scope.

We discuss each of these situations below.

Data and databases are typically of no use on their own. Instead, they are used as input for various products or services. These products or services may be supplied by other firms than the one generating/collecting and controlling the data. If both the data controller firm and some other firms can provide competing services, the data controller firm (a data monopolist) may not have incentives to share the data it controls, leading to weakening of competition in the market of the final product or services. If these adverse incentives are reinforced by protection rights provided by the Database Directive, the adverse effects on competition are reinforced.[88]

Aftermarket services relying on data (and databases) controlled by firms in the primary market offer a good example. For example, providers of repair and maintenance services for aircrafts (or any other machine) rely on sensor data placed on the aircraft (that is controlled by the manufacturer of the aircraft) to provide high quality services.[89] If the manufacturer of the aircraft also provides repair and maintenance services, it will have less incentives to grant easy access to the sensor data collected from the operation of the aircraft and this will lead to the weakening of competition (higher prices, lower quality) in the repair and maintenance aftermarket. If it can refer to protection rights conferred by the Database Directive, the competition problem in the aftermarket becomes stronger.

Furthermore, lack of access to data or databases, also supported by some protection rights offered by the Database Directive, may hamper innovation if data is an input into the innovation process. [90] In many instances, the analysis of key data may structure the innovation process by ruling out non-promising avenues of research[91] or, analysis of key data may make testing of innovative ideas more efficient. In such instances, lack of access to data would result in suboptimal level of innovation.

Finally, as indicated by economic theory, data that is aggregated from multiple sources may give rise to economies of scope and be more valuable than the sum of values derived individually from the multiple single sets of data. It may therefore be the case that monopolistic

---

87 See Stucke, M. E. (2018). "Should we be concerned about data-opolies?".

88 See Duch-Brown et al. (2017) and Martens et al. (2020), "Business-to-Business data sharing: An economic and legal analysis."

89 See Rengasamy, D., Morvan, H. P., & Figueredo, G. P. (2018, November). "Deep learning approaches to aircraft maintenance, repair and overhaul: a review". In 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 150-156). IEEE.

90 See Duch-Brown et al. (2017) and Stucke, M. E. (2018). "Should we be concerned about data-opolies?".

91 For example, in the pharma industry, access to various types of data may help pharma companies to invest in compounds that are expected to have more favourable toxicity and inhibition characteristics and thar would most likely be approved in clinical trials and have low production costs.

data markets which entail exclusive private control over data exhibit a market failure as such (super-additive) aggregation is not possible. And in the case of a market failure with limited access for re-use and economies of scope of data, the monopolistic data market situations can lead to lower welfare benefits.[92]

While the Directive may not provide the best incentive system to achieve the optimal level of generation, organization and sharing of data, the issue of ownership rights linked to data further exacerbates the problem. One key issue is when the party ending up with controlling some data (or database) claims ownership over such data and it cannot be 'forced' to open up that data. This creates even more distortions when data is generated as a by-product of some other economic activity.[93]

In such cases, the marginal cost of 'producing' the data – often in an automated process, using hardware and software – is very low, at least compared to the various cost elements related to the primary economic activity, of which data is a 'by-product'. [94,95] Clearly, this will have an impact on the incentive to generate, collect and organize such data. In the context of various protection rights, the discussion revolves around the so-called 'by-product theory' which states that a legal protection, such as 'ownership' or, indirectly via the databases containing data, the *sui generis* right, may not be needed. [96] An interesting situation arises when the marginal cost of a by-product data is low, but the economic benefits of the value they offer is high, for example, it may soften competition in a primary market (for some durable goods with extensive secondary or aftermarkets) and lead to high rents for data holders.[97]

Data collected by agricultural machines provide a good illustration of many of the points discussed above. Structural characteristics on the data driven agriculture markets suggest that a market failure could be evident. With the current technological developments, the agriculture business has been able to adopt sophisticated technologies that have helped realize productivity gains through precision farming. Manufacturers of the agriculture machines design these machines such that they retain exclusive access to the data. This allows the manufacturers to foreclose downstream markets that can potentially realise additional productivity gains from data. There are also lock-in effects on the market due to the absence of clear mechanisms that force these companies to transfer data between providers if farmers

---

92 See Duch-Brown et al. (2017) and Acquisti, A. (2010). "The economics of personal data and the economics of privacy".

93 One of the first papers discussing this in a competition policy context is Cremer et al. (2019), "Competition policy for the digital era", https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf.

94 One example could be data on geothermic conditions in oil rigging explorations, see Gal, M S and D L Rubinfeld (2019), "Data standardization", NYU Law Review

95 For the economic implications of data being generated as a by-product of some economic activity, see Jones, C I and C Tonetti (2020), "Nonrivalry and the Economics of Data", American Economic Review, vol. 110(9), pp.2819-2858, Farbodi M and L Veldkamp (2021), "A Growth Model of the Data Economy", NBER WP 28427, de Corniere A and G Taylor, "Data and Competition: A General Framework with Applications to Mergers, Market Structure and Privacy Policy", TSE Working Paper, n. 20-1076 and Carrière-Swallow Y and V Haksar (2019), "The Economics and Implications of Data An Integrated Perspective", IMF WP No 19/16

96 See Kerber (2016).

97 Aircraft manufacturing provides a good example.

want to switch companies. As a consequence of possible foreclosure and lock-in effects, economic theory suggests that prices may increase, which can lead to farmers not being able to reap additional potential welfare gains. [98]

2.3 Why should the EU act?

Free flow of data in the EU is an important driver of future economic growth and societal progress. It enables creation of data-driven products and services. These are often developed based on data from different Member States and later commercialised across the EU.

A coherent legal framework applicable across the EU is a precondition for a flourishing internal market. Fragmentation from adopting national rules will lead to increased transactional costs, lack of transparency and legal certainty. Each of these consequences will limit the ability for data to flow freely between Member States. As an example, we can look at industrial data value chains which naturally operate cross-border. Fragmentation creates difficulties in addressing issues arising in these. This could be fairness related to contractual rules on data sharing, access or use at the level of the Member States.

Striking the right balance will allow the EU to benefit from the scale of the internal market. Unlocking this benefit requires policy makers to balance between 1) allowing data from the public sector, businesses and citizens to be accessed and used in the best possible manner and 2) protecting rights in relation to investments made into their collection.

Therefore, it is necessary and proportionate for the EU to act now to prevent future divergences between legislative frameworks that may otherwise be adopted for MGD at the national level and hamper the internal market. As such, legal fragmentation across Member States should be prevented by the creation of a harmonized approach to the protection of MGD in the EU internal market.

[98] See Atik & Martens (2020), "Competition Problems and Governance of Non-personal Agricultural Machine Data: Comparing Voluntary Initiatives in the US and EU"

3 Policy objectives and policy options

The following chapter includes a description of the policy objectives and policy options as well as a legal assessment of policy options not discarded at the early stage.

3.1 Policy objectives

According to the Better Regulation "Toolbox", we frame the general and specific objectives of the policies linked with the problems and causes which are summarised in the objective tree shown in Figure 6 below.

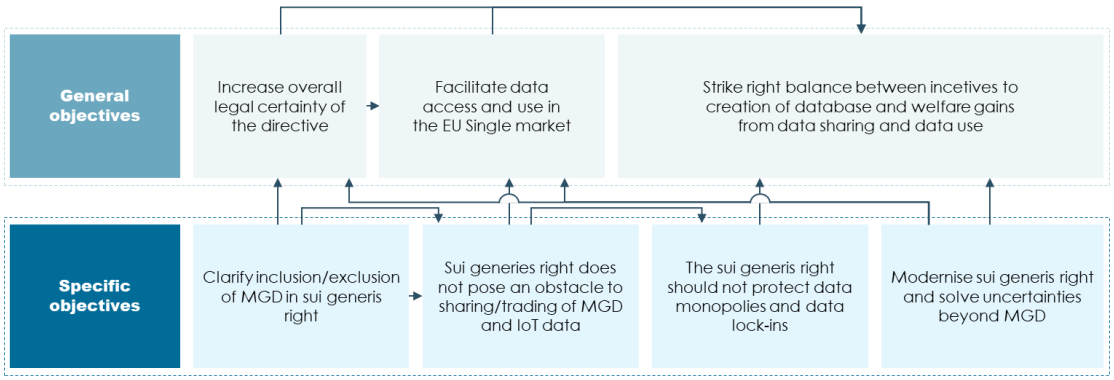
3.1.1 General objectives

The general objective of the review of the Database Directive is, in the context of the Data Act, to increase legal certainty and facilitate data access and use in the EU Single market. This must be achieved by striking the right balance between the protection needed to incentivise the creation of database and the societal benefits derived from unleashing the full potential of data sharing and data usage. This is considered essential for stimulating innovation and growth in a data-driven society.

3.1.2 Specific objectives

The main specific objective is to bring clarify on the status of MGD databases in relation to the Database Directive with a specific focus on the *sui generis* right and to ensure that the Database Directive does not contribute to data lock-ins, in particular in the context of sharing of and trading in MGD and data generated in IoT environment. The *sui generis* right should not facilitate the creation and subsistence of data monopolies and data lock-ins that may curb competition and innovation in the EU market. Another specific objective of the study is to address the other identified legal uncertainties linked with the Directive that go beyond MGD and IoT to ensure legal coherence and stimulate the data economy.

Figure 6: Objectives' tree



3.2 Policy options

In this section we present the list of policy options considered at different level of depth in the present study. The options can be divided in three main groups: i) baseline scenario policy option (on which all other options can be compared against) and the complete repeal of the *sui generis* right, ii) policy options specific to the relation between the Database Directive and MGD core of the present study and iii) supplementary options for a wider review of the Database Directive, thus with application to all databases protected by the *sui generis* right, and that can be adopted in combination to the second group of policy options related to MGD.

After an initial scoping phase, some options were either discarded or not explored in the depth while the others were retained and further assessed.

First, we present the list of policy options that have been explored in depth in the Study. Second, we briefly mention the options that have been either discarded or not further explored after the scoping phase. Missing policy numbers in the list of retained policy options below were initially assigned to discarded or not further explored options.

3.2.1 Policy option 0: Baseline scenario

Under this scenario, the Database Directive will remain as it currently is, with no action at EU level.

3.2.2 Policy option 2: Options that exclude MGD from the scope of application of the *sui generis* right

Under this option the scope of the *sui generis* right will be amended in a narrow sense to make sure that it does not apply to MGD/IoT data. This option could be achieved a) directly through an explicit exclusion of MGD or b) indirectly by adapting existing legal concepts such as “substantial investments”, “obtaining”, etc.

This option will be presented in the form of four sub-options depending on whether and how the exclusion of MGD is complemented with other ways to incentivize the creation and access of MGD database.

***Policy option 2a*: Exclusion of MGD from the scope of application of the *sui generis* right**

Under this sub-option, MGD will be excluded from the scope of *sui generis* right without introducing any other modification.

Policy option 2b:* Exclusion of MGD from the scope of application of the *sui generis* right and *substitution of the current infringement with a more flexible infringement test related to economic detriment

Under this sub-option, MGD would be excluded from the scope of *sui generis* right and a more flexible test would be created drawing from the recent *CV v Melon* case[99] on the other databases under the scope of *sui generis* protection. The new infringement test will also have the effect of overcoming the unpractical distinction between insubstantial and substantial taking. The test could include a weighing of interests within a fairness or economic detriment test in the infringement analysis.

Policy option 2c: Exclusion of MGD from the scope of application of the sui generis right and introduction instead of an alternative control mechanism applied only to databases containing MGD

Under this sub-option, instead of the *sui generis* right, an alternative protection mechanism will be introduced to provide some control over MGD database for the database maker. However,

[99] C762/19 CV-Online Latvia v Melons

the protection would be limited and targeted against some unauthorised unfair competitive/abusive uses by third parties.

3.2.3 Policy option 3: Options that include MGD in the scope of application of the *sui generis* database right

Under this option the scope of the *sui generis* right will be amended to make sure that it does apply to MGD/IoT data. This option could be achieved a) directly through an explicit inclusion of MGD or b) indirectly by adapting existing legal concepts such as “substantial investments”, “obtaining”, etc.

This option was originally considered in the form of two sub-options depending on whether and how the inclusion of MGD is complemented with other ways to incentivize the access to MGD database. However, only the policy option 3a (only inclusion of MGD) was retained.

Policy option 3a: Inclusion of MGD in the scope of application of the sui generis database right

Under this sub-option 3a, MGD will be included in the scope of *sui generis* right without introducing any other modification.

Under this sub-option a specific access regime will be introduced to allow users access (in certain circumstances) to MGD databases.

3.2.4 Supplementary policy options for a potential wider review of the Database Directive

This set of options are treated separately as their scope is different to the main options that deal with machine generated data and the database right. These, on the other hand, would apply to all databases in the scope of the *sui generis* right. The supplementary options can be adopted in combination to any of the options related to MGD. The possible supplementary options were defined jointly with the Commission upon a scoping phase.

Supplementary policy option S1: Exceptions to sui generis right would be expanded in line with broader general copyright exceptions.

Under this supplementary option, limitations that will be in the future introduced into copyright will be extended to the database right by default[100]. As to existing copyright limitations, this policy option extends them to the *sui generis* right. This avoids new uncertainties and fragmentation. Many of the copyright exceptions may not prove very useful in the database context (e.g. reporting), but they probably do not harm either (e.g. parody exception as applied to databases). A few selected exceptions (private use, quotation and reporting) will be discussed.

[100] This has been already done, for example, with the new Text and Data Mining (“TDM”) exception introduced in the Directive on Copyright in the Digital Single Market (“DSM Directive”) of 17 April 2019 (Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC)

Supplementary policy option S2: Strengthening of research exception of the database right.

Under this supplementary option, an expanded research exception in Art. 6 and Art. 9 of the Database Directive could be introduced with the following characteristics: a) it could be made mandatory[101], b) it could encompass the re-utilisation and the reproduction of a database as a whole for scientific purposes, c) it could apply also to infrastructure operators of non-commercial scientific databases and d) it could be extended for commercial research (under certain conditions). The new exception should not contradict or change the existing TDM exceptions in the DSM Directive.

Supplementary policy option S3: Public bodies would be excluded from the sui generis right.

Under this supplementary option, the *sui generis* right will make clear that public bodies do not hold *sui generis* rights.

3.2.5 Policy options either discarded or not further explored

As part of the scoping phase, the following policy options where also considered but then discarded or not further explored:

- Policy option 1: Repeal of the *sui generis* right in the Database Directive
- Policy option 2d: Exclusion of MGD from the scope of application of the *sui generis* right coupled with the introduction of a specific access regime
- Policy option 3b: Inclusion of MGD in the scope of application of the *sui generis* database right coupled with the introduction of a specific access regime
- Supplementary policy option S4: Introduce compulsory licencing for sole-source databases
- Supplementary policy option S5: Introduce a user's minimum right/consultation right to databases
- Supplementary policy option S6: Alter or render more flexible the duration of the protection as well as clearer rules on renewal terms
- Supplementary policy option S7: Introduce new exceptions specific to the search engines and web-scraping.

More elaboration on these policies is presented in the Annex.

[101] As it is already the case for example for the TDM exception.

3.3 Elaboration of policy options and legal assessment

Below we elaborate on the legal implementation and assessment of the policy options and sub-options that have been retained after the initial scoping phase and for which we provide an in-depth analysis.

3.3.1 Policy option 0: Baseline

As part of Impact Assessment studies, the baseline scenario should always be considered as a policy option. Under this scenario, the Database Directive would remain as it currently is, with the implications as described in the Evaluation 2018 study. Existing uncertainties would continue to exist, as would possible obstacles the *sui generis* right poses to data sharing.

3.3.2 Policy option 2: Options that exclude MGD from the scope of application of the *sui generis* right

This option would include the amendment of the scope of the *sui generis* right in a narrow sense to make sure that it does not apply to MGD/IoT data. In the aftermath of the *Fixtures* decision of the CJEU in 2004, there was intensive discussion among experts on whether the strong confinement of *sui generis* protection to data collection, as opposed to data generation, would also exclude MGD from protection. According to Leistner "many authors have derived that in typical big data scenarios, the investments of 'producers' of sensor or machine-generated data of all kinds will be excluded from the *sui generis* right because in most practical cases, such investments would have to be regarded as investments in the "creation" of data."[102]Gervais also suggests that "machine produced outputs (such as new data corpora) based on analyses of Big Data are neither "obtained" nor "collected"; they are generated".[103]

However, Leistner challenges this view and suggests caution in reaching this conclusion too quickly. Firstly, as explained in Sections 2.1. and 2.2 and further below, the dichotomy creation/collection put forward by the CJEU is not always straightforward in Big Data scenarios and sensors that can observe and collect (pre-existing) information from the environment (e.g. road traffic, blood pressure, soil nutriments, genomic data). In this regard, Hugenholtz (2017) argues that the distinction between "observed" and "created" data depends on the type of data processed by the machines: the author brings the example of sensor data produced by radar system or observation satellites as "observed" data of which investments should be relevant for the database right compared to purely "created" computer-generated airline schedules[104].

---

[102] Leistner (2018), p. Leistner in his article cites, as example, Herbert Zech (2015), "Industrie 4.0" – Rechtsrahmen für eine Datenwirtschaft im digitalen Binnenmarkt' [2015] GRUR 1151, 1157 and similarly the Commission Staff Working Document on the free flow of data and emerging issues of the European data economy, SWD(2017) 2 final, 20.

[103] Gervais (2019), p.10.; see also J. Drexl, Design competitive markets for industrial data – Between propertisation and access, in JIPITEC, 2017, p.268.

[104] Hugenholtz, Bernt P. 2017. Data Property: Unwelcome Guest in the House of IP. Paper presented at Trading Data in the Digital Economy: Legal Concepts and Tools, Münster, Germany.

However, as expressed by the Commission in its Staff Working Document for the 2018 evaluation of the Directive: " in the context of automated data collection by sensor-equipped, connected Internet of Things objects it becomes increasingly difficult to distinguish between data creation and obtaining of data when there is systematic categorisation of data already by the data-collecting object (e.g. industrial robots)".[105]

Secondly, even if data is created the requirement for sui generis protection can still be fulfilled by substantial investments in "verification" or "presentation" of the content. Verification but especially the presentation of the content may be quite resource-intensive in the IoT context, where, as evidenced in the technical discussion in Section 2.1, the data producer needs to invest resources in pre-processing, cleaning, aggregating and further curating the flow of unstructured information from the sensor network before insights can be derived from it. Falce (2020) agrees with Leistner´s view on the importance of substantial investments in verification and presentation in the Big Data scenario[106]. In another recent study, Noto La Diega discusses the issue in the context of AI-powered databases reaching similar conclusions on the possibility that MGD databases could benefit from sui generis right by "substantial" investments in AI technologies and algorithms necessary to handle and process the large volume of unstructured data in order to make them usable.[107]

[105] European Commission (2018), Evaluation of the Directive 96/9/EC on the legal protection of databases [SWD(2018) 147 final], p.15.

[106] Falce, V., 2020. Uses and abuses of database rights. In Kritika: Essays on Intellectual Property. Edward Elgar Publishing.

[107] G. Noto La Diega, Artificial Intelligence and databases in the age of big machine data, in AIDA 2018, p.119.

With regards to case law, on the one hand, case law has provided some direction to the creation/collection dichotomy – though not specific to MGD – and national case law seems to have followed the CJEU rulings of 2004.[108] On the other hand, the CJEU has not explicitly described what constitutes an investment in "obtaining". While the 2004 Fixture cases can be seen as clear-cut cases of data "created" or "invented", it is still not clear whether recording/measuring existing information counts as "obtaining".[109] For instance, the Court of Appeal of England and Wales held that investments necessary to record live information of football matches (e.g. scores, fouls etc.) can be viewed as an investment in "obtaining" data – thus eligible for database right.[110] The same was held in a case by the Austrian Supreme Court who considered the collection of results of football matches as related to the collection of data.[111] Similar activity, however, was considered as "created" in Germany according to representatives of the German sport industry interviewed in the Evaluation 2018. Another case that can be seen to indirectly adopt the view of considering recorded data as "obtained" data is the *Freistaat Bayern v Verlag Esterbaue GmbH* case over the protection under the sui generis right of topographical maps.[112] While the case is mostly referred to in the context of what makes a database, Burdese (2021) notes that "[i]t is significant that neither the parties nor the judge who referred the question to the CJEU suggested that topographical data might be considered as created rather than obtained".[113]

---

[108] See European Commission (2018), Evaluation of the Directive 96/9/EC on the legal protection of databases [SWD(2018) 147 final], p.15.which brings as examples in France Précom, Ouest France Multimedia v Direct Annonces, Court of Cassation (Cass.), 1st civ., 5 March 2009; in Spain Ryanair v Atrapalo Case STS 572/2012 (9 October 2012); in Germany Zweite Zahnarztmeinung II BGH (2011) GRUR 724. Furthermore, also in the well-known Ryanair case, the database containing content linked to the operations of the company was found not to be protected by the *sui generis* right. As additional examples, other two cases in Germany where the national court closely follow the CJEU 2004 decisions were BGH 1.12.2010 I ZR 196/08, Zweite Zahnarztmeinung II (an online evaluation portal containing evaluation tests by employees) and OLG Hamburg 8.6.2017, 5U54/12 (concerning database of medical classification of diagnoses and procedures). The OLG Hamburg 8.6.2017, 5U54/12 indicates that the investment in the determination of existing elements for compilation in a database is protectable. Investments in the procurement and collection of existing elements, their search, locating, recording, processing and the type of provision that justify the ancillary copyright under Section 87b (1) sentence 1 UrhG. On the other hand, the means used to produce the works or elements found in the database are not to be equated with the investments associated with obtaining the content of this database. As a result, they cannot be taken into account in assessing whether the investments involved in creating this database are significant.

[109] This is a question beyond MGD. See definition of recorded data from E. Derclaye, The legal protection of databases: a comparative analysis, p., 98-99 as: "data [that] occur in nature or in time and are generally recorded by instruments of measure in order for them to be intelligible by man. Examples include results of sport competitions, meteorological, astronomical data and genomic data".

[110] Football Dataco Ltd v Stan James Ltd (No 2) [2013] EWCA Civ 27. European Commission SWD(2018) 147 final, also cite British Sky Broadcasting plc v Digital Satellite Warranty Cover Ltd [2011] EWHC 2662 as another case with similar approach to recording data.

[111] Austrian Supreme Court, Dec. of Mar. 24, 2015, 4 Ob 206/14v.

[112] C-490/14, Freistaat Bayern v. Verlag Esterbauer Gmbh, "Verlag Esterbauer case". In this case the CJEU was asked to clarify the definition of "database" in the case between the Federal State of Bavaria and the Austrian publisher Verlag Esterbauer. The Austrian publisher has allegedly infringed database rights by scanning topographic maps from the Land of Bavaria's database.

[113] Burdese (2021). "AI-Generated Databases. Do the Creation/obtaining Dichotomy and the Substantial Investment Requirement Exclude the Sui Generis Right Provided for under the EU Database Directive?

A case especially relevant to the current discussion around the creation vs collection of data in the context of the MGD database is the German "Autobahnmaut" case[114] where sui generis protection was given to the MGD database of toll data. In that case, the company Toll Collect operated a charge system for lorries for the use of the motorway and built a dynamic database for billing operators based on lorry traffic data collected during the operation of the main activity. The German Federal court considered the investments into the obtaining of the data, i.e. terminals that register the lorry traffic, as relevant for the sui generis protection - thus considering the data about the traffic as pre-existing. Moreover, the Court also considered an investment of the company in software for processing data as relevant for the verification and presentation of the content. The case is also important as it can be seen as a classic situation when the database is a spin-off of the main activity. It could be assumed that the company did not need extra incentives (in the form of sui generis protection) in order to collect the data and maintain the database.[115] Some authors are concerned that if future cases were to follow the approach undertaken in the Autobahnmaut case, this would expand the scope of protection of MGD and challenge the narrow interpretation derived from the CJEU 2004 rulings.

The main objective of this option would be to bring legal clarity. Sub-options will be also examined to possibly substitute the sui generis right with an instrument likely better suited to the specifics of MGD. This could ensure that the sui generis right, as an additional layer of protection of data, does not interfere with the objective of the Commission to incentivise free flow of data, as inter alia expressed in the Inception impact assessment of the Data Act.[116] Moreover, the option would help limiting the risk that protection is granted to investments in "generation" of data rather than the production of databases which is the original objective of the Directive.

This option is presented in the form of three retained sub-options as summarised in the table below.

Table 1: Sub-options for the exclusion of MGD from the scope of the *sui generis* right

POLICY OPTION	DESCRIPTION	NOTE
Sub-option 2a	Exclusion of MGD from the scope of application of the *sui generis* right	Only exclusion of MGD from the scope of *sui generis* right without the introduction of any other modification.
Sub-option 2b	Exclusion of MGD from the scope of application of the *sui generis* right and substitution of the current infringement with a more flexible infringement test related to economic detriment	Exclusion of MGD and introduction of a new infringement test that would include elements of fairness and economic detriment for the databases still under scope of the *sui generis* right (inspired by recent *CV v Melon* case).

Reflections and Proposals" , WIPO Academy, University of Turin and ITC-ILO - Master of Laws in IP - Research Papers Collection - 2019-2020.p.7.

[114] Autobahnmaut, BGH I ZR 47/08 (25 March 2010).

[115] See Drexl (2018), p.72-73 for a more elaborate discussion and economic perspective on the Autobahnmaut case.

[116] Inception impact assessment Data Act (including the review of the Directive 96/9/EC on the legal protection of databases) (2021): Inception impact assessment -Ares(2021)3527151.

		This will likely limit the protection of the *sui generis* right for both MGD databases and other databases.
Sub-option 2c	Exclusion of MGD from the scope of application of the *sui generis* right and introduction instead of an alternative control mechanism applied only to databases containing MGD	Complement exclusion of MGD with another, lighter, control mechanism for MGD databases. This would represent a way of compensating MGD producers for stripping them from protection in order to incentivize the data producer/de facto controller to share its MGD databases (balancing the risks that database producers would turn to contractual or technical restrictions).

Source: Copenhagen Economics

Possible legal approaches to implement the option

Before presenting the different sub-options, the first main issue is the choice of the legal means to achieve this option, i.e. excluding MGD from the scope of the *sui generis* right. This could be achieved using two general approaches: directly through an explicit exclusion of MGD or indirectly by adapting existing legal concepts such as "obtaining" and "substantial investments".

The first general approach could be implemented by modifying the definition of databases protected by the *sui generis* right, i.e. databases as a collection of data and other elements, explicitly excluding MGD (accompanied definition). For example, statutory definition Art. 1(2) Directive could become:

"For the purposes of this Directive, "database" shall mean a collection of independent works, data or other materials, excluding MGD, arranged in a systematic or methodical way and individually accessible by electronic or other means."

The policy could use the definition of similar concepts such as co-generated data or more strictly SGD that may be provided in the context of the Data Act.

However, it would be advisable to avoid the risk that the national legislators do not start to install their own protection mechanisms for MGD that could lead to an unharmonized landscape of different protection measures that would run counter to the objectives of EU policies in this field.[117] To that purpose, one of the options would be to explicitly state this in a separate provision such as: "no protection shall be granted to MGD".[118]

[117] See Derclaye, E & Husovec, M (2021): Sui Generis Database Protection 2.0: Judicial and Legislative Reforms, SSRN Working Paper 18 November 2021, p. 11-13, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3964943.

[118] Another solution would be to rely on regulation as a regulatory instrument to avoid this issue. Several legal experts during the workshop supported the possibility relying on regulation. See the discussion of why a Regulation is the preferable regulatory instrument for EU-wide experimentation regarding the protection of property in Husovec, M (2020): The fundamental right to property and the protection of investment: how difficult is it to repeal new intellectual property rights?, in Geiger, C (ed.) (2020): Research Handbook on Intellectual Property and Investment Law, Research Handbooks in Intellectual Property, Edward Elgar Publishing Ltd, Cheltenham, UK, 385 - 405

An explicit exclusion has the advantage of clarity (assuming MGD can be defined precisely). Building on the definition provided in Section 2.1 and technical discussion in Section 2.1 and Annex, the exclusion may target mainly raw sensors data. There should be explicit mention of operational data generated by the machine. While challenges remain to delineate what is raw data, the definition may include some pre-processing activities that (sometimes) are done directly by the sensor such as data compression, data encoding, or transmission of raw data directly to the cloud structure. Data already structured in data warehouses and ready to be used for deriving insights might not be included in the definition of MGD for exclusion (thus still possibly be protected). Examples such as weather data where physical external phenomena are recorded from multiple external sources and further processed with data/mathematical modelling to predict atmospheric events may fall outside of the definition.[119] There, some significant human intervention and curation still may take place. On the other hand, real-time unstructured streams of data such as in-vehicle data may be excluded if not yet aggregated, processed and arranged in static databases to be used for analysis.[120]

The second general approach, i.e. adapting existing legal concepts in the Database Directive, could be implemented by:

i. directly excluding investments in the generation of data assuming MGD is "generated" and not "collected" or
ii. specifying the requirement of substantial investment and establishing some minimum standard that will be hard to overcome by MGD.

Both alternatives in the second general approach would raise serious issues on implementation.

As to alternative i., it is freighted with all uncertainties that have been discussed in connection with the *Fixtures* cases in 2004, and, for example, the more recent *Autobahnmaut* case[121] in Germany and which have been elaborated more specifically above and in Sections 2.1. and 2.2. The *Fixtures* cases excluded investments in the generation of data as opposed to collection. One could argue that most investment in the context of MGD goes into the generation rather than the collection of data – especially for operational data specific to the functioning of the machine/device. In reality, the separation between the two phases appears difficult in most cases.

The survey results give us a mixed view. On one side, if we consider only IT experts respondents' answers to the question: "if MGD is collected by your company, who invests and takes the economic risks of setting up and maintaining the database?" the majority (n=19) reported "my company", thus not allowing a clear distinction of the two phases. On the other

---

119 "The data on the current state of the atmosphere come from a network of ground measuring stations that measure wind speed, temperature, air pressure and humidity, as well as precipitation levels. Data from radiosondes, weather satellites, commercial aircraft and weather ships are also used." See European Commission (2020), "Trends and Developments in Artificial Intelligence: Challenges to the Intellectual Property Rights Framework". Another discussion, however, may arise to determine whether databases generated by AI and machine learning algorithms may count as new "invented"/"generated" data or as processed data. The investment undertook with the AI and machine learning should be towards the database rather than towards the data itself for further ("infinite") re-use down the data value chian.

120 See Drexl (2018) p.76 for a discussion on dynamic databases and sui generis protection.

121 Federal Supreme Court (Bundesgerichtshof) of 25 March 2010, Case I ZR 47/08, Autobahnmaut

side, however, if we look at all respondents of the survey, the findings suggest that the distinction between generation and collection of information would be easier to handle in practice. In this case, the survey shows that in most cases the person generating the data is different from the one running the database. Among all respondents (n=51) only 18 responded "my company" to the question above.

This could suggest that, despite the difficulties in drawing a clear delineation between generation and collection of data elaborated on above, in some cases, the generation of data and the establishment and operation of the database could be distinguished in practice, and respective investments be allocated. More specifically it suggests that the different forms of processing and value-adding on raw data could be separated to some degree.

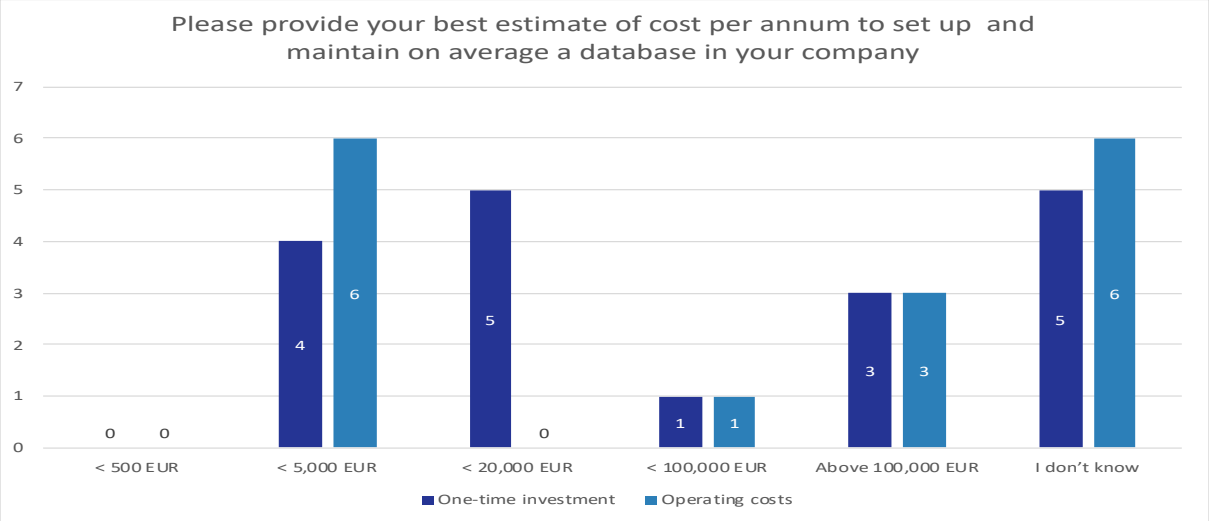
In a second step, the pertinent investments could be allocated either to the generation or the collection of data for the database. Only the investments in the latter would be relevant for protection. Those investments would be done by the company that is running the database. The example of the connected car below supports this notion. However, the basis of our survey is rather narrow and in practice many situations will remain where the delineation will be hard to draw.

Moreover, this alternative could resort to qualitative aspects of data processing and limit relevant investments on certain value-added forms of processing. This could include annotating the data, reformatting, curation of "historic" data (as opposed to real-time data streams), cleansing the data if done in direct connection with inserting the data into the database. While this would create new problems of delineation it could raise the threshold in a meaningful way to exclude the generation of MGD.

As to alternative ii**.,** case law has not managed to stipulate a clear threshold for substantial investment in 26 years. It will be difficult to find appropriate criteria for that. Responses from the survey conducted in the present study suggest that the amount of investments put into databases is quite broad[122]: the lowest investment costs were less than €5,000 and the higher investment cost exceeded €100,000 (this is, of course, highly dependent on the size of the respondent).

[122] See Figure in Annex : "Average cost per annum of set up and maintaining a database".

Figure 7: Average cost per annum of set up and maintaining a database



Source: Survey for the Database Directive Review

This may work against establishing an absolute quantitative minimum amount for substantial investments in favor of relating substantiality to the type of data and the size of the database (this approach will be explained below). Moreover, it would be difficult for a plaintiff to prove substantial investment in MGD separately in cases of mixed databases or for a defendant to reject that notion.

A possible way to introduce a regime that creates a threshold for databases with MGD, is the introduction of a substantiality requirement of the average investment per data element. Currently, *sui generis* right requires a substantial investment in the full contents of a database. This criterion tends to favor larger databases, since they are more likely to command a substantial investment than smaller ones. This issue is even more prominent as currently the "thresholds" of substantial investments are set at low levels by national courts.[123] This appears to also favor databases with MGD since the efficiency and economies of scale of a machine enable the construction of large database.[124] This works against the aim of stimulating the

[123] See Evaluation Study 2018;. Both CJEU and national courts set rather low levels in order not to undermine the Database objectives. (C-46/02 Oy Veikkaus, Opinion of AG Stix-Hackl, para 49; C-338/02 Svenska Spel, Opinion of AG Stix-Hackl, para 39; C-444/02 OPAP, Opinion of AG Stix-Hackl, para 55).

[124] Network systems of sensors are capable of generating enormous amount of data. The Large Hadron Collider (LHC) at CERN is able to generate 1 Petabyte per second of collision data (see https://home.cern/news/news/computing/cern-data-centre-passes-200-petabyte-milestone) , while an autonomous car can generate from 5 to 32 Terabyte per day (see https://blocksandfiles.com/2020/01/17/connected-car-data-storage-estimates-vary-widely/). However, as noted in the technical discussion in Section 2, it might not be so straightforward to argue that databases including MGD are larger than others. There may be more data points (content), but typically the IoT databases which provide time series use NoSQL databases, and those are more flexible and unstructured (they do not have a typical table structure). Relational databases, on the other hand, such as Web of Science, or LinkedIn (more than 700 million registered users) , Facebook (2.9 billion users) are likely very large and possibly larger as many databases containing sensors. The examples above, however, would still likely be considered automatically generated databases without human intervention (although with human-related information). Additionally, going back to the example of vehicle data, it should be considered that a large portion of vehicle-

sharing of MGD. This raises the question of how the criterion could be adapted to better reflect an allocation of protection that stimulates sharing.

A typical example would be a producer of an IoT device or another connected device such as a car. Sensors in the device capture data and the data are sent to a cloud server managed by the manufacturer. Since the IoT device is more often than not mass produced, the manufacturer receives data not from one, but from many, potentially millions of devices. The collection of the data from so many sources requires a large storage capacity and the sheer volume allows the manufacturer to claim substantial investment in the collection of the data. The cloud server is separated from the devices so that it is clear that the manufacturer collects existing data. The manufacturer thus isolates himself from any uncertainties about the legal relevance of investment in the data capture of the devices. Even if data capture in a device is seen as data creation (and thus not contributing to substantiality of investment required for *sui generis* protection) and even if the data capture is seen as an investment by the user of the device, the manufacturer's *sui generis* protection would not be affected, because the manufacturer does not need to rely on any investment in data capture by the device. They rely on the investment in the subsequent collection (and perhaps verification and presentation) of existing data received from the devices, completely in line with the *Fixtures* decisions of the CJEU in 2004. Therefore, size and/or volume is thus reason to be concerned about *sui generis* rights in MGD databases.

A relatively simple adaptation of the criterion would be to require that the average investment per data element is substantial. That would put large databases with MGD at a disadvantage. The average investment tends to be low, since machines are efficient (at least more efficient than humans) and the large size of the database would give rise to benefits from economies of scale. Smaller databases with much human input would be favored by the adapted criterion.

This criterion does not exclude all large databases from *sui generis* protection. For example, a large database with legal information (case law, journal articles, etc.) that requires human editing of each new element with a view to uniform presentation within the database would likely still be protected, since the per element investment would still be substantial. However, databases automatically filled with electronic sensor data would be created so efficiently that per element the investment would not be substantial anymore. By having 'per element substantial investment' as the criterion, it is irrelevant that the sum of insubstantial per element investments would amount to a substantial investment in the entire database (as is relevant under the current criterion).

The criterion may be operationalised by requiring the database holder to show what the overhead costs are in the production of the database and over how many elements these costs can be spread, next to costs that are specifically made to collect, verify, or present an individual element.

This would be an adaptation that holds across the board for all databases. Therewith the *sui generis* right would be kept relatively simple. No separate regime would be introduced within the already specialized regime of *sui generis* protection.

An additional advantage of this implementation alternative in the second general approach is that it would include only a minimal set of new concepts that might raise questions or lead to

generated data are primarily of a technical nature. They exist only temporarily and are used locally within vehicle systems and are never stored.

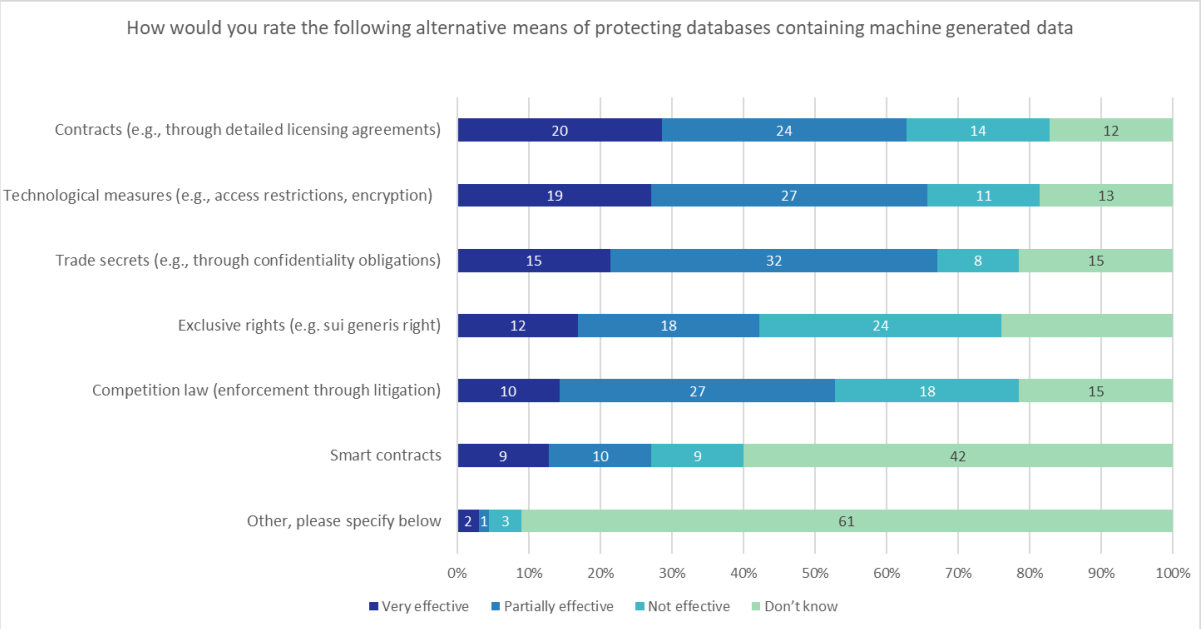
disputes about their application. The only new element is the introduction of the concept of average investment, but that is relatively straight forward. One may need to elaborate on which version of a database to count. It might stand to reason to take the version for which protection is sought, e.g. the state of the database at the date(s) of infringement. This will probably raise less questions than the introduction of a separate regime for databases with MGD.

A drawback is that the current criterion of substantial investment has never become completely clear, even after 26 years. This may have a knock-on effect on the adapted criterion (average substantial investment). Average investment is the investment in the entire database divided by the number of elements. This may be partially overcome if it is clarified in the recitals of the Directive that this change is meant to exclude large databases filled with the extensive deployment of automated means. Such will give practitioners a useful indication of how the criterion (average and substantial) should be applied.

Relation with other protection instruments

After having discussed possible approaches to exclude MGD databases from the scope of the *sui generis* right, it must be noted that both the protection of data and the propensity to data sharing depend on the mix of instruments for protection: exclusive *sui generis* right, unfair competition, contract and technology. Results from the survey conducted in the present study show that trade secrets and Technical Protection Measures ("TPMs") are the most effective alternatives in view of respondents.[125]

Figure 8: Rating of alternative means of protecting databases containing MGD



Source: Survey for the Database Directive Review

[125] See Annex.

The instruments have different properties, catering to different protection needs. Nonetheless, and to a certain extent, they may function as substitutes for each other. The policy option changes the mix legally and possibly factually/in an economic sense. From a factual/economic perspective, the question is whether database possessors will miss the specific protection properties of the exclusive right and whether they will seek (and find) compensation in the other instruments. For data users, the question is whether the absence of *sui generis* rights will make data use appreciably easier. If database possessors consider switching to other protection instruments this will affect the ease of data use as well. This effect – sitting at the end of a longer causal chain – may however be difficult to assess.

Below we discuss the different sub-options related to the exclusion of MGD.

3.3.2.1 Policy 2a: Exclusion of MGD from the scope of application of the sui generis right.

Option 2a would be limited to excluding MGD from the application of the database right by means described above without adding additional protective instruments.

This option could address several challenges related to MGD and the Database Directive.

First, one of the main advantages of this option is to provide clarity and avoid any opportunistic claims about the possible protection of MGD by the *sui generis*. In other words, it will avoid cases where data holders that share their data under the provision of service contracts may claim in an unspecified way *sui generis* right/IP rights on the data they pass on as an extra standard safeguard clause. The validity of this claim is dubious and left for litigation which may increase transaction costs[126]. While the implementation challenges addressed above apply also here, the section focuses on i) possible issues related to the case of mixed databases and ii) the relationship to other existing means of protection.

Second, as discussed above when assessing policy option 2 in general, investments undertaken in the context of MGD are seen by some as relating to the "creation" of new data. This is especially true for raw data collected in an unstructured way via a network of sensors integrated into connected machines and devices.[127] Most of these sensors run in the background of the operation of the machines. The large amount of data continuously harvested may or may not be used to produce insights[128]. Depending on the situation and use applications, data can be adjusted, combined, structured and analysed differently. Therefore, the investment in the data infrastructure is done with a view on the data as such rather than the database that is used to present the content. This contradicts the main original objective of the Database Directive which was to stimulate the database industry and protect investments in databases rather than the content of the database. As also reaffirmed in the reasoning of the CJEU in 2004: "[t]he purpose of the protection by the *sui generis* right provided for by the directive is to promote the establishment of storage and processing systems for

---

126 See for example current study by DG JUST on fairness in data transaction contracts.

127 See Gervais (2019), p.10.; J. Drexl, Design competitive markets for industrial data – Between propertisation and access, in JIPITEC, 2017, p.268; Farkas, T.J., 2017. Data created by the Internet of Things: the new gold without ownership. Rev. Prop. Inmaterial, 23, p.5.

128 Companies are often unaware of all the data (and possible uses) being generated as a by-product of the main economic activity. See Gimpel, G. (2020). Bringing dark data into the light: Illuminating existing IoT data lost within your organization. Business Horizons, 63(4), 519-530.

existing information and not the creation of materials capable of being collected subsequently in a database".[129] Finally, the European Commission in its communications have also stressed the original objective of the Directive, with regards to the lack of database right protection of non-processed MGD noting that: "the Database Directive did not intend to create a new right in the data. The CJEU thus held that neither the copyright protection provided for by the Directive nor the *sui generis* right aim at protecting the content of databases. Furthermore, the ECJ has specified that the investment in the creation of data should not be taken into account when deciding whether a database can receive protection under the *sui generis* right."[130]

Third, partially linked to the argument above, some view MGD databases typically as spin-off databases or in general by-product/incidental to the main activity of the database maker. The *Autobahnmaut* case discussed above can be seen as a classic example of a spin-off database as the traffic data were collected in the main activity of charging lorries for the use of the motorway. In the IoT context in general, an important part of data, e.g. operational data, are collected/generated while the sensor-equipped machines are operated for e.g. farming or automated line production.[131] According to the "spin-off" doctrine originated in the Dutch jurisprudence, spin-off databases (e.g. event schedules, television or radio programmes, train or plane timetables, telephone subscriber data), should not be given *sui generis* protection regardless of whether there was any substantial investment.[132] An important argument to support this theory was that when databases are created quasi-automatically from a different core activity "there is no need for legal protection as an incentive to produce such databases, as they are produced anyway".[133] The objective to stimulate the production of databases was the very reason why the Database Directive was introduced. A second important, although separate, argument linked with the spin-off theory is that spin-off databases are often single source and may thus be monopolized. Therefore, database protection may risk exacerbating the situation and increase barriers to entry in the primary and secondary market of the sole-source database maker. The classic example is operational data related to the performance of the specific connected device for which there is no alternative source available.

It must be noted, however, that the spin-off theory has not been formally adopted by the CJEU in its rulings. The CJEU focuses on the distinction between generation and obtaining. On the one hand, if the primary activity of the firm generates/creates the data contained in the spin-

---

[129] Judgment of the Court of 9 November 2004, ecli:eu:c:2004:695, para 31. - British Horseracing Board Ltd.

[130] European Commission (2017a): Staff Working Document on the free flow of data and emerging issues of the European data economy, SWD(2017) 2 final.

[131] The online context is also full examples concerning information generated by customers and/or users as they interact and transact on the website (e.g. the data on market prices of eBay).

[132] See Derclaye, E., 2008. The legal protection of databases: a comparative analysis. Edward Elgar Publishing.p.94 for a general discussion; Hugenholtz, P.B., 2005. Abuse of Database Right: Sole-source information banks under the EU Database Directive. Antitrust, Patent and Copyright.

[133] Beunen, A.C., (2007). Protection for databases: The European database directive and its effects in the Netherlands, France and the United Kingdom. Leiden University, p.113; Duch-Brown, N. (2017): The economics of ownership, access and trade in digital data, JRC. In the context of our study, 17 out of 22 IT experts that responded to the survey stated that "Generating and using [MGD] data for internal use only (e.g., optimisation of processes)" was very important/somewhat important. In those cases, where MGD is collected/generated to optimize internal processes and performance, it can be assumed that the organization does not need extra incentives such as legal protection in order to collect MGD data.

off databases, the investment in the primary activity would not be relevant for the protection both under the CJEU interpretation and the spin-off doctrine. On the other hand, as discussed above in the general policy option 2 assessment, spin-off databases might still be protected if substantial investments were proven for the "verification" and "presentation" of the spin-off database[134].

Finally, further care should be taken when applying the spin-off doctrine in the IoT and data-economy context since the core activity of a firm is an evolving notion. New, data-centric business models are developing and gaining relevance with the consequence that applications and development of data services that were a by-product or spin-off of the main activity might become more and more central to the competitive strategy of the firm in the future.[135]

Mixed databases

In this scenario, a problem may arise from a practical side, especially for "mixed" databases including MGD and other types of data. In these cases, separate investment in the collection, verification and presentation of the database with respect to MGD has to be verified to be able to discount it from the overall investment into the database. This also includes the designation of data as MGD in the specific case. Possible legal uncertainties as to the delineation of MGD may add to the problem. This issue may also continue in the case of contracts about the use of data and databases as MGD would be free to use whereas contractual restrictions could only be based on the *sui generis* right as to those data that are not MGD. Therefore, it might be needed to prove in certain situations that some data are not MGD but within the scope of the database right. Finally, it should be stressed that the extent to which mixed databases as a challenge would occur will greatly depend on the definition of MGD that the EU legislator will opt for.

The same problem may arise if data is used by third parties without permission or license. In the case of mixed databases, the alleged infringer could claim that the data she or he used would consist of MGD although stemming from a protected database. The right-holder in such a case may need to prove that the data allegedly misappropriated would not constitute MGD.

---

[134] Leistner (2018, footnote.10) notes in cases: "where a machine 'produces', stores and transmits real-time operational data which is vital to the very functioning of the machine. In such cases, indeed, it would not be far-fetched to argue that such data are 'created' by the very operation of the machine if and to the extent that the operation cannot be separated from the measuring, storing and transmitting of the data and if such data are not available by any other means than the very operation of the machine. Such situations, indeed, are situated on the thin red line between the BHB v Hill doctrine and the spin-off doctrine (which latter the ECJ essentially rejected). Hence, in such cases legal uncertainty prevails, until the ECJ further clarifies the scope of the sui generis right in a future judgment". See also: Beunen (2007), p. 125; Derclaye, (2008), p. 94 and Ducato, R., 2013. "Adiós sui generis": A Study of the legal Feasibility of the Sui Generis Right in the Context of Research Biobanks on further arguments on possibility for spin-off databases to be protected via investments in verification and presentation of content.

[135] Tombal (2020) provides the example of Monsanto, historically considered as an agriculture and bioengineering business, that is now "heading towards becoming an information broker with the acquisition of Climate Corp". Another example is the automotive sector where car makers "might attempt to argue that in the near future, with the advent of autonomous cars, they will strive towards becoming "mobility data companies" rather than simple "car builders". See Tombal, T., (2020). Economic dependence and data access. IIC-International Review of Intellectual Property and Competition Law, 51(1), pp.70-98. See also Beunen, A.C., (2007) for additional criticalities in seprating primary and secondary activity and for the use of internal databases.

A possible solution may be to include a provision that a database containing mostly MGD or where MGD is inextricably linked with other data should be excluded from protection. An analogy can be drawn with the 2019 Commission guidance on the Regulation on a framework for the free flow of non-personal data in the European Union. Regarding mixed datasets combining personal and non-personal data. In the guidance, the Commission states: "if the non-personal data part and the personal data parts are "inextricably linked", the data protection rights and obligations stemming from the General Data Protection Regulation fully apply to the whole mixed dataset, also when personal data represent only a small part of the dataset". By analogy, if MGD and other data are inextricably linked in a database, the database could be fully excluded from protection. In addition, a rebuttable presumption could be adopted according to which mixed databases are excluded from protection, unless the database holder can show that the database mostly consists of non-MGD. This would also provide incentives for database makers to separate data and clarify the status of the data to still benefit from protection for the non-MGD part. As a result of these incentives, the impact of this policy option on non-MGD may be limited.

Of course, the size of the problem depends on how common the case of "mixed" databases would be in reality. Generally, also based on the technical discussion presented in the background chapter and Annex, it can be assumed that the closer to the raw data collected by the sensors, the less likely the database would be mixed as the first step in the data value chain usually entails pre-processing (standard data cleaning etc.) and fusion with other sensors data gathered in the IoT network.

Evidence from the in-depth interviews conducted in the present study, on the other hand, indicate that mixed databases can be frequent.

A European association of independent automotive aftermarket distributors stated that the type of data (MGD vs non-MGD data) depends on where the data is accessed: in-vehicle access to data provides access to pure MGD, while off-board access might aggregate both types of data. Another association of authorized automotive dealers and repairers reported that data are getting increasingly mixed: "If you look at Repair Maintenance Information (RMI) data it's just generated data, and if you look at in-vehicle data, it is machine-generated data. As such it's mixed. At the end of the day, we pull it together, and above it you've got artificial intelligence data produced as well, that can be seen as machine-generated data based upon human-generated data… it's getting more and more mixed, and it comes from the need to have more predictive maintenance which requires both kind of data. Based on these, and through AI and machine learning, you can predict the next maintenance for instance". Similar common existence of mixed databases and challenges related to separation MGD and non-MGD were reported by a large industrial technology database maker and user. Finally, a sport data database maker indicated that differentiation between machine and human in the dataset would be difficult in practice.

By excluding mixed databases from protection, however, incentives are created for industry to invest in technologies that can categorize and track data after collection. Even though some databases may eventually be qualified as mixed once the data is further processed, separating MGD and non-MGD for as long as possible benefits those who wish to rely on protection for non-MGD. Depending on the costs and practicalities of classifying data and separating databases, the entry into force of the proposed regulatory approach could have the effect that problems associated with mixed databases will become less pronounced in the longer term.

Relation with other protection instruments

As to the means of protection in the mix of instruments no considerable change should be expected. The lack of exclusive rights on data will avoid several problems associated with data as a special kind of subject matter. The problems are: risk of too broad protection, difficulties in definition and specification of the protected subject matter, and other information and transaction cost problems.[136] Based on the assumption that no additional incentives are needed for the production of data, data holders will keep relying on contracts as well as trade secret law and technical measures.

The main question concerning the mix of instruments is situated in the premise that producers of databases with MGD need protection, while protection in the form of a *sui generis* right is excessive. The added value of the *sui generis* right is that it can be invoked against anybody, also persons with whom the database producer does not have a contractual relationship. If the *sui generis* right were to be removed, the database maker would fall back on trade secrecy, unfair trade practices and general tort law (and to some extent contract networks) to act against third parties. Trade secrets could be a type of protection with an effect coming close to an IP right. However, it depends on keeping the information secret. In a networked data economy this becomes increasingly difficult (even though MGD will often be sole-sourced). Secrecy will mostly rely on technical protection of the data and especially the use of encryption technology will be increasingly applied in this field.[137] While it could retain secrecy to a certain degree the increased need for exchange, use and value adding of data will work against maintaining secrecy.

These are instruments that are currently already used in addition to the *sui generis* right. These instruments all have their limitations compared to the *sui generis* right. The question is whether enough protection remains for companies to engage in data sharing and to fend off competition.[138]

Contract law seems to be a viable alternative to legal control[139], and with standard contracts and contract networks, it might be possible to achieve (to a certain degree) a legally balanced situation concerning data governance. Additional back-up measures might be needed in the form of best practices and model contracts. Moreover, a related measure might include amending the unfair contract clauses Directive with Business to Business (B2B) application. We will not elaborate and assess further these additional measures as the introduction of standard mandatory contractual clauses and a possible model contract is one of the possible focuses of the Data Act.

---

136 See Leistner, M. (2020)

137 Two examples of a database maker applying encryption to data have also been reported in the Evaluation 2018, p.112 in the case of car industry and wind turbines.

138 We note that with more data sharing the risk of data ending up in the hands of parties with whom no contractual relation exists becomes bigger. But since this is a new situation, it is hard to assess now whether the remaining instruments will prove adequate.

139 Another alternative to ensure control over the data is represented by TPMs. Nonetheless TPMs could be seen at best partially as legal control - the circumvention of TPM may play a role in trade secret protection, criminal offense of hacking.

3.3.2.2 Policy 2b: Exclusion of MGD from the scope of application of the sui generis database right and introduction of a more flexible infringement test related to economic detriment

Sub-option 2b extends 2a by introducing a new infringement test on the database right under the protection of the Database Directive after the exclusion of MGD. This can be achieved by codifying the key statements of the *CV v Melons* case in the Directive whilst providing further clarifications on the test. It is important to note though that the interpretation regarding the *sui generis* right in this ruling presents such a new line of thought that it needs to be considered as a significant source of supplementary sources of the EU law. It means that it already forms part of acquis communautaire and, in practice, the new interpretation will be observed by the national courts of the Member States in similar cases limiting the need to codify it in the secondary law.

Advocate General Szpunar proposed in a recent Opinion in the *CV v Melons* case to interpret the Database Directive as meaning that the database maker has a right to prevent the extraction or the reutilisation of the whole or a substantial part of the contents of that database only on condition that such extraction or reutilisation "adversely affects its investment in obtaining, verifying or presenting those contents".[140] Following this Opinion, the CJEU argued in its June 2021 judgment in *CV v Melons* case that the interpretation of the scope of protection offered by the Database Directive requires a fair balance to be struck between "on the one hand, the legitimate interest of the makers of databases in being able to redeem their substantial investment and, on the other hand, that of users and competitors of those makers in having access to the information contained in those databases and the possibility of creating innovative products based on that information".[141] In the view of the CJEU: "the main criterion for balancing the legitimate interests at stake must be the potential risk to the substantial investment of the maker of the database concerned, namely the risk that that investment may not be redeemed".[142]

This would rather complement the scope of protection with an unfair trade practice approach resorting to principles of slavish imitation. The holding may be open to different interpretations. Under the wording of the Directive, the re-utilisation of substantial parts would be enough to constitute infringement. The decision would add a weighing of interest and a criterion of detriment to the investment into the infringement analysis.[143]

Beyond the case, however, this element is worth consideration if implemented by statutory amendment. This would combine the flexibility of an unfair trade practices approach with the legal security of statutorily defined IP right and still fit the objective of protecting the investment.

Consequently, the approach in the *CV v Melons* Judgment, which proposes an economic risk or detriment test, could be implemented by amendment of Art. 7 of the Database Directive. In

---

[140] Opinion in Case C-762/19, ECLI:EU:C:2021:22, par. 59

[141] Case C-762/19, ECLI:EU:C:2021:434, par. 41

[142] Case C-762/19 ECLI:EU:C:2021:434, par. 44

[143] During the workshop, the experts also pointed out that applying this type of test in the infringement analysis may reduce the need for the exceptions' regime (which in some cases can create complexities and difficulties in the interpretation of the Directive) as legitimate and public interests would be considered in the balancing act.

this option, MGD are now excluded from the *sui generis* right. Therefore, MGD databases would not be subject to the infringement test. We note that legal experts participating in the workshop, appeared to favour the application of the *CV v Melons* approach to all databases protected by the Database Directive.

The option introduces a flexible test that includes a weighing of interests within a fairness or economic detriment test which clears the way for the development of case groups in case law. There is still some uncertainty, according to the legal experts present in the workshop, on how to apply the Latvian case decision and the exact criteria to use in the test. Therefore, the option should consider how much freedom to give to national courts in operationalizing the test. Such guidance may come in the form of elements that may or may not be considered. Hereinafter four elements are discussed: economic detriment, purpose of use, substantiality of taking and duration of the terms of protection.

Economic Detriment

The CJEU describes the main criterion for balancing the legitimate interests at stake as the potential risk to the substantial investment of the maker of the database concerned, namely the risk that that investment may not be redeemed. This policy option would still require a substantial investment. All difficulties in assessing the substantiality of investment would still be present. The experts in the workshop pointed to the difficulties to claim (the sufficient level of) investments in this type of setting where the process is very opaque and no registry of investments in databases is contemplated. A question that would need an answer is whether the detriment test should be allowed to revive the spin-off theory which asks whether there is a relevant, separate investment in the creation of a database. Given that it is hard to allocate costs to a certain purpose, a negative answer would be preferable.

The main part of the detriment element is harm, i.e., the risk that investment may not be redeemed. In the workshop, legal experts expressed the need to specify whether this would apply only to competitors, and how those would be defined considering whether active in downstream markets or not, or also other users. A preliminary question would be whether there is one answer that fits all cases. In some industries, protection covering a windfall – for example a market for data that was never anticipated – may be needed to redeem investment. In other industries, it may be sufficient that the *sui generis* right – as readjusted by the detriment test – provides some shielding from competition, but there is no need to construe non-competing uses of a database as detriment. Resort may be taken to the combination of “high/low risk” and “high/low gain”.[144] For instance, in cases where the risk to the investment of the database maker is low detriment should not weigh in heavily.

The purpose of the taking

The CJEU describes this element as relating to the interests of users and competitors of the database maker in having access to the information contained in those databases and the possibility of creating innovative products based on that information. During the workshop, the legal experts suggested emphasizing the follow-on innovations aspect as mentioned by the Court. A follow-on innovation of significant economic importance would certainly weigh in heavily as a legitimate purpose but should not be construed as a necessary condition. Legislation could provide some guidelines favouring value-added uses as well as uses in the

[144] See Derclaye and Husovec (2021)

public interest. Follow-on innovation would weigh in heavily where it is dependent on a sole source database.

Substantiality of the taking

Theoretically, the substantiality of the taking could be considered as an element that could be included in the balancing. However, one may wonder whether this element does not hamper follow-on innovation too much, since innovation relies increasingly on large data volumes. Application of the current requirement of extraction/reutilization of a substantial part of the database has proven difficult to apply. Hence, it is advisable to exclude this element from the weighing of interests. By doing so, this policy option can reduce the complexity of the Database Directive and remove concepts from it, the interpretation of which may give rise to disputes.

3.3.2.3 Policy 2c: Exclusion of MGD from the scope of application of the sui generis right and introduction instead of an alternative control mechanism applied only to databases containing MGD

Under this sub-option, instead of the *sui generis* right, an alternative protection mechanism would be introduced in order to provide some control over MGD databases for the database maker. However, the protection would be limited and targeted. This option is carefully analysed below. However, there are certain counterarguments to this policy option that are also summarised in the analysis.

The idea for this option is to provide possible assurance to MGD database makers to trade/provide access to their database in a way that nevertheless does not amount to a fully-fledged property right that is generally taken to limit data sharing. Therefore, a control mechanism similar to a "defensive right" or a limited protection right would be given to MGD database makers that protects against some unauthorised unfair competitive/abusive uses by third parties. This defensive right might be included in the Database Directive as a reformed *sui generis* right that only applies to MGD and may require some proactive steps to be undertaken by the MGD database maker in order trigger the right.

In the following section we present and assess in detail a possible defensive right inspired by trade secret protection. Further we present and briefly assess an alternative defensive right inspired by an unfair competition practice instrument recently introduced in Japan.

Defensive right specific to MGD

One way to create a defensive right is to draw inspiration from how unlawful uses are defined in the context of trade secret protection, such as the definitions of unlawful and lawful acquisition, use and disclosure in the EU Trade Secret Directive. The idea is that some unauthorised third party uses that do not harm the legitimate interest of the MGD database maker would not be caught by the *limited* protection offered by the defensive right but be lawful use. Instead of subjecting MGD to exclusive rights, the approach would be to vest contractual restrictions with a limited third-party effect as does the protection of trade secrets. The right could declare the obtaining, use and transmission of data from a specific database unlawful if the person knows or should have known at the time of obtaining, use or transmission, that the person from which it obtained the data was wrongfully using or forwarding the data. This concept was included in a draft of Art. 34 the ALI-ELI Principles for a Data Economy[145] (and

[145] See ALI-ELI (2021): Principles for a data economy: Data rights and transactions

included in a Report of the German Commission on Data Ethics in 2019[146]). Wrongfulness of use as the key notion would mostly draw on contractual prohibitions but could also include other illegal activities like hacking. Beyond those, other types of permission could be included as was done with the permission of reverse engineering as well as reference to free speech in the Trade Secret Directive. This limited protection would apply to any database without an additional threshold of qualification, but protection would be limited to takings that misappropriate the data from the database in a "wrongful" way that allows a weighing of interest and further limitations.

This concept would keep the focus on contractual practice as the current standard and preferred method of protection in the industry while giving some limited protection extending over the contract relationship to cases of unlawful use by third parties, but in a limited form consistent with competition principles. While currently, industry seems to cope well with a network of contractual restrictions or the standard terms applied to data platforms, situations may arise in the future where data is transmitted to persons not bound by the restrictions who could use the data for their competitive advantage without proper remuneration. [147] To avoid this, a protection drafted similar to the scope of trade secret protection could provide effective safeguard without the need to establish a new IP right that would be overly broad and detrimental to innovation and the need for stimulating flow of data and data sharing. The proposed option would limit protection in degree and scope to what is necessary to protect data against misappropriation without unduly restricting competition.

The proposal goes beyond the protection of contracts against third parties as it is now established in general tort law or unfair trade practices law in Europe. On the other hand, deviating from trade secret law a de minimis rule could be introduced to limit the scope of protection. The ALI-ELI Principles for Data Economy introduced such a rule to avoid overly deterring transactions due to the cumulation of relevant restrictions. This rule could be framed as excluding protection where the wrongfulness was not material under the circumstances and could not reasonably be expected to cause material harm to the party protected. Generally speaking, a case-by-case assessment needs to be made to decide whether or not a violation or breach on the part of the supplier is material. In doing so, it is suggested to take into account the significance of the duty breached for the legitimate interests of the protected party and whether the supplier was acting purposely, recklessly, negligently or innocently. As far as contractual restrictions are concerned, analogies could be drawn to doctrines relating to material breach of contract. Unauthorized access would usually be considered to be material, but there may be exceptions, for example, where security measures taken are very weak and it would not be justified to assume downstream third-party effects. In assessing potential harm to the protected party, an objective standard is suggested. Another limitation could be introduced in case the recipients' interests clearly outweigh the interests of a protected party.

This option would be flexible enough to i) allow for exclusion from the protection sole source databases seen as *de facto* data monopolies and ii) integrate into the analysis of the legitimate interests of the parties an economic harm test and/or fairness notion. On this weighing of interests, examples could be included in the statutory provision or in the recitals to give guidance for the courts in a specific case. Both aspects could be inspired by the framework discussed in sub-option 2b.

---

[146] See Datenethikkommission (2019): Gutachten der Datenethikkommission, p. 144

[147] See Leistner, M. (2020), p.248.

The option would apply to all MGD databases. Databases under contractual restrictions would be a major field of application as the prevalent industry practice right now. Contractual or technical measures used to limit access could also be interpreted as kind of proactive steps to trigger the right. By giving a limited third-party effect to contractual restrictions, it supports the contractual freedom and gives more assurance to the data holder. The data holder will be more ready to share the data with this additional protection. On the other hand, the protection would not be too strong to impede the sharing of data. It would create a more flexible set of instruments than an IP-based exclusive right.

The Report of the German Data Ethics Commission in 2019 provided a similar example for a defensive right below the level of trade secrets.[148]

One possible concern of applying the defensive right to databases under contractual or technical locks is that it actually deters innovators from using data whose provenance is not crystal clear and therewith, curbing data use for innovative purposes. An alternative, lighter, criterion to mark databases subject to the defensive could cover MGD databases, the content of which, the producer does not (or refuse to) make available under open-source conditions. In this case, the database would not be confidential but there would still be a clear intention on the part of the producer to protect his/her investment.[149]

Recent contributions in the literature have addressed a similar defensive right also inspired by the legal regime established by the Trade Secret Directive initially proposed in the context of the Commission's Communication on "Building a European Data Economy" of 10 January 2017.[150] Also in that case, the right would not reach the status of full exclusivity right but still allow the right-holder to sue other parties in case data was misappropriated. An important distinction is that the defensive right originally proposed would protect raw non-personal MGD instead of databases as in the present case. Studies such as Drexl (2018)[151], Kim (2018)[152] and Banterle (2019)[153] raised concerns over this proposal pointing out that the right-holder would be able to benefit from this right by having mere *de facto* possession over the raw MGD without even the need to fulfil any threshold for protection such as reasonable security measures and secrecy in the context of the Trade Secret Directive. This may tilt the balance of interests too much in favour of the data holder and would disproportionally discourage re-use of MGD. Drexl (2018), indeed noted that, since under the defensive right the right holder would not have to show the information kept secret had economic value, unlike for trade

148 See Datenethikkommission (2019): Gutachten der Datenethikkommission

149 Tying it to an open access condition would come close to a compulsory license but without remuneration.

150 European Commission (2017a): Staff Working Document on the free flow of data and emerging issues of the European data economy, SWD(2017) 2 final, 20

151 See Drexl, J. (2018): Data access and control in the era of connected devices. Report for BEUC. Drexl also use the recent Autobahnmaut case to analyze the different application of the three protection regimes, i.e. sui generis right, trade secret and the defensive right proposed by the Commission in 2017, concluding that the first two would give excessive protection to the data holder.

152 See, Kim, D. (2018): No one's ownership as the status quo and a possible way forward: A note on the public consultation on Building a European Data Economy

153 Banterle, F. (2020): Data ownership in the data economy: a European dilemma. In EU Internet Law in the digital era (pp. 199-225)

secrets, there would be no balancing of interests' exercise between commercial interest of the data holder and freedom of information present under the Trade Secret Directive. Another point of discussion was the difficulties in assigning the right over raw data in the context of co-generated data where multiple actors contribute.

The proposed defensive right presented here may address to some extent the concerns put forward by the authors. First, as mentioned before, to benefit from the defensive right, the database maker may still need to use contracts for the use and access of its database to third parties or undertake certain efforts to put in place sufficient TPMs. This would mean that the right would not be assigned by mere *de facto* possession of the data. Second, the implementation of defensive rights may integrate a fairness test to be carried out in case of infringements so as to guarantee a weighting of interests' exercise. Third, while the approach proposed by the Commission in 2017 was focusing on raw MGD, the current defensive right would be in the Database Directive framework, thus having MGD databases as subject matter which may make a difference and alleviate the problem of identifying the right holder.

Finally, we note that some legal experts expressed some doubts during the workshop on the idea of introducing, within a *sui generis* right regime, another kind of *sui generis* right regime specific for MGD. Indeed, it has to be clearly acknowledged and underlined that such a specific additional regime may seriously risk creating additional complexity and uncertainty. As a result, it may undermine the objective of the Database Directive review and may contradict the goals of the Commission policy as stipulated in the planned Data Act. If the policy is implemented, the experts suggest beginning with a trial phase, putting an expiration date on the legislation and mandating the legislators to prolong it if proven effective.[154] This would make the intervention and its implementation very complicated but could potentially avoid that the instrument will survive by inertia even if not achieving its desired objectives. The question is, however, whether a defensive right with a sunset clause will produce any discernible effect since companies may be reluctant to react to this policy option if it is uncertain whether it will survive. This question and related uncertainties raise doubts concerning the usefulness of this option.

Defensive right inspired by the Unfair Competition Prevention Act in Japan.

Initially, another approach drawing from unfair trade practices concepts was explored. Japan recently introduced in its Unfair Competition Prevention Act the concept of "shared data with limited access", defined as (i) any technical or business information that is accumulated in a reasonable amount by electronic or magnetic means (ii) provided to specified persons on a regular basis and (iii) managed by electronic or magnetic means. The aim is to protect data gaining value through accumulation that is intended to be widely shared with specific parties under particular conditions. Unauthorized acquisition, use, or disclosure of such data falls under unfair competition and is subject to civil law remedies. A similar design could be adapted for the new alternative measure of protection. This approach is resembling more an IP approach as certain data are defined as a subject matter that will be protected against certain acts of usage, acquisition, and disclosure.

---

[154] See as example of sunset clause implemented in Europe the Council Directive 1999/85/EC of 22 October 1999 which amended Directive 77/388/EEC to allow a reduced level of VAT to be applied to labour-intensive services on an experimental basis. Recitals (2) and (3) of the Directive state that the objectives of the experiment are to increase employment and reduce the black economy. The experiment was strictly limited in time and concerned only the services described in the new Annex K added to Directive 77/388/EEC.

However, this approach appears to be too broad and establishes a new property right on data itself. While the infringing acts could be defined in parallel to the Trade Secret Directive, new problems may arise by limiting the subject matter in a way different from trade secrets. The Japanese approach is directed at aggregated data that are regularly provided to specified persons. While this definition is not very clear, the limitation to aggregated data could exclude part of the MGD and add another layer of complexity. Hence, the Japanese approach would have to be adapted which could be done by focusing protection on MGD within our definition. However, such an approach would virtually protect all MGD which would provide a much broader protection of the data as such than if MGD were included in the *sui generis* right. Moreover, it is hardly conceivable which notion should substitute the requirement of secrecy in such a scheme. The "Japanese" requirement of "being provided to specified persons on a regular basis" seems to be unclear and not suited.

Hence, this approach is not really convincing unless some qualification would be added which could be found in the aggregation of MGD providing for some level of value. However, this qualification would be close to the existing substantial investment requirement and would not raise the level of clarity.

Finally, the concerns of legal experts interviewed for the study presented above in the context of the defensive right inspired by trade secret protection hold also in this case.

3.3.3 Policy option 3: Options that include MGD in the scope of application of the *sui generis* right

This option would include the amendment of the scope of the *sui generis* right in a broader sense, ensuring that it applies to MGD/IoT data. This option would bring about clarity as to the legal status of machine-generated data. Some respondents to the OPC consider MGD (including data generated by sensor-equipped objects connected to the Internet-of-things objects) already covered by the *sui generis* right[155]. The opinion is shared widely by representatives of publishers.

In the case of inclusion of MGD in *the sui generis* protection without any additional access right, the use of MGD by third parties would often infringe on the database right as users would often extract or re-use the whole database containing MGD. [156] The serious concerns on data access due to MGD being included in the scope of the *sui generis* right and thus covered by an exclusive right become even more relevant in instances where (equivalent) data is not easily available from an alternative source for the user. This is common with operational data generated by sensors in a device.

The problem might become more visible and prevalent when the database right is more acknowledged in practice and might become an impediment to the data economy if applied more frequently. This effect might be detrimental to users, other database makers and for the general interest in competition and innovation. By contrast, there is certain possibility that MGD database makers might have more incentives to make their databases public as the exclusive right would protect them against misuse of their data by third parties.

---

[155] This is also in line with the academic debate presented in the assessment of policy 2 which implied the possibility for MGD to be covered by the sui generis right.

[156] See Leistner, M. (2020)

Similar to what was discussed in policy option 2a, there could be multiple approaches to implement this option. One approach would still be to rely on a definition of MGD – in this case to state in the Directive that MGD (as described in the definition) are included in the scope of protection. This would amount to an indication that the same rules apply to databases with MGD and databases containing other data.

A different approach would be to overcome the limitation of the concept of "obtaining", by including in the requirement of substantial investment also investments related to the generation of data. This would avoid the elaboration of a workable and stable statutory definition of MGD databases, which, as previously discussed, may be difficult to achieve[157]. The lack of a specific definition of MGD would also bypass the issue described in policy 2 related to mixed databases containing both MGD and non-MGD data, which, according to interviews conducted for the present study, appear common.

However, a potential source of uncertainty may be present also with this approach related to where to draw the line of investments to be considered in the generation of data. A large MGD database maker-user interviewed for the present study stressed that the relevant investments now go increasingly into the phase before the generation of MGD.[158] Therefore, the question may arise whether the investments and costs to develop and manufacture the device that collects the data should be included to assess substantiality. Likewise, as at present the generation and collection of MGD could be seen as a by-product of the main economic activity of a firm, e.g. producing connected vehicles, the spin-off problem reappears in a different context.

A possibility would be to limit the investment related to generation of data to the equipment of sensors that serve to collect the data.

This solution would also mitigate the risk that the inclusion of investments in data generation may have repercussions on other fields beyond MGD such as football fixture lists and horse racing where data may become protected again – thus going against the CJEU 2004 judgments. Another option to possibly limit the risk of repercussions for other types of data would be to limit the inclusion of generation of data to MGD and clarify that in this case only investments in sensors could be included.

Another possible problem related to policy option 3a would be to define co-ownership in legislation in a way suitable to the specific situation with MGD. This problem was introduced in Section 2.2.2.4 and elaborated by various authors.[159] One example of co-created data that could give rise to joint-ownerships is represented by the data-holder equipment manufacturer (e.g. producer of smart tractor for smart farming) and the user of the equipment (the farmer) which could also collect the data with a view to the database. Generally, both Drexl and Leistner assume that in this case the data-holder equipment manufacturer would be the one

---

[157] Although, the current review of the Database Directive is undertaken under the Data Act initiative, which itself may provide a definition of similar concept as MGD such as co-generated data.

[158] These investments may include the efforts made in optimally designing the device in order to best equip it with sensors as well as the sensors, the efforts required to set up the infrastructure able to handle the flow of data, the design of the various interfaces as well as the algorithms directly linked with the raw data

[159] See for example Drexl (2018) and Leistner (2018).

taking the initiative to set up the infrastructure to collect and store the data in a database[160] and thus would be the one entitled to the *sui generis* right. Other examples of a similar type could be seen in the consumer IoT services.[161] There the user of a smart speaker produced by company A may activate an application for streaming music offered by company B using an integrated voice assistant developed by company C. The operating system of the device, as well as the IoT application and the voice assistant may be interested in claiming/using the data generated during the operation of the speaker related to the performance of the application, the device, the assistant and the behaviour of the user.

More closely related to the production of databases is the example of data collected in distributed databases in the industry where different companies contribute data to the same database.[162] In general, besides establishing a minimum threshold of investment for qualification as co-owner and determining internal and external relations of the co-owners, problems may also arise with respect to the application of the database right, as in networked IoT environments there may be a large group of or even unidentified co-owners that would require alternative ways of rights management.

Finally, this option could be theoretically extended in the same fashion as sub option 2b, by introducing an economic detriment test based on the *CV v Melons* ruling that in this case, would apply also to MGD databases.

3.3.4 Possible supplementary policy options for wider review of the Database Directive

This set of options should be understood as applying to all databases in the scope of the *sui generis* right with no limitation to MGD. The existence of the *sui generis* right may cause uncertainties for data users in various ways that are not captured by the preceding policy options, such as exceptions that are not aligned with general copyright law, limited research exceptions, and exercise of *sui generis* rights held by public bodies. These issues are addressed here and may be relevant in combination with any of the policy options described above.

3.3.4.1 Supplementary option S1: Exceptions to sui generis right would be expanded in line with broader general copyright exceptions

The set of exceptions to the *sui generis* right is much smaller than the set of exceptions to copyright law. For example, Art. 5 Directive on Copyright in the Information Society[163] ("InfoSoc") contains a large set of exceptions applicable to copyright, many of which do not have an equivalent in *sui generis* law. Since both copyright and *sui generis* right may be

[160] The manufacturer has also higher incentives and returns to scale since it can collects and aggregate all data from all devices it sells (from which it can extract better insights) while the user would get only individual data from its own device(s).

[161] See European Commission (2021): Preliminary report – sector inquiry into consumer internet of things, SWD (2021) 144

[162] In our survey most respondents stated that they cooperate with other companies in establishing and running databases.

[163] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society

applicable to the same database (be it to different aspects of it), the lack of exceptions to *sui generis* right might unduly restrict the use of a database. It may be logical to harmonise these instances of exceptions, especially where an exception that has corresponding equivalent in copyright law is missing in the *sui generis* right. This option may reduce uncertainty and stimulate the use of databases, since users will not need to assess whether their intended use is covered by differently formulated and diverging exceptions under copyright and *sui generis* right.

In the first place, limitations that will be introduced into copyright in the future should be extended to the database right by default. For example, this has been done with the new Text and Data Mining (TDM) exceptions in Art. 3 and Art. 4 of Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market (the DSM Directive).As to existing limitations in general copyright law, in the interest of clarity a preferred solution would be to adapt any limitations to database copyright in Art. 6 as well as that in Art. 9 of the Database Directive. This could hold for any limitations to be introduced in the future. Clarity would be provided by this option as no special rules for database protection have to be reviewed and a unitary system of specific exemptions could be established all over the protection for material in copyright and neighbouring rights.

However, the risk to all stakeholders is that this would neglect the special situation of databases that deserve some special consideration, as has been the case for copyright in computer programs. E.g., a general limitation for private copying might distort a balanced protection to the detriment of right-holders. For that reason, specific exceptions were provided for database copyright that restricted the use of traditional exceptions like private use concerning the high risk of piracy in electronic databases. Exceptions to the database right were provided with respect to the special objective of protection of the database right and its specific design specially tailored.

However, in light of other means or protection available for digital goods and files like technical protection measures and trade secrecy protection, it is questionable whether these special considerations still apply to the same extent as they did 26 years ago. As to Art. 6(2)(d) of the Database Directive the legislator already opened the gate to apply other limitations in copyright on a national level broadly. As to neighbouring rights, national law includes exceptions to a divergent extent, sometimes limited to certain exceptions, sometimes with broad reference to general copyright limitations.

One option here is to open the database copyright as well as the database right for general copyright limitations. Reference to Art. 5 of the InfoSoc could be made rendering any limitation applicable to the extent provided for by national law. While not all these limitations seem to be relevant for database copyright and *sui generis* right, no detrimental effect would arise from it and there may be exceptional cases when even more rarely used limitations could be applied.

A simple reference to Art. 5 of the InfoSoc would leave discretion to the Member States which limitations to adapt in general as well as in the case of databases. This could open the gate for diverse situations in different Member States to the detriment of clarity and harmonization in the digital single market. The other approach would be to make all or certain limitations mandatory in general copyright as well as in the interest of harmonization and reduction of barriers to trade within the Single Market. While copyright is connected to different cultural backgrounds more than other protection schemes this does not appear to be so relevant anymore as it has strongly developed a commercial character that would weigh in favour of harmonised limitations.

If an adaptation of the InfoSoc (making all exceptions mandatory) could only be done within the context of a broader revision of copyright law, an intermediate solution may be to apply all copyright exceptions to databases, leave the exceptions for the time being optional, but with the provision that Member States must make the same choice for databases as for other works. Hence, a Member State could only adopt an exception for both databases and other works or choose to not have the exception for both databases and other works.

A less encompassing approach would be to look at the specific limitations as to the extent they fit the special characteristics of databases and the two tiers of protection.

Some exceptions of Art. 5 InfoSoc lend themselves better to adoption under the *sui generis* right than other exceptions. The three most promising exceptions are analysed below. These are: the private use exception, the quotation exception, and the reporting exception. The exception for scientific research is dealt with under option 4b since this exception could be reformed extensively.

Private Use

Current copyright exception

Art. 5(2)(b) InfoSoc: "2. Member States may provide for exceptions or limitations to the reproduction right provided for in Article 2 in the following cases: […] (b) in respect of reproductions on any medium made by a natural person for private use and for ends that are neither directly nor indirectly commercial, on condition that the rightholders receive fair compensation which takes account of the application or non-application of technological measures referred to in Article 6 to the work or subject-matter concerned."

Context

Data are becoming ever more relevant for various aspects of societal life. It is important that everybody is able to familiarize themselves with data and its analysis, also as a means to prepare oneself for public discussion and debate. To prevent data illiteracy, there must be a free space to play with data. An exception from the extraction right for private use, can contribute to creating the room for experimentation and toying and open up possibilities for full participation in the public debate.

Relation to existing exceptions

Art. 9 Database Directive: "Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorization of its maker, extract or re-utilize a substantial part of its contents: (a) in the case of extraction for private purposes of the contents of a non-electronic database; […]". This exception to the *sui generis* right has important limitations. It is only available for lawful users and it is only applicable to non-electronic databases.

The copyright exception of art. 5(2)(b) InfoSoc is bound to the condition that the right-holders receive fair compensation which takes account of the application or non-application of technological measures.

Conditions

The condition of a lawful user may be removed from the wording of the private extraction exception. However, the condition would therewith not disappear completely. The CJEU

decided in the C-435/12 - *ACI Adam and Others* case [164]that a private copy from an" illegal" source requires permission and does not fall under the copyright exception. A proposal would be to make clear in a recital that this finding of the CJEU would be mutatis mutandis applicable to the *sui generis* variant of the exception. So, a core of the condition of the lawful user would live on.

A more drastic change is to open up the private extraction exception for electronic databases. This change is unavoidable if the adaptation of the exception should have an appreciable effect on the private use of data. The copyright exception of art. 5(2)(b) Directive 2001/29/EC requires a fair compensation for the rights holders, taking account of TPMs. In copyright laws of many Member States, a fair compensation had to be introduced already before the enactment of Directive 2001/29/EC, because the private copying of music and movies was done so vastly that the statutory exceptions for private copying no longer complied with the three-step test of art. 9(2) Berne Convention. Art. 9(2) Berne Convention reads: 'It shall be a matter for legislation in the countries of the Union to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author.' If the use that is made of an exception unreasonably prejudices the legitimate interests of the author, a fair compensation for excepted uses may resolve the issue. The three-step test was later copied into art. 5(5) Directive 2001/29/EC. Should a fair compensation also be required for the private copying exception to the *sui generis* right? According to art. 8(2) of the Database Directive, a lawful user of a database which is made available to the public in whatever manner, may not perform acts which conflict with normal exploitation of the database or unreasonably prejudice the legitimate interests of the maker of the database. It appears that at least the last two steps of the three-step test do apply under the *sui generis* right. Does this imply that a compensation for private copying exception under the *sui generis* right is also needed? That is not necessarily the case.[165] That an exception is subject to (the last two steps out of) the three-step-test does not require a compensation per se. A compensation is only required if the use that de facto is made of the exception gives rise to conflict with a normal exploitation or is prejudicial to the justified interests of the maker.[166] The frequency and intensity with which beneficiaries of a private copying exception under the *sui generis* right will use it, has to be awaited. The nature of a database differs from music and movies. Therefore, it is not very well possible to extrapolate data about the frequency and intensity of private use of music and movies to databases. Hence, there is not a priori reason to require a compensation for the private use of databases. Moreover, there are no international conventions in the field of intellectual property that apply to the *sui generis* protection of databases and that would require a fair

[164] Case C-435/12 - ACI Adam and Others

[165] See the last sentence of recital 35 Directive 2001/29/EC: In certain situations where the prejudice to the rightholder would be minimal, no obligation for payment may arise.

[166] THE CJEU in its case law consistently links the fair compensation to the harm suffered by rights holders through private copying of their works without authorisation. See CJEU 21 April 2016, C-572/14, ECLI:EU:C:2016:286, Austro-Mechana Gesellschaft zur Wahrnehmung mechanisch-musikalischer Urheberrechte GmbH v Amazon EU Sàrl, Amazon Services Europe Sàrl, Amazon.de GmbH, Amazon Logistik GmbH, Amazon Media Sàrl, at 19: '[...] fair compensation must be regarded as recompense to rightholders for the harm suffered by them (see, to that effect, judgments of 21 October 2010 in Padawan, C-467/08, EU:C:2010:620, paragraph 40; 16 June 2011 in Stichting de Thuiskopie, C-462/09, EU:C:2011:397, paragraph 24; 11 July 2013 in Amazon.com International Sales and Others, C-521/11, EU:C:2013:515, paragraph 47; 10 April 2014 in ACI Adam and Others, C-435/12, EU:C:2014:254, paragraph 50; and 5 March 2015 in Copydan Båndkopi, C-463/12, EU:C:2015:144, paragraph 21).

compensation. Finally, a fair compensation would likely come in the form of a levy system. Levy systems are a form of 'rough justice'. The obligation to pay levies will necessarily apply to some extent to individuals that do not engage in private copying. Therefore, levy systems are not without legal concerns and should not be introduced unless absolutely necessary.

Quotation

Current copyright exception

Art. 5(3)(d) InfoSoc: "3. Member States may provide for exceptions or limitations to the rights provided for in Articles 2 and 3 in the following cases: […] (d) quotations for purposes such as criticism or review, provided that they relate to a work or other subject-matter which has already been lawfully made available to the public, that, unless this turns out to be impossible, the source, including the author's name, is indicated, and that their use is in accordance with fair practice, and to the extent required by the specific purpose;[…]".

Context

In a world in which statements, unsupported by evidence, gain traction and statements supported by evidence are increasingly called into question and qualified as fake news it is important that IP rights such as the *sui generis* right do not stand in the way of substantiation of statements and opinions with data that can support the facts underlying these. Therefore, it is relevant that data from databases can be quoted in order to support opinions and statements. It contributes to the quality of public debate. In the *Spiegel-Online v Beck* case[167], the CJEU decided that, in the context of the exception in copyright law, a quotation can take place in the form of a hyperlink. If this policy option is adopted, also databases can be quoted and the possibility to quote by hyperlink makes quotation of databases into something that can be done in practice.

Relation to other exceptions

The reporting exception is limited to the press or the reporting of current events. Public and scientific discussion is not limited to the press or news reporting and therefore the reporting exception does not cover all needs for substantiation with data. A quotation rights gains in relevance in particular if the distinction between (extraction and reutilization of) insubstantial and substantial parts is abolished. In such case, a quotation would no longer fall outside the exclusive rights and an explicit quotation exception would be needed.

Conditions

Apart from obvious changes (database for work etc), the conditions as stated may remain in place.

[167] Case C-516/17 - Spiegel Online

Reporting

Current copyright exception

Art. 5(3)(c) InfoSoc: “3. Member States may provide for exceptions or limitations to the rights provided for in Articles 2 and 3 in the following cases:

(c) reproduction by the press, communication to the public or making available of published articles on current economic, political or religious topics or of broadcast works or other subject-matter of the same character, in cases where such use is not expressly reserved, and as long as the source, including the author's name, is indicated, or use of works or other subject-matter in connection with the reporting of current events, to the extent justified by the informatory purpose and as long as the source, including the author's name, is indicated, unless this turns out to be impossible”.

Context

If in the context of reporting on current events databases are communicated, it can be relevant that other outlets reporting on current events can re-communicate the same databases to the extent justified by the informatory purpose. The second prong of the copyright exception (starting from “or use of works…") is formulated rather broadly. The first prong has a reciprocity element in it. It is for the press, by the press. A similar condition is lacking in the second prong in an online environment where news reporting is not confined to specific professional parties, such a reciprocity requirement would not very well be possible. However, it risks giving the exception a wide scope.

Relation to existing exceptions

The Database Directive has no exception to the *sui generis* rights in relation to news reporting.

Conditions

If a news reporting exception is to be introduced, it could be considered to include the condition that use for reporting on current events is not reserved, also in the second prong of the exception, in light of the fact that the second prong appears to have a much wider scope than the first prong.

3.3.4.2 Supplementary option S2: Strengthening of research exception of the database right

Besides aligning limitations of the database right with copyright law, reference to the special situation of databases not only includes reviewing existing limitations but also the inclusion of new limitations or expansion of existing ones specifically tailored to database protection in copyright and *sui generis* and to enhance competition and innovation.

Here a proposal for an extended exception for research purposes is further elaborated and legally assessed.

Research has demonstrated that the non-mandatory character of the research exception in the InfoSoc as well as in Art. 6 and 9 of the Database Directive led to divergent rules in the Member States and the narrow scope of this exception may have negative consequences for

research activities[168]. Adding to that is the uncertainty related to the prolonging of the term of protection through value-added activities especially relevant for derivative databases. Hence, making the research exception mandatory may be considered to increase harmonisation of the legal framework applicable to research activities in a digital environment that extends across borders. It is however important to consider in this respect that the 2019 DSM Directive has already introduced harmonised exceptions applicable to text and data mining in the area of research which are mandatory and apply to the *sui generis* database right.

An example of the needs of the scientific community in an electronic environment may be found in the establishment of an e-research infrastructure as represented by the European project OpenAIREplus. Within OpenAIREplus, a complex database of records of publications and research data is supposed to be created.[169] The aim of the project is not to gain new knowledge through basic research in the area of hard or social sciences but to create a complex database as a research infrastructure for all information related to scientific publications resulting from EU-funded research, complemented by research data and research information. Thus, OpenAIREplus is not just a tool to query other databases, but a complete database that collects data, especially metadata, about every kind of scientific publication.

Within the OpenAIREplus infrastructure, metadata of publications and research data will be openly accessible to different kinds of people such as users and administrators on an open access basis. Parts of the research data that are accessible via OpenAIREplus are taken from scientific databases. These databases are generally protected by the *sui generis* database right. Repositories usually operate on an open access basis.

At the outset, the consent of the respective rightholder would be required as far as quantitatively or qualitatively substantial parts of the data of a database are used within such an infrastructure. In many countries of the EU, the extraction, but not the re-utilisation, of substantial parts of a database does fall within the scope of the scientific research exception.

Since infrastructures like OpenAIREplus aim to provide a comprehensive database of metadata of publications and related scientific research data as far as possible, substantial parts of other databases will likely have to be used.[170] Even if the used data of the individual acts of use is not of a substantial nature, there is a strong cumulative effect, since many such acts of use of insubstantial parts are carried out within OpenAIREplus, which added together potentially lead to the use of substantial parts. This would raise the need for a licence either on a commercial basis or open access.

The research data is to be accessible on an open access basis to many users over information networks. In many countries, such a making available/re-utilisation does not fall within the scope of the scientific research exception. This leads to the result that it is impossible to create

---

168 Guibault, L., & Wiebe, A. (2013): Safe to be open: Study on the protection of research data and recommendations for access and usage. Universitätsverlag Göttingen.p. 37 et seqq.; Reichman, J. H., & Okediji, R. L. (2012). When copyright law and science collide: empowering digitally integrated research methods on a global scale. Minnesota Law Review, 96(4), 1362; Derclaye, E. (Ed.). (2009). Research handbook on the future of EU copyright. Edward Elgar Publishing.

169 See Guibault, L., & Wiebe, A. (2013): Safe to be open: Study on the protection of research data and recommendations for access and usage. Universitätsverlag Göttingen.p. 18.

170 See the different usage scenarios in Guibault, L., & Wiebe, A. (2013): Safe to be open: Study on the protection of research data and recommendations for access and usage. Universitätsverlag Göttingen.pp. 118 et seqq.

an e-infrastructure without the consent of the respective rightholder of the database right. The TDM exception does not cover this as the data is not collected for the purpose of text and data mining analysis but to be reproduced and re-utilised within the infrastructure and then by the users.

The example of such an electronic research infrastructure supports the need to consider the introduction of an extended and mandatory limitation for the scientific use of databases. The scope of privileged uses should be enlarged, explicitly encompassing the reproduction and making available of a database as a whole or substantial parts for scientific purposes.

In most European countries copying has to be carried out by a person for their own scientific use to fall within the scope of the exception for scientific research. Within such an infrastructure the copying is done automatically and not pertaining to the needs of individual researchers, so the scope of the exception should be extended to encompass infrastructure operators of non-commercial scientific databases. By thus extending the permitted acts of extraction and re-utilization beyond only the purpose of text and data mining analysis, the scope of such an exception could complement the TDM exception. The TDM exception is not applicable in the case of infrastructure operators as the documents are reproduced and re-utilised for re-use and not for analysis such as in TDM. As the existing limitations stem from the age of print and entertainment industries many commentators see the need for a broader adaption to e-science.[171]

Hence, Art 9(b) and 6(2)(b) of the Database Directive may need to be extended to encompass extraction and the making available right for scientific purposes. Moreover, the privilege should also be applied to infrastructure operators of non-commercial scientific databases.

As universities increasingly are engaged in commercial activities or ends and do R&D projects with companies that have commercial purposes, a broad exception overcoming the distinction between commercial and non-commercial uses was proposed.[172] Such a broad exception seems to be a too heavy burden for right-holders and could have negative effects on competition as it may be difficult to confine the implementation of the exception to research purposes. More in line with the *CV v Melons* case, an extension to include commercial purposes could be considered limited by the requirement that no economic harm is done to the database holder. This would open up the re-use of databases for companies not in direct competition but could lead to new problems of delineating what could be considered as constituting scientific purpose in a commercial environment. An alternative approach would be to limit the exception to research institutions, thus mostly favouring universities and their activities as was done with the TDM exception in Art. 3 DSM Directive 2019.

---

[171] Reichman & Okediji, 96 Minn. L. Rev. 1362, 1433 (2012), who favour an "automatic fair use provision" , proposed by David, 29, and ultimately plead for a broad unlimited research exception allowing for use of digital materials for any scientific purpose; Max-Planck-Institute, Hilty et al., EUROPEAN COMMISSION–GREEN PAPER: COPYRIGHT IN THE KNOWLEDGE ECONOMY–COMMENTS BY THE MAX PLANCK INSTITUTE FOR INTELLECTUAL PROPERTY, COMPETITION AND TAX LAW (2008), p. 12, available at https://www.researchgate.net/publication/43234255 .

[172] Reichman & Okediji, 96 Minn. L. Rev. 1362, 1433 (2012); Max-Planck-Institute, Hilty et al., EUROPEAN COMMISSION–GREEN PAPER: COPYRIGHT IN THE KNOWLEDGE ECONOMY–COMMENTS BY THE MAX PLANCK INSTITUTE FOR INTELLECTUAL PROPERTY, COMPETITION AND TAX LAW (2008), p. 9, available at https://www.researchgate.net/publication/43234255.

3.3.4.3 Supplementary option S3: Public bodies would be excluded from the sui generis right.

The policy discussion on public bodies' database right goes back a long way. A major legislative change took place after the 2018 evaluation when the Open Data Directive (the reform of the PSI Directive) curtailed in relevant ways the public sector's potentially access-restricting use of the *sui generis* right. While Art. 1(6) of the Open Data Directive (ODD) goes a long way to prevent public bodies from overly curtailing availability of public information under open principles, it may not have resolved all tension with access rights, hence this proposed targeted option to deal with the residual issues.

Currently, the Database Directive is indifferent to the private or public nature of the producer of a database. This gap in the Directive has been subject to criticism from the beginning.

In case law an analogy to general copyright law has often been drawn for the *sui generis* right. Copyright links the exemption to creations authored by public bodies, if the very policy or legal purpose of the creation is to be disseminated to the public to the maximum extent. Under the same conditions (i.e. legal or policy interest in maximum dissemination of certain data), an exemption for the *sui* generis right could be framed relating to public bodies as database makers.

If a policy interest in maximum dissemination is not present it would mean that access to data of public bodies would be subject to the rules of the Open Data Directive ("ODD")[173] unless its application is excluded by Art. 1(2) ODD. The ODD still allows for contractual restrictions on the re-use of information in accordance with Art. 8 et seqq. ODD.

Excluding public bodies from ownership of *sui generis* right would corroborate the pertinent policy decisions made in the Open Data legislation of the EU. Open Data legislation includes the option not to strip public bodies of the *sui generis* right, but to restrict the implementation of the right to uses to the limits set by the ODD (Art. 1(6) ODD) while keeping the possibility of holding database rights in principle untouched. While the provision of databases in public hands is governed by the ODD it still leaves room for restrictions on the re-use in accordance with Art. 1(6) ODD.

The distinction between the existence of IP rights and their use has long been established in European competition law.[174] Following that distinction, the ODD severely restricts the use of the database right by public bodies but does not strip them of holding this right. The ODD still leaves room for different licensing practices in the Member States.[175] The UK developed the first own Open Government License where the Crown completely discarded *sui generis*

---

[173] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)

[174] See CJEU 262/81 - Coditel II

[175] See Solda-Kutzmann, D. (2011): Public Sector Information Commons. Informatica e diritto, (1-2), 199-217; as to Creative Commons see Papadopoulos and Bratsos, Openness/open Access for Public Sector information and works - the Creative Commons Licensing Model, 105.

rights.[176] France followed this example with “License Ouverte”.[177] In Germany, the "Datenlizenz 2.0" allows for free use but requires a designation of source and changes.[178] National divergencies have the potential to considerably impede the creation and distribution of value-added products and services between Member States. This could be avoided by an exclusion of public bodies from protection.

While Art. 1(6) ODD goes a long way to prevent public bodies from overly curtailing the availability of public information under open principles, interpretation of this limitation leaves some leeway and may lead to different practices in the Member States. A more far-reaching option would be to exclude public bodies from the *sui generis* right to make clear that they do not hold *sui generis* rights. This takes away a potential source of uncertainty and prevents the impression that the body could impose restrictions going further than permitted under Sec. 1(6) ODD (which will be less likely if any rights are excluded outright). This could help avoiding the process of challenging in each case whether the limits are kept and thus reduces transaction costs. Moreover, if the information had been provided without specific licensing provisions pursuant to Art. 8 ODD, the public body could still try to impose restrictions based on its database right at a later stage resulting in difficult legal evaluations. As licensing restrictions are only valid between parties this problem may also arise in cases where the information has been used in value-added processes down the information life cycle by companies not bound by contractual limitations. Based on the database right as an absolute right, claims could still be made against such companies and extensive legal conflicts may arise as to the scope of Art. 1(6) ODD in these cases.

Although occurrence of such cases may not appear very likely, they could not be excluded in the first place. The danger is relevant especially since many public bodies are still very reluctant to provide information for free as they claim to have some ownership rights. Hence, exclusion was also demanded by many respondents in the 2018 Evaluation Study on the Database Directive and recommended in the Study save for public bodies that are self-funded.[179]

Excluding ownership of database rights for public bodies in the first place could create clarity and avoid any later dispute as mentioned. It may also be regarded as good regulation not only aligning database law to general copyright but also to place exceptions in the pertinent legislative context for the sake of clarity and legislative and systematic coherence, as was done, e.g., with the decompilation provision in Art. 6 Computer Program Directive 2009/24/EC that is essentially a competition law provision. Clear and systematically appropriate legislation becomes ever more important the more different pieces of European legislation overlap or complement each other.

Tak the case of a researcher at a public institution like a university who establishes a database to be filled with empirical data she collected. While Art. 1(1)(d) ODD expressly states that raw data are covered by the ODD, Art. 10 ODD expressly opens provisions of such data to

---

[176] See De Filippi, P., & Maurel, L. (2015). The paradoxes of open data and how to get rid of it? Analysing the interplay between open data and sui-generis rights on databases. International Journal of Law and Information Technology, 23(1), 1-22.

[177] See Buri (2012), Accessing and Licensing Government Data under Open Access Conditions, p. 75.

[178] See Richter, H. (2018). Informationsweiterverwendungsgesetz (IWG). CH Beck., § 4 n. 142.

[179] See Evaluation 2018 Legal Analysis, pp. 131, 133.

commercial and non-commercial purposes. Implementation of this open principle in this context, however, depends on the implementation of the ODD in the Member States. The German government, e.g., practically refused to implement that with the argument that this open principle does not really fit the research landscape.[180] In the final version of the enactment public research institutions were practically obliged to provided research data, but this was restricted by “warranted business interests or knowledge transfer activities”.[181] Moreover, there is a lot of unwillingness by administrations to really implement the open principles. The public body could also hint to Art. 10(1) where intellectual property rights should be considered. Excluding IP in research data could prevent them putting it forward as an excuse. However, Recital 27 ODD makes very clear that provision of this data is important for innovation in society. This example shows that from the practical side, leaving this aspect wholly to the ODD is suboptimal as compared to the explicit exception in the Database Directive. The exception will not by itself remove the obstacles to actively provide data open, but it could support this and take away a purported obstacle or argument against it.

An exclusion could be even more relevant as to the area that is not covered by the ODD. On the one hand this would be the exceptions laid down in Art. 1(2) ODD, e.g., databases erected outside the scope of the public task of the public body (Art. 1(2)(a) ODD). Looking at a case of outsourcing of public tasks to private enterprises like in "Sächsischer Ausschreibungsdienst"[182] (German Supreme Court, GRUR 2007, 500) where a private publisher took publishing procurement documents in the course of fulfilment of a public task. The German Supreme Court regarded the documents as free of protection under copyright because of the function of the private entity as being "official" regardless of whether performed by public or private entity. If the same notion would be applied to the exception on the database right ODD and protection of database right would be congruent. Conversely, excluding any database produced or held by public bodies from protection by the database right would extend further than ODD and leave those databases produced outside the public task free from restrictions by the ODD. In times of ubiquitous digitisation it is not so unlikely that databases may also be produced by public bodies beyond their core tasks.

Clearly outside the scope of ODD are public undertakings under the conditions laid down in Art. 1(2)(b) ODD. A company like Juris GmbH that runs databases with court decisions and other legal texts may be owned by public bodies to more than 50% but may be regarded as an entity running business under market conditions and exposed to competition and hence not covered by the ODD. Application of an exception for public bodies to be introduced would then hinge on the extent to which private entities would be regarded as public bodies in the sense of the exception. The delegation of public tasks to private bodies does not change the validity of the notions underlying the exceptions if they perform public or general interest functions which should serve as a limiting factor in cases of public undertakings.

On the other hand, the ODD is limited to the “"re-use" of public documents while acts infringing on *sui generis rights* may also relate to access to public documents that do not necessarily amount to "re-use". Here the unclear delineation between access and re-use in a theoretical as well as practical sense comes into play. This is especially relevant as to access rights. Access to documents is not covered by the ODD but subject to national information access

[180] See Fortschrittsbericht Open Data, BT-DRs. 14140, p. 37

[181] See (§ 2(2) Datennutzungsgesetz of 16.07.2021 BGBl. I S. 2941, 2942, 4114

[182] See German Supreme Court, GRUR 2007, 500

rules. Limited access under these rules may lead to also excluding right to re-use pursuant to Art. 1(2)(d) ODD. Here, it is unclear whether the public body may refuse access based on its own rights with the consequence of also negating the right to re-use. While systematic construction may lead to a conclusion that their own rights may not be stipulated to avoid access[183] it is far from clear that such a construction would hold in all Member States alike. To avoid any disputes on the scope of the ODD, as a limiting factor to *sui generis* rights, the complete exclusion of public bodies from being right holders provides a clear and certain solution that helps avoiding additional transaction costs.

Introducing a separate provision into the Database Directive is warranted in light of the rationale of open data as well. Public documents are financed by public money and should be freely available to the public. As the *sui generis* is an economic right protecting investment, the general exclusion of public bodies is in line with open data policies.

This would benefit users, re-users as well as the public. No special considerations apply to MGD in this context.

4 Assessment of the policy options

In this chapter we present the assessment of the options related to MGD and the possible supplementary options for a wider review of the Database Directive. The assessment criteria covered are i) effectiveness in achieving the policy objectives, ii) efficiency in terms of costs and benefits from the option, iii) impact on fundamental rights, and iv) coherence of the option with existing policy initiatives and legal frameworks.

The assessment is based on limited and mostly qualitative evidence. The main sources for the assessment are legal analysis, literature review and desk research, answers to the survey on Database Directive Review, workshop with legal experts and interview with companies and business associations, and answers to the OPC-Data Act (although those were not an integral part of the support study as the OPC was running in parallel to the study).

4.1 Policy option 0: Baseline scenario – no action at EU level

4.1.1 Effectiveness

Without any EU action, the Directive will remain as it is. The current situation goes against the general objective to create legal clarity of the Database Directive.

In terms of specific objectives, the interpretation of the criteria of substantial investments of the CJEU in 2004, which excluded investments in the creation of data, and the lack of many infringement cases related to MGD, (coupled with little use of *sui generis* right by MGD database makers), seem to indicate that, at the moment, the Database Directive does not pose a serious obstacle to the sharing and usage of MGD. However, this might change in the future as MGD databases become more important as input for innovation and competition allow database makers to exploit the legal uncertainty of the directive to obtain a broader interpretation of the exclusive right resulting in a suboptimal level of overprotection.

[183] See Richter, H. (2018). Informationsweiterverwendungsgesetz (IWG). CH Beck. §1 n. 361.

Moreover, the directive currently may present some legal uncertainties beyond the MGD and data in IoT context. Elements such as a narrow list of exceptions compared to copyright law may further create obstacles to the objective of promoting the sharing and use of data.

4.1.2 Efficiency

4.1.2.1 Direct costs

Regulatory charges

No regulatory charges (e.g. taxes, levies, etc) are required by the Directive.

Substantive compliance costs

Around half of the database makers and user makers participating in the 2018 survey have experienced no or low additional compliance costs as a consequence of the *sui generis* right. By contrast, participants in a workshop organised as part of the 2018 evaluation underlined that makers of databases for public use incur costs to overcome the *sui generis* right (e.g. by subscribing CC0[184] licenses) and allow the extraction of the entire database.

Administrative costs

No administrative costs stem from the Directive.

Enforcement costs

More than 60% of the database makers and user-makers participating in the 2018 survey have experienced **no or low costs to enforce the *sui generis* right** (e.g. costs of monitoring activities and equipment, legal advice, litigation costs). This figure might be explained by the fact that the *sui generis* right is not so frequently applied, as other forms of protection are preferred. For example, among the respondents to the survey conducted for the present study (n=70), only 21 respondents have used or plan to use the *sui generis* right to protect their databases containing MGD, in contrast with 33 using contracts, 29 using technological measures, and 22 using the trade secret. Likewise, the majority of users and user makers responding to the 2018 survey maintained that they bear no or low costs linked to litigation. A large MGD database maker and user interviewed for the study, however, has stated that while it has not been important for their activities so far, the rise of these new technologies and data might imply a change in the future.

4.1.2.2 Indirect costs

Most users and user makers participating in the 2018 survey did not experience additional costs related to the search, access and use of alternative data or of databases that are not protected as a consequence of the *sui generis* right. Nor database makers incurred more costs in the creation of the databases as a consequence of the *sui generis* (e.g. to identify potential customers).

By contrast, some participants in the workshop organised as part of the 2018 evaluation highlighted that **legal uncertainty** affects in particular makers and users of **databases containing MGD** as it is not clear how the latter are covered by the Directive. This legal

[184] Creative Commons.

uncertainty has been reaffirmed in the open public consultation on the Data Act. Around half of the respondents (n= 279) declared themselves uncertain as regards the relation between MGD and the Database Directive and more than half of them (n=287) think that it is necessary to clarify the scope of *sui generis* right provided by the Database Directive in relation to the status of MGD.

The 2018 evaluation suggests that this uncertainty leads to **legal costs to stipulate contractual agreements** between makers, user makers and users. While MGD database makers can benefit from the flexibility and discretion ensured by contractual terms, users, especially SMEs, are at a disadvantage when negotiating with large organisations.[185] At the same time, the legal uncertainty concerning MGD databases does not ensure an appropriate protection to MGD database makers either. According to previous studies[186], in the absence of clear data ownership rights, data holders grant access to their data through bilateral contracts, which, however, do not give them any leverage in case of data leaks by third parties. These risks may reduce data collectors' incentives to make data available for re-use and be more restrictive in granting data access. Half of the organisations responding to the survey of the present study (n=40), declared that they have encountered problems when trying to obtain access to databases containing MGD. However, the reasons did not lie in the fact that the database was legally protected and there was no licensing available, but rather that either i) the data in the database were kept secret, or ii) the costs to obtain access to the database were too high, or iii) the database was protected with technical measures, or iv) there was no interoperability.

Stakeholders in the research field, consulted as part of the 2018 evaluation, indicated that the strict rules on the re-using of data for research purposes, limited to certain categories of research institutions (preamble 51), may **create friction to achieve innovations with societal impact,** since the *sui generis* right can create obstacles to collaborations and data exchange between research organisations, citizens, researchers outside academia, SMEs, spin-off etc. For example, it was mentioned that the *sui generis* right has constrained the sharing of data necessary to develop innovative, more eco-friendly solutions (e.g. in the sector of the renewable energy and energy efficiency), thus having potential indirect impacts on the environment. Moreover, the availability of quality data, or lack thereof, may also influence policies or projects, which are having a direct impact on the environment.

Finally, previous studies[187] have highlighted that the Database Directive may have negative societal impacts insofar it **is in contrast with policies incentivizing the re-use of public sector information**. Re-using public sector information has indeed societal benefits in terms of creation of jobs in the data economy, increasing the transparency of the public sector and the accessibility of information for citizens.

4.1.2.3 Direct benefits

Whilst one third of respondents were **uncertain about the benefits** stemming from the Directive, **database makers** participating in the 2018 survey generally reported **no or low benefits** experienced from the *sui generis* right. In particular, the majority of the database

[185] Deloitte (2017): Study to support the review of Directive 2003/98/EC on the re-use of public sector information

[186] Martens, B. (2020)

[187] Deloitte (2017)

makers responding to the 2018 survey contended that **the *sui generis* right did not positively influence the level of investments** made in databases in the EU nor it contributed to increase the number of databases produced. On the other hand, more than one third of respondents participating in the 2018 survey claimed that they have benefitted from **improved legal certainty** and **better protection of databases** against their unauthorised use by third parties compared to the previous situation. According to the database makers and user-makers, the Directive brings additional legal certainty compared to the previous situation and offers protection of investment where other means of protection (licenses, technological means, etc.) cannot help. Moreover, some participants to the workshop organised as part of the 2018 evaluation underlined that the harmonisation of European legal framework has reduced administrative costs linked to the stipulation of contractual agreements with all third parties. From the users' perspective, the main benefits stemming from the Directive concerned the improved certainty about the legality of the use of databases and about the identification of the owner, leading to the reduction of legal costs within the European data market.[188] However, this benefit does not materialise for MGD databases, for which there is not clarity about the legality of use of databases and for which the identification of the owner is not trivial.

However, asked specifically about the usefulness of the *sui generis* right, around one-third of the respondents to the survey conducted for the present study (n=38) did not express any opinion, 14 respondents found it not useful at all, 5 respondents think it is potentially useful, but they have not used it in practice, and 5 think that it is useful.

4.1.2.4 Indirect benefits

The access to databases that would not have been available or created without the *sui generis* right, is deemed a high or moderate benefit by 40% of the users participating in the 2018 survey. However, more than half of the responding database makers contended that the *sui generis* right did not positively influence the level of investments made in databases in the European Union and the number of databases produced. Therefore, the link between this indirect benefit and the Directive is uncertain.

Impact on competitiveness, functioning of the internal market and competition

According to experts consulted as part of the 2018 evaluation, the Directive increases the barriers to entry for potential competitors as it increases the costs for third parties to access and use data. This particularly concerns the use scenario in which competitors need access to complete, aggregated data sets to access the primary market or certain entirely new, complementary or aftermarkets. This effect hinders competition and innovation in sectors where new solutions could be developed if data were available (e.g. sensor-generated turbine data in the wind sector). Academic experts were concerned that the *sui generis* right can worsen lock-in problems. Moreover, it was observed by consumer organisations consulted in the 2018 evaluation that the *sui generis* right has strengthened the position of the market leader, raising costs to the detriment of consumers.

4.1.2.5 Summary on efficiency of the option

In general, the cost/benefit impact of the Directive in the broader data economy context remains unknown as its relevance is limited and contractual agreements usually overcome its

[188] Survey conducted as part of the 2018 evaluation.

provisions. The **uncertainty about the costs and benefits associated with the Database Directive** emerged as a key finding in the 2018 evaluation.

As noticed by the 2018 evaluation "most stakeholders have experienced low, if any, benefits from the Database Directive except in terms of improved legal certainty. However, the associated costs have not been significant either". At the same time, the uncertain relation between MGD and the Database Directive seems to have negative repercussions in terms of reduced access to databases containing MGD – usually protected through other measures - and legal costs to stipulate contractual agreements. Moreover, the current provisions of the Directive have negative indirect impacts when it comes to the collaboration among different actors for research purposes, to the access to databases held by public bodies and, ultimately, to competition within the internal market, as extracting or re-using substantial parts or entire databases could be crucial to develop new products and services.

4.1.3 Impacts on fundamental rights

4.1.3.1 Fundamental right to protection of personal data

The fundamental right to protection of personal data is enshrined in Art 8 EU Charter of Fundamental Rights and envisages that (1) everyone has the right to the protection of personal data concerning him or her; (2) personal data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law; everyone has the right of access to personal data and the right to have it rectified. Finally, (3) an independent authority shall control compliance with these rules.

Databases in many cases contain personal data in the sense of the General Data Protection Regulation (GDPR). State of the art legal scholarship agrees that the concept of personal data in law is very broad and it is increasingly difficult to distinguish personal from non-personal data,[189] according to some, to the point of all data becoming personal.[190] Providing access to and sharing databases containing personal data constitutes processing personal data. Therefore, the impact of the reform on the fundamental right to protection of personal data must be assessed.

Although the fundamental right to protection of personal data is enshrined in the EU Charter of fundamental rights, secondary legislation, currently the GDPR, gives effect to the right, and "the protection of personal data shall be exercised in accordance with the conditions and limits defined by the measures adopted to give effect to it."[191] Therefore the analysis of the impact

---

[189] E.g. Zwenne, GJ (2013): Diluted Privacy Law. Inaugural Lecture (University of Leiden, 2013) 33 et seq., Purtova, N. (2018): The Law of everything. Broad concept of personal data and the future of EU data protection law. Law, Innovation, and Technology 10(1); Graef et al (2019): Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation. European Law Review 2019, vol. 44 no. 5, p. 605-621; Lynskey, O. (2019): Criminal justice profiling and EU data protection law: precarious protection from predictive policing. International Journal of Law in Context 15(2) pp. 162-176; Lynskey, O (2020): Delivering Data Protection: The Next Chapter. German Law Journal 21(1) 80-84

[190] Purtova, N (2018)

[191] EU Network of Independent Experts on Fundamental Rights Commentary of the Charter of Fundamental Rights of the European Union, at 95 (June 2006), http://ec.europa.eu/justice/fundamentalrights/document/index_en.htm

on the fundamental right to protection of personal data will be done with reference to the GDPR.

According to Art. 13 of the Database directive, the instrument is without prejudice to provisions concerning data protection. The GDPR and the Database Directive, although may apply simultaneously to the same situations, in principle are two parallel legal regimes that serve different purposes: the former to protect fundamental rights and interests requiring protection of personal data, and the latter – to protect industrial property rights. Yet, the Directive does have a relationship of significance with some aspects of data protection rights.

Impact on the right to data portability.

The Database directive is of significance to the right to **data portability** under Art. 20 GDPR. The right is composed of two elements: (1) the right to receive personal data in a structured, commonly used and machine-readable format and (2) the right to transmit those data to another controller[192] or have the data transmitted to another controller directly where technically feasible.[193] Under Art 20(4) of GDPR, the right "shall not adversely affect the rights and freedoms of others". These rights and freedoms of others include IP rights such as the *sui generis* database right of a database maker.[194]

The effects of the *sui generis* database rights on the right of the data subject *to receive* his or her personal data (1st element) seem to be marginal. Portability for personal use is unlikely to encroach on the *sui generis* database rights of a database maker and will likely override any IP interests on balancing. The GDPR does not offer explicit guidance on how the right to data portability and IP rights should be reconciled in case of conflict. However, Recital 63 GDPR contains such guidance concerning the right of access. As the right to receive data under Article 20 GDPR is similar to the right of access, in particular, the right to obtain a copy of data undergoing processing (Article 15(3) GDPR), Article 29 Working Party applies this guidance to the context of data portability. Using the language of Recital 63 GDPR, Article 29 Working Party explains that even when conflicting IP rights are considered before answering a portability request, "the result of those considerations should not be a refusal to provide all information to the data subject".[195] In other words, a database maker acting as a data controller should make steps to satisfy the data subject's portability request and provide the data in a form that does not infringe on the *sui generis* right.

The right *to transmit or have the data transmitted* from the database maker to another controller (2nd element) can be affected more seriously as porting to actors other than data subjects *in specific cases* is more likely to raise concerns of possible violations of *sui generis* database

---

192 Art 20(1) GDPR

193 Art 20(2) GDPR

194 Graef et al (2018): Data Portability and Data Control: Lessons for an Emerging Concept in EU Law. German Law Journal 19(6) 1367; Elfering, S. (2019): Unlocking the Right to Data Portability: An Analysis of the Interface with the Sui generis Database Right, p. 48

195 Art. 29 Data Protection Working Party, Guidelines on the Right to Data Portability, 16/EN WP 242 rev.01 (Apr. 5, 2017), p. 12, referring to Recital 63 GDPR concerning right of access which is similar to the right to receive personal data in a structured, commonly used and machine-readable format, the 1st element of the data portability right.

rights.[196] For instance, this could be the case when collective (automated) portability requests by the data subjects are organized or done on behalf of the data subjects in bulk by subsequent controllers, e.g. using the Application Programming Interfaces (APIs) provided by the database maker to facilitate data portability. Such requests are likely to occur, among others, with further development of data intermediaries as an industry, e.g. data brokers or providers of data management services who target large numbers of data subjects. The resulting transfers in bulk would likely be of a substantial part of the original controller's database and/or systematic and hence at risk of violating *sui generis* database rights. In specific cases where the requested transmission of data amounts to a violation of the *sui generis* database right, the two rights will have to be balanced against each other, and a request to transmit the data to another controller may be refused, if the *sui generis* database right overrides the portability right. As Article 29 Working party explains, "[t]he right to data portability is not a right for an individual to misuse the information in a way that could be qualified as an unfair practice or that would constitute a violation of intellectual property rights".[197]

Importantly, the right to data portability does not enjoy any priority or a higher status in relation to a *sui generis* database right and will not always override the *sui generis* database right on balancing. The latter also enjoys the status of a fundamental right under Article 17(2) of the EU Charter. Both fundamental rights to data protection and intellectual property are not absolute and have to be considered in relation to their social function.[198] There is an academic opinion that "data protection automatically trumped other interests and could not be traded off for economic benefit".[199] The CJEU has on several occasions prioritized data protection interests of data subjects over economic interests of data controllers.[200] However, these instances involved considerations of circumstances of a particular case and degree of interference with the data protection rights rather than a general principle that a fundamental right to data protection always overrides economic fundamental rights. Under Article 52(1) EU Charter, both fundamental rights to data protection and IP can only be limited while, among others, their essence is respected. Hence, one of the two rights will override the other automatically only when its essence would otherwise be violated by the exercise of the other right. Given that the understanding of what constitutes the essence of the fundamental right to data protection has not yet fully formed and the Court so far has construed this concept very narrowly,[201] it is problematic to argue at this point that the right to data portability belongs to

---

196 Graef et al (2018): Data Portability and Data Control: Lessons for an Emerging Concept in EU Law. German Law Journal 19(6) 1380 et seq.

197 WP 242 rev.01, p. 12

198 Concerning the right to data protection see Recital 4 GDPR and CJEU, Joined cases C-92/09 and C-93/09, Volker und Markus Schecke GbR and Hartmut Eifert v. Land Hessen [GC], 9 November 2010, paras. 47–52, 58, 66–67, 75, 86 and 92. Concerning IP rights see Case C-200/96 at para. 21, Metronome Musik and more recently CJEU, Case C-70/10, Scarlet Extended, ECLI:EU:C:2011:771, Judgment of Nov. 24, 2011, para. 43.

199 E.g. Feretti, F. (2014): Data protection and the legitimate interest of data controllers: Much ADO about nothing or the winter of the right? 51 CML Rev 843, 852

200 Case C-131/12, Google Inc. v Agencia Española de Protección de Datos and Mario Costeja González, ECLI:EU:C:2014:317 [81]

201 See e.g. Brkan, M. (2019): The Essence of the Fundamental Rights to Privacy and Data Protection: Finding the Way Through the Maze of the CJEU's Constitutional Reasoning. German Law Journal, 20, pp. 864–883 ("the Court often perplexingly portrays data protection as a minimalistic right limited to security measures, in particular, in the context of blanket retention of data", e.g. in Joined Cases C-293/12 & C-594/12, Digital

that essence. Hence the right to portability cannot be said by default to override the fundamental right to protection of intellectual property. Even more so, doubts have been raised in the literature whether or not the GDPR right to data portability is even a part of the fundamental right to data protection under Article 8 of the EU Fundamental Rights Charter.[202] Unlike the right of access (Article 15 GDPR), it is not explicitly mentioned in Article 8 EU Charter. Koops argues that data portability belongs more to the consumer or competition law.[203] As Ausloos et al. argue, it may be difficult to claim - following Article 52(3) EU Charter that connects the EU Charter to the European Convention of Human Rights - that it is necessary in a democratic society to limit data protection rights "representing essential values in a democratic society"[204] to serve economic interest alone. Yet, the right to data portability itself seems to be an economic right and it is unclear. In any case, it will have to be demonstrated how the right to transmit data to another controller serves democratic values. Finally, the presence of a dedicated legislative provision restricting the exercise of the right to data portability on the ground of rights and freedoms of others, including protection of *sui generis* database rights (Article 20(4) GDPR) makes it more justifiable for a controller in light of Article 52(1) EU Charter not to grant a data portability request purely based on the conflict with his *sui generis* right.[205]

The outcome of the balancing of economic rights such as *sui generis* rights with the right of access (Article 15 GDPR) is likely to be different and in favour of the right of access. This is because of a special role the right of access plays in the system of data protection. Under Recital 63 GDPR, access is essential for the ability of the data subjects to exercise control over processing of their data: specifically, it serves to make data subjects aware of, and enable them to "verify, the lawfulness of the processing". In light of this role, it is unlikely that limiting this right in favour of the economic right to protection of intellectual property will be considered necessary in a democratic society. Accordingly, Recital 63 GDPR concerning the right of access explicitly instructs that the considerations of protecting intellectual property rights should not lead to "a refusal to provide all information to the data subject". Recital 68 GDPR concerning data portability contains no similar language.

The outcome of the balancing between the *sui generis* database right and the right to have personal data transmitted to another controller will depend on the circumstances of each case, in accordance with the principle of proportionality (Article 52(1) Charter). Since there is no EU Court of Justice case law where an IP right and a data portability right are balanced against each other, the outcome of such balancing is hard to predict.

Rights Ireland, at para. 40). Neither is the essence of the right to protection of intellectual property is clear (see Husovec, M. (2019): The Essence of Intellectual Property Rights Under Article 17(2) of the EU Charter. German Law Journal, Volume 20, Special Issue 6: Interrogating the Essence of EU Fundamental Rights , September 2019 , pp. 840 – 863, DOI: https://doi.org/10.1017/glj.2019.65)

202 Graef et al (2018)

203 B.J. Koops (2014): The trouble with European data protection law. International Data Privacy Law Vol. 4, No. 4, 258

204 Ausloos et al (2019): Getting Data Subject Rights Right A submission to the European Data Protection Board from international data rights academics, to inform regulatory guidance. 10 JIPITEC 283, p. 306

205 Ausloos et al (2019), p. 307

The impact of the Database directive on data protection is **limited by a narrow scope of the right to data portability.** The right to data portability can be invoked only when the data is processed on the ground of consent or performance of a contract (Art 20(2) GDPR), which are grounds suitable for use in the context of consumer-oriented connected devices such as connected cars and domestic appliances and consumer-oriented eHealth wearables. Consent and contract are not suitable as grounds of data processing in the context of employment (e.g. data generated by wearables worn for reasons of workplace safety or by connected machinery in the context of production). Personal data processed on other grounds, such as legitimate interest, also is not subject to mandatory portability. In addition, only personal data "provided" by the data subject to the controller falls within the scope of the data portability right.[206] Provided data includes the "data actively and knowingly provided by the data subject", e.g. by filling in forms, and "observed data provided by the data subject by virtue of the use of the service or the device",[207] e.g. eHealth wearables or IoT devices, but excludes data that is "inferred" and "derived", i.e. created by the controller via an analysis of provided data, e.g. like assessments, profiles, scores, etc.[208]

4.1.3.2 Impact on the right to property and the freedom to conduct a business

The right to property and the freedom to conduct a business are historically closely related.[209] The freedom to conduct a business concerns to an appreciable extent the freedom to dispose of property in the context of maintaining an enterprise and the freedom to contract (e.g. with respect to property). Therefore (and in order to prevent repetition) the text below will foremost deal with the property right and deal with the freedom to conduct a business where such is needed.

Neither the right to property, nor the freedom to conduct a business are absolute rights.[210] Art. 52(1) of the EU Charter of Fundamental Rights (CFR) indicates under what conditions the rights can be limited: "

1. Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others."

Many policy options that have been discerned in this report potentially limit data producer's enjoyment of the right to property and the freedom to conduct a business. Art. 52(1) CFR provides the framework for assessing whether such limitations are allowed.

---

206 Art 20(1) GDPR

207 WP 242 rev.01, p. 10

208 WP 242 rev.01, p. 10

209 Groussot et al (2017): Chapter 14: Weak right, strong Court – the freedom to conduct business and the EU Charter of Fundamental Rights, in: Sionaidh Douglas-Scott and Nicholas Hatzis (eds.), Research Handbook on EU Law and Human Rights, Research Handbooks in European Law series.

210 C-70/10, Tiscali v. SABAM

The right to property

The right to property is laid down in Art 17 CFR: "

1. Everyone has the right to own, use, dispose of and bequeath his or her lawfully acquired possessions. No one may be deprived of his or her possessions, except in the public interest and in the cases and under the conditions provided for by law, subject to fair compensation being paid in good time for their loss. The use of property may be regulated by law in so far as is necessary for the general interest.
2. Intellectual property shall be protected."

Art.17(1) CFR gives its own rules for the circumstances under which a deprivation of the right can take place. These rules do not completely coincide with the rules that art. 52(1) CFR gives for allowing for (other) limitations of the right (than deprivation). This will be taken into account when assessing the policy options.

It is however not completely clear whether the norms laid down in paragraph 1 for physical property are mutatis mutandis applicable to intellectual property. Paragraph 2 only declares that intellectual property shall be protected. We take as a starting point that also for a deprivation of intellectual property stricter rules apply than for any other limitation.

It is also assumed here that the *sui generis* right is intellectual property.[211] The handbook on intellectual property usually qualifies the *sui generis* right as intellectual property (and also as industrial property).

The EU Charter of Fundamental Rights has a close relation to European Convention on Human Rights. Art. 52(3) CFR formalizes and describes this relation:"

3. In so far as this Charter contains rights which correspond to rights guaranteed by the Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down by the said Convention. This provision shall not prevent Union law providing more extensive protection."

A corresponding right to the CFR's right to property can be found in Art. 1 protocol 1 European Convention on Human Rights (ECHR), entitled "Protection of property": "Every natural or legal person is entitled to the peaceful enjoyment of his possessions. No one shall be deprived of his possessions except in the public interest and subject to the conditions provided for by law and by the general principles of international law".

The preceding provisions shall not, however, in any way impair the right of a State to enforce such laws as it deems necessary to control the use of property in accordance with the general interest or to secure the payment of taxes or other contributions or penalties.

Art. 1 of Protocol No. 1 ECHR is applicable to intellectual property, as is clear from the Anheuser-Busch case.[212] This case concerned a trademark (actually an application for a

---

[211] Theoretically, an argument for the opposite position could be that a sui generis right does require any creativity on the part of the maker, only substantial investment.

[212] C-245/02 - Anheuser-Busch , § 72.

trademark). Neither the fact that a trademark is an industrial property, nor that the trademark application in question belonged to a corporate entity was a ground to deny protection under human rights. On this basis, it is assumed here that also a *sui generis* right held by a corporate entity finds protection under art. 1 protocol 1 ECHR.

The freedom to conduct a business

Article 16 CFR recognises the freedom to conduct a business

The freedom to conduct a business in accordance with union law and national laws and practices is recognised. This is the freedom to operate a business or engage in enterprise without unnecessary state intervention.

With regards to the impacts on the freedom to conduct a business and the right to property, unlike under other options, in the baseline these rights will not be limited by legislative intervention, hence no impact on these rights is expected.

4.1.4 Coherence

The *sui generis* right seen as a broad exclusive right on the (re) use of databases could be considered conflicting to some extent with the policy priorities formulated in the European strategy for data such as “the data can flow within the EU and across sectors”. Moreover, the current legal uncertainty around the *sui generis* right may not be coherent with the provision of the European strategy for data stating: “the rules for access and use of data are fair, practical and clear”. Other elements addressed such as not aligned exceptions with existing copyright law as well as exclusive rights to public bodies indicate low legal coherence with respect to other existing legal frameworks.

4.2 Policy option 2a: Exclusion of MGD from the scope of application of the *sui generis* database right

4.2.1 Effectiveness

This option could help achieving the general objective of providing legal clarity on database rights as the legal uncertainty present in the baseline scenario on the application of Database Directive to MGD will be resolved. Furthermore, by changing the scope of protection of the *sui generis* right in a narrower sense, i.e. to not cover MGD databases, the option will contribute to the specific objective of avoiding the Database Directive to become a potential obstacle in sharing and trading MGD and data generated in the IoT environment. As this type of data/databases are becoming increasingly relevant in the data economy this will contribute to fostering overall data access and usage in the Digital Single Market.

The real effectiveness, however, will also depend on two important factors: a) how the exclusion of MGD is implemented in the Database Directive, either via a direct exclusion of explicitly defined MGD databases or indirectly via the modification of the relevant investments and thresholds for the “substantial investment” criteria, b) the risk, on one side, to reduce incentives to create databases (thus increasing the available data in the economy) and, on the other side, that users would lose the use regime granted by the *sui generis* right as MGD database makers would resort to other means of protection.

Regarding the first point, on the one hand, the introduction of a new concept will inevitably carry a certain degree of uncertainty. With the concept of MGD the risk is potentially higher as

human intervention is difficult to completely rule out, as it was also mentioned by legal experts. On the other hand, the exclusion of MGD via exclusion of investments in "generation" of data or a change in the substantiality criteria will start from an already existing concept for which, however, clarity is still missing after many years, and for which agreed standards and thresholds have been proven difficult to establish. Either way, under the current legal uncertainty, not acting now to try to exclude MGD from the scope of *sui generis* right, would pose the high risk that, increasingly relevant MGD and big data scenarios could be captured by a broader interpretation of the *sui generis* right as suggested by experts in the literature.[213] Moreover, as noted in the legal assessment, in the context of MGD and IoT, the *sui generis* right might be used to protect investments in raw MGD as such, thus going against the original intention of the Directive to protect investments in databases rather than the generation of the elements contained in them.

Regarding the second point, the previous evaluation and the evidence presented in the efficiency assessment e.g. statement from the IP Federation Data Committee, suggest that *sui generis* right protection of the investment in databases has no or little positive effect on incentivizing databases creation. This is even more true, for MGD which in most cases are generated as a spin-off or a by-product of other main economic activities, e.g. in vehicle data. Moreover, database makers already rely extensively on other measures of protections such as contracts, TPMs and Trade Secrets (see results from the survey reported below in the section assessing indirect costs), and in general MGD databases already tend to be not publicly available,[214] so less in need of *sui generis* right protection. Therefore, an exclusion of MGD would unlikely trigger more contractual restrictions and the adoption of other means of protection. On the other hand, it would reduce the risk that, as MGD become even more relevant, the *sui generis* exclusive right might start to be claimed by MGD database makers seeking extra protections beyond *factual* control, becoming a possible obstacle to newly introduced access and usage rights[215]. This will present the risk of creating further data monopolies (as most MGD related to the operations and performance of the machine/device itself could be seen as sole sources) and reduce ability for users of devices to make full potential of their device and third parties to rely on data to develop and create new products in downstream markets.

Therefore, while this discussion suggests that the exclusion of MGD would unlikely have a major, immediate effect in sharing and usage of MGD, the option will likely have a forward-looking positive effect, especially in combination with other possible actions related to data access and usage rights that might be considered in the Data Act. This conditional on implementation of the provision to exclude MGD via a definition (either of MGD, or substantial investments) that is clear and stable.

Finally, 2a would reduce the possible opportunistic litigation and thus reduce transaction costs by stripping away the possibility of spurious baseless litigations of third-party data use. This will avoid uncertainty from the side of the user on the validity of possible IP claims made by MGD database makers.

---

213 See Leistner (2020)

214 See Evaluation 2018.

215 Drexl (2018), Leistner (2020)

4.2.2 Efficiency

4.2.2.1 Direct costs

Regulatory charges

No regulatory charges are envisaged by this option.

Substantive compliance costs

Businesses might incur compliance costs in case this option requires to distinguish MGD from non-MGD in "mixed" databases. Asked about this type of databases, interviewees from different sectors revealed that databases tend indeed to merge both types of data. In one of the interviewees' opinion, separating the two types of data should be feasible, although would require additional data elaborations. Another interviewee, instead, claimed that it is not feasible to separate the two types of data. Apart from the technical obstacles, interviewees underlined the difficulties to clearly identify which data are MGD and which are not within the same database. In their opinion, applying a single approach (i.e. that of MGD) to the entire mixed database seems a more efficient solution.

Administrative costs

This option does not entail any administrative costs since no information obligations are required.

Enforcement costs

The enforcement costs linked to this option mainly depend on the approach chosen to exclude MGD from the *sui generis*.

The first approach (i.e. providing a definition of MGD and excluding them explicitly) carries less risks of legal uncertainty. However, depending on the clarity of the definition, disputes might arise on its interpretation before national courts and, eventually, the CJEU. On the other hand, the second approach (i.e. changing the substantial investment as a requirement for database right) while keeping the *sui generis* right relatively simple (i.e. minimal introduction of new concepts that might raise questions or lead to disputes about their application), it brings major difficulties to the implementation. Indeed, it will be difficult to find appropriate criteria for stipulating a clear threshold for substantial investment. Moreover, it would be difficult for the plaintiff to prove substantial investment in MGD separately in cases of mixed databases or for the defendant to reject that notion.

In addition, if the option consists only in excluding MGD from the *sui generis* and does not envisage a separate provision explicitly stating that "no protection shall be granted to MGD", enforcement costs may arise if any Member States approve legislation that protects MGD with other means. In this scenario, indeed, the national court would need to try to interpret the national law in line with the Database Directive and, eventually, can refer the case to the CJEU.

4.2.2.2 Indirect costs

By excluding databases containing MGD from the *sui generis*, the access and re-use of (parts of) databases would be regulated mainly by contractual agreements. The costs of protecting MGD databases with contracts are perceived differently by the respondents to the survey conducted for the present study (n=70): 14 respondents consider the costs of protecting MGD databases with contracts as "high" or "very high", 16 consider them as "considerable", 15 as "no costs" or "minor", 25 respondents did not express an opinion. Among the respondents,

legal experts and R&D experts with legal background considered the costs linked to contracts as "considerable", "high" or "very high" (n=40). However, the 2018 evaluation found that, in practice, contractual terms are often used even when the *sui generis* applies. Coherently, the results of the survey conducted for the present study (n=70) show that less respondents (21) have used or plan to use the *sui generis* right compared to contracts (33), technological measures (29), and trade secret (22). As a consequence, there should not be significant additional costs compared to the baseline. Members of the IP Federation industry' Data Committee confirmed that data holders primarily choose to control the use of their dataset with third parties under contractual terms; they are not usually reliant on database rights (*sui generis* or copyright) and the availability of rights under the Database Directive does not affect their dataset licensing strategy or decisions.

By excluding databases containing MGD from the *sui generis*, makers of MGD databases would lose protection against other third parties, which was considered one of the main benefits of the Directive according to the respondents to the 2018 evaluation survey.[216] An interviewee (database maker) explained that the protection of the *sui generis* is seen as a guarantee that they will be able to commercialise the database or the data analytics that rely on the database. The database maker will likely fall back on trade secrecy, unfair trade practices and general tort law (and to some extent contract networks) to act against third parties. However, these instruments all have their limitations compared to the *sui generis* right. Coherently, almost one third of respondents to the survey carried out for the present study and expressing an opinion (7 out of 20) maintained that they would be negatively affected by excluding MGD databases with regard to "legal protection from employees and other third parties" and 8 out of 32 with regard to "costs of protection". One MGD database maker/user interviewed, mentioned that lack of structured protection such as *sui generis* right to MGD would reduce their incentives to share and produce data themselves due to uncertainty and lack of trust. Despite this shortfall, exclusive rights as the *sui generis* are considered less effective in protecting MGD databases[217] and there is uncertainty among survey respondents on the usefulness of the *sui generis* right to protect MGD databases.[218]

In principle excluding MGD databases from the *sui generis* removes a layer of protection from them. However, in practice, several studies have shown that, if no provision explicitly grants fair access to MGD (or at least certain types of MGD), exclusion alone may potentially still not be sufficient enough to entirely remove restrictive access to databases.

First of all, part of the literature[219] and the stakeholders consulted within the context of the 2018 evaluation noted that, as a consequence of the *Ryanair case*, databases not protected under the copyright or *sui generis* right do not enjoy the exceptions and authorisations included in the Database Directive (i.e. the rights of lawful users, the exception for private purposes, for teaching and for scientific research, for public security or administrative or judicial procedure, as well as the exceptions introduced by the 2019 DSM Directive and amending the Database

---

216 See Evaluation 2018, p.52

217 See responses to the question "E2. How would you rate the following alternative means of protecting databases containing MGD?".

218 See responses to the question "Question E6. To what extent is sui generis database right useful for the purposes of your data protection compliance, e.g. to protect personal data from unauthorized access, etc.?"

219 See Borghi, M., & Karapapa, S. (2015): Contractual restrictions on lawful use of information: sole-source databases protected by the back door. European Intellectual Property Review, 37(8), 505-514

Directive, such as the TDM exception). However, these exceptions apply only to users and not to third parties. As observed by Borghi and Karapapa (2015), “makers of non-protected databases are not bound by any obligation to ensure that lawful users can access the content without restrictions”.

A JRC study[220] focused on the agricultural sector, highlighted the existence of lock-in situations when contracts are the main instruments for protecting databases. The study demonstrated that contractual negotiations prevail in cases of voluntary agreements and the database maker exploits its advantaged position.[221] The same situation was depicted by two interviewees, active respectively in the automotive and the industrial technology sector, who claimed that the exclusion from the *sui generis* would trigger protective behaviour on the side of the database maker. In the view of one of the interviewees, the *sui generis* constitutes an essential common ground of protection, which shelters especially smaller companies. Without the *sui generis*, according to him, bigger companies would easily prevail over smaller companies in negotiations to reuse their data, while the former ones would refrain as much as possible from allowing access and re-use of their databases.

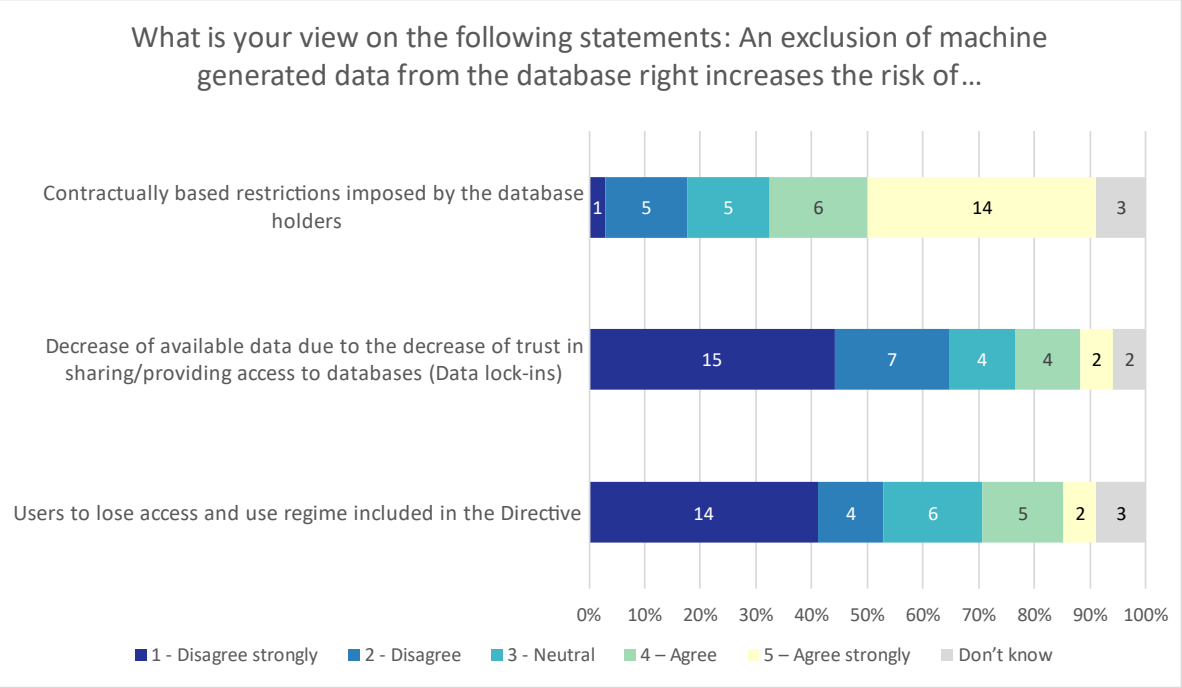
The results of the survey conducted for the present study, however, reveal a more mixed opinion on the risks arising from excluding MGD databases: although the majority of respondents (20 out of 34) think that there will be more contractually-based restrictions imposed by the database holders, they do not think it will translate in lock-in situations or in less access to data for users.[222] Moreover, the Data Act proposal may include rules on contracts dealing with data-sharing that could limit the possible increase in contractually based restrictions by database holders after the exclusion of MGD.

220 Atik & Martens (2020)

221 Atik & Martens (2020)

222 By having a closer look at the responses of R&D stakeholders with legal expertise, they do not think that exclusion of the MGD from the database right would decrease available data due to the decrease of trust in sharing/providing access to databases, thus producing data lock-ins (only one thinks so). They rather think (4 out of 8) that it would bring to contractually based restrictions imposed by the database holders.

Figure 9: Risk from exclusion of MGD from *sui generis* right



Source: Survey for the Database Directive Review

4.2.2.3 Direct benefits

By excluding databases containing MGD from the *sui generis*, the expected direct benefits are:

- Bringing legal certainty with respect to the baseline. At the moment, there is uncertainty as to whether databases containing MGD are protected by the *sui generis* or not.[223] Indeed, around half of the respondents to the survey conducted for the Data Act Study (n=279) declared to be not sure about how the database directive applies to MGD databases. Moreover, more than half of the respondents (n=287) think that it is necessary to clarify the scope of *sui generis* right in relation to the status of MGD. 26 of the 35 respondents to the survey conducted for this study maintain that excluding MGD will have a positive effect on obtaining legal certainty. Nevertheless, the materialisation of this benefit largely depends on the clarity of the provisions to identify the object of exclusion (by definition or through the substantial investment criterion). Both approaches to exclude MGD databases suffer the intrinsic difficulty of defining concepts on which there is not an agreement after years of discussion.
- Ensuring that the *sui generis* right, as an additional layer of data protection, does not interfere with the objective of the Commission in the data policy to incentivise data sharing (especially considering the potential of MGD). However, less than 20% of the respondents (n= 275) of the survey conducted for the Data Act Study declared that the

[223] Evaluation 2018.

sui generis database right or possible uncertainties with its application created difficulties and prevented him/her from seeking access or use data.[224]

- Reduction of information and transaction costs. According to Leistner 2018, it is often very difficult to identify the relevant right holder, i.e. the database maker. This raises serious information and transaction cost problems which substantially aggravate the existing and acknowledged access problems which already follow from factual control over the data source. Moreover, even if the database maker is properly identified, it is often very difficult to design comprehensive use contracts in actual practice because, in many cases, complex network infrastructures with different inputting and consuming participants will be concerned. By excluding MGD and making use of contract networks would have the potential effect to efficiently assign the property right to the person which is most central to the investment effort thereby generating a cost saving for the users and avoiding misappropriation of *sui generis* right by unlawful makers. In this view, excluding MGD databases would reduce the number of possible opportunistic litigations and the linked transaction costs by stripping away the possibility of spurious baseless litigations of third-party data use. However, in order to make contract-based solutions effective to overcome transaction costs and information asymmetry, additional legal guideposts might be needed.
- Allowing database makers and users to rely on the most effective tools. Asked about rating the different means of protecting MGD databases, more respondents to the survey consider means such as contracts, technological measures, trade secrets and competition law as very or partially effective compared to exclusive rights as the *sui generis*. In particular, for legal experts participating in the survey the most effective means to protect databases containing machine-generated data are trade secrets and contract obligations. Both options are at or above 80% of agreement. The same opinion is shared by R&D experts with legal background. This subgroup seems considering the *sui generis* right as rather 'not effective' to protect MGD datasets (7 out of 11 respondents). Giannopolou (2019) underlines the benefits for the industry in "relying solely on contract law, being that it allows for the flexibility and modularity needed in providing dynamic access and usage rights depending on the nature of the data set and the purpose of the use". [225]

4.2.2.4 Indirect benefits

By removing a layer of protection to MGD databases, this option might contribute to maximising the impact of MGD on European economy.[226]

In particular, the majority of respondents to the survey conducted for this study (n=35) see positive effects for innovation and research activities and for revenues generated from the production and/or exploitation of databases as a consequence of excluding MGD databases.

---

[224] However, the majority of them do not know.

[225] See Giannopoulou, A. (2019): Access and reuse of MGD for scientific research. Erasmus L. Rev., 12, 155.

[226] A Deloitte study indicates that the seamless transfer of non-personal MGD between devices and different actors would have a positive impact of around 1.4 trillion EUR to European Union GDP in 2027. Deloitte (2018): Realising the economic potential of machine-generated, nonpersonal data in the EU.

However, as noted in the section on indirect costs above, there are elements that can jeopardise this effect (e.g. protective behaviour by database makers).

Impact on competitiveness, functioning of the internal market and on competition

Excluding MGD databases from the *sui generis* might have relevant positive impacts on competition in different sectors, as it can ease access to complete datasets for market entrants, who might use these data to develop innovative products. Leistner (2017)[227] observes how, in typical big data settings, users will need, in most cases, complete data sets in order to develop value-added new products or services. In this view, Leistner maintains that "the limitation to the use of substantial parts of a database […], is not an efficient means to protect the freedom of competition and to prevent leveraging potentials in respect of typical big data". More than half of the respondents to the survey (n=35) share Leistner's opinion and believe that excluding MGD from the *sui generis* will have positive effects in terms of entering new markets (18 out of 35) and compete against other firms and developing new/value-added products (23 out of 35). As remarked by more than one respondent from the automotive industry participating in the survey conducted for the present study: "Excluding MGD from the *sui generis* right and easy access to such data would foster innovation and competition with regard to data driven business models".

It should be noted that given the amount of data necessary to perform big data analytics, this mechanism is particularly relevant for MGD databases. Similarly, the 2018 evaluation of the Database Directive highlighted the barriers to entry for potential competitors due to the *sui generis* right, in particular in use scenarios in which competitors need access to complete, aggregated data sets to access the primary market or certain entirely new, complementary or aftermarkets. Access to sensor-generated turbine data were mentioned as examples. Although excluding MGD from the *sui generis* will not address possible issues linked to the existing exclusive position of the manufacturer, it will facilitate the application of competition rules e.g. in case of sole-source databases. As already mentioned, some view MGD databases typically as spin-off databases (or in general by-product/incidental to the main activity of the database maker spin-off databases), which are often single source and may thus be monopolised. Therefore, this option is seen to reduce the barriers to entry in the primary and secondary market of the sole-source database maker.

4.2.2.5 Summary on efficiency of the option

Assessing the efficiency of this option presents challenges as long as i) there are some technical aspects that are not defined yet and may affect the judgment on the overall option, such as the approach to exclude MGD, the treatment of mixed databases, the definition of MGD adopted; ii) there is few tangible evidence available. The assessment is mainly based on the anticipated possible effects reported in selected studies and academic works and on the opinion of survey's respondents. Despite these challenges, the following findings can be considered to have an indication of the efficiency of the option.

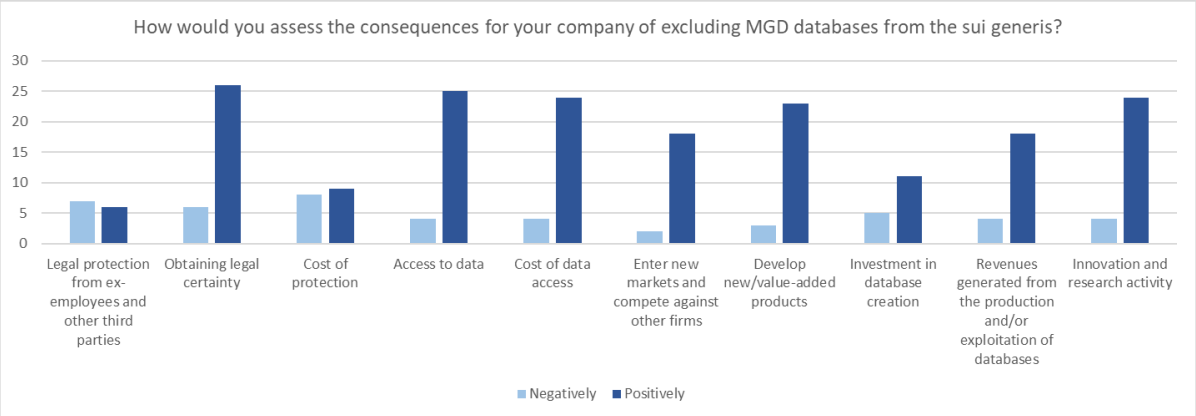
The option brings no significant additional direct costs for the stakeholders although some implementation and enforcement costs may materialise due to the difficulties in having a stable and shared definition of MGD or of "substantial investment" (depending on the approach chosen). Whilst these elements seem to be the premise for the good functioning of the option,

[227] Leistner (2018): Big Data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform, in Sebastian Lohsse, Reiner Schulze and Dirk Staudenmayer (eds), Trading Data in the Digital Economy: Legal Concepts and Tools (Nomos 2017) 27

the majority of respondents to the survey expect positive effects from it in terms of more legal certainty with respect to the baseline. This should be reflected in reduced information and transaction costs, with less room for opportunistic litigations.

As for the indirect costs possibly caused by excluding MGD from the *sui generis*, two aspects should be mentioned. First, database makers of MGD databases (and of mixed databases if the exclusion applies to them as well) will lose protection against third parties. Secondly, users of MGD databases will not enjoy the exceptions and limitations granted by the Database Directive ex. Article 8 and 9 [nor the TDM exception]. In this respect, however, it is noteworthy that in the current situation (baseline) the *sui generis* is not generally applied to MGD databases and it is considered by respondents to the survey as a less effective means of protection compared to contracts, trade secrets, competition law or technical measures. Some scholars have underlined the risks associated with a scenario in which database access is mostly regulated by contracts, warning about the power imbalances between data holders and users and the possible lock-ins' effects. However, concerned stakeholders consulted do not share this opinion and doubt that it will translate to lock-in situations or in less access to data for users. Moreover, issues related to contract provisions impacting access and use of data should be considered also in light of other regulatory initiatives, that can potentially introduce balances (e.g. the Data Act). Most of the enterprises participating in the survey, recognise that the exclusion of MGD from the *Sui generis* may lead to restrictions imposed by data holders via contractual agreements, but this will not necessarily lead, in their opinion, to a decreased access to databases or lock-in situations. By contrast, excluding MGD from the *sui generis* is expected to have positive effects on competition, as it will facilitate the entry in new markets and development of new/value-added products.

Figure 10: Assessment of consequences of excluding MGD from the *sui generis* compared to the baseline

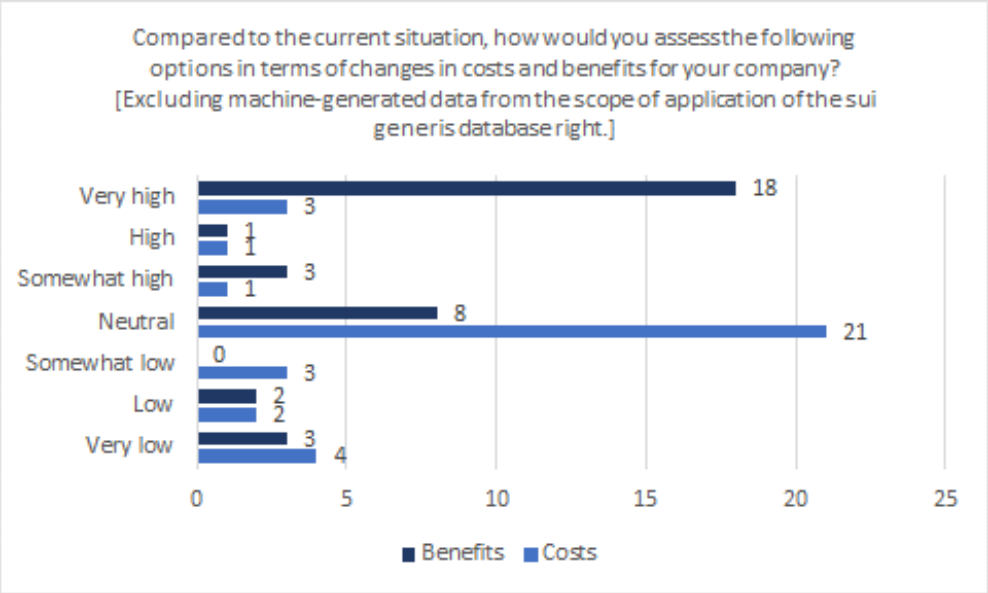


Source: Survey for the Database Directive Review

Despite the low number of responses to the survey conducted for this study, the results show that respondents expect this option to bring great benefits and no additional costs compared to the baseline. This opinion is prevalent also within the subgroup of legal experts participating in the survey.

Benefits compared to the baseline include: bringing legal certainty, ensuring that the *sui generis right* does not interfere with the objective of the Commission in the data policy to incentivise data sharing, reducing transaction costs and allowing database makers and users to rely on the most effective tools to manage extraction from and reutilisation of databases.

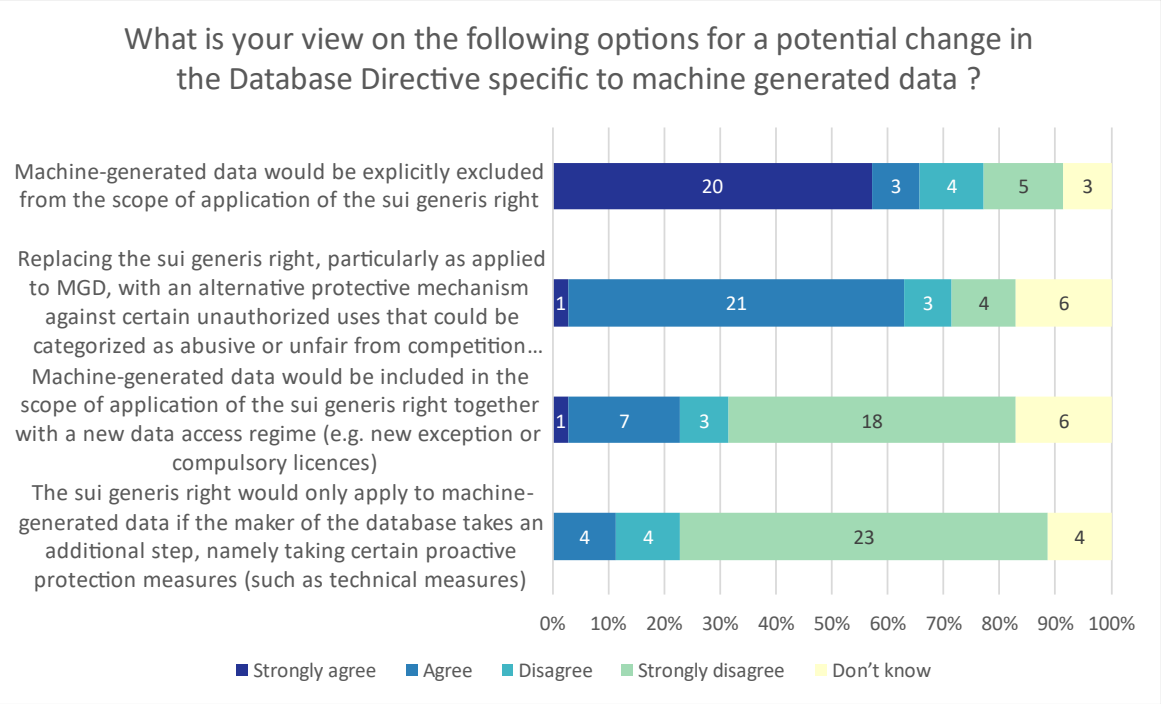
Figure 11: Assessment of costs and benefits of excluding MGD from *sui generis* compared to the baseline



Source: Survey for the Database Directive Review

In line with the above findings, the exclusion of MGD from the scope of the *sui generis right* appears to be the potential change (related to MGD) with the highest support by the respondents of the survey conducted for this study -20 out of 35 respondents reported to "strongly agree" with the change.

Figure 12: Agreement about potential change in the Database Directive specific to MGD



Source: Survey for the Database Diretive Review

Finally, to assess the overall efficiency of the option we should also consider that databases containing MGD will increase in the future. The exclusion of these databases from the *sui*

generis implies that this provision will be relevant for a very limited number of databases. This might undermine the efficiency of the Directive as a whole and might trigger its revision in the future.

4.2.3 Impacts on fundamental rights

4.2.3.1 Fundamental right to protection of personal data

Impact on the right to data portability.

In the specific cases where the *sui generis* database rights conflict with and limit the right to data portability under policy option 0, exclusion of the MGD from the scope of application of the *sui generis* database right will impact the right to portability, especially the right to transmit personal data to another controller, positively with regard to MGD.

Machine-generated personal data will fall within the scope of the right to data portability as "provided data" when it is "observed data provided by the data subject by virtue of the use of the service or the device,"[228] e.g. a wearable eHealth device, IoT home appliance or a connected car. The positive impact will occur since removal of the MGD from the scope of *sui generis* database right will leave no intellectual property right to be balanced against a data protection right to data portability. The positive effect of exclusion of the MGD from the scope of *sui generis* database right will likely be more substantial for the right *to transmit or have the data transmitted* from the database maker who is the original controller to another controller. As discussed under policy option 0, when a request to port data to another controller conflicts with a *sui generis* database right, the two rights must be balanced against each other and if the *sui generis* right overrides, the request can be refused. With the *sui generis* database right not applicable to the MGD, this ground of refusal of portability requests will no longer exist in relation to such data.

4.2.3.2 Fundamental right to right to property and to the freedom to conduct a business

The freedom to conduct a business

Under this policy option, databases with MGD are excluded from the *sui generis* right. For the analysis undertaken here, it is assumed that databases with MGD can be protected under the *sui generis* right in current law. The policy option would deprive the MGD database producer from the property protection he now still enjoys and would make conducting a business more burdensome, since an instrument to act against free riders is lost. Unlike under the right to property, it is not necessary to distinguish between the different forms in which this policy option is introduced (such as with retroactive effect or prospectively affecting existing databases or prospectively only affecting new databases). Hereinafter it is analysed whether the abolition of the *sui generis* right for databases with MGD is provided for by law, respects the essence of the freedom and is proportional.

The limitation is provided for by law. It does concern here the removal of a protection in law. Such removal is (obviously) subject to the same democratic guarantees as the enactment of law that introduces a new instrument and provides a comparable level of clarity. Any questions of law about the nature or extent of the removal can be resolved by the CJEU.

[228] WP29 on data portability, p 10

Respect the essence of freedom to conduct a business. Businesses are not deprived of each and every means to protect their investment against free riding. They can still use contract or technical means. Furthermore, they can rely on flanking laws, like trade secret protection.

For the assessment of proportionality, we need to look at the aim necessity and effectiveness of the proposed measure and see to what extent it is reasonable means to the end.

What is the aim, necessity and effectiveness of the measure? The aim is to bring data sharing to a higher level to ensure that the potential of data for innovation, economic growth and the public interest is used. The *sui generis* right creates uncertainty under data users about the existence and extent of protection of databases with MGD. This lack of clarity has a deterrent effect on data usage. Policy option 2 reduces uncertainty by excluding MGD databases from the *sui generis* right.

Is the measure a reasonable means to achieve the aim? Thereto, we first look what problem it raises in terms of the freedom to conduct business. Database producers are deprived of a legal instruments suitable to act against persons or entities free riding on investments in the creation of databases and therewith, the exposure of database producers to such free riding is potentially enlarged. Secondly, we look why the measure is nonetheless a reasonable means to achieve its aim. Thereto we observe that there is an important empirical/causal element in the reasoning underlying the claim that policy option 2 impinges on the freedom to conduct business, viz. that the loss of the *sui generis* right leads to greater exposure to free riding. It is unclear whether this indeed the case. A database producer has other effective instruments to protect himself against free riding: contract and technology. The *Ryanair* case provides anecdotal evidence that companies see contract as the primary instrument to protect data. Ryanair was in fact prepared to forego *sui generis* protection in exchange for broader contractual protections. The survey indicates that the *sui generis* right is little known and used by producers of databases with MGD. In conclusion, objections against abolition of the *sui generis* right for databases with MGD do not appear to weigh in heavily. Thirdly we assess whether taking away *sui generis* protection is the lightest instrument to achieve the aim. The alternative would be to continue the existing situation in which the *sui generis* right (as per assumption above) also covers databases with MGD. Relief for data users would then be created via access rights (option 3) or new exceptions (option 4). However, under these options the burden to show that the terms for invoking such access rights or exceptions are present rests on the data user. This might introduce new uncertainties. Moreover legally, rules that are exceptions tend to be interpreted narrowly, which would also weigh against data users in case of adoption of the alternatives. Finally, effects of the basic policy option 2 on data producers could be reduced by introducing a defensive right. The need for a defensive right would however depend on the extent to which we may be able to show that in the current situation producers of MGD rely on the *sui generis* right. If this reliance is low, it is not necessary to introduce a defensive right. In fact, then it will prove difficult to find a defensive right that is less effective than the current *sui generis* right.

In conclusion, the freedom to conduct a business would most likely allow for the introduction of the second option.

4.2.4 Coherence

This option does not conflict with existing legal frameworks and EU initiatives. Rather, it is complementary to the parallel initiative in the Data Act which might introduce measures of access and usage rights to data. By excluding MGD from the Database Directive, possible

conflicts between these new access regimes and the *sui generis* right protection can be prevented.[229]

Other regimes beyond the Database Directive and the future Data Act may apply in parallel to MGD, such as the Trade Secret Directive for commercially valuable information that is kept secret and the GDPR for personal data. However, the exclusion of MGD from the Database Directive does not change the situation with regard to the parallel application of these legal regimes, which already exists now in the current status quo.

4.3 Policy option 2b: Exclusion of MGD from the scope of application of the *sui generis* database right and introduction of a more flexible infringement test related to economic detriment

4.3.1 Effectiveness

This option has the same effectiveness as 2a with regards to MGD. Moreover, by establishing in the Database Directive a "fair use" or "fairness test" in the infringement analysis inspired by the recent *CV vs Melons* case, where the protection would be weighed against the societal benefits and innovation generated from the use of database made by third party, the policy aims to contribute to the objective of fostering the access and re-use of data in the EU Digital Single Market. In turn this will work as a basis to promote innovation and competition beyond MGD as this will be applied to all databases protected by the Database Directive.

Additionally, this option will help achieve the general objective of bringing legal clarity in the Database Directive by removing the distinction between substantial and insubstantial takings which have been a major source of uncertainty so far as identified in the background assessment. Moreover, the new test may also consider the duration of the protection based on the time to redeem the investment, thus removing the fix duration and the renewal terms.

Nevertheless, the need to implement the CJEU ruling through legislation may not be pressing considering that the ruling already forms part of *acquis communautaire* and, in practice, should be observed by the national courts of the Member States in similar.

Moreover, as explained further below, the effectiveness of the policy carries a risk in its implementation. The test proposed in the recent Latvian judgement remains quite vague. If implemented in the Database Directive, the criteria for the test must be further defined in order to provide clarity to the national courts that will need to apply it, otherwise further legal uncertainty and fragmentation at the national level could arise. Furthermore, we note that the uncertainty on the side of the database maker – due to the fact that the *sui generis* protection would depend on other contextual factors – may disincentivize them to make the data public, thus restricting *de facto* ability of users to access them.

4.3.2 Efficiency

Since this option builds on option 2a, all the costs and benefits identified under option 2a as a consequence of excluding MGD databases from the *sui generis* are valid for option 2b as well. In order to avoid repetitions, this section focuses on the costs and benefits linked to the introduction of a fairness test for databases covered by the *sui generis.* Nonetheless, the overall assessment on efficiency takes into account the impacts identified in option 2a.

[229] Drexl, J. (2018)

4.3.2.1 Direct costs

Regulatory charges

No regulatory charges are envisaged by this option.

Substantive compliance costs

Compared to the baseline, more costs will arise for the right holders since they will need to prove that there has been a significant detriment to their investment or a risk of such detriment. All difficulties in demonstrating the substantiality of investment would still be present.

Administrative costs

No administrative costs should arise from this option.

Enforcement costs

Important enforcement costs will arise for national courts as they will need to define on a case-by-case basis whether the database is entitled to be covered by the *sui generis*. This situation can lead to national courts referring more questions to the CJEU or the development of a body of nation case law (with the risks of enhancing fragmentation among Member States).

Implementation costs can also arise depending on whether the test should apply only to competitors (and how these are defined) or also to other users. This element has the potential to jeopardise a uniform application of the rule in different sectors, thus increasing implementation costs.

4.3.2.2 Indirect costs

Because of the different elements of the test, the database maker would not know a priori whether its database will be likely to pass the infringement test, since it depends not on its investment but on external factors (i.e. the benefits for the competitors and the society to access the database). Facing this uncertainty, the database makers might choose to turn into other forms of protections.

4.3.2.3 Direct benefits

With respect to the baseline, where the substantial/insubstantial investment threshold focuses only on the right of the database maker, without considering the contextual factors and the interests of the other party nor of the society, the option will create a more flexible test to strike a fair balance between, on the one hand, the legitimate interest of the makers of databases in being able to redeem their substantial investment and, on the other, that of users and competitors of those makers to have access to the information contained in those databases and the possibility of creating innovative products based on that information.

It can also be argued that the option can bring simplification with respect to the duration of protection and renewal terms. Indeed, as already mentioned, the fairness/infringement test would allow for linking the duration of protection with the time required to redeem the investment in the specific case, possibly using the notion of amortisation, without the need to have fixed terms of duration.

4.3.2.4 Indirect benefits

Adopting such a flexible solution will have positive impacts on the sustainability of the legislation, disregarding the type of data included in the databases and the tools with which they have been collected.

Impact on competitiveness, functioning of the internal market and competition

Encouraging competition and innovation in the internal market are the main objectives of this option. The legitimate interest of competitors and other third parties, as well as other societal interests, are indeed the counterbalancing factors to the right of the database maker to redeem its investment. As a consequence of this option, "very innovative products producing strong consumer benefits or socially important re-use of public data by journalists that does not have significant impact on the investments made are likely to prevail". [230]

4.3.2.5 Summary on efficiency of the option

As illustrated in option 2a, excluding MGD databases from the *sui generis* would bring no significant additional direct costs for the stakeholders, although some implementation and enforcement costs cannot be excluded (for instance, they can arise from technical and legal aspects such as the definition of MGD). As for the indirect costs, first, database makers of MGD databases will lose protection against third parties. Secondly, users of MGD databases will not enjoy the exceptions and limitations granted by the Database Directive ex. Article 8 and 9 (although rarely applied). Despite the concerns raised by these two aspects, there is uncertainty as to the general usefulness of the *sui generis* to protect MGD databases while other more effective measures could be adopted.

Nonetheless, the benefits in terms of legal certainty, increased competition, reduction of transaction costs and potential to contribute to the EU data economy were pointed out.

Compared to the baseline and option 2a, this option, establishing an infringement test on databases that could be protected by the *sui generis*, would have positive impacts in terms of flexibility and competition. However, as it is currently envisaged, the option presents important compliance and enforcement costs. Although the option resembles the CJEU opinion, many obstacles to its implementation can be anticipated, with the risk of bringing additional legal uncertainty and fragmentation among Member States.

Further analysis would be needed to assess the efficiency against the baseline: as a matter of fact, the EC ruling *CV v Melon case* is very recent and the effects it will have on the interpretation of the current substantial insubstantial criterion are still uncertain. Should it set a precedent for the upcoming rulings at European and national level, this option might serve to consolidate the case law.

4.3.3 Impact on Fundamental Rights

4.3.3.1 Impact on the Fundamental Right to Protection of Personal Data

Impact on the Right to Data Portability

As this policy option also involves exclusion of the MGD from the *sui generis* database right, it also potentially impacts the exercise of the right to data portability positively as described for

[230] Derclaye and Husovec (2021)

policy option 2a. However, similarly to policy option 2c, the exact degree of that impact will depend on the exact configuration of the test. Here, as it is tailored for specific non-data protection purposes (protect the database maker's investment and the database maker from economic detriment), it would not impact data protection rights beyond the effects following the exclusion of MGD from the scope, as discussed under policy option 2a.

4.3.4 Coherence

This option does not appear to be conflicting with other existing legal frameworks.

4.4 Policy option 2c: Exclusion of MGD from the scope of application of the *sui generis* right and introduction instead of an alternative control mechanism applied only to databases containing MGD

4.4.1 Effectiveness

This option aims to achieve the general objective of promoting access and re-use of data in the EU Digital Single Market. It aims to do so by striking the right balance between, on one side, limiting exclusive rights on MGD databases deemed to strengthen data monopolies and under-use of data at the expenses of innovation and competition, while, on the other, still giving MGD database makers a sort of control mechanism to protect them from unlawful misappropriation and use of their databases by third parties. Thus, the option has a potential to still incentivise database makers to share their data. As this option focuses on the introduction of a new "defensive right" specific to MGD databases, it contributes to the specific objective of avoiding the Database Directive to become an obstacle in sharing and trading of MGD databases and data in the increasingly relevant context of IoT.

While the defensive right could be effective in fostering data sharing, the right may not be effective in incentivising and stimulating production of new MGD databases, and thus more data in the economy. This because of the same reason mentioned throughout the report that internal incentives for firms to generate and collect MGD in their main economic activities are already present.

Finally, as explained in the efficiency assessment below and stated by legal experts in the workshop, there is the risk that this option would not achieve overall more legal clarity. While MGD would be clearly excluded from the scope of the existing *sui generis* right, conditional on implementing and effective and stable definition of MGD, the definition and especially implementation of the new alternative right would still carry uncertainty in its introduction. It would require a case-by-case approach in the assessment of balancing interests between the database market and third parties and general public interests

4.4.2 Efficiency

4.4.2.1 Direct costs

Regulatory charges

No regulatory charges are envisaged by this option.

Substantive compliance costs

In order to benefit from the defensive right, the database maker should rely on contracts for the use and access of its database to third parties or should have undertaken any efforts to put in place sufficient TPMs. However, as described under option 2a, contracts and other TPMs

are already used to protect MGD databases. Therefore, no relevant additional compliance costs are expected.

Administrative costs

No administrative costs should arise from this option.

Enforcement costs

The possible enforcement costs are mainly linked to the specific elements and provisions that will be introduced to limit the scope of protection. Depending on the clarity of definitions and ease of application of the different elements, they may create obstacles to the implementation of this option. For instance, enforcement costs can materialise if a case-by-case assessment is needed to decide whether a violation or breach on the part of the supplier is material.

4.4.2.2 Indirect costs

An important unintended effect of this option seems to be the creation of legal complexity. According to legal experts consulted in the context of this study, another kind of *sui generis* right regime specific for MGD may create additional legal complexity and to some degree uncertainty, especially if the requirements for protection are too vaguely defined and difficult to prove in practice. Especially the criterion of "wrongfulness" and the associated weighing of interests will be difficult to evaluate by the parties in advance. Whereas one of the objectives of revising the Database Directive is to address the uncertain regime of MGD databases under the current regime, according to the consulted legal experts, this option would contribute to that objective only to some extent, i.e. by clarifying the regime applicable to MGD, but with unintended negative consequences due to the uncertain application of the defensive right.

Asked about his view on the introduction of a defensive right for MGD databases, an interviewee representing a big company acting as user maker, was sceptical about its efficiency due to the complexity of the scheme and the difficulties, for the database maker, to prove the materiality of the breach. On the other hand, he underlined the need for database makers to rely on a structured protection scheme (such as the *sui generis*).

In addition, as observed in the discussion of impacts of option 2a, the exceptions provided by Art. 8 and Art.9 do not apply to databases excluded by the *sui generis*. Therefore, depending on whether any exception to the defensive right will be introduced, there could be inconsistency between the rights of lawful users of databases covered by the *sui generis* and lawful users of databases covered by the defensive right.

4.4.2.3 Direct benefits

The main benefit of this option is granting protection to MGD databases against third parties once they are formally excluded from *sui generis* right. The loss of legal protection from ex-employees and third parties is in fact the main negative consequence of the exclusion of databases containing MGD from the *sui generis* right as expressed by the survey respondents. A limited defensive right would support contractual freedom, without impeding data sharing.

4.4.2.4 Indirect benefits

Indirectly, by replacing a broad and unspecified *sui generis* right (as in the baseline) with a limited protection, will reduce information and transaction costs for third parties, as it should be easier for them to check the legality of re-using (parts of) databases. As a matter of fact,

36% of user-makers responding to the survey of the 2018 evaluation reported important or very important difficulties in determining when a database is protected by the *sui generis* right.

In addition, such a defensive right seems an effective forward-looking tool against misappropriations in situations where data is transmitted to persons not bound by the restrictions who could use the data for their competitive advantage without proper remuneration.

Impact on competitiveness, functioning of the internal market and competition

Descending from the increased assurance regarding protection against third parties, the new protective regime for MGD databases is expected to discourage database makers from resorting to trade secret protection or other strict forms of protection, hence favouring competition. However, at this stage, there is inconclusive evidence on the extent to which this option will actually create incentives to increase sharing and re-use of MGD. One big company in the interview stated that protection was necessary to enhance data sharing. However, the defensive right may only supplement trade secret protection as it is not strong enough to provide exclusive protection and trade secret protection would still be used to avoid the data to be released to the public domain.

Concurring with the arguments made by Drexl (2018)[231], Kim (2018)[232] and Banterle (2019)[233] against a possible similar defensive right applied to raw data, one may argue that this option, by *de facto* introducing another kind of protective regime specific for MGD, will discourage re-use of MGD databases while benefitting data holders.

4.4.2.5 Summary on efficiency of the option

Before comparing the costs and benefits brought about by this option (for which, however, few or no tangible evidence is available), the assessment of the efficiency should acknowledge the risk underlined by the consulted legal experts that this policy option might be counterproductive with regard to the objective of a REFIT initiative[234]: bringing simplification and improved efficiency**.** In the experts' opinion, indeed, creating a defensive right applicable only to MGD databases, which would run in parallel to the *sui generis* – from which MGD databases would be excluded – would create serious additional complexity and uncertainty, leaving issues such as the definition of MGD and the treatment of mixed databases open.

As for the other objectives of the review of the Database Directive[235], this option seems to contribute to striking the right balance between the incentives to create and share databases and the welfare gains from access and use of MGD databases. However, the Evaluation 2018 concluded that exclusive rights, such as the *sui generis*, do not seem to affect the willingness

---

[231] See Drexl, J. (2018). Drexl also use the recent Autobahnmaut case to analyze the different application of the three protection regimes, i.e. sui generis right, trade secret and the defensive right proposed by the Commission in 2017, concluding that the first two would give excessive protection to the data holder.

[232] See, Kim, D. (2018), pp.154-165.

[233] Banterle, F. (2020), pp. 199-225.

[234] European Commission's regulatory fitness and performance programme (REFIT)

[235] See Objectives Trees in Chapter 3.

to create databases, pointing to a lack of effectiveness of this kind of measure. Although not favouring the creation of the databases, this option is still expected to positively affect competition, as database makers will be assured as to the protection of their databases against third parties and will thus refrain from resorting to trade secret protection or other strict forms of protection. However, there is no evidence that this mechanism will materialise, and some authors have argued the opposite, i.e. that the defensive right, by creating a layer of protection to MGD databases, will discourage their use.

The defensive right is conceived as a limited right; on the one hand, these limitations will be crucial to ensure that the right balance is found. On the other hand, depending on the clarity of the provisions and on the possibility of applying them in a uniform manner (e.g. on the requirements to prove the materiality of the breach), they risk creating implementation and enforcement costs.

To conclude, the risk of negative impacts on the overall clarity and functioning of the Directive seems to supersede the possible benefits. Nevertheless, it is worth noting that when asked about the possibility of replacing the *sui generis* right, particularly as applied to MGD, with an alternative protective mechanism against certain unauthorized uses that could be categorized as abusive or unfair from a competition perspective, 22 out of 35 respondents agree or strongly agree.

4.4.3 Impacts on fundamental rights

4.4.3.1 Fundamental right to protection of personal data

Impact on the right to data portability

As this policy option also involves exclusion of the MGD from the *sui generis* database right, it also potentially impacts the exercise of the right to data portability positively as described for policy option 2a. However, the exact degree of that impact will depend on the exact configuration of any alternative control mechanism, such as “defensive rights”, introduced for the databases containing MGD. Any alternative control mechanisms will likely have a similar restrictive effect on the exercise of the right *to transmit or have the data transmitted* from the database maker who is the original controller to another controller. However, if the scope of those alternative mechanisms is narrower than the scope of the current *sui generis* database right, the restrictive impact will also be narrower. For instance, if the database makers who are holders of sole-source databases are disqualified from having “defensive rights” they will not be able to invoke those to prevent data from being ported.

4.4.3.2 Fundamental right to property and freedom to conduct business

The Right to Property

As indicated under the policy option of merely excluding MGD databases from *sui generis* protection, some variants of excluding MGD from the right may require compensation of database producers. The introduction of a defensive right may take away part of the loss suffered as a consequence of the termination of the *sui generis* right. Given the rationale of stimulating data sharing by weakening the rights of database producers, it stands to reason that any defensive right that is introduced is weaker than the *sui generis* right that is taken away. Therefore, it is not certain that the introduction of a defensive right would take away all losses.

4.4.4 Coherence

The option appears overall coherent with the intentions set in the European strategy for data to enable data access and use. There may be concerns that the defensive right would risk overlapping with the protection granted to trade secrets. As was pointed out in the elaboration and legal assessment of policy option 2a in Chapter 3, however, trade secret protection for data is rather hard to maintain. Even if possible, the defensive right would provide an alternative measure of protection while at the same time creating incentives not to keep secrecy but to share the data which would be in the interest of innovation. It could also be regarded as providing an instrument supplementing trade secret protection.

With respect to compatibility with possible future access regimes introduced by the Data Act, the option does not establish exclusive rights but a relative form of protection that does not interfere with access rights of third parties.

4.5 Policy option 3a: inclusion of MGD in the scope of application of the *sui generis*

4.5.1 Effectiveness

While this option would potentially increase legal certainty on the scope of *sui generis* right in MGD, it would add a layer of exclusive right for database makers and could row against the main vision of the European strategy for data and the main objective of the review of the Database Directive to support data sharing within EU and across sectors. As in most cases no new incentives are needed to produce MGD databases, inclusion of MGD would not stimulate more production of MGD databases and might risk overly restricting access and use of MGD, to the detriment of other database makers and users and the general competitive interest in creating innovative products and services. For other database makers it would be more difficult and costly to use data from other databases to add value to their own database users. The fact that often MGD databases could be regarded as source databases, thus being *de facto* data monopolies, their protection may further reduce competition and aggravate possible power imbalances between large data holders and smaller users.

However, we note that the inclusion would in theory still provide an advantage in the form of legal certainty and relatively clear rules for transactions and protection. While the protection may not incentivise further production of MGD, it could be argued that it may encourage database makers to make their data public as the *sui generis* right would protect them from unauthorised use of the data by third parties. This may encourage database makers to contract with users licencing terms for the use of their data thus also promoting data sharing in a way. A large MGD database make user interviewed for the study stated that *sui generis* right would be helpful as a complementary protection tool in the creation and functioning of data exchange platforms as makers would feel more secure, especially small ones with less resources to protect in other ways their databases. By contrast, asked about what would be the benefits of an exclusive right that covers databases containing MGD and that could be exercised against anybody (n=72, multiple choice question), most respondents did not see any benefits of having such a right. While 21 respondents agreed that it offers protection against third-party infringements (i.e. unauthorised use of MGD) which is greatly needed, 17 agreed that it offers protection without triggering unnecessary costs and 17 agreed that it offers the opportunity to better regulate the relationship with clients, including licences.

Moreover, inclusion of MGD in database right would avoid the specification and attribution problems inherent in establishing a new exclusive right in data as such. It could also avoid, to a certain extent, the issue of “mixed databases” (as both type of data, MGD and non-MGD would be protected) which interviewees for this Study have reported to be quite common.

However, if MGD would be included by overcoming the distinction between generation and collections the uncertainties connected to the database right in general would remain and be aggravated due to the delineation problem of what to include in the generation of data.

The effectiveness of this option could be further strengthened by introducing, via the new infringement test explained in policy 2b, a weighting of interests in the infringement analysis. As discussed in policy option 2b, the new test, in this case also applying to MGD, could provide more flexibility to the benefit of competition and innovation.

4.5.2 Efficiency

4.5.2.1 Direct costs

Regulatory charges

No regulatory charges are envisaged by this option.

Substantive compliance costs

Including MGD databases in the scope of the *sui generis* is likely to increase the costs of data access. Almost two-thirds of the respondents to the survey of the present Study (n=35) maintain that their companies will be negatively affected by the inclusion of MGD databases with regard to the cost of accessing data, while 13 out of 35 respondents think that it will ultimately negatively affect their access to data – this aspect is indicated as a negative consequence by more than 50% of the legal experts participating in the survey (n=19). Respondents are also concerned by the negative effects on the cost of protection and on their investment in database creation. The responses on these two aspects, however, were probably influenced by the fact that the question envisaged, along with the inclusion of MGD databases, a new access regime (e.g. exception or compulsory licensing). By having a closer look at the responses of the R&D stakeholders with legal expertise we notice that their opinion follows the pattern of the wider sample: they fear that the inclusion of MGD in the *sui generis* right would have a negative or very negative effect particularly on the access to data, the costs of protection, and the costs of data access.

Administrative costs

No administrative costs should arise from this option.

Enforcement costs

Compared to the current situation, this option does not entail additional enforcement costs since it will simply expand the current regime to all databases, including those containing MGD.

4.5.2.2 Indirect costs

Indirectly, by including MGD databases within the scope of the *sui generis*, this option would broaden the protection for MGD database makers and might hamper access to this type of data. However, this view is not shared by an interviewee representing a big company acting as user-maker. By drawing a parallel with Intellectual Property Rights, such as patents, he explained that a structured common protection for both MGD and non-MGD databases would encourage makers to publish data and thus will positively contribute to data sharing. To him, when databases are open and on the market, one can better chose and negotiate for access.

An argument to support the negative impact on the cost and the possibility of accessing MGD if this option would be implemented is provided by those who view MGD databases typically as spin-off databases. According to them spin-off databases are often single source and may thus be monopolized. Therefore, data-based protection to MGD may risk exacerbating the situation and increase barriers to entry in the primary and secondary market of the sole-source database maker. It must be noted, however, that the spin-off theory has not been formally adopted by the CJEU in its rulings.

4.5.2.3 Direct benefits

Compared to the current situation, the main advantage would be to provide relatively clear rules for transactions and protection. The definition of MGD themselves would create less obstacles in the implementation, since MGD databases will not be differentiated from other databases. Half of the respondents to the survey carried out for this Study (n=35) think that they would be positively or very positively affected by the inclusion of MGD databases (with new access regime) as regard obtaining legal certainty.

More respondents participating in the survey (n=35) think that they would be positively rather than negatively affected by the inclusion of MGD database within the *sui generis*, coupled with a new data access regime, in terms of:

- entering new markets and competing against other firms;
- developing new / innovative products;
- accruing revenues generated from the production and/or exploitation of databases;
- innovation and research activity.

Interviewees supporting the inclusion of MGD databases highlighted:

- the difficulty to differentiate between MGD and non-MGD data, as the majority of databases are “mixed”;
- the investment and necessary human intervention at the basis of MGD databases (as well as other databases);
- the advantages, from a database maker and re-user perspective, of having a structured protection, so that companies have an incentive to open their databases, discuss licensing and negotiate access;
- the benefits, especially for smaller companies, of having a structured protection, that shelter them in the negotiation process.

4.5.2.4 Indirect benefits

Impact on competitiveness, functioning of the internal market and on competition

There are opposite views on the impact of enlarging the scope of the *sui generis* on competition. On the one hand, a great part of the stakeholders consulted as part of the 2018 evaluation addressed the *sui generis* as an obstacle to competition, especially in the context of MGD / IoT databases, since competition would greatly benefit from accessing entire databases. On the other hand, 13 out of 34 participants in the survey (n=34) think that the

inclusion of MGD databases (coupled with an access regime) would have positive effects on their capacity to enter new markets and on competing with other firms.

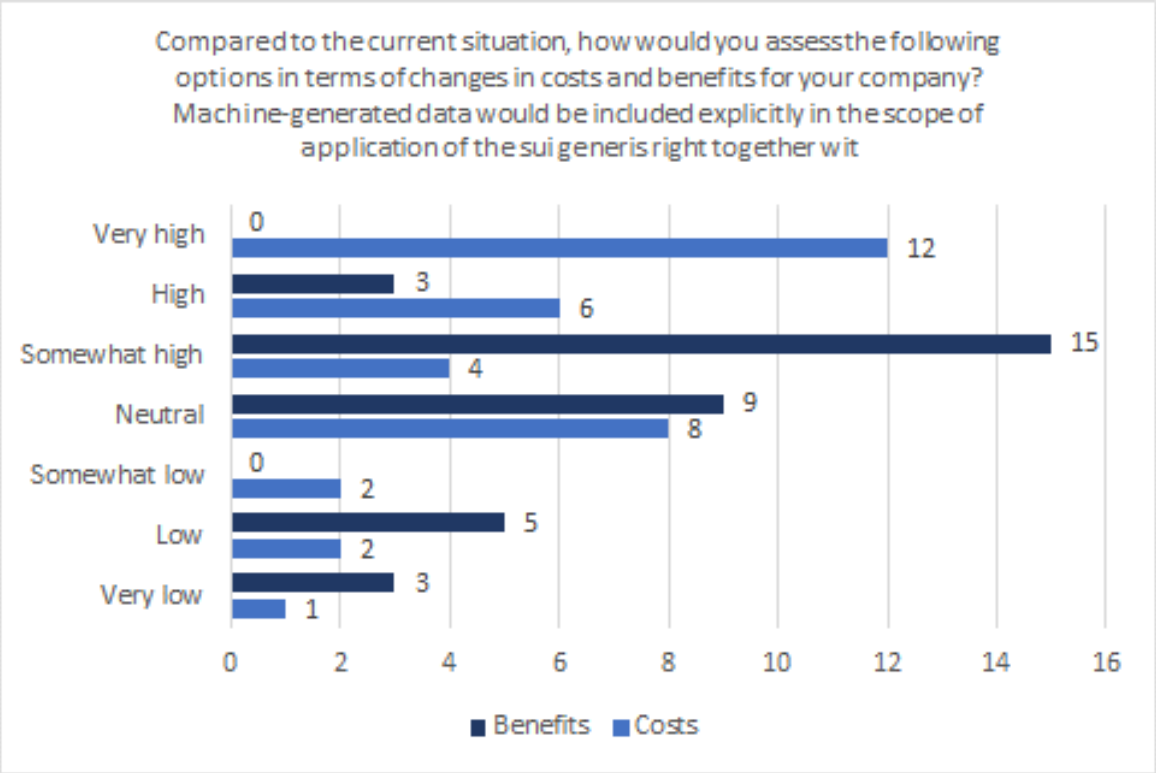
Likewise, a large MGD database maker – user interviewee supporting this option, explained that, from a re-user or intermediary perspective, a structured protection serves to establish a common ground to discuss licencing, compare databases based on their quality and pricing, and chose the best ones. In this way, the inclusion of MGD databases within the scope of the *sui generis* would favour competition. In his view, without a structured protection, database makers would tend to protect their databases as much as possible, leaving less choice for users.

4.5.2.5 Summary on efficiency of the option

The assessment on the efficiency of this option relies on two considerations. First, the costs deriving from this option concern the cost and the possibility itself of accessing data. This option could jeopardise the main objective of the review of the Database Directive (and Data Act initiative), i.e. enable data to flow within the EU and across sectors.

Secondly, apart from the improved legal clarity, there are contrasting views on the benefits brought about by this option. Asked about their views on possible changes to the scope of the *sui generis*, 21% of the respondents to the Open Public Consultation launched for the Data Act (n=276) were in favour of explicitly including MGD databases in the scope of the *sui generis*, whereas 14% of them were in favour of excluding MGD from its scope. However, by far largest group (43%) had no opinion or did not know. The evidence collected by the survey conducted for the present study points to a different outcome. Compared to the current situation, 22 out of 35 respondents think that including MGD in the scope of the *sui generis* (with an access regime) will entail high or very high costs, whereas half of the respondents (n=35) see somewhat high or high benefits as well. The survey suggests therefore that costs would slightly exceed the benefits. Likewise, in the opinion of the legal experts participating in the survey (n=19) this option would bring no or low benefits and somewhat high costs.

Figure 13: Assessment of costs and benefits of including MGD databases within the *sui generis* (with an access regime) compared to the baseline



Source: Survey for the Database Directive Review

4.5.3 Impacts on fundamental rights

4.5.3.1 Impact on the Fundamental Right to Protection of Personal Data

Impact on the Right to Data Portability

As this option explicitly includes MGD in the scope, the impact is comparable to the impact of policy option 0. Indeed, because the second part of the data portability right (transmission of the data to another controller) risks violating the *sui generis* database right, both rights would need to be balanced. As discussed under policy option 0, when a request to port data to another controller conflicts with a *sui generis* database right, the two rights have to be balanced against each other and if the *sui generis* right overrides, the request can be refused.

4.5.3.2 Fundamental right to property and freedom to conduct business

The Right to Property

The third policy option concerns the explicit inclusion of databases with MGD under the *sui generis* right. Currently, there is uncertainty about the applicability of the *sui generis* right to databases with MGD since the cost of the actual machine generation of data may not contribute to the substantiality of investment as interpreted in the *BHB v Hill* case of 2004. Therewith, the policy option would extend the *sui generis* right to MGD databases or retain the status quo (where MGD databases are already protected). Since neither property rights are taken away, nor the exercise and enforcement are limited, there are no impacts under the right to property that need discussion.

4.5.4 Coherence

This option seems to be coherent with other existing forms of protection. Clarifying the application of the SGD to MGD does not give rise to new conflicts in this respect. Even though it may provide protection to more sole source databases compared to the exclusion of MGD this could trigger the application of competition law that systematically sets limits to IP law in general.

Possible conflicts may arise with a future introduction of an access right in the upcoming Data Act. The protection of MGD by the *sui* generis right may pose an impediment to data sharing as right-holders may be more reluctant to provide data. However, this may be alleviated by the horizontal nature of access regimes that will override direct or indirect legal protection applicable to data on a legal level.

4.6 Supplementary option S1: Exceptions to *sui generis* right would be expanded in line with broader general copyright exceptions

4.6.1 Effectiveness

The set of exceptions to the *sui generis* right is much smaller than the set of exceptions to copyright law protecting authorial works in general[236]. For example, Art. 5 InfoSoc contains a large set of exceptions applicable to copyright, many of which do not have an equivalent in *sui generis* law. Since both copyright and *sui generis* right may be applicable to the same database (be it to different aspects of it), the lack of exceptions to *sui generis* right might unduly restrict the use of a database as well as provide legal uncertainty on the scope of protection. This could be felt as a missed opportunity, especially where an exception to the *sui generis* right is missing that has corresponding equivalent in copyright law. Therefore, the policy option aims at achieving the general objective of fostering use and access to data, as the expansion of the menu of exceptions to the currently narrow set of the *sui generis* right will likely increase the ability of certain users to (re)-use databases protected by the *sui generis* right. Moreover, this will also increase the general objective of legal clarity as the larger and established body of case law applied in copyright could be used as reference for parties when dealing with *sui generis* right exceptions. This policy could be effective in contributing to the specific objectives of decreasing legal uncertainties and increase use of data beyond MGD databases. This would be particularly important in the context of big data use cases, as highlighted by Leistner (2020) where data users may need complete databases (not just insubstantial takings as now in the exceptions of the *sui generis* right) from multiple sources in order to employ analytics and create value added services.

During the workshop, legal experts expressed the need to align the exceptions to the InfoSoc Directive, though there was no clear view on how to implement the option, e.g. which individual exceptions to introduce in the *sui generis* right. They agree that it would require detailed research to identify which exceptions should be more applicable to database right and even so it would be difficult to predict how their effect would evolve in the future as new concepts and applications of databases arise. (The current study's focus did not allow for carrying out such robust research). The results of the survey conducted for the present study present neutral results that do not provide straightforward answer to the need for action: 14 respondents agreed or strongly agreed on expanding the exceptions compared to 13 which disagreed or strongly disagreed.

[236] This does not refer to copyright protected databases (i.e. Chapter II of Database Directive).

Finally, we note, also based on discussion during legal expert workshop, that the effect of this option will likely depend on whether combined with option 2b: exclusion of MGD from the scope of the application of the *sui generis* database right and introduction of a more flexible infringement test related to economic detriment. The economic detriment test, if properly applied, may permit extraction and re-utilization of protected databases in cases of e.g. public interest without the need to resort to an exception.

4.6.2 Efficiency

The option will probably entail low direct costs even though this should be caveated until further detailed research on cost is carried out. While administrative and regulatory costs are not existent, compliance costs mainly refer to the possible loss of revenues of database makers, limited to the cases in which the revenues are accrued from a reuse of a database protected by *sui generis* and falling under one of the applicable exceptions under Art 5 of InfoSoc. Considering that database makers not often rely on the *sui generis* right,[237] these costs should be very limited. As for possible enforcement costs, considering the aim of the option (i.e. aligning the exceptions to the *sui generis* with the traditional copyright rules), its implementation should not require much work from the Member States, as they can simply expand the existing national rules on copyright to the *sui generis*.

By leaving discretion to the Members States on choosing the exceptions to adopt (under the condition that they are equally applicable to copyright and *sui generis*), this option will be in line with the baseline and will not contribute to further harmonisation in the Digital Single Market. Should some or all limitations be introduced as mandatory, the Digital Single Market would benefit from more harmonisation compared to the existing situation (baseline), in which copyright exceptions applicable to databases depend on national law.

Nonetheless, aligning the exceptions in the database right to those of the general copyright right would:

- Increase clarity, as a unitary system of specific exemptions could be established all over the protection for material in copyright and neighbouring rights.
- Avoid risk of regulatory arbitrage deriving from the opportunistic use of the *sui generis* to avoid exceptions granted under traditional (national) copyright rules.[238]
- Improve balance between *sui generis* right and the public interest of extracting (a substantial part of) data from the database.[239]

Moreover, considering the criticisms expressed in the 2018 evaluation by users and user-makers (especially SMEs) and legal experts on the way the *sui generis* right has been hampering competition, the expansion of exceptions is likely to have positive repercussions on the competition of sectors affected by the exceptions (e.g. publishing). As a matter of fact, the alignment with the provisions of the InfoSoc Directive will remove some conditions to the exceptions that had undermined their usage and effectiveness, in particular with regard to,

---

[237] See findings from the 2018 evaluation: opinion of members of the IP Federation Data Committee

[238] Hugenholtz (2016): Something Completely Different: Europe's Sui generis Database Right, in Frankel, Gervais (2016): The Internet and the Emerging Importance of New Forms of Intellectual Property

[239] Evaluation 2018, p. 59

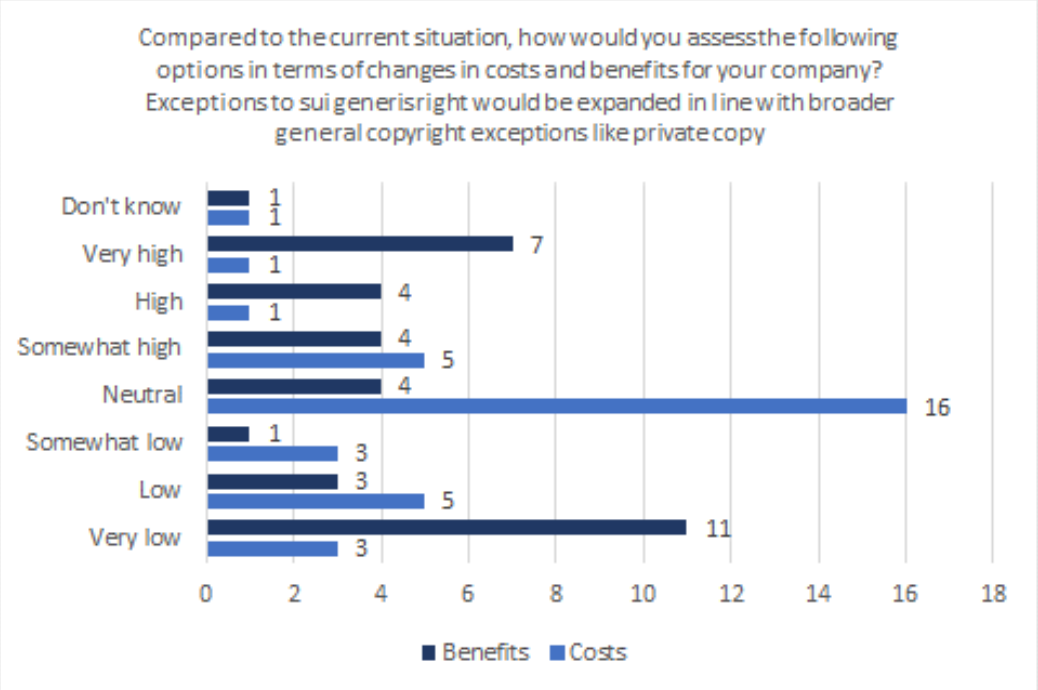
first, the fact that exceptions of Art. 8 are limited to “lawful users”; and, secondly, the possibility of copy for private purposes only for non-electronic databases.

The introduction of an exception for information purposes – copy of a published database from the press or media – as well as for quotation (including through hyperlink) can bring positive societal impacts. These exceptions might indeed contribute to an ecosystem in which citizens, who are constantly exposed to news and fake news, can easily check facts.

In conclusion, although the Evaluation 2018 did find strong evidence pointing to regulatory burdens stemming from the misalignment of the InfoSoc and the Database Directive, compared to the current situation (baseline), the efficiency of this sub-option can be positively assessed considering that the additional costs would be minimal, whereas it would bring benefits in terms of increased clarity and uniformity of rules between exceptions applicable to copyright and *sui generis* right, improved balance between *sui generis* right and the public interest of extracting (a substantial part of) data from the database.

According to the respondents of the survey (n=35) the benefits of expanding the exceptions to *sui generis* right in line with broader general copyright exceptions would exceed the costs, which instead, are considered on average in line with the baseline. According to the legal experts participating in the survey (n=19), this is the supplementary option with the associated lowest costs but considerable benefits.

Figure 14: Assessment of costs and benefits of including MGD databases within the *sui generis* (with an access regime) compared to the baseline



Source: Survey for the Database Directive Review

4.6.3 Impacts on fundamental rights

4.6.3.1 Fundamental right to protection of personal data

Impact on the right to data portability

Extending scope of exceptions and limitations of the *sui generis* database right applied to all databases will facilitate the exercise of the data portability right by the data subjects in relation to any type of personal data as it will limit the scope of the *sui generis* right potentially conflicting with the portability right. To illustrate, an exception permitting for private copying on any medium will eliminate any potential of restricting the right of a data subject to receive the personal data which he or she has provided to a controller, in a structured, commonly used and machine-readable format (the first element of the right to data portability), even though currently this right also does not seem to be affected significantly by the *sui generis* database right (see analysis of the baseline). Extending the exceptions from the *sui generis* right to the taking of the whole database rather than of substantial parts, as it stands now, will facilitate the second element of the portability right. That is to transmit or have the data transmitted to another controller directly which is now more seriously obstructed in case of the multiple portability requests by the data subjects organized or facilitated by subsequent controllers aiming to attract large numbers of data subjects and port their data where the resulting transfers are of a substantial part of the original controller's database and/or systematic.

4.6.3.2 *Fundamental right to property and freedom to conduct business*

The Right to Property

Under option 4a, exceptions to *sui generis* right would be expanded in line with broader general copyright exceptions. This has a potential impact on the enjoyment of a *sui generis* right, since a rights holder may have to accept certain uses of a database that pre-implementation of this option might have been under his exclusive legal control. This does not amount to a deprivation of a property right. The approach is to extend all copyright exceptions to the *sui generis* right. Most of them, may not be very useful for databases and have no impact on the right of property. In Chapter 3, a few copyright exceptions have been singled out that may have an impact. Under the respective headings, "context", the general interest served with extending the exceptions to the *sui generis* right have been stated. Under the respective headings, "conditions", limitations have been introduced that circumscribe the exceptions in order to limit their impact on the *sui generis* right to that which is strictly needed to serve the stated general interests.

4.6.4 Coherence

The policy will be coherent with the vision of the European strategy for data to increase access and use of data within EU and across sectors. The policy is in line with harmonising the legal framework for copyright in the Digital Single Market. Coherence and alignment to be established with Copyright Law. However, the mode in which the policy will be implemented, i.e. whether making exceptions mandatory or leave their adoption to Member States will leave some uncertainty on the alignment with InfoSoc.

4.7 Supplementary option S2: Strengthening of research exception of the database right

4.7.1 Effectiveness

The starting point ought to be the TDM exception for research that was passed by the EU legislator in the DSM Directive in 2019 that is currently being transposed by Member States. It seems advisable that any new exception to the Database Directive should consider this major development and be compatible with it, especially until the TDM exception is fully transposed and operational.

This option proposes the following potential areas to consider for improving the research exception: a) consider making mandatory the exception, b) could extend permitted acts to include re-utilisation, c) could cover also e-research infrastructures and d) could overcome the criteria of "non-commercial" purposes. The proposed exception could be effective in promoting research and innovation with public interest to the benefit of society.

This supplementary option would in particular make the research exception in the Database Directive mandatory to be adopted in national legislations (similar to the recently introduced TDM exception which applies to all Member States). This will improve harmonization across Member States compared to the current status quo in the baseline option where, as evidenced by prior studies[240] , every Member States may interpret the Database Directive differently (e.g. granting the exception to certain types of organizations and not others). The harmonization would foster "legal interoperability"[241] between different databases and data infrastructures across EU, thus help fostering the dissemination and use of data for research purposes across the EU single market to the benefit of EU economy and society.

As explained above and in the legal assessment, the current research exception is seen as an impediment to research in an electronic environment. This could be improved upon by expanding the permitted acts to include also re-utilisation of the material for scientific purposes to allow documents to be reproduced and re-used for e.g. teaching or further research. This would enlarge possibilities to access, use and exchange data for scientific research and innovation.

Furthermore, as described in the efficiency section, a revision of the exception that overcome the restriction of "non-commercial purposes" from a lighter criterion of e.g. no economic harm, would allow research institutions and projects partially founded by commercial companies to access and use data for their research without the risk of infringing the *sui generis right.* This could effectively be achieved also by limiting the exception to the entity, i.e. research institution, rather than the purpose.

4.7.2 Efficiency

Representatives of the research community, consulted as part of the 2018 evaluation, indicated that the strict rules on the re-using of data for research purposes negatively affects their activities. They explained that due to the *sui generis* right, opening data entails a higher cost than it should have due to the incompatibility of database licences. As a matter of fact, researchers who make intensive use of databases need to reach many different contractual

[240] See Guibalt, L. and Wiebe, A. (ed) (2013): Save to be Open

[241] See Guibalt, L. and Wiebe, A. (ed) (2013), p.12

agreements, which represents a high cost for the research institution. Also, according to the 2018 evaluation, *sui generis* right may create friction to achieve innovations with societal impact since it can create obstacles to collaborations and data exchange between research organisations, citizens, researchers outside academia, SMEs, spin-off etc.

Commenting on the current exception for research purposes to (i) extraction (as opposed to reutilisation) and (ii) non-commercial research, a library representative, LIBER, explained: "First, limiting this exception to extraction of data creates issues for researchers as they also need to be able to re-use data, even if only to analyse it, in order to conduct meaningful research. Second, limiting the exception for non-commercial use goes against the public interest. This limits any potential knowledge transfer, especially seeing that research activities are nowadays more and more conducted in the context of public private partnerships. Moreover, this limitation fails to recognise the fact that research has also become more reliant on private funding due to budget cuts. These restrictions thus render this exception useless in practice and negatively affect the re-use of data." COMMUNIA raised similar points.

The terminology of the current exception and criticisms thereof are also discussed in the literature. For instance, Derclaye[242] argues that the limitation to extraction renders the exception virtually unusable: "The corresponding exception in the sui generis right chapter is therefore far more restricted and in effect quasi unusable since to teach and research one almost always has to communicate to the public."[243] In fact, some Member States have implemented it in different ways or opted not to transpose this optional exception.

Compared to the current situation, the strengthening of the exception could improve the effectiveness of the exception. This could possibly have positive societal impacts as they will ensure that substantial parts or entire databases can be extracted and re-used by a larger cohort of research actors, thus stimulating the cooperation among them. The extension of permitted acts could favour the establishment and functioning of cross-border electronic research infrastructure (such as OpenAIRE) and possibly avoid a too heavy burden on right holders by leveraging a "do not harm" principle introduced to EU law by the recent *CV v Melon* judgment. This is of course, conditional on the uptake and influence of the "do not harm" principle derived from the *CV v Melon* judgment will exert on the practice, which is still too early to tell. While extension to commercial uses may possibly avoid difficult problems of delineation and thus reduce transaction costs, the introduction of the "do not harm" principle may limit the exception and translate into a burden for the database makers to demonstrate that the extraction and reutilisation of the database prevent them to redeem the investment (in line with the *CV v Melon* judgment). This appears to strike a proper balance in the interest of innovation. The positive effects of a mandatory extension for extraction and re-use of databases scientific research were already suggested in the 2018 evaluation. In the context of the OPC-Data Act, two respondents representing a wide range of stakeholders supporting open science repeat that the teaching and scientific research exceptions should cover the acts of re-utilization and the private use exception in Article 7(a) should not be limited to non-electronic databases.

[242] 'The Database Directive' in Stamatoudi and Torremans, EU Copyright Law: A Commentary (ed), 2014.

[243] Ibid.. Similarly critical assessments in the literature: Guibault and Wiebe (2013), Reichman (2002).

4.7.3 Impacts on fundamental rights

4.7.3.1 Fundamental right to protection of personal data

Impact on the right to data portability

The impact is comparable to the impact of supplementary option S1.

4.7.3.2 Fundamental right to protection of personal data

The Right to Property

An extended research exception may have a potential impact on the enjoyment of a *sui generis* right, since a rights holder may have to accept certain uses of a database that pre-implementation of this option might have been under his exclusive legal control. In circumscribing the new exception, several limitations such as an economic harm test, or a limitation of beneficiaries to research institutions, may balance the negative impact on rightsholders.

4.7.4 Coherence

The proposed exception is coherent with the general copyright exceptions but transcend the limitation in Art. 5(3)(a) InfoSoc on non-commercial use. However, the InfoSoc Directive does not apply to databases. The expanded research exception would not modify but rather complement the exception for TDM in the DSM Directive.

4.8 Supplementary option S3: Public bodies would be excluded from the *sui generis* right

4.8.1 Effectiveness

The analysis of this supplementary option should recognise the level of harmonisation achieved by the Open Data Directive (ODD) (and Data Governance Act), which for many occurrences resolved the tension between access and database protection for public sector databases. Therefore, this option should be analysed through the prism of a significantly narrower problem than what existed prior to the Open Data Directive.

Currently, the database directive is indifferent to the private or public nature of the producer of a database with fragmentation at national level. Excluding public bodies from the *sui generis* right would make clear that they do not hold *sui generis* rights. This takes away a potential source of uncertainty regarding the permissibility of using the contents of the databases held by public bodies. This policy will, therefore, contribute to the general objective of increasing legal clarity of the Database Directive. Furthermore, the policy aims at achieving the general objective of fostering data access and usage since, without *sui generis* rights protection, users will likely better able to (re)-use databases provided by public bodies (or private bodies with public tasks). The use of this data may serve to support innovation and the development of new products and services. This policy will contribute to achieving the specific objective of addressing legal uncertainties not linked specifically with MGD and data in the IoT environment. Legal experts participating to the workshop supported unanimously the adoption of this option. The degree of effectiveness will depend on how the policy will be implemented as explained in the efficiency section below.

4.8.2 Efficiency

The Evaluation 2018, that was carried out prior to the adoption of the ODD (former the PSI Directive), observes that many respondents to the consultation — including users and research

bodies — "think that [the] *sui generis* right clashes with the PSI directive … [and] that publicly-funded databases should be excluded from the *sui generis* right protection as official works under the copyright regime"[244]. The evaluation then concludes: "There is strong evidence that there is no coherence, a clash or no clarity or uncertainty as regards the relationship between the Database Directive or at least the *sui generis* right and the PSI directive and open access policies. The *sui generis* right is seen by many as a barrier to innovation and knowledge exchange and thus to economic growth as research and public data cannot be reused either at all (if refusal to license), or less fast or at a greater cost". [245] On the other hand, one big company stressed the point that universities may rely on income from the *sui* generis right and exclusion of public bodies may devaluate universities, also as a business partner.

From the time of the Evaluation 2018, the PSI Directive has been replaced by the ODD[246], which includes a provision specifying that the *sui generis* right cannot be exercised to override the rules contained in the ODD (Art. 1.6). This policy option could either align the Database Directive with the ODD provisions, by adding an exception reflecting the scope of the ODD (i.e. that applies to databases held by public bodies and containing data and documents listed in Art. 1 ODD), or expand the scope of the exception envisaged in the ODD to databases erected outside the scope of the public task of the public body and including not only "re-use" but also "access" to databases.

Moreover, a simple and clear exclusionary rule placed in the Database Directive would enhance legal certainty for public bodies as well as the private sector, reduce existing uncertainty, and, moreover, support transparency by regulating in the appropriate systematic context.

Considering that, in practice, this option would not create new requirements for stakeholders, the costs associated with it are almost null. Since the option would simply expand to the *sui generis* what is already envisaged in general copyright rules and in the ODD, it is reasonable to expect very low enforcement costs from the transposition of the new exception in the Member States' legislation and its application.

By contrast, this sub-option will bring significant efficiency gains compared to the baseline by reducing the number of disputes arising on possible inconsistencies and the related transaction costs. [247] As a consequence of the increased certainty, more users are likely to use databases owned by public bodies. According to the Evaluation 2018, one-third of potential users refrain from using a database when they are uncertain as to whether a database is protected or not.[248] Indirectly, the Database Directive will contribute to the objectives of the ODD, namely: to promote the use of public sector data and stimulate innovation in products and services based on public information. 65 per cent of users responding to the survey conducted for the Evaluation 2018 have indeed declared that they rely on public sector databases. Coherently,

---

[244] Evaluation 2018, p.120.

[245] Evaluation 2018, p.121.

[246] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)

[247] Van Eechoud (2021): A Serpent Eating Its Tail: The Database Directive Meets the Open Data Directive. Springer. See also case law on pag. 119 of 2018 Evaluation and Leistner (2020).

[248] Evaluation Study 2018, p. 68.

the 2005 evaluation had found that there was an "increasing demand for consumer access to information contained in databases owned by public bodies, such as weather data, maps and statutory registers".

Finally, excluding public-sector databases from the *sui generis* right would ease access to markets. Literature and case law have indeed reported abuses from public bodies (e.g. public trade registries) that claim database right to create market power and protect revenue streams from (semi)commercially exploited public sector databases. This danger is still not excluded with passing of the ODD as public bodies may still use existing database rights in neglect of the obligation to keep within the limitations of the ODD. This could change where a clear rule for exclusion is implemented and communicated.

The possibility of treating databases held by public authorities differently than other type of databases under the Database Directive was positively welcomed by more than half of the respondents participating in the OPC-Data Act (n=287). Among those against the exclusion of public bodies from ownership of *sui generis* right, there are some universities. Although universities generally aim at open research infrastructures, some of them believe there are several juridical and ethical grounds that may restrict free use and re-use of research data. Therefore, databases held by public authorities like universities should retain their rights to control the use of their data.

In conclusion, compared to the current situation (baseline), the efficiency of this sub-option can be positively assessed considering that it brings almost no additional costs and, on the contrary, is likely to reduce transaction costs for the stakeholders and have wider positive effects on the Data Economy. Nevertheless, it is worth noting that the Evaluation 2018[249] pointed out the risks linked to the exclusion from the *sui generis* of public bodies that are self-funded, i.e. not funded by the state but relying on the *sui generis* right for their very existence. This option could cause a loss of revenues for them that might have repercussions on the future availability of their databases. This risk might materialise especially if the option does not explicitly exclude private organisations active in competitive environments.

4.8.3 Impacts on fundamental rights

4.8.3.1 Fundamental right to protection of personal data

Impact on the right to data portability

The impact of this option on the exercise of the right to data portability would be limited, because public bodies seldom process personal data on the grounds of consent or contract and hence the data they process often will not fall into the scope of the data portability right anyway. Given that in various Member States applications of the grounds of contract and interpretation of what a public body is can be different, to the extent the right to data portability does apply, the impact of this option on it will be positive and comparable to the impact of option 2a but magnified to all databases regardless of whether MGD is processed.

[249] Evaluation Study 2018, p. 121.

4.8.3.2 Fundamental right to property and freedom to conduct business

The Right to Property

It appears that public entities only in a very limited set of circumstances can benefit from the rights enshrined in the ECHR and the Charter of Fundamental Rights (CFR). According to Art. 34 ECHR, governmental organisations have no standing and cannot bring a case to the European Court for Human Rights. The concept "governmental organisations" does comprise – apart from central government - also many lower governmental organisations, such as municipalities.[250] The concept also comprises legal entities which participate in the exercise of governmental powers or run a public service under government control. The Charter of Fundamental Rights of the EU does allow public entities under certain very limited circumstances to benefit from the Charter.[251] It would need further analysis to determine whether there are relevant entities that could benefit from the rights enshrined in the ECHR and the Charter of Fundamental Rights.

4.8.4 Coherence

This option is coherent with the clear goal to stimulate and allow access to and use of databases controlled by public bodies in the Government to Business (G2B) data sharing framework. More specifically, the option is in line with the objectives of ODD, the Data Governance Act and the ODD. Uncertainties could arise with regard to the different scope of the exclusion proposed here and Art. 1(6) ODD. However, both approaches have the same goal and no conflicts in practice could be expected with different results of application. Moreover, regulating from two sides of the coin is not uncommon in IP law. For instance, with respect to competition law, respective policy decisions are sometimes regulated in IP law as well. One example for this would be Art. 6 of the Computer Program Directive that has a competition law background.

[250] European Court of Human Rights (2021): Practical Guide on Admissibility Criteria.

[251] Julicher et al (2019: Protection of the EU Charter for Private Legal Entities and Public Authorities? The Personal Scope of Fundamental Rights within Europe Compared. Utrecht Law Review.

5 Comparison of policy options

The following section presents a comparison of the policy options related to MGD in light of the assessment undertaken in Chapter 4.

The comparison will be limited to the options related to MGD: policy options 2a, 2b, 2c and 3a. By contrast, the supplementary options, dealing with different topics that are not the main focus of this review and are not mutually exclusive, thus can only be assessed with respect to the baseline option of no action as already done in Chapter 4.

The options related to MGD are compared against the three main criteria explored in Chapter 4: 1) effectiveness in achieving the policy objectives, 2) efficiency in terms of costs and benefits from the option, and 3) coherence of the option with existing policy initiatives and legal frameworks.

Table 2 provides a summary of the performance of the options against the four criteria. For each policy and criteria, a qualitative score is assigned which summarise the performance of the option against the baseline. The score in the individual criteria goes from "+++" to "---": "+++" signifies higher positive performance than the baseline, "++" moderate positive performance than the baseline, "+" slight positive performance than the baseline, "+/-" neutral, "-" slight negative performance than the baseline, , "—" moderate negative performance than the baseline, and "---" signifies higher negative performance than the baseline. The total is the net of the scores over the three criteria and can exceed the three symbols. For the efficiency criteria the different signs of the scores should be read as benefits and costs.

Table 2: Overall impact of the policy options related to MGD

Criteria	2a: exclusion of MGD from the *sui generis* right	2b: exclusion of MGD from the *sui generis* right and introduction of flexible infringement test	2c: exclusion of MGD from the *sui generis* right and introduction of an alternative control mechanism	3a: inclusion of MGD in the *sui generis* right
Effectiveness	++	++	+	+/-
Efficiency	+/-	-	-	-
Coherence	++	++	+/-	+/-
Total	++++	+++	+/-	-

Note: Impact of the various policy options. The symbols "+" and "-" indicate respectively positive and negative performance compared to the baseline. The number of symbols is the summary of individual ratings and indicates the size of the change: "+++,---" high, "++,--" moderate, "+,-" slight, "+/-" neutral. The total is the sum (net) of the three criteria.

Based on the qualitative and quantitative assessment carried out in Chapter 4, the sub-option 2a excluding MGD appears to be the best choice based on the evidence gathered, followed by policy option 2b then 2c and last 3a. Below we elaborate on the comparison of the options on each criterion.

5.1 Effectiveness

The effects strived for are legal clarity and achieving a higher level of data sharing.

Legal clarity

First, an important source of lack of clarity derives from literature in which the applicability of *sui generis* right to databases with MGD is contested. In terms of effectiveness, all four policy options could be seen as effective in bringing clarity to the relation of Database Directive with MGD, improving the current situation in the baseline. However, in the process new sources of uncertainty would be introduced (e.g. what is the exact definition of MGD? What is a database with MGD? What is an average substantial investment? What is the wrongfulness criteria for the defensive right?). This seems to cloud the clarity effect somewhat. This is mainly relevant for option 2.

Other points of uncertainty that have been subject to discussion would be acknowledged. This included the uncertain distinction between substantial and insubstantial taking that is proposed to be abolished in option 2b.

Stimulation of data sharing

Option 2 and the option 3 start from diametrically opposed approaches to the stimulation of data sharing. The option 2 is based on the idea that users would be stimulated to use MGD because the *sui generis* right would not be applicable to databases with MGD. A barrier is taken away. Option 3 is based on the idea that an exclusive right is a stimulus to create databases in the first place, that reliance of database producers on an exclusive right is creating less barriers to data sharing than reliance on contract and technology for protection of the database, and that an exclusive right stimulates sharing because it gives database producers a means to act against data leaks (which may occur more frequently if data are shared amongst more parties). Conventional theories of intellectual property law aim at finding the right balance between stimulus for creation through protection and access and distribution. Academic literature so far suggests that for data the stimulus through protection may be not (or little) needed as data will also be produced without it, while the possibility to give access to data (conductive to innovation) is of superior importance. This would be equally true for database rights that provide at least an indirect protection for data.

In accordance with that, although the evidence base is thin, the arguments underlying option 3 appear relatively weak. It is factually doubtful that a *sui generis* right stimulates creation of databases. Contract and technology are now already used next to the *sui generis* right and seen as important protection instruments. Whether the *sui generis* right would be missed as means to act against data leaks (if one of the options 2 is chosen) is difficult to prove because this can be evaluated only in the course of future practice.

How do legal clarity and the stimulation of data sharing relate to each other?

If strong rights are given to database producers, the consequences of lack of legal clarity will be mainly borne by data users as the weaker party. If rights are taken away from database maker, the effects of legal unclarity will be borne to a larger extent by the database maker as the (now) relatively weak party. In our analysis of stimulation of data sharing, option 2 appears to perform better. Given that option 2 strengthens the position of data users this would somewhat mitigate the effects of unclarity of concepts like MGD or average substantial investment at least for data users. It is mainly the uncertainty under data users in the status quo that is perceived as an obstacle to data sharing.

With regards to the extra measures of options 2b and 2c, in option 2b the position of the data user is further strengthened compared to 2a. The economic detriment test would of course only apply to other databases rather than MGD databases. However, since mixed databases may be difficult to categorize the detriment test may still have a positive effect on the sharing of MGD databases, it strengthens the position of data users in situations in which the status of a database is not readily apparent. Yet, the effect would be reduced if mixed databases would be by default excluded from protection. In option 2c, the effect of removing *sui generis* protection for MGD databases is weakened (or offset) by the introduction of a new control mechanism. This option may be considered if database makers prove to need some means to act against data leaks. However, since the position of the database maker would still be relatively strong the effects of legal unclarity would again come to rest more on the shoulders of data users. This weakens the appeal of this policy option.

Overall, policy options 2a and 2b are expected to rank first against the effectiveness criterium with a moderate positive performance compared to the baseline. Following, there is option 2c (ranked third) which would still be slightly more effective than the baseline, while the inclusion of MGD would be neutral compared to no action.

5.2 Efficiency

In the current baseline scenario, the uncertainty surrounding the relation between the *sui generis* right and MGD may reduce the benefits that could be derived from increased access to MGD databases and, at the same time, increase transaction costs for users to understand whether a database is protected by the right, this may also increase legal costs related with contractual agreements.

Table 3 provides a breakdown of the scores between direct and indirect benefits and costs for the efficiency criteria. The table should be read as the overall table above except that the different signs of the scores represent benefits and costs.

Table 3: Efficiency impact of the policy options related to MGD

Cost/Benefits	2a: exclusion of MGD from the *sui generis* right	2b: exclusion of MGD from the *sui generis* right and introduction of flexible infringement test	2c: exclusion of MGD from the *sui generis* right and introduction of an alternative control mechanism	3a: inclusion of MGD in the *sui generis* right
Direct cost	-	---	-	-
Indirect costs	--	---	---	--
Direct benefits	++	+++	++	++
Indirect benefits	+	++	+	+/-
Total	+/-	-	-	-

Note: Impact of the various policy options. The symbols "+" and "-" indicate respectively benefits and costs performance compared to the baseline. The number of symbols is the summary of individual ratings and indicates the size of the change: "+++,---" high, "++,--" moderate, "+,-" slight, "+/-" neutral. The total is the sum (net) of the three criteria.

In policy option 2a, the exclusion of MGD from the scope of the *sui generis* right is not expected to generate significant additional added costs compared to the current situation. Enforcement costs may arise if a clear and stable definition of MGD and/or modifications in legal requirements cannot be found, However, those are expected to be more than offset by the direct benefits coming from lower transaction costs in the form of lower risk of opportunistic litigations brought forward by MGD database makers on the basis unfounded exclusive rights claims. Policy option 2a could generate indirect costs compared to the baseline in the form of i) lack of protection for MGD database makers against unauthorised use of their database by third parties, ii) inability for users to benefit from exception and limitations currently present in the Directive, iii) greater ability for large data holder to rely on contracts to exercise their power over data users. The indirect costs would be partially compensated by the indirect benefits derived from increase competition and innovation from broader access and use of data. Overall, the net sum of costs and benefits of policy 2a would be more or less equal to the current situation of no action.

In policy option 2b, the balance of costs and benefits related to MGD databases will be similar to 2a. For other non MGD databases, the additional introduction of an infringement test inspired by the recent ruling in the *CV v Melon case,* could lead to higher direct and indirect benefits by narrowing the scope of protection and promoting innovation. Yet, the uncertainty on its application increases the expected costs associated to it, making the option slightly less efficient than the baseline and policy 2a.

In policy option 2c, the exclusion of MGD from the *sui generis* right and the introduction of a lighter right to protect database makers from misuse of their data from third parties would reduce the indirect cost (i) highlighted in policy 2a. At the same time, it would incentivise database makers to share their data with the implied indirect economic and societal benefits derived from increased availability of data in the economy. Nevertheless, the risk of the new right becoming an additional layer of protection reduces the benefit of this option. Moreover, the introduction of a completely new type of right only for MGD database is riddled with uncertainties on its implementation, rising the expected enforcement costs and indirect costs. Therefore, the policy option 2c with its additional layer of uncertainty, is expected to be slightly less efficient than the baseline.

In policy option 3a, the inclusion of MGD is expected to generate less implementation costs than the other policies as it requires a smaller change in the existing Directive. Though it risks increasing direct costs of access to data for users compared to the other policies analysed, thus reducing ability to access data and the indirect economic benefits derived from it. These effects may be accentuated by the sole-source nature of part of MGD databases, which make them hard to be substituted with other sources. This could be counterbalanced by benefits derived from more willingness of makers to sell and/or licence their data to the public compared with option 2a and 2b where the lack of protection would expose them to third parties' misuses. Overall, however, the efficiency of the option is expected to be only slightly negative compared to the baseline.

To summarize, the most efficient policy option, based on the evidence gathered, would be 2a, the simple exclusion of MGD - though the net balance of costs and benefits would be just equal to the current baseline. All the other remaining options entail, for different reasons, slightly more costs than benefits compared to the status quo.

5.3 Coherence

Coherence is assessed with an emphasis on the effect on existing forms of protection as well as the compatibility with a future Data Act aiming at improving data sharing.

Options 2a and 2b would not create any conflicts with other legal instruments for protecting data directly or indirectly as they explicitly exclude the application of *sui generis* right to MGD. Policy option 2c possibly could interfere with protection granted by trade secrets by overlapping to a certain degree. However, this does not appear to be a major problem assuming that the relevance of trade secret law in a networked data economy will be decreasing. Policy option 3a also seems to be coherent with other existing forms of protection as it would only clarify the application of the already *existing sui generis* right to MGD, however policy option 3a may not be coherent with the original objective of the Database Directive insofar investments directed directly to the raw MGD data as such are protected rather than investments in the database per se.

Possible conflicts may arise in case an access or usage right would be proposed in the upcoming Data Act, mostly with respect to option 3a. The protection of MGD by *the sui* generis right may pose an impediment to data sharing as right-holders may be more reluctant to provide data or appeal to the exclusive right.

Policy option 2c would not establish exclusive rights and thus seems to establish no obstacle to access rights and data sharing, however, the risk for the new right to work as an additional layer of protection, as voiced by academics, remains present. By contrast, policy options 2a and 2b, appear better tailored to avoid the *sui generis* right to become an obstacle for data sharing and newly introduced access regimes as they would exclude protection of MGD.

Overall, policy option 2a and 2b seem more coherent than the other two options in light of the two main areas of conflict emphasized here.

5.4 Refit Cost Savings

In line with the Better Regulation Toolbox, the REFIT analysis is reported in what follows. In order to assess the extent to which the option ranked first (option 2a) would improve simplification and reduce the regulatory costs stemming from the Database Directive, it should be noted that the Evaluation 2018 found very low costs linked to the Directive. The Directive does not require regulatory charges nor administrative costs. At the same time, the majority of database makers, user-makers and users participating in the survey carried out for the purposes of the Evaluation 2018 declared that they had incurred no or low compliance or enforcement costs. The main costs reported by the Evaluation 2018 were linked to the legal uncertainty surrounding the databases containing MGD, from which litigation and transaction costs could arise.

The table below outlines the main cost savings deriving from the option 2a compared to the baseline. Due to the general lack of data and unwillingness of stakeholders of sharing this type of information, it is not possible to quantify the cost savings.

Table 4: Refit cost savings

REFIT COST SAVINGS – HIGHEST RANKING OPTION		
DESCRIPTION	**AMOUNT**	**COMMENTS**
Reduced litigation costs The option would bring legal clarity by providing a clear and stable definition of MGD or explicit criteria for exclusion of MGD databases from the *sui generis*.	Quantitative estimates cannot be established	Affected stakeholders: Database makers and users
Reduced information and transaction costs Excluding MGD databases and making use of contract networks would have the potential effect to efficiently assign the property right to the person which is most central to the investment effort, reducing the number of possible opportunistic litigations and the linked transaction cost by stripping away the possibility of spurious baseless litigations of third-party data use	Quantitative estimates cannot be established	Affected stakeholders: Database makers and users

Note: Estimates are with respect to the baseline of the unchanged legislation.

6 Conclusions

Technological developments continue to change society and sustain growth in an increasingly knowledge-based economy. A key role in this transformation is played by data and the endless insights that can be derived from it to answer the current challenges of our time. The EU aims at becoming a leader in the data economy by stimulating the availability of data for use and reuse to unlock the full potential of data-driven innovation. Fostering data flow within EU and across sectors has become a policy priority for the EU.

Particular attention is placed on data generated/collected by sensors and machines in the IoT environment which are increasingly adopted across many sectors. Under the Data Act initiative, the EU aims, among other things, to increase fairness and provide clear rules for business to business (B2B) sharing and access to this type of data. This is seen as a necessary condition to foster data sharing and create more competitive markets, in general. As part of the Data Act initiative, a review of the Database Directive was also announced.

The Database Directive aims to promote the production of databases in EU by granting exclusive right, i.e. *sui generis* right, to database-makers that undertook a substantial investment in obtaining, verifying, or presenting the contents of a database. Currently, there is legal uncertainty around the scope of the Directive, in particular in relation to machine-generated data. Therefore, some aspects of its application are seen as not in line with recent technological developments in the data economy. While currently underused, the legal uncertainty of the *sui generis* right may create frictions in the market for data with the risk of limiting the access and use of MGD.

The study presents evidence on the impact of possible options to review the Database Directive, specifically the *sui generis* right, and ensure the instrument is fit for the data economy. One of the main objectives of the review is to clarify the scope of the Directive in relation to databases containing MGD and avoid that it becomes an obstacle for the sharing and trading of MGD.

Several policy options have been elaborated and assessed based on information gathered through extensive desk research, literature review, a stakeholder consultation undertaken specifically for the present study between June and August 2021 as well as the results of the Open Public Consultation on the Data Act. [252] The stakeholder consultation consisted of:

- an online survey that resulted in 114 respondents across multiple countries active in:
  - sectors that are either using already substantially IoT or are likely to do so in the near future
  - legal professional services
  - research organisations
- five in-depth interviews with business and business organization dealing with MGD
- one workshop with legal experts in the field.

It should be noted that, due to the current limited awareness and use of the Database Directive among MGD stakeholders, the evidence collected during the consultation is limited and mostly

[252] The Open Public Consultation was not integral part of this support study but rather a complementary source.

qualitative. Thus, the assessment of the policy options relies mainly on views of legal experts in industry, research, and academia as well as legal practitioners.

The study initially defined a list of 13 options/sub-options for the review of the Directive, both in relation to MGD and to all protected databases.

After the scoping phase, four policy options related to MGD have been further developed and analysed while two were not further explored[253] as they would entail forms of data access and/or usage regimes potentially overlapping with measures that may be developed under the Data Act.

The four policy options related to MGD retained are:

- Policy option 2a: exclusion of MGD from the scope of the *sui generis* right
- Policy option 2b: exclusion of MGD from the scope of the *sui generis* right and substitution of the current infringement test with a more flexible test related to economic detriment
- Policy option 2c: Exclusion of MGD from the scope of the *sui generis* right and introduction instead of an alternative control mechanism applied only to MGD databases
- Policy option 3a: inclusion of MGD in the scope of the *sui generis* right.

A legal analysis explored various legal means for the implementation of the policies. highlighting the challenges present in various approaches. Approaches investigated included:

a) direct exclusion/inclusion of MGD via statutory definition,

b) indirect exclusion/inclusion of MGD by distinguishing between investments made in the creation vs the collection of MGD and

c) the definition (for the exclusion) of a threshold of substantial investments that would be hard to overcome by MGD databases.

The exclusion of MGD may be better implemented via the introduction of a statutory definition. By contrast, including the generation of data in the substantial investment (currently not relevant for the protection) may be a better approach for the inclusion of MGD. All alternatives, however, present their difficulties linked to the technical and legal challenges in i) separating the different activities (and investments) undertaken by database makers in the collection and processing of raw data, and ii) distinguishing between generated and a collected/observed data.

In general, all four options would bring legal clarity to the status of the Directive in relation to MGD. Policy options 2a and 2b would ensure that the *sui generis* right does not become an additional layer of protection for MGD that could stand in the way of possible obligations to allow access to such a data. Policy option 2b, while behaving the same as 2a for MGD

[253] The complete repeal of the sui generis right was also discarded at early stage.

databases, would create a more flexible test drawing from the recent EU Case law[254] for other databases still under the scope of *sui generis*. The test could include a weighing of interests with an emphasis on the ability to redeem the risk of investment by the database maker as well as the purpose of the taking by the user (e.g. follow-on innovation). Nevertheless, the uncertainty around the actual implementation of the test poses risks on the effectiveness and clarity of the option. Policy options 2c would still grant a degree of protection to MGD databases by introducing a new, lighter, regime limited and targeted against unauthorised uses by third parties. This would provide assurance for MGD database makers to share and trade their databases. However, policy option 2c would risk increasing uncertainty and complexity due to the need to introduce a completely new protection regime that would go in parallel with the *sui generis* right. Finally, policy option 3a could incentivise database makers to make their databases public, without resorting to other protective measures or rights, as the exclusive *sui generis* right would protect them against misuse of the third parties. However, existing literature suggests that the protection incentive needed to produce data is not strong.

The study assessed and compared of the four options against three criteria: i) effectiveness in achieving the policy objectives, ii) efficiency in terms of costs and benefits from the option, and iii) coherence of the option with existing policy initiatives and legal frameworks. Furthermore, the study explored the impact of each policy on the fundamental rights to protection of personal data as well as to property and freedom to conduct business without finding major conflicts in any policy.

Based on the results of the assessment and the evidence gathered for the present study, option 2a is according to this study the most suitable option: excluding MGD for the scope of the *sui generis* right. An important direct benefit of this option is expected from reduced information and transaction costs for database users due to less room for opportunistic litigation on third party data use. Moreover, the exclusion of MGD would limit the risk that investments in generation of raw data rather than databases as originally envisaged by the Directive, would benefit from exclusive rights. Stakeholders that participate in the survey generally support the option, expecting more benefits than costs. However, there were also interviewees which expressed a preference for the inclusion of MGD as a structured protection that would incentivise database makers to make their data more public. In general, the exclusion of MGD present a stronger alignment with the Data Act and the intention to stimulate access and usage of data across sectors. With regards to the other options, 2b ranked second, according to the assessment, followed by 2c and 3a as the least performing option against the criteria chosen.

The study also explored a set of six supplementary options applying to all databases in the scope of the *sui generis* right, which are offered to the consideration of the Commission should the review of the Database Directive go beyond the issue of MGD. These supplementary options cover different issues thus are not always directly comparable and mutually exclusive. After a scoping phase, three supplementary options were prioritised and further explored. The prioritised options are:

- alignment of exceptions with copyright law (option S1),
- introduction of an extended exception on research, (part of option S2),
- exclusion of public bodies from the scope of the *sui generis* right (option S3),

[254] C762/19 CV-Online Latvia v Melons

Generally, benefits are expected compared to the current status quo for all the options above. Option S1 is expected to have minimal additional costs, whereas it would bring benefits in terms of increased clarity and uniformity with copyright law. In option S2, the new expanded exception on research is expected to provide positive societal impacts while possibly reducing transaction costs for researchers that struggle to understand the conditions at which they can access and use protected databases for scientific purposes. As regards to option S3, excluding public bodies from ownership of *sui generis* right would corroborate the pertinent policy decisions made in the Open Data legislation of the EU, however, the expected benefits are likely to be limited due to the recent adoption of the ODD.

References

Abrahamson, Z (2014): Essential Data , Yale Law Journal

ALI-ELI (2021): Principles for a data economy: Data rights and transactions

Acquisti, A. (2010): The economics of personal data and the economics of privacy

Aggarwal, Ch. (editor) (2013): Managing and mining sensor data, Springer

Angelina Zier (2021), Schutz von Maschinendaten, PhD Dissertation, mimeo.

Ashton, K. (2009): That "internet of things IoT" thing, RFID Journal, available http://rfidjournal.com/article/print4986

Aszendorf, Z., Pratt, G. (2021): Future looks uncertain for EU database right, https://www.lexology.com/library/detail.aspx?g=62868ff0-c7ec-45cb-838c-0b851fee8d63

Atik & Martens (2020): Competition Problems and Governance of Non-personal Agricultural Machine Data: Comparing Voluntary Initiatives in the US and EU

Ausloos et al (2019): Getting Data Subject Rights Right A submission to the European Data Protection Board from international data rights academics, to inform regulatory guidance. 10 (2019) JIPITEC 283, p. 306

Banterle, F. (2020): Data ownership in the data economy: a European dilemma. In EU Internet Law in the digital era. Springer, Cham.

Ben Gris, B. and Ashall, S. (2020): European Union and United States: Antitrust and Data

Beunen, A.C., (2007). Protection for databases: The European database directive and its effects in the Netherlands, France and the United Kingdom. Leiden University, p.113

Bird & Bird (2021): Commission Inception Impact Assessment on a proposed Data Act: What does it mean for IP owners? Lexology

B.J. Koops (2014): The trouble with European data protection law. International Data Privacy Law Vol. 4, No. 4, 258

Borghi, M., & Karapapa, S. (2015): Contractual restrictions on lawful use of information: sole-source databases protected by the back door. European Intellectual Property Review, 37(8), 505-514

Buri, I. (2012): Accessing and Licensing Government Data under Open Access Conditions.

Brkan, M. (2019): The Essence of the Fundamental Rights to Privacy and Data Protection: Finding the Way Through the Maze of the CJEU's Constitutional Reasoning. German Law Journal, 20, pp. 864–883

Burdese (2021). "AI-Generated Databases. Do the Creation/obtaining Dichotomy and the Substantial Investment Requirement Ex-clude the Sui Generis Right Provided for under the EU Database Directive? Reflections and Proposals" , WIPO Academy, University of Turin and ITC-ILO - Master of Laws in IP - Research Papers Collection - 2019-2020.p.7.

Carrière-Swallow Y and V Haksar (2019): The Economics and Implications of Data An Integrated Perspective", IMF WP No 19/16.

Christian Frodl, C (2015): Commercialisation of Sports Data: Rights of Event Owners over Information and Statistics Generated About Their Sports Events", Marquette Sports Law Review, 55.

Colangelo, G. & Maggiolino, M (2017): Big data as misleading facilities , European Competition Journal

Corrales Compagnucci, M. (2020): Big Data, databases and "Ownership" rights in the cloud, Springer

Cremer et al. (2019): Competition policy for the digital era

CREATE IoT Project (2017): CROSS FERTILISATION THROUGH ALIGNMENT, SYNCHRONISATION AND EXCHANGES FOR IoT, D05.03 IoT Data Value Chain Model. p. 24-25.

Danson, A., Dunn, E., Bond, T. (2021): Sports data rights in 2021: the outlook, Blog MediaWrites

Datenethikkommission (2019): Gutachten der Datenethikkommission

Datennutzungsgesetz of 16.07.2021 BGBl. I S.

Davilla, M. (2017): Is big data a different kind of animal? The treatment of big data under the EU competition rules

de Corniere A and G Taylor (2020): Data and Competition: A General Framework with Applications to Mergers, Market Structure and Privacy Policy, TSE Working Paper, n. 20-1076

De Filippi, P., & Maurel, L. (2015): The paradoxes of open data and how to get rid of it? Analysing the interplay between open data and sui-generis rights on databases. International Journal of Law and Information Technology, 23(1), 1-22.

Derclaye E., (2008) The legal protection of databases: a comparative analysis, p. 98-99.

Derclaye, E. (Ed.). (2009). Research handbook on the future of EU copyright. Edward Elgar Publishing.

Deloitte (2017): Study to support the review of Directive 2003/98/EC on the re-use of public sector information

Deloitte (2018): Realising the economic potential of machine-generated, nonpersonal data in the EU

Deloitte (2018): The internet of Things: A technical primer

Derclaye and Husovec (2021): Access to information and competition concerns enter the sui generis right's infringement test – The CJEU redefines the database right

Derclaye, E & Husovec, M (2021): Sui Generis Database Protection 2.0: Judicial and Legislative Reforms, SSRN Working Paper 18 November 2021, p. 11-13, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3964943.Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)

Drexl, J. (2017): Designing Competitive Markets for Industrial Data , JIPTEC

Drexl, J. (2018): Data Access and Control in the Era of Connected Devices , Study commissioned by BEUC

Drexl, J. (2020): Connected Devices - an unfair competition law approach to data access rights of users

Duch-Brown, N. (2017): The economics of ownership, access and trade in digital data, JRC.

Ducato, R., 2013. “Adiós sui generis”: A Study of the legal Feasibility of the Sui Generis Right in the Context of Research Biobanks on further arguments on possibility for spin-off databases to be protected via investments in verification and presentation of content

Elfering, S. (2019): Unlocking the Right to Data Portability: An Analysis of the Interface with the *Sui generis* Database Right, p. 48

Erikson Mobility Report (2021): 5G on the road to mass market

European Commission (1996): Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases

European Commission (2005): First evaluation of Directive 96/9/EC on the legal protection of databases. DG Internal Market and Services Working Paper

European Commission (2012): Commission Staff Working Paper, Impact Assessment accompanying the document Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) and Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data” SEC(2012) 72 final, Brussels , p. 28

European Commission (2015): European Commission Communication, A Digital Single Market Strategy for Europe, COM (2015) 192 final, 6 May 2015

European Commission (2016a): Staff Working Document on the modernisation of EU copyright rules, SWD(2016) 301 final, part 1/3

European Commission (2016b): Key principles for comparison tools

European Commission (2017a): Staff Working Document on the free flow of data and emerging issues of the European data economy, SWD(2017) 2 final, 20

European Commission (2017b): The European Data Market Study Update

European Commission (2018), Evaluation of the Directive 96/9/EC on the legal protection of databases [SWD(2018) 147 final], p.15.

European Commission (2020a): The European data market monitoring tool: key facts & figures, first policy conclusions, data landscape and quantified stories. D2.9 Final Study Report

European Commission (2020b): Communication from the Commission, A European strategy for data, COM(2020) 66 final, 2

European Commission (2020c): European Commission Communication, Making the most of the EU’s innovative potential, COM (2020) 760 final

European Commission (2020d): The European Data Market Study Update

European Commission (2020e), “Trends and Developments in Artificial Intelligence: Challenges to the Intellectual Property Rights Framework”.

European Commission (2021a): Work Programme 2021: https://ec.europa.eu/info/publications/2021-commission-work-programme-key-document

European Commission (2021b): Preliminary report – sector inquiry into consumer internet of things, SWD (2021) 144

Falce, V. (2020): Uses and abuses of database rights , Kritika: Essays on Intellectual Property. Edward Elgar Publishing

Farbodi M and L Veldkamp (2021): A Growth Model of the Data Economy", NBER WP 28427

Farkas, Th. (2017): Data created by the internet of thingsIoT: the new gold without ownership. Rev. Prop. Inmaterial, 23, p.5.

Feretti, F. (2014): Data protection and the legitimate interest of data controllers: Much ADO about nothing or the winter of the right? 51 CML Rev 843, 852

Fortschrittsbericht-Open Data, BT-DRs. 14140, p. 37

Freitas A., Curry E. (2016): Big Data Curation. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer

Frodl, C. (2015): Commercialisation of Sports Data: Rights of Event Owners over Information and Statistics Generated About Their Sports Events, 26 Marq. Sports L. Rev. 55

Garadin, D. (2020): Access to in-vehicle data by third-party service providers: is there a market failure and, if so, how should it be addressed?

Geiger, C (ed.) (2020): Research Handbook on Intellectual Property and Investment Law, Research Handbooks in Intellectual Property, Edward Elgar Publishing Ltd, Cheltenham, UK, 385 - 405

Gervais, D. (2019): Exploring the Interfaces Between Big Data and Intellectual Property Law, JIPITEC

Giannopoulou, A. (2019): Access and reuse of MGD for scientific research. Erasmus *L. Rev.*, *12*, 155.

Gils, Th. (2017): Blockchain and Law: An Analysis of Blockchain Technology Under the Database Directive 96/9 , SSRN

Gimpel, G. (2020). Bringing dark data into the light: Illuminating existing IoT data lost within your organization. Business Horizons, 63(4), 519-530.

Graef, I. (2016): Data as essential facility, PhD Thesis, KU Leuven

Graef, I. (2016), EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility (Kluwer Law International), p. 269-276.

Graef et al (2018): Data Portability and Data Control: Lessons for an Emerging Concept in EU Law. German Law Journal 19(6) 1367

Graef et al (2019): Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation, European Law Review 2019, vol. 44 no. 5, p. 605-621

Groussot et al (2017): Chapter 14: Weak right, strong Court – the freedom to conduct business and the EU Charter of Fundamental Rights, in: Sionaidh Douglas-Scott and Nicholas Hatzis (eds.), Research Handbook on EU Law and Human Rights, Research Handbooks in European Law series.

Guibault, L., & Wiebe, A. (2013): Safe to be open: Study on the protection of research data and recommendations for access and usage. Universitätsverlag Göttingen.p. 37 et seqq

Gupta (2017): Footprints of Feist

Herbers, B., Matthiesen, R. (2021): European Commission moves ahead with proposed 'Data Act' regulating access to data in B2B and B2G relationships, https://www.lexology.com/library/detail.aspx?g=874bda55-5681-43b5-a6ab-bf6497ba786c

Herbert Zech (2015), '„Industrie 4.0" – Rechtsrahmen für eine Datenwirtschaft im digitalen Binnenmarkt' GRUR 1151, 1157.

Hervey, M. (2020): EU report on AI-assisted creativity and invention, Blog Gowling WLG, https://www.lexology.com/library/detail.aspx?g=6904f6ac-8973-4afd-8494-01e148111329

Hugenholtz, P.B., 2005. Abuse of Database Right: Sole-source information banks under the EU Database Directive. Antitrust, Patent and Copyright

Hugenholtz (2016): Something Completely Different: Europe's Sui generis Database Right, in Frankel, Gervais (2016): The Internet and the Emerging Importance of New Forms of Intellectual Property

Hugenholtz, P.B. (2017a): Data Property in the System of Intellectual Property Law

Hugenholtz, Bernt P. (2017b): Data Property: Unwelcome Guest in the House of IP. Paper presented at Trading Data in the Digital Economy: Legal Concepts and Tools, Münster, Germany

Husovec, M. (2019): The Essence of Intellectual Property Rights Under Article 17(2) of the EU Charter. German Law Journal, Volume 20, Special Issue 6: Interrogating the Essence of EU Fundamental Rights, pp. 840 – 863, DOI: https://doi.org/10.1017/glj.2019.65)

Husovec, M (2020): The fundamental right to property and the protection of investment: how difficult is it to repeal new intellectual property rights? in Geiger, C (ed.) (2020): Research Handbook on Intellectual Property and Investment Law, Research Handbooks in Intellectual Property, Edward Elgar Publishing Ltd, Cheltenham, UK, 385 – 405.

IDC (2021): Global Datasphere And StorageSphere Forecasts IDC/Lisbon Council

Inception impact assessment Data Act (including the review of the Directive 96/9/EC on the legal protection of databases) (2021): Inception impact assessment -Ares(2021)3527151

JIIP and Technopolis (2018): Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases, Report for the European Commission

JIIP, IViR – University of Amsterdam (2020): Trends and Developments in AI – Challenges to the IPR framework, Report for the European Commission

Julicher et al (2019): Protection of the EU Charter for Private Legal Entities and Public Authorities? The Personal Scope of Fundamental Rights within Europe Compared. Utrecht Law Review

Jones, C I and C Tonetti (2020): Nonrivalry and the Economics of Data, American Economic Review, vol. 110(9), pp.2819-2858

Kasrin, N. et al (2021): Data-sharing markets for integrating IoT data processing functionalities, CCF Transactions on pervasive computing and interaction, 3:76-93

Kerber, W. (2016a). A new (intellectual) property right for non-personal data? An economic analysis. An Economic Analysis (October 24, 2016). Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil (GRUR Int), 11, 989-999.

Kerber, W. (2016): Digital markets, data, and privacy: competition law, consumer law and data protection. *Journal of Intellectual Property Law & Practice*, 11(11), 856-866

Kim, D. (2018): No one's ownership as the status quo and a possible way forward: A note on the public consultation on Building a European Data Economy. Journal of Intellectual Property Law & Practice, 13(2), pp.154-165.

Kop, M. (2020): Machine Learning & EU Data sharing practices , SSRN

Krishnamurthi, R., Kumar, A., Gopinathan, D., Nayyar, A., & Qureshi, B. (2020): An overview of IoT sensor data processing, fusion, and analysis techniques. Sensors, 20(21), 6076.

Lambrecht, A., & Tucker, C. E. (2015). Can big data protect a firm from competition?. Available at SSRN 2705530.

Leistner, M. (2018): Big Data and the EU Database Directive 96/9/EC: current law and potential for reform, SSRN

Leistner, M. (2020): The existing European IP rights system and the data economy – An overview with particular focus on data access and portability. Available at SSRN 3625712

Lynskey, O. (2019): Criminal justice profiling and EU data protection law: precarious protection from predictive policing, International Journal of Law in Context 15(2) pp. 162-176

Lynskey, O (2020): Delivering Data Protection: The Next Chapter. German Law Journal 21(1) 80-84

Martens, B. (2018): The impact of data access regimes on artificial intelligence and machine learning, JRC

Martens, B. et al. (2020): Business-to-Business data sharing: An economic and legal analysis, JRC working paper 2020-05

Max-Planck-Institute, Hilty et al. (2008), EUROPEAN COMMISSION–GREEN PAPER: COPYRIGHT IN THE KNOWLEDGE ECONOMY–COMMENTS BY THE MAX PLANCK INSTITUTE FOR INTELLECTUAL PROPERTY, COMPETITION AND TAX LAW, p. 9, available at https://www.researchgate.net/publication/43234255.

McGrath et al (2013): Sensor Technologies: Healthcare, Wellness and Environmental Applications, APress.

McKenna, C., Olswang, N. (2021): Embracing open data is now more important than ever (open data note 2 of 2) https://www.lexology.com/library/detail.aspx?g=ac07421e-3c10-465b-98ea-e93b6e08b1e7&filterId=4a372f96-ad34-46bf-8be0-11151ef52b15

Myška M., Harasta J. (2016): Less is more? Protecting databases in the EU after Ryanair, Masaryk University Journal of Lawe and Technology, 10(2), 170-199. DOI 10.5817/MUJLT2016-2-3

Neto, J. A. (2021): The five V's of Big Data

Noto La Diega, G. (2019): Artificial Intelligence and databases in the age of big machine data, AIDA Annali Italiani del diritto d'autore della cultura e dello spettacolo.

OECD (2015): Data-Driven Innovation - Big Data for Growth and Well-Being

Papadopoulos, M., & Bratsas, C. (2015): Openness/Open Access for Public Sector information and works—the Creative Commons licensing model. Aristotle University of Thessaloniki: Thessaloniki, Greece.

Purtova, N. (2018): The Law of everything. Broad concept of personal data and the future of EU data protection law, Law, Innovation, and Technology 10(1);

Reichman, J. H., & Okediji, R. L. (2012): When copyright law and science collide: empowering digitally integrated research methods on a global scale. Minnesota Law Review, 96(4), 1362.

Reichman, J. H., Dinwoodie, G. B., & Samuelson, P. (2007): A Reverse Notice and Takedown Regime to Enable Pubic Interest Uses of Technically Protected Copyrighted Works. Berkeley Tech. LJ, 22, 981.

Rengasamy, D., Morvan, H. P., & Figueredo, G. P. (2018): Deep learning approaches to aircraft maintenance, repair and overhaul: a review. In 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 150-156). IEEE.

Richter, H. (2018). Informationsweiterverwendungsgesetz (IWG). CH Beck.

Sattler (2020) Rechtshandbuch Industrie 4.0 und Internet of Things, in Sassenberg/Faber (Hrsg.)

Schweitzer, H., Peitz, M. (2017): Datenmärkte in der digitalisierten Wirtschaft: Funktionsdefizite und regelungsbedarf? ZEW Discussion paper 17-043

Solda-Kutzmann, D. (2011): Public Sector Information Commons. Informatica e diritto, (1-2), 199-217.

Stepanov, I. (2020) Introducing a property right over data in the EU: the data producer's right – an evaluation, International Review of Law, Computers & Technology, 34:1,65-86, DOI: 10.1080/13600869.2019.1631621

Stamatoudi and Torremans, (2014):'The Database Directive', EU Copyright Law: A Commentary.

Stucke, M. E. (2018). Should we be concerned about data-opolies?.

Tombal, T., (2020). Economic dependence and data access. IIC-International Review of Intellectual Property and Competition Law, 51(1), pp.70-98

Tsvetanov, F.A. (2020): Storing data from sensor networks, IOP Conf. Series: Materials Science and Engineering, 1032

Van Eechoud (2021): A Serpent Eating Its Tail: The Database Directive Meets the Open Data Directive. Springer

Weiser, M. (1991). The Computer for the 21st Century. Scientific American, [online] 265(3), pp.94-104. Available at: https://www.ics.uci.edu/~corps/phaseii/Weiser-Computer21stCentury-SciAm.pdf

Wiebe, A (2014): Der Schutz von Datenbanken – ungeliebtes Stiefkind des Immaterialgüterrechts, CR 30,1

Wiebe, A (2016): Protection of industrial data – a new property right for the digital ecoomy?, GRUR Int.

Wiebe, A (2017): Schutz von Maschinendaten durch das sui-generis Schutzrecht für Datenbanken, GRUR, 338

Wiebe, A (2017): Protection of industrial data - a new property right for the digital economy? J.Int.Prop Law & Politics

Zech, H. (2016): A legal framework for a data economy in the European Digital Single Market: rights to use data. Journal of Intellectual Property Law & Practice, Vol 11, 6

Zeitlin, M (2018): Everything counts in large amounts. Protection of the big data under the Database Directive, Master Thesis University of Uppsala

Zentralverband Elektrotechnik- und Elektronikindustrie (ZVEI) (2016): Die Elektroindustrie als Leitbranche der Digitalisierung.

Zwenne, GJ (2013): Diluted Privacy Law, Inaugural Lecture (University of Leiden, 2013) 33 et seq.

Other information sources

Advanced technologies for industry, https://ati.ec.europa.eu/

https://aws.amazon.com/data-warehouse

https://behrtech.com/blog/top-10-iot-sensor-types/

Cision PR Newswire (2019): Global IoT Sensors Market Analysis, Trends, and Forecasts

https://de.rs-online.com/web/generalDisplay.html?id=ideen-und-tipps/sensoren-leitfaden

European Commission (2017), The European Data Market Study Update**.** https://digital-strategy.ec.europa.eu/en/library/european-data-market-study-update.

European Commission (2020), The European Data Market Study Update, https://digital-strategy.ec.europa.eu/en/library/european-data-market-study-update.

European Court of Human Rights (2021): Practical Guide on Admissibility Criteria

Heger, S (2021): Einheitliche Datenstruktur: Ohne nahtlose Zusammenarbeit kein IoT

Hugenholtz (2019): The New Copyright Directive: Text and Data Mining (Articles 3 and 4)

Husovec, M., Derclaye, E. (2021): Access to information and competition concerns enter the sui generis right's infringement test – The CJEU redefines the database right, Kluwer Copyright Blog. Available at: http://copyrightblog.kluweriplaw.com/2021/06/17/access-to-information-and-competition-concerns-enter-the-sui-generis-rights-infringement-test-the-cjeu-redefines-the-database-right/

Datalandscape (2018): The European Data Market Monitoring Tool

https://blocksandfiles.com/2020/01/17/connected-car-data-storage-estimates-vary-widely/

https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation-evaluation-directive-969ec-legal-protection-databases

https://home.cern/news/news/computing/cern-data-centre-passes-200-petabyte-milestone

https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0760&from=EN

https://www.industry-of-things.de/einheitliche-datenstruktur-ohne-nahtlose-zusammenarbeit-kein-iot-a-1013344/

https://www.ip-insider.de/ein-blick-auf-die-gefahrenlage-im-internet-der-dinge-a-691062/

https://www.maschinenmarkt.vogel.de/opc-router-steigert-automatisierungseffekt-a-815154/

https://www.opc-router.de/maschinendatenerfassung/#EdgeFog

Statista (2021): Size of the database management system (DBMS) market worldwide from 2017 to 2020

Case law

BGH 1.12.2010 I ZR 196/08, Zweite Zahnarztmeinung II

BGH GRUR 2007 , 500 - Sächsischer Ausschreibungsdienst

BGH 25.04.2010 I ZR 47/08, Autobahnmaut

British Sky Broadcasting plc v Digital Satellite Warranty Cover Ltd [2011] EWHC 2662

C-7/97, Bronner, ECLI:EU:C:1998:569

C-30/14, Ryanair v PR Aviation

C-46/02, Fixtures Marketing Ltd v. Oy Veikkaus AB

C-70/10, Scarlet Extended, ECLI:EU:C:2011:771, Judgment of Nov. 24, 2011

C-92/09 and C-93/09, Volker und Markus Schecke GbR and Hartmut Eifert v. Land Hessen [GC], 9 November 2010

C-93/09, Volker und Markus Schecke GbR and Hartmut Eifert v. Land Hessen [GC]

C-131/12, Google Inc. v Agencia Española de Protección de Datos and Mario Costeja González, ECLI:EU:C:2014:317

C-200/96, Metronome Musik

C-202/12, Innoweb BV v Wegener ICT Media BV and Wegener Mediaventions BV

C-203/02, The British Horseracing Board Ltd & Ots v. William Hill Organization Ltd

C-245/02, Anheuser-Busch

C-262/81, Coditel II

C-293/12 & C-594/12, Digital Rights Ireland

C-304/07, Directmedia Publishing

C-338/02, Fixtures Marketing Ltd v. Svenska Spel AB

C-360/10, Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v Netlog NV

C-435/12, ACI Adam and Others

C-444/02, Fixtures Marketing Ltd v. Organismos Prognostikon Agonon Podosfairou

C-462/09, EU:C:2011:397, Stichting de Thuiskopie

C-463/12, EU:C:2015:144, Copydan Båndkopi

C-467/08, EU:C:2010:620, Padawan

C-490/14, Freistaat Bayern v. Verlag Esterbauer Gmbh

C-521/11, EU:C:2013:515, Amazon.com International Sales and Others

C-435/12, EU:C:2014:254, ACI Adam and Others

C-516/17, Spiegel Online

C-572/14, ECLI:EU:C:2016:286, Austro-Mechana Gesellschaft zur Wahrnehmung mechanisch-musikalischer Urheberrechte GmbH v Amazon EU Sàrl, Amazon Services Europe Sàrl, Amazon.de GmbH, Amazon Logistik GmbH, Amazon Media Sàrl

C-762/19, CV-Online Latvia v Melons

C-762/19, ECLI:EU:C:2021:434

Case STS 572/2012 Ryanair v Atrapalo

Court of Cassation (Cass.), 1st civ., 5 March 2009, Précom, Ouest France Multimedia v Direct Annonces

Football Dataco Ltd v Stan James Ltd (No 2) [2013] EWCA Civ 27.

OLG Hamburg 8.6.2017, 5U54/12

T-201/04, Microsoft, ECLI:EU:T:2007:289

A Annex A: Options either discarded or not further explored

The following appendix provides a brief high-level discussion of each policy and supplementary policy option that has been either discarded at early stage or for which no further in-depth assessment has been conducted after the initial scoping phase.

A.1 Policy option 1: Repeal of the *sui generis* right in the Database Directive

This option was also discussed in the 2018 evaluation study after the study did not find conclusive positive effects of the Database Directive. While the evaluation study suggested that the abolition would not increase more the already high legal uncertainty, we understand that the implementation of this options may be very unlikely, partly because of the issues already noted in the 2018 evaluation study.

This option is discarded at early stage. The main reason being that repeal of the Directive would have a de-harmonising effect resulting in Member States setting up their own *sui generis* regime or going back to unharmonized protection against unfair trade practices in the form of slavish imitation, if no provision is made barring Member States from applying such regimes. Moreover, apart from the case of databases with MGD, there is insufficient proof of harm by the *sui generis* right to justify repeal.[255]

A.2 Policy option 2d: Exclusion of MGD from the scope of application of the sui generis database right coupled with the introduction of a specific access regime

Under policy option 2, MGD would not be protected by the *sui generis* right. As a result, the access regime would then require the construction of a new legal framework. The introduction of access rights should be independent from rules on databases so as not to be bound by any definition of what is a database. This type of exercise is currently explored within the Data Act initiative; therefore, this option will not be assessed in-depth. However, below we provide high level suggestions on how this access regime could be designed.

The policy option could be designed to include two types of access:

1. **An access, use and portability right for the user (lawful customer) of a connected device** of data generated by (sensors of) that device. This would apply to individual device data, e.g. data related to the performance and operations of the device, and would be granted to the legitimate user of the device. This would relate to concept of “lawful user”. The legal implementation could start from Art. 8(1) and 15 of the Database Directive to create minimum rights comparable to Art. 5(1) of the Computer Programs Directive [256] that “go” with the user. This approach would be a reasonable way to strike a balance between protection of database and access. The focus of this minimum right would be on access and usage data. As MGD databases would be excluded from protection, the right would apply to unprotected data in the sense of the *Ryanair*

[255] See overview of previous evaluations and public consultations in Section “Evaluation of the Database Directive”.

[256] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs

Decision[257]. The minimum rights could be extended to use of the whole database which might be especially important for MGD. There would also be a need to specify which persons should be regarded as legitimate users - going beyond buyer and licensee to persons having control over the specific IoT device. The specific access may not include a remuneration for the data holder, assumed to be the device manufacturer, since the manufacturer could factor in the price of the device the cost to provide access to data.[258]

Moreover, the portability part of the access will enable users to grant access to data of their machine to repairs and other providers to stimulate competition and growth in secondary markets. This access may only apply to databases produced in the context of sensors and IoT, where the user is co-producer of data. To be workable in a dynamic IoT context, the access would need to be of a real-time, continuous nature, and not merely be interpreted as a one-off, static type of portability. The access right could be limited to the access of certain type of MGD, e.g., data about the performance of the machine, rather than observed data on the environment in which the machine operate. The former is more linked with the concept of sole-source database, while for the latter there could be substitute/alternative means to collect the same information. The distinction between personal and non-personal data is relevant here, as the GDPR has already created a right to data portability for natural persons under Art 20.

The question remains to what extent the provision of minimum rights will create effective access for users in practice. Even when included in the Database Directive, users will still have to claim and enforce their rights. Depending on the willingness of database holders to proactively facilitate the exercise of the minimum rights, users may need to start legal proceedings to have their rights implemented.

2. **A compulsory-licensing regime for access of databases containing (aggregated) MGD by other firms/competitors for commercial use**. This approach would provide for access and use of MGD databases even if these were excluded from database protection as a special compulsory licensing regime could be established. This would connect to respective plans for sole source databases in the original Draft of the Directive in 1992 and would avoid the problems that occurred in connection with the legislative solution to limit the scope of protection to substantial parts. MGD might include problems similar to those of sole-source databases and hence deserve similar treatment.

   The criteria to define when this access right would apply may be drawn from competition law, i.e. in line with the prohibition on abuse of dominance under Article 102 TFEU[259]. The right will be designed in using the so-called essential facilities doctrine of Article 102 TFEU (i.e. the data should be indispensable, implying that: (1) there are no alternatives readily available on the market for the requested data, and (2) it is not feasible for the access seeker to create a similar dataset itself). The licenses may be given under fair, reasonable,

---

[257] Case C-30/14 Ryanair v PR Aviation

[258] However, this could be left to the courts. Introducing such a minimum rights provision has the potential to be extended to databases in general and could create clarity with respect to the still dubious "consultation" right from the *BHB v Hill* case 2004 and could be extended to data sets in databases that are not protected to take care of the *Ryanair* Decision of the CJEU.

[259] Article 102 of the Treaty on the Functioning of the European Union. See : https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12008E102

and non-discriminatory terms (FRAND). It is expected, however, the implementation of this access right system to be quite complex and costly. Moreover, compatibility with trade secret protection as well as data protection legislation would have to be provided for[260].

In the context of data sharing in data-driven markets, Prüfer & Graef (2021) have suggested three possible governance mechanisms to implement a form of mandatory data sharing.[261] One is to centralize investigations and enforcement in a European Data Sharing Agency and to provide national competition authorities with decision-making power. The second approach is to set up a Data Sharing Cooperation Network coordinated through a European Data Sharing Board, and with the national competition authority best placed to run the investigation adjudicating and enforcing the mandatory data-sharing decision across the EU. This approach combines features of the current enforcement mechanisms of EU consumer and data protection law. A third approach is to mix both governance mechanisms and to ask national competition authorities to investigate and adjudicate, and let enforcement take place at the EU level through the European Data Sharing Agency. A system of compulsory licensing of MGD databases could be based on similar ideas.

This type of access rights, to be effective, may require an understanding of the context. Therefore, sector-specific regulations/instruments may be better suited to include access regimes rather than a general instrument such as the Database Directive.

A.3 Policy option 3b: Inclusion of MGD in the scope of application of the *sui generis* database right coupled with the introduction of a specific access regime

Under this sub-option a specific access regime would be introduced to allow users access (in certain circumstances) to MGD databases.

The difference from policy option 2d is that MGD would be included in the scope of protection. The access regime could be implemented as part of a revised Database Directive by including provisions that would impose additional requirements on MGD. The access regimes could be designed in a similar vein as described in policy option 2d. However, as stated in policy option 2d above, the possibility of introducing an access regime for MGD data may be left to the Data Act, thus this option will not be assessed further.

A.4 Supplementary policy option S4: Introduce compulsory licencing for sole-source databases.

This supplementary option would fill a gap that has been perceived since the Database Directive was enacted. It has many positive effects but may face problems of implementation. The urgent need to implement it could be reduced by a possible implementation of access rights in the upcoming Data Act. Therefore, this option will not be explored further in-depth. Nevertheless, a general elaboration and discussion on how to deal with issue of sole-source databases is provided below.

---

[260] See Leistner, M (2020): The Existing European IP Rights System and the Data Economy–An Overview With Particular Focus on Data Access and Portability for a discussion on a differentiation for databases which are additionally protected as trade secret.

[261] Graef, I., & Prüfer, J. (2021): Governance of data sharing: A law & economics proposal. Research Policy, 50(9), 104330.

Supplementary option S4 could have consisted in introducing compulsory licensing of sole-source databases (whether or not MGD databases) to balance the objective to incentivize creation of databases and the objective to foster wider data access and usage. Similar to what mentioned in sub-option 2d, this option would connect to respective plans for sole source databases in the original Draft of the Directive in 1992 and would avoid the problems that occurred in connection with the legislative solution to limit the scope of protection to substantial parts. The impact of this option might have depended on whether it was implemented in combination with option 2a, or not. In combination with 2a, this new access regime would have applied only to non-MGD sole-source databases as MGD databases would be outside of the scope of protection of the *sui generis* right. While if MGD were not excluded from the *sui generis* right, the impact would have been on all sole-source databases regardless of whether primarily MGD or non-MGD.

As mentioned above, the obligation to licence sole-source databases may interfere and/or overlap with access regime that could be possibly introduced by the Data Act.

The concern related to sole-source databases is that third parties can be excluded access by a dominant firm to the data they need to operate on the market. For defining the scope of sole-source databases, inspiration can be drawn from how competition cases have interpreted the notion of indispensability under the essential facilities doctrine of Article 102 TFEU. A key condition to hold refusals to give access to data of dominant firms abusive is that the data is indispensable, implying that: (1) there are no alternatives readily available on the market for the requested data (actual availability of alternatives), and (2) it is not feasible for the access seeker to create a similar dataset itself (potential availability of alternatives).[262] These elements can also be relevant for establishing a defensive right for MGD in a way that prevents competition concerns from arising due to the sole-source nature of certain datasets.

The notion of indispensability has been interpreted according to different standards in competition cases. A strict interpretation was applied by the CJEU in the *Bronner* case. Oscar Bronner was refused access to Mediaprint's nation-wide home-delivery system for newspapers in Austria. With regard to the indispensability of Mediaprint's home-delivery system, the Court argued that several alternatives were available for the distribution of Oscar Bronner's daily newspaper, such as delivery by post and sales in shops and at kiosks, even though they may be less advantageous.[263] In relation to the potential availability of alternatives, the Court noted that there do not appear to be "any technical, legal or even economic obstacles capable of making it impossible, or even unreasonably difficult" for another publisher to set up its own home-delivery system.[264] In this regard, the Court stated that it was not enough for Oscar Bronner to argue that it is not economically viable to set up its own system due to the small circulation of its own newspaper.[265] For such access to be capable of being regarded as indispensable, it would be necessary, in the Court's view, to establish "at the very least [...] that it is not economically viable to create a second home-delivery scheme with a circulation

[262] See for instance the analysis in: Graef, EU Competition Law, Data Protection and Online Platforms: Data as Essential Facility (Kluwer Law International, 2016), p. 269-276.

[263] Case C-7/97, Bronner ECLI:EU:C:1998:569, par. 43

[264] Case C-7/97, Bronner ECLI:EU:C:1998:569, par. 44

[265] Case C-7/97, Bronner ECLI:EU:C:1998:569, par. 45

comparable to that of the daily newspapers distributed by the existing scheme".[266] Considering that the Court used the scale of activities of the dominant firm as benchmark, it set a high threshold in *Bronner* for determining when inputs qualify as indispensable under Article 102 TFEU.

This was different in the *Microsoft* case[267] where the General Court accepted lower standards for fulfilling the indispensability requirement under Article 102 TFEU. One of the abuses in the *Microsoft* case concerned a refusal by Microsoft to license interoperability information that providers of non-Microsoft work group server operating systems needed to allow their systems to run on Microsoft's dominant client PC operating system Windows. In its judgment, the Court argued that it was necessary for competitors to be able to interoperate with Windows on an equal footing in order to compete viably on the market.[268] This approach deviates from the statement of the Court of Justice in *Bronner* that access is not indispensable if alternatives are available, even if they are less advantageous.[269] As a result, the requirement of indispensability under Article 102 TFEU is not one-dimensional and can be adjusted to the circumstances at stake.

A similar tailoring of standards would have to be done to determine when a database has a sole-source nature. For instance, not all sensor- or MGD databases would meet the indispensability requirement. In cases where sensors are collecting internal operational data of machines, there will be no realistic actual or potential alternatives to create an equivalent so that such databases can be considered as sole-source databases. In other situations where data is sensored from the outside world, there may be ways for other market players to create a similar database. This means that the notion of MGD is broader than the concept of sole-source databases in the sense that not all databases consisting of MGD will qualify as sole-source. During the legal expert workshop, an expert, while generally favouring the idea of excluding sole-source databases, has also questioned whether this would really solve the competition problem, since a lot of databases may not be sole source but still raise competition concerns.[270]

A.5 Supplementary policy option S5: Introduce a user's minimum right/consultation right to databases

This supplementary option would create a minimum right that could provide clarity and promote innovation. However, if the recent *CV v Melons* case will be consolidated, the need for a minimum right may be reduced as the acts covered could be exempted under that doctrine. Moreover, part of these acts may be covered by possible measures proposed within the upcoming Data Act. Hence, this option will not be explored further in-depth and no immediate action would be advised and this. As for the other non-priority policy options, a high level elaboration of the policy is provided below.

---

266 Case C-7/97, Bronner ECLI:EU:C:1998:569, par. 46

267 Case T-201/04, Microsoft v Commission.

268 Case T-201/04, Microsoft v Commission. ECLI:EU:T:2007:289, par. 421

269 Case C-7/97, Bronner ECLI:EU:C:1998:569, par. 43

270 The expert provided the classic example of Google's quote: "competition is one click away".

Similar policy option is discussed in sub option 2d applied only to MGD databases. In legal terms this approach could start out with Art. 8(1), 15 of the Database Directive and create minimum rights comparable to Art. 5(1) of the Computer Programs Directive: In the absence of specific contractual provisions, reproduction and alteration do not require authorisation by the right-holder where they are necessary for the use of the computer program by the lawful acquirer in accordance with its intended purpose, including for error correction. These minimum rights would "go" with the user and establish a reasonable way to strike a balance between protection and access. The focus of this minimum right would be on use of data. From a legal point of view the minimum rights could be extended to the use of the whole database or the part of the database that the user should have access to, given the relation he has with the database producer. There would also be a need to specify which persons should be regarded as lawful users, e.g. owner of the IoT device. However, this could be left to the courts. Such a minimum rights provision could create clarity with respect to the still dubious "consultation" right from the *BHB v Hill* case in 2004 and could be extended to data sets in databases that are not protected to take care of the *Ryanair* Decision of the CJEU. Further legal evaluation would have to be done to certify that minimum rights may also be established in relation to unprotected subject matter as there may be constitutional issues imposing a duty of providing access on database holders of unprotected databases.

Experience from the Computer Program Directive suggests that it could be helpful to further specify the acts covered by the minimum rights as in case law and legal discussion uncertainty remains in this respect. Access to the data would be the baseline, and different forms of usage could be included, e.g., for repair purposes or value-added uses. During the workshop, moreover, legal experts observed that the minimum right in the Computer Program Directive, on which the implementation of this option would be based, does not work well as it remains difficult for users to circumvent technical protection measures (TPMs) imposed by software maker. If the right would involve a right to access, the burden to grant the access should be on the database makers, otherwise there is risk that the policy would not be effective as for small users would be too costly to seek enforcement of the right e.g. by suing not complying database maker.

A.6 Supplementary policy option S6: Alter or render more flexible the duration of the protection as well as clearer rules on renewal terms

This supplementary option dealt with the relevant issue of renewal terms. However, the proposed solution initially considered in the scoping phase was uncertain as to its efficiency and its effective implementation – no further in de-depth exploration was pursued.

The original motivation for the option was the duration of the database *sui generis* right protection considered by some views as too long. This is especially relevant for MGD whose economic relevance may be shorter lived than other types of data. However, introducing a shorter term just for MGD would enhance the complexities of the duration of protection and entail the need to separate MGD from other types of data. Alternatively, a shorter term for database rights should be considered in general. Linking it to the duration of redeeming investment would be in line with the rationale of the database right and the *CV v Melons* case as well (this solution will be integrated in the elaboration of policy option 2b). Alternatively, a fixed term of, e.g. 10 years would provide certainty and avoid unnecessary disputes over the issue of amortisation.

Another point of great uncertainty is the renewal of the term of protection in cases of a new database emerging from inserting new data into the database or maintenance operations. The goal should be to clarify that the new term does not cover the whole database resulting in possibly eternal protection, which could be especially detrimental for scientific research. An

option here is to keep the renewal term but to explicitly provide a rule that only those data is in the scope of protection that has not been added more than 15 years ago. This could be documented by technical means, e.g., a time stamp on the data. The addition of a time stamp may involve considerable cost for database makers. More specific research would be needed to determine how this could be realised in practice. Furthermore, a provision on transparency with respect to the timestamping could be considered. E.g. lawful users of the database should be able to inspect the timestamp, and/or query the database for data with older time stamps in order to select data falling in the public domain. An alternative, more simple and convincing approach, would be to limit renewal term to one year and limit this to substantial changes in the database while excluding any maintenance activities on the database that could easily be used to create "eternal" protection (Zier 2021)[271]. The one-year renewal term would then apply to the whole database and no efforts would have to be made to identify the data that has been added or refurbished. This proposal could create a practically working solution to the renewal problem.

A.7 Supplementary policy option S7: Introduce new exceptions specific to the search engines and web-scraping.

Under this supplementary option, new limitations or expansion of existing limitations specifically tailored to database protection in copyright and *sui generis* were proposed with the objective of enhancing competition and innovation.

Initially, two new exceptions for the operation of search engines and for web-scraping activities were considered.

The new exceptions for search engines and web-scraping proposed initially appeared to be less pressing than the expanded exception on research because, among other factors, the recent *CV v Melons* decision already paves the way for exempting these techniques from the *sui generis* right. Therefore, after the scoping phase, these two new exceptions were not explored further. A high-level elaboration of the two exception is presented below.

New exception for search engines

Search engines are indispensable for the functioning of the internet. Search engines are also at the basis of comparison tools, increasingly used by consumers to get information in an efficient way on product availability and prices from multiple providers.[272] For this reason, there could be benefits from exempting them from legal restrictions that could seriously hamper their functioning and thus prove detrimental to innovation.

The operation of search engines has been proven problematic in case law as evidenced by the *Innoweb* case of the CJEU[273] where meta-search engine was found to be infringing the database right. The scope of the *Innoweb* decision remained unclear, e.g., whether it would be limited to meta search engines or apply to search engines in general. However, the Court in this case found even a very limited search to be infringing on the *sui generis* right as a use of the whole database which leaves room for the interpretation that any search engine operation would be infringing. This issue may become even more relevant considering the

[271] Angelina Zier (2021), Schutz von Maschinendaten, PhD Dissertation, mimeo.

[272] See European Commission (2016b): Key principles for comparison tools

[273] C-202/12 - Innoweb BV v Wegener ICT Media BV and Wegener Mediaventions BV

notion that websites may be protected as databases as well which could render a bigger part of search engine operations infringing. To alleviate this problem and for sake of clarity a general exclusion of the operation of search engines was considered in the initial scoping phase.

New exception for web-scraping

Another activity which may be affected by the protection granted by the Database Directive is web-scraping. Web-scraping is an increasingly relevant means to access and use information online [274], develop AI tools and digital services. Web-scraping refers to the extraction, copying, storage and re-use of web content which includes legitimate purposes like web crawling and substitution of web services as well as misappropriation of website content by electronic or other means without the consent of the right-holder. This type of activity is mostly not covered by the recently introduced TDM exception of the DSM Directive, thus it could be considered as complementary to the TDM exception. Art. 3 of the DSM Directive 2019 which allows for the reproduction of protected material for the purposes of pursuing text and data mining in the interest of scientific research. TDM is not reproduction itself, but analysis of protected materials that has been copied before to extract information and facts. The TDM exception does not allow, however, for using this "corpus" of collected materials freely in competition of another person´s commercial interest. Art. 3(2) DSM Directive emphasizes this point by limiting access to the copies made in accordance with the exception and allowing it to be retained only for purpose of scientific research.

However, not to encroach on innovation through overly broad limitations a further distinction may be made. Search engines as well as web scraping is open to different legal assessments depending on whether they are pursued in favour of some general interest like the functioning of internet communication that is dependent on the operation of search engines. However, if the instruments are mostly used to build up competing databases and to take unfair advantage of the efforts of the database maker, this should be legally restricted in the interest of competition and innovation. This distinction seems to be at the core of the *CV v Melons* case as well. Hence, if the case will become established case law or option 2b will be implemented, no further specific exceptions will be needed. For these reasons the introduction of these new exceptions is not further explored.

B Annex B: Technical aspects and examples of sensor data

This annex is divided into two sections: the *first* section presents an elaborated technical discussion on sensor data, its transmission and data management, the *second* section present few examples in detailed.

B.1 Technical primer on sensor data, its transmission and data management

The technical literature uses the terms “machine generated data” and “sensor data” interchangeably.

The term "machine data" summarises all data that can be retrieved by a machine, typically at runtime. A distinction can be made between simple data points and complex data sets.

[274] See Evaluation Study 2018.

Simple data points are recorded continuously or whenever there is a change. These data are recorded and stored with a current value and a time stamp as reference. This allows the status of the machines to be analysed via the historical data series.

More complex data from the machine is recorded as contiguous data sets. In addition to the optional time stamp, these data primarily contain unique references to data objects to which they provide associated data such as material numbers, machine numbers, batch numbers, etc.[275]

A sensor is a technical component that independently detects certain physical or chemical variables and properties in the environment or of objects. An important feature is the conversion of measurement data into signals or visual displays. This happens on a 'sensor node' which contains several components with a microcontroller, transceiver, external memory (storage unit), a power source (e.g., a battery) and one or more sensors. IoT sensors are furthermore associated with a wired or wireless communication interface; communication protocols such as Wi-Fi, Bluetooth, 4G, 5G etc. can be used for the IoT sensor device communication.[276]

Communication between machines as common in IoT is equally enabled through sensors and connected non-IP based processors. The connection among machines or industrial plants happens through bus systems or through a hardware gateway.

Connected to a network, they can share data with other connected devices and management systems. There are a number of different sensor types such as[277]:

- Temperature sensors – they measure the amount of heat in a source (such as a machine or device), detect changes and convert this information into data.
- Proximity sensors – through electromagnetic fieelds or beams of radiation, these sensors are used for car parks, indicating availability. They can also be used in retail and assembly lines.
- Gyroscope sensors – measure the angular rate of velocity. They can be found e.g. in navigation and electronic stability control systems in cars or in motion sensing for the game industry.
- Infrared sensors – sense their surroundings by infrared radiation. They are used in healthcare (monitoring blood flow), TVs to interpret the signals from a remote controls, or art historians to detect hidden layers of paint.
- Optical sensors – can be found in cars to recognise signs and obstacles, in smart phones or TV sets for ambient light, or in the health sector to monitor heart-rates.

---

[275] https://www.opc-router.de/maschinendatenerfassung/#EdgeFog

[276] Krishnamurti, R. (et al) (2020): An overview of IoT sensor data processing, fusion, and analysis techniques, in: Sensors, 20, 6076

[277] Examples are taken from : https://behrtech.com/blog/top-10-iot-sensor-types/, https://de.rs-online.com/web/generalDisplay.html?id=ideen-und-tipps/sensoren-leitfaden

- RFID sensors278 – are example of machine-to-machine sensors.

In an IoT environment, but also through smart consumer devices and the ever-increasing data volume, sensor data collection poses challenges. These are due to natural errors and incompleteness in the collection process. Sensor data is inherently noisy and uncertain and therefore requires 'cleaning'. Given the data volumes, issues on storage and processing need to be addressed - it may be impractical or too costly to store the entire raw data; dropping or compression of data needs to be decided.

The data the sensor node or network generates, is stored in its memory. For further use, a first step of the sensor data acquisition needs to decide if the data can be acquired through pull-based or push-based approaches. In the pull-based approach, data is only acquired at a user-defined frequency whereas in the push-based approach, sensors will only send data based on an agreed behaviour between sensor and base station, for example only deviating values are transmitted. Given the mentioned noise and erroneous sensor values, data cleaning is needed. This is done for example through the development of model-based approaches. An objective of this approach is to process queries by accessing minimal amounts of data, to handle missing values. Data compression then aims at eliminating redundancies.

An important aspect for the acquisition of data concerns costs and energy consumption. Most sensors are battery-powered and have a limited energy resource. Changing the battery would be a rather costly procedure. The limited energy source and the needed energy to communicate the sensed values to the base station are constraining factors. Therefore, minimizing the number of samples to be obtained from a sensor is key. These factors are built in the model that needs to be developed.[279]

From a technical perspective, sensor data as such requires strategies to obtain, clean and validate. This is achieved with data models that need to be developed, i.e. programmed in the sensor node. An important aspect of the data value chain process is called data integration by data analysts, and it starts *before* any data gets stored in a database. Data integration of (or data fusion) starts by integrating data from different/multiple sensors as this is often required to improve the accuracy in various applications. Data integration also includes integration of information from different sources and can be seen in all kinds of complex systems, from ships, cars, production techniques, medical diagnosis, satellites etc. For the scope of the current study, however, as long as data integration relates to integration of multiple and different sensors, the dataset containing this information would still be considered a fully MGD databases rather than "mix database". For the database to be considered "mixed", it would need to integrate data not directly generated by sensors/machines.

The combination of two or more data sources (also from third-parties) is used to obtain more consistent and better estimates in dynamic systems. "The combination provides better results than if the sensors were used individually."[280] However, "consuming third-party data comes

[278] Radio-frequency identification (RFID).

[279] See Aggarwal, Ch. (editor) (2013): Managing and mining sensor data, Springer. Chapter 2 includes various models for data acquisition, data cleaning, query processing, and data compression

[280] Rajalakshmi et al (2020)

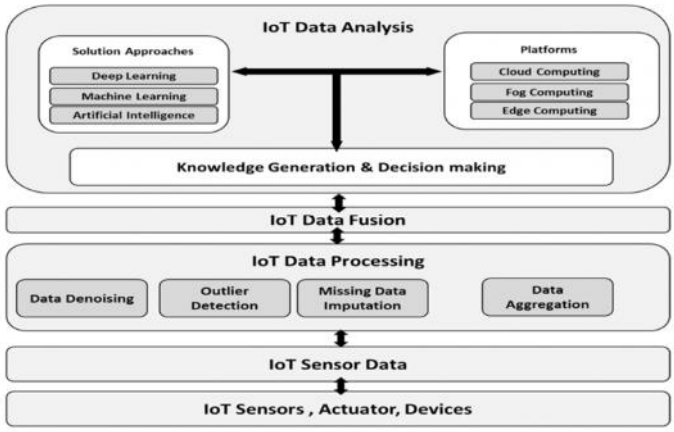
with the intrinsic cost of repurposing, adapting, and ensuring data quality for its new context."[281] The heterogeneity of sources and formats provides challenges to process and inter-operate between them.[282]

Wireless transmission of sensor-based data bears some risks. With the increase of functions and use cases developed in an IoT world, cyberattacks on production plants, infrastructures or end-user devices become more frequent. The risks include manipulated control systems, falsified data transfer, or interception of the communication between microcontroller and transmitter. To avoid attacks on sensors and the wireless data transfer, their protocols need to include defence mechanism such as data encryption.[283]

The actual sensor data processing suggests that there is a cost factor to be considered. While the cost of a sensor may be very small, its limited life span and potential high opportunity cost to replace sensor nodes as well as data security measures require the investments in data acquisition strategies, cleaning, end-to-end encryption, and other data related tasks. Some of the costs may be one-off investment costs and can be linked to the property of the sensor.

Once this has been achieved, data analytics takes place in an architecture that can possibly be labelled a database. In an IoT environment these are stored in servers (join server, network server, application server) and platforms (cloud or edge computing). To be used efficiently in networked production, machines themselves require corresponding data input (e.g., for parameterisation, for individual work steps or for comparison of the throughput with data from the production control system).

Figure 15: The basic architecture for IoT sensor data processing, data fusion and data analysis.



Source: Krishnamurthi et al. 2020

Machine generated data in an IoT environment is, as mentioned above, either connected through processors or sensors. Communication between different machines is enabled

[281] Freitas, A., Curry, E. (2016): Big data curation, in: Cavanillas, J.M. et al (eds.): New horizons for a data-driven economy., pp 87-118

[282] Kasrin, N. et al (2021): Data-sharing markets for integrating IoT data processing functionalities, CCF Transactions on pervasive computing and interaction, 3:76-93

[283] https://www.ip-insider.de/ein-blick-auf-die-gefahrenlage-im-internet-der-dinge-a-691062/

through communication standards like OPC (Open Platform Communication) ensuring the similar and manufacturer-independent exchange of data. The Unified Architecture (UA) specifics ensures independence of the platform (such as Amazon AWS or Microsoft Azure). Every machine, sensor or other system that provides OPC UA data can be connected in the same way. Due to the mechanisms defined in OPC UA for describing data structures and data points, machine data acquisition can be set up independently of the underlying machine types and proprietary protocols. OPC UA also specifies how data point hierarchies can be set up and searched.[284]

How is sensor data organized/structured in databases?[285] There are three main models that can be distinguished with either storage and processing of the data outside of the sensor network, distributed storage and processing in the network, or combined solutions. The first model is suitable mainly for low volumes of sensor data since transmission requirements from cameras or acoustic sensors to the outside of the sensor network easily exceeds the available power source. In the second model, the sensor network works as a distributed database that supports requests for data and its processing. In such as network, the sensor node stores its data locally. If it receives a request, it can send processed data rather than raw data.

In the case depicted in the figure below, data from several sensors are sent to the network coordinator that pre-processes the data[286]. Only then it is sent on to the cloud for processing. It can then be viewed, manipulated, or modified on client devices. Pre-processing by the coordinator (gateway) can include several tasks such as data compression, data encoding, or transmission of raw data directly to the cloud structure. The cloud system receives the data from the coordinator and records it. Each data has a timestamp. Within the cloud the tasks such as data validation, cleaning, transformation etc. can be executed.

In the case of storage of sensor data in an external cloud, the data from the sensor(s) can be stored as a relational database, yet, while they can store time-series data, they are limited in terms of processing ordered sets of time-series elements. An alternative in big data environments are non-relational time series databases (NoSQL). They are software systems optimised to handle arrays of time-indexed numbers. Depending on the database, different properties are supported, such as encoding additional metadata along the timelines.

Processing of the enormous amounts of sensor data – in industry and in consumer realms alike – favours another type of processing space with “data warehouses”. These are central repositories which can contain individual relational databases, or are fed through databases and other data sources. It is the place where data analytics takes place. The data warehouse includes tiers: a front end – where the user can access analysis, dashboards, reports; the middle tier where the analytics are prepared; and the back-end, the database server, where the data is loaded and stored.[287]

---

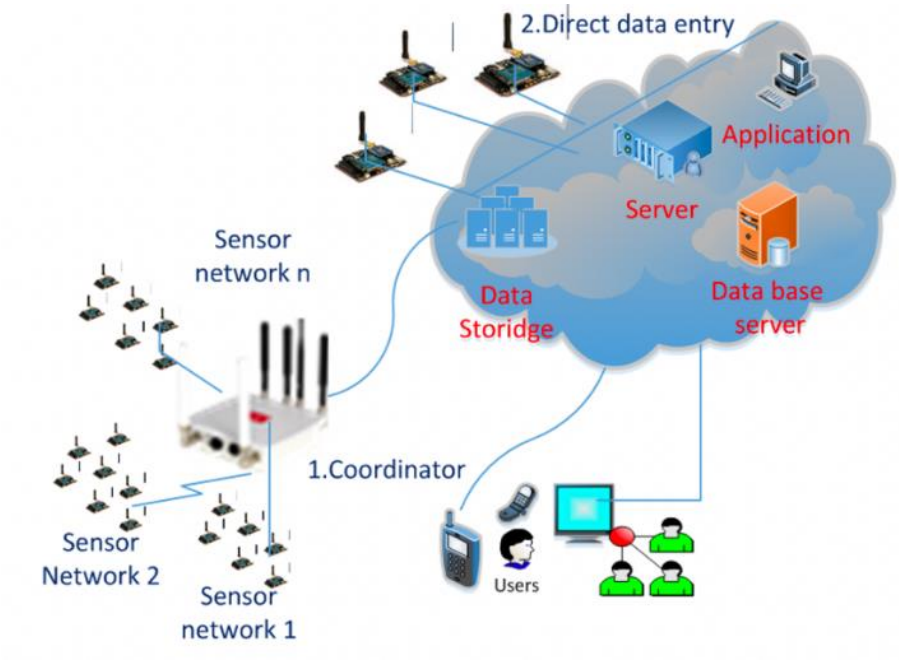
[284] See : https://www.opc-router.de/maschinendatenerfassung/#EdgeFog

[285] The following details are taken from Tsvetanov, F.A. (2020) : Storing data from sensor networks, IOP Conf. Series : Materials Science and Engineering, 1032

[286] In European Commission (2021) : Preliminary report – sector inquiry into consumer internet of thing SWD (2021) 144, survey respondents speak about ‘data lakes’ for the pre-processing whereas processing of data happens in structured ‘data warehouses’.

[287] See : https://aws.amazon.com/data-warehouse/

Figure 16: Sensor network with a cloud structure



Source: Tsvetanov 2020

The middle tier is the area where big data and machine learning have a role. Using algorithms and self-learning systems, insights are gained from the large amount of data collected. This can be used to optimise processes and detect emerging error situations (predictive maintenance).

Machine data is normally stored in order to generate data series that can be analysed. Where and how the data is stored depends on the structure and the application. The most sensible way to store machine data is in a central database. Classic relational databases are often used for this purpose. In these databases (e.g. Microsoft SQL Server, Oracle, MySQL), the data is stored in fixed table structures and can be made available from there to different target systems.

The NoSQL databases mentioned above, where data is not stored in tables but in loose structures, allows a completely new type of access, larger amounts of data and a high degree of flexibility with regard to the structure of the data. Special database types are available for different types of data (document-oriented, graph databases, time series, key-value, multi-value).

B.2 Examples of use cases of sensors and MGD

B.2.1 ABB industrial robots

What is highlighted in the literature and by experts alike are the opportunities and new business models which are linked to systematic data analysis and intelligent data fusion. Technology-based service models can be identified in a number of industries. An example mentioned in the literature is the "Remote service for robots" by ABB.[288] Already since 2007, ABB offers monitoring and optimization of its robots on the premises of clients. The signals from the robot controller are read out and sent from the production site to ABB. ABB than predicts from the transmitted data when individual components will lead to errors in the production and reports to the client. Specific sensor data from the engine or other components are used for this analysis. For this 'condition monitoring', data from the robots is transmitted to ABB at regular intervals. Customers can log on to the ABB Internet portal and check data on their robots. The portal shows how many robots are operating without restrictions and which ones will soon require maintenance. ABB customers can choose to monitor the data themselves or to have it done by ABB experts. Different response times can be defined in the service contract so that the monitoring and reporting of critical events can be adapted to production conditions.

B.2.2 Producer of chainsaws in machinery sector

STIHL, a producer of high-end chainsaws, hedge trimmers and brush cutters has, like many other firms in the machinery sector its own range of automatic stations, handling and transport equipment, robot stations and automatic testing machines allowing for a relatively high automatization degree. These machines use a range of components by different providers including programmable logic controllers (PLC) of the Simatic series and CNC systems (Siemens), robots of different producers, sensors and actuators. In order to bring these different components to work optimal in a highly automated process and to have highest quality standards, the company uses a data management and visualisation tool. In this case, the company needs an event-controlled reading of process values from a PLC which is then transferred to an analytics' software. This is channelled through an OPC router. This architecture is the same in all of the companies geographically spread production plants. Every single assembly line communicates through the OPC router with a local SQL server, which receives and keeps all the information necessary for production and that accumulates during the processes. Central databases work on a higher level, combining information from all assembly lines (here in a manufacturing management system (MES)).

External data such as order information is connected through an enterprise resource planning software (ERP) of SAP enabling the linking of the ERP system with STIHL's SQL server of any assembly line, and from there through the OPC router to the local database of the assembly line and then to the MES.

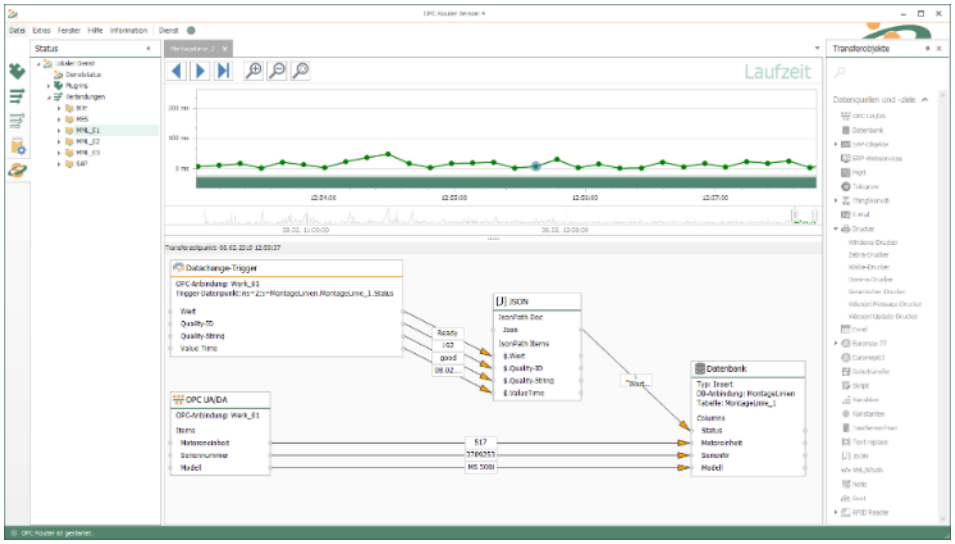
Information from the assembly line starts with the individual series number of a product, which is attached to any process within the production including quality control. All is documented in an individual data batch which is integrated in the assembly line's individual database.[289]

---

[288] Zentralverband Elektrotechnik- und Elektronikindustrie (ZVEI) (2016): Die Elektroindustrie als Leitbranche der Digitalisierung.

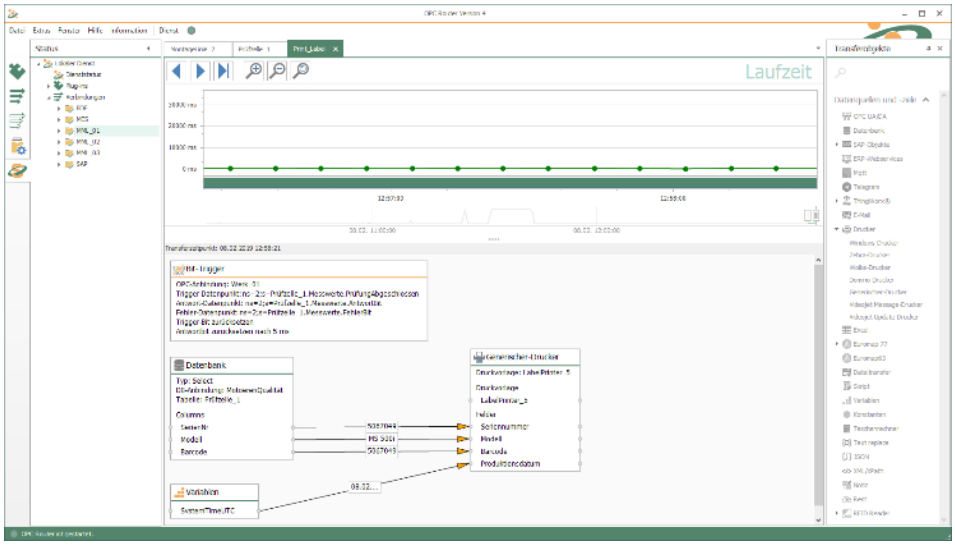
[289] See : https://www.maschinenmarkt.vogel.de/opc-router-steigert-automatisierungseffekt-a-815154/

Figure 17: Graphic interfaces (1)



Source: see https://www.maschinenmarkt.vogel.de/opc-router-steigert-automatisierungseffekt-a-815154/

Figure 18: Graphic interfaces (2)



Source: see https://www.maschinenmarkt.vogel.de/opc-router-steigert-automatisierungseffekt-a-815154/

B.2.3 Chemical industry

In the chemical industry many companies are realizing the significant value of IoT in their organization. Connected operations begin by connecting existing devices and assets to take advantage of these newly connected things. This enables the companies to have a better overview of every part of the operations from the collection of data from sensors that can help optimize operations. There are generally two main ways of introducing IoT in the chemical industry. First, remote operations which allow the company to monitor its equipment. Previously, a company might have had to send employees to investigate a malfunctioning piece of equipment, now many problems can be resolved directly. Second, advanced analytics which help the company understand what is coming from the sensors. Advanced analytics can help raise the understanding what happens in a chemical plant. Within advanced analytics, predictive maintenance is commonly used. Predictive maintenance analyses the historical

performance data of production units and their machinery to forecast when equipment is likely to fail.

B.2.4 Consumer Internet of Things

A current study by DG Competition addresses the consumer IoT.[290] The main difference to the technical functions of sensors and data transmission is that data is collected either automatically in the background (similar to the IoT sensors), through direct manual input (for example when a user sets up an account or registers a device), and through the use of the device or service. Thus, personal, behavioural, and user data can be distinguished.

Consumer devices are full of sensor nodes and as such include all the features to store and transmit data. Consumer devices need power supply – they either work with batteries or through a power cable. This is maybe a key difference to those sensors in harsh or remote environments, for which a strategy is needed to optimize data transfer and the life of the sensor. For consumer devices, this is much less of a concern, and this also enables a sensor node to collect and transmit data permanently. Sensors on devices may also have sufficient storage and processing capacity allowing smart watches and fitness bracelets for example to process and store data on the device and thus provide a 'desired user experience'.[291] Since sending of data from a sensor to an external network is rather resource intensive, many devices process and store the data on the device. This data can be transferred from the smart device to other processing locations such as a cloud or a company platform. This tends to be done when processing requires complex computation for which the external device would not be equipped for.

B.2.5 Sport data

One classical case where now more automated/machine data collection is adopted comes from the sports industry. The commercial exploitation of collected data by sports governing bodies, the marketing of sports data to betting and media organisations, licensing to the gaming industry, and partnering of sports event owners with IT companies to develop software to collect, manage, and disseminate sports data have important economic impacts on the sports business.[292]

The sector relies to a large extent on the protection rights defined and regulated by the Database Directive. It is a good example to demonstrate the variety of data and data collection means. Sports data can categorised in event data and performance data. Event data may include information on weather, temperature or attendance, while performance data related to single events may include points, fouls, goals, assists, corners, etc. as well as individual athletes' performance (e.g. distance covered, time in match, speed of service). Event data is collected by manual research and observation of the sport event. Performance data is collected through camera-based systems and software that transforms the information into statistics or sensors woven into the fabrics of the athletes' wearables. The data cells transmit the information wirelessly to a computer, the software displays the performance data. A further

---

[290] European Commission (2021) : Preliminary report – sector inquiry into consumer internet of things, SWD (2021) 144.

[291] Ibid.

[292] See Christian Frodl, C (2015) : Commercialisation of Sports Data : Rights of Event Owners over Information and Statistics Generated About Their Sports Events.

distinction concerns raw and refined data. In this example, the raw data may refer to the single event and single performance data collected while refined data relates to aggregated and cumulated information in form of statistics and its visualisation. In case of refinement of data, each step may create new proprietary rights. Protection of databases containing sports data may therefore change depending on the level of refinement of the raw data and its visualisation.

C Annex C: Survey

Since one of the key options concerns legal clarity on the inclusion or exclusion of machine-generated data (MGD) in the context of IoT, the survey aimed to address a wider range of industries which are either already in the core of IoT or expected to be so in the near future. A key source on the uptake of IoT was the survey results from the "Advanced Technologies for Industries" project (see: https://ati.ec.europa.eu). Based on a range of key indicators, the ATI project has a geographic focus on seven EU countries with Denmark, Italy, Germany, France, Poland, Spain, and Sweden. These countries, plus Romania and the Netherlands were initially selected for an industry survey for the present study.

In a second step, a matching of the key industry sectors as used in the ATI project with the NACE classification was done. Based on Eurostat's structural business statistics, the selected countries' numbers of companies by size class in the relevant sectors was used to populate the survey sample.

Given the specificity of the subject, it was decided to purchase emails from the B2B service provider Leadiro plc. The provider offers emails from LinkedIn. Search by function/role/title of the job were limited to "legal" and "IT". Per company, one relevant email was selected.

The survey was created using LimeSurvey due to its advanced branching options. The survey was initially open in 2021 until June 30 but given the low response rate, extended to July 31.

A total of 1781 persons were contacted directly through Limesurvey on 9/6/2021. About 25% of all emails bounced back while 0.45% were opting out. An additional 240 were sent through a dedicated email (databasedirectice@technopolis-group.com) on 14/06/2021. Reminders to the first batch were sent on 22.06.2021 and 09.07.2021, reminders to the second, smaller batch were sent on the 28/06/2021 and on 12.07.2021.

In order to increase the survey response rate further, the Directorate-General for Research and Innovation (DG RTD) contacted their stakeholders o 24/25.07.2021 and invited them to participate.

The study team then provided translated surveys and sent out another batch of 620 emails on 28.07.2021 with personalised emails and the surveys in the different languages.

In order to promote the survey, a number of efforts were undertaken:

- The survey link was sent out to the approximately 5,000 members of the German Association for the Protection of Intellectual Property (GRUR) on 23/06/2021.
- The survey was promoted through nine different posts (3 different texts with 3 different images) to a targeted LinkedIn audience (again, the identified industries in combination to legal and IT positions between 23-29/06/2021. This generated a total of 7'297 views. However, of this total, only 14 people clicked through to the survey (with no guarantee of completing it).
- The survey was further advertised on Twitter, the Technopolis Group site on LinkedIn, and the homepages of the consortium partners.
- The European Commission reached out to BDI, Business Europe, Orgalim, and EEN asking for support in diffusing the survey among members.

Overall, 347 surveys were started, but only 114 were answered completely or to a workable degree.

C.1 Survey questionnaire

Survey to support an Impact Assessment for the review of the Database Directive

The European Commission, **DG CNECT is currently preparing an impact assessment on a potential revision of the Database Directive**, which was adopted in 1996. The assessment aims to understand if the Directive is still fit for purpose in the data economy, and in particular the relationship of the Directive with machine generated data.

Under EU law, protection of databases against unauthorised use may take two forms with Copyrights – which protect the database structure (a creative act) and the Database rights (or 'sui generis' right). This right protects the investment in either obtaining, verifying or presenting the database contents, but not the content as such. **The current survey concerns only the database (sui generis) right.**

On behalf of the European Commission, **Copenhagen Economics and Technopolis Group conduct this industry survey** as part of a study in the context of the review of the Database Directive. **It aims to obtain insights on the use (and usefulness) of the Directive, views on possible revisions (policy options) and their impacts.** Given the specificity of the topic, we would expect that senior legal and/or IT managers will be best placed to respond to the questions. Filling out the questionnaire should not take longer than 15 minutes.

Please see also our letter of recommendation here (https://www.technopolis-group.com/wp-content/uploads/2021/06/Letter-of-recommendation-for-the-Database-Directive-study.pdf).

We would appreciate very much your support in improving this legislation to the benefit of the business sector.

Please send any queries to the following email address: databasedirective@technopolis-group.com (mailto:databasedirective@technopolis-group.com).

In accordance to GDPR requirements, the survey results will only be used in aggregated form (at country or industry level). For more information on how we treat the collected data, please see our privacy policy (https://www.technopolis-group.com/privacy-policy/).

There are 54 questions in this survey.

I. YOUR ORGANISATION AND YOUR PROFILE

In which country is your organisation located?
(country of residence of operating unit)
*

Please choose **only one** of the following:

- ◯ Austria
- ◯ Belgium
- ◯ Bulgaria
- ◯ Croatia
- ◯ Cyprus
- ◯ Czechia
- ◯ Denmark
- ◯ Estonia
- ◯ Finland
- ◯ France
- ◯ Germany
- ◯ Greece
- ◯ Hungary
- ◯ Ireland
- ◯ Italy
- ◯ Latvia
- ◯ Lithuania
- ◯ Luxembourg
- ◯ Malta
- ◯ Netherlands
- ◯ Poland
- ◯ Portugal
- ◯ Romania
- ◯ Slovakia
- ◯ Slovenia
- ◯ Spain
- ◯ Sweden
- ◯ Other

Approximately how many people are currently employed (full-time or part-time) in your organisation, including all branches, divisions and subsidiaries? *

ℹ Choose one of the following answers
Please choose **only one** of the following:

- ◯ 10-49
- ◯ 50-249
- ◯ 250-499
- ◯ More than 500
- ◯ Don't know

Which of the following industries best describes your organisation's primary business? *

Choose one of the following answers
Please choose **only one** of the following:

- ○ Agriculture
- ○ Banking
- ○ Insurance
- ○ Business or professional services, excluding IT and Legal services
- ○ IT services
- ○ Legal services
- ○ Healthcare
- ○ Process manufacturing
- ○ Discrete manufacturing
- ○ Retail trade
- ○ Wholesale trade
- ○ Telecommunications
- ○ Media
- ○ Transport and logistics
- ○ Utilities
- ○ Oil and gas
- ○ Other

Which of the following best describes your position within your organisation? *

Choose one of the following answers
Please choose **only one** of the following:

- ○ IT expert
- ○ Senior data engineer
- ○ Legal expert
- ○ Senior Legal Officer/ Counsellor
- ○ Other

Please note: "others" will be seeing both the IT and the legal part of the questionnaire. There is always an option to respond with "no answer/don't know".

With regard to databases and machine generated data (see below), please select what describes you and your organisation best! *

Check all that apply
Please choose **all** that apply:

- ☐ User of databases including primarily machine generated data
- ☐ User of databases including primarily other type of data than machine generated data
- ☐ Maker of databases including primarily machine generated data
- ☐ Maker of databases including primarily other type of data than machine generated data
- ☐ Don't know / Not applicable

What do we mean with machine generated data:

Machine generated data (MGD) are data recorded, collected, or generated independently of direct human intervention by:

- sensors processing information received from equipment, software or machinery, whether virtual or real
- computer processes, applications or services.

Please note that MGD includes machine generated data that are the result of observation of (acts of) humans (e.g. which pages within a website a person visits). However, data that are the result of humans consciously providing information/choices are excluded (persons typing in their name, address etc. to create an account).

Which of the categories would further describe your organisation?

Check all that apply
Please choose **all** that apply:

- ☐ Data holder (organisations competent to decide about the use of data regardless of who collected, stored, processed or disseminated the data, e.g. private sector companies, which are the IoT product/service providers)
- ☐ Data co-producer (organisations which are IoT product/service users, e.g., transport companies, airlines)
- ☐ Data re-user (players interested in accessing data and/or re-using the data from a data holder, e.g. data analytics companies, users and repairers of smart devices)
- ☐ Data intermediary (organisations that enable data holders to share their data, e.g., data marketplaces or industrial data platforms which enable sharing of data)
- ☐ Don't know / Not applicable

II. QUESTIONS ON THE TECHNOLOGICAL-COMMERCIAL CONTEXT OF DATA AND DATABASES IN YOUR COMPANY

How important are the following use cases of machine generated data in your organisation? *

Only answer this question if the following conditions are met:
Answer was 'Other' *or* 'Senior data engineer' *or* 'IT expert' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	Very important	Somewhat important	Neutral	Rather unimportant	Not important	Don't know
Generating and using data for internal use only (e.g., optimisation of processes)	◯	◯	◯	◯	◯	◯
Selling/licensing data to third parties for a fee (commercialising data)	◯	◯	◯	◯	◯	◯
Providing free services to third parties (e.g., clients, suppliers) based on data	◯	◯	◯	◯	◯	◯
Providing fee-based services to third parties (e.g., clients, suppliers) based on data analytics	◯	◯	◯	◯	◯	◯
Obtaining data from third parties through contracts/licensing for own data analytics or optimisation of production processes/devices	◯	◯	◯	◯	◯	◯
Obtaining data from device manufacturers through contracts/licensing to provide aftermarket services such as repair and maintenance and development of applications	◯	◯	◯	◯	◯	◯
Sharing data for free on a mutual basis with business partners on a platform	◯	◯	◯	◯	◯	◯
Using data to design innovative solutions/products	◯	◯	◯	◯	◯	◯
Using big data for training AI	◯	◯	◯	◯	◯	◯
Other, please specify below	◯	◯	◯	◯	◯	◯

What is your opinion on the use of machine generated data? *

Only answer this question if the following conditions are met:
Answer was 'Other' *or* 'Senior data engineer' *or* 'IT expert' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	Strongly agree	Agree	Disagree	Strongly disagree	Don't know
Data from our internal processes are valuable only for organisation-internal deployment	◯	◯	◯	◯	◯
My organisation invested substantially in our data infrastructure (IT system, database structure, protection, training, etc.)	◯	◯	◯	◯	◯
Selling or licensing our data to third parties provides extra profits	◯	◯	◯	◯	◯
Selling or licensing our data to third parties endangers our exclusive data analytics insights	◯	◯	◯	◯	◯
Selling or licensing our data to third parties enables the wider user/consumer benefits	◯	◯	◯	◯	◯
Our organisation benefits from buying third party data	◯	◯	◯	◯	◯
Sharing data for free on a mutual basis with business partners/on a platform benefits our organisation	◯	◯	◯	◯	◯
Our organisation experienced difficulties when requesting access to other companies' data	◯	◯	◯	◯	◯
Other, please specify below	◯	◯	◯	◯	◯

Please specify:

Only answer this question if the following conditions are met:
Answer was 'Other' *or* 'Senior data engineer' *or* 'IT expert' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please write your answer here:

III. GENERATION AND COLLECTION OF DATA AND DEVELOPMENT OF DATABASES

Which type of database for machine generated data do you use from a technological point of view? *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Check all that apply
Please choose **all** that apply:

- [ ] SQL
- [ ] NoSQL
- [ ] In-memory
- [ ] Distributed databases
- [ ] We are not running databases ourselves but using cloud or other data services
- [ ] Don't know
- [ ] Not applicable
- [ ] Other:

Are you cooperating with other companies in establishing and running databases?

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Choose one of the following answers
Please choose **only one** of the following:

- ( ) Yes, to a wide extent
- ( ) Yes, in many cases
- ( ) Yes, only in few cases
- ( ) No, and we do not plan to in the next 12 months
- ( ) No, but we plan to
- ( ) Don't know

Are these cooperating partners…

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'Yes, only in few cases' *or* 'Yes, in many cases' *or* 'Yes, to a wide extent' at question '11 [C2]' (Are you cooperating with other companies in establishing and running databases?)

Check all that apply
Please choose **all** that apply:

- [ ] Firms within the same industry
- [ ] Firms from other sectors
- [ ] Suppliers
- [ ] Customers
- [ ] Other:

If you collect machine generated data, where is the data stored? *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Choose one of the following answers
Please choose **only one** of the following:

- ( ) Directly in a database in my company
- ( ) In a database in a partner company
- ( ) In the cloud
- ( ) In an intercompany network
- ( ) Data remain in IoT device, but can be read from a distance at any moment
- ( ) Don't know / Not applicable
- ( ) Other

If your company provides supplementary data analytics, will the resulting data be stored in: *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Choose one of the following answers
Please choose **only one** of the following:

- ( ) The original database
- ( ) A new database
- ( ) Don't know / Not applicable
- ( ) Other

Is machine generated data collected in your company considered as a by-product of your primary activity?

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

❗ Choose one of the following answers
Please choose **only one** of the following:

- ◯ Yes, it is central for the good functioning of our operation
- ◯ Yes, it is secondary to our business
- ◯ No, but it is becoming increasingly relevant in our operation
- ◯ No, it isn't
- ◯ Don't know / Not applicable

If machine generated data is collected by your company, who invests and takes the economic risks of setting up and maintaining the database?

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

❗ Choose one of the following answers
Please choose **only one** of the following:

- ◯ My company
- ◯ Manufacturer of sensors
- ◯ Operator of devices (e.g. in tractors, autonomous cars)
- ◯ Participants in a network (i.e. different stakeholders contributing to database)
- ◯ Don't know / Not applicable
- ◯ Other [ ]

Please provide your best estimate of cost per annum to set up and maintain on average a database in your company: *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	<500 EUR	< 5,000 EUR	< 20,000 EUR	< 100,000 EUR	Above 100,000 EUR	I don't know
One-time investment	◯	◯	◯	◯	◯	◯
Operating costs	◯	◯	◯	◯	◯	◯

Please elaborate as you wish:

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please write your answer here:

If the machine generated data is generated/collected by another person or company (controller of the collection device), does the data holder grant you:

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

❗ Check all that apply
Please choose **all** that apply:

- ☐ Right to access/visualise the data
- ☐ Right to download/ use the data
- ☐ Ownership on data
- ☐ Right to re-use/modify the data
- ☐ Access was denied
- ☐ Data was shared on a mutual basis
- ☐ Don't know / Not applicable
- ☐ Other: [ ]

Would you be able to estimate the costs to establish a contract (e.g., data sharing agreement) to access the data?

(contracting costs may include internal and external resources such as legal service, if your organisation contracted in multiple instances please provide an estimation for an average contract, organisation can incur costs even when access was denied)

*

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

❗ Choose one of the following answers
Please choose **only one** of the following:

- ◯ Less than 1,000 EUR
- ◯ 1,000 – 5,000 EUR
- ◯ 5,000 – 10,000 EUR
- ◯ 10,000 – 20,000 EUR
- ◯ More than 20,000 EUR
- ◯ Don't know
- ◯ Other [ ]

IV. SHARING AND ACCESS TO DATA AND DATABASES

In the last 12 months, how often has your company shared its databases containing machine generated data or requested access to/use of another company's databases? *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	On a daily basis	Several times	Only a few times	Not applicable	Don't know
Sharing databases	◯	◯	◯	◯	◯
Accessing other databases	◯	◯	◯	◯	◯

In the last 12 months, has your company denied a request of access for databases containing machine-generated data? *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

❗ Choose one of the following answers
Please choose **only one** of the following:

- ◯ Yes
- ◯ No
- ◯ Don't know
- ◯ Not applicable

If Yes, for which reason?

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'Yes' at question '22 [D2]' (In the last 12 months, has your company denied a request of access for databases containing machine-generated data?)

❗ Check all that apply
Please choose **all** that apply:

- ☐ Too costly infrastructure investments (allowing APIs, define protocols,...)
- ☐ Potential loss of competitive advantage from exclusive access to data
- ☐ Potential loss of revenues of exclusive rights on data
- ☐ Other: [ ]

Companies may become the sole source of certain data contained in protected databases establishing a de facto monopoly. Please indicate, to which statement you agree or disagree. *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	Strongly agree	Agree	Disagree	Strongly disagree	Don't know
For databases containing machine-generated data I cannot obtain this or similarly useful data from other sources	◯	◯	◯	◯	◯
For other type of databases I cannot obtain this or similarly useful data from other sources	◯	◯	◯	◯	◯

When you tried to obtain access to databases containing machine generated data, did you encounter any problems?

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Choose one of the following answers
Please choose **only one** of the following:

- ◯ Yes
- ◯ No
- ◯ Don't know/ Not applicable

If Yes, for which reason?

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'Yes' at question '25 [D4]' (When you tried to obtain access to databases containing machine generated data, did you encounter any problems?)

Check all that apply
Please choose **all** that apply:

- ☐ No market for the type of database needed
- ☐ The data in the database was kept secret
- ☐ The database was legally protected and there was no licensing available
- ☐ Too high costs to obtain access to the database (licensing costs, infrastructure developments,…)
- ☐ The database was protected with technical measures
- ☐ Lack of interoperability
- ☐ Don't know / Not applicable
- ☐ Other:

Do you think there should be mandatory access e.g through a licence, open access, etc. to machine generated data: *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

	...for the co-producer	…for the re-user	…for the intermediary	…for none of them	I don't know
All machine generated data should be accessible…	☐	☐	☐	☐	☐
Certain data or databases should be accessible, depending on their nature (e.g. of specific public interest, such as mobility data needed for decisions on the extension or reduction of the road infrastructure)…	☐	☐	☐	☐	☐
Certain data or databases should be accessible, depending on the business situations (e.g.sole source database, where data is not available from other sources) …	☐	☐	☐	☐	☐
There should not be mandatory access…	☐	☐	☐	☐	☐

What are the reasons why a mandatory access should not be granted?

Only answer this question if the following conditions are met:
(((A4.NAOK (/admin/questions/sa/view/surveyid/726265/gid/382/qid/4878) == "-oth-" or A4.NAOK (/admin/questions/sa/view/surveyid/726265/gid/382/qid/4878) == "A1" or A4.NAOK (/admin/questions/sa/view/surveyid/726265/gid/382/qid/4878) == "A2") and (D5_R004_C001.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R004_C002.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R004_C003.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R001_C004.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R002_C004.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R003_C004.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900))))

Check all that apply
Please choose **all** that apply:

- ☐ it would affect my business interest
- ☐ it would affect my freedom to contract too much
- ☐ Other:

If you think there should be access to databases containing machine generated data, how should access be granted?

Only answer this question if the following conditions are met:
(((A4.NAOK (/admin/questions/sa/view/surveyid/726265/gid/382/qid/4878) == "-oth-" or A4.NAOK (/admin/questions/sa/view/surveyid/726265/gid/382/qid/4878) == "A1" or A4.NAOK (/admin/questions/sa/view/surveyid/726265/gid/382/qid/4878) == "A2") and (D5_R004_C004.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R001_C001.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R001_C002.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R001_C003.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R002_C001.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R002_C002.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R002_C003.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R003_C001.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R003_C002.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900) or D5_R003_C003.NAOK (/admin/questions/sa/view/surveyid/726265/gid/385/qid/4900))))

Choose one of the following answers
Please choose **only one** of the following:

- ◯ As a free access
- ◯ As a compulsory license on FRAND terms
- ◯ Don't know
- ◯ Other

Suppose databases containing machine generated data would be mandatorily accessible. Could you provide an estimate of the potential gains and losses in revenues for the data holder? *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	Revenues not affected	Less than 5% of the total revenues	5-10% of the total revenues	10-20% of the total revenues	more than 20% of the total revenues	Don't know / Not applicable
In case of expected gains	◯	◯	◯	◯	◯	◯
In case of expected losses	◯	◯	◯	◯	◯	◯

How would you rate the immediate impacts on your business of introducing a specific access regime? (Please select on a 1-5 scale from 1 – Very negative and 5 – Very positive) *

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	1 - Very negative	2- Negative	3- Neutral	4 - Positive	5- Very positive	Don't know
Access right only on databases of public interest	◯	◯	◯	◯	◯	◯
Access right for scientific research	◯	◯	◯	◯	◯	◯
Access right to databases necessary for developing new innovative products/new markets	◯	◯	◯	◯	◯	◯
Access right to databases for providers of aftermarket services such as repairing and maintenance etc.	◯	◯	◯	◯	◯	◯
Access right only on sole source databases (i.e. data not available from other sources)	◯	◯	◯	◯	◯	◯
Other, please specify below	◯	◯	◯	◯	◯	◯

Please specify:

Only answer this question if the following conditions are met:
Answer was 'IT expert' *or* 'Senior data engineer' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please write your answer here:

V. SUI GENERIS AND OTHER TYPES OF DATABASE PROTECTION MEANS

EU legal protection of databases: the *sui generis right*:

The sui generis right is a special intellectual property right protecting substantial investment in either obtaining, verifying or presenting the database contents under the EU's Database Directive.

Under this right, substantial extractions and re-use are subject to rightsholders' prior consent.

Under the Directive, some contractual restrictions on the actions of lawful users of a database are ineffective. The sui generis right protects against taking of substantial parts as well as insubstantial parts that may cause damage to the investment due to use in competition. However, if a database is not protected by the sui generis right (and by copyright), the operator of the database is free to impose whatever contractual restrictions are permitted under the relevant national law. Further protection may arise at national level, mainly via unfair competition. This sui generis right is not granted to entities operating outside the EU. So, an investment by a firm based in the United States in production of a database does not give rise to sui generis right in the EU (and there is no such right in the US).

How familiar are you with the Database Directive? *

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

ⓘ Check all that apply
Please choose **all** that apply:

- ☐ I have been subject to a legal claim due to a database right
- ☐ I have enforced the database right myself
- ☐ I am familiar with the Database Directive
- ☐ I am aware of it but I am not familiar
- ☐ I do not know the Database Directive
- ☐ Other:

How would you rate the following alternative means of protecting databases containing machine generated data:
*

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	Very effective	Partially effective	Not effective	Don't know
Exclusive rights (e.g. sui generis right)	◯	◯	◯	◯
Competition law (enforcement through litigation)	◯	◯	◯	◯
Contracts (e.g., through detailed licensing agreements)	◯	◯	◯	◯
Trade secrets (e.g., through confidentiality obligations)	◯	◯	◯	◯
Technological measures (e.g., access restrictions, encryption)	◯	◯	◯	◯
Smart contracts	◯	◯	◯	◯
Other, please specify below	◯	◯	◯	◯

Please specify:

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please write your answer here:

How costly do you assess the following means of protecting databases containing machine generated data to be for the data holder: on a scale from 1 (no cost) to 5 (very high cost)? *

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	1 - No costs	2 - Minor	3 - Considerable	4 - High	5 - Very high	I don't know
Exclusive rights (e.g. sui generis right)	◯	◯	◯	◯	◯	◯
Competition law (enforcement through litigation)	◯	◯	◯	◯	◯	◯
Contracts (e.g., through detailed licensing agreements)	◯	◯	◯	◯	◯	◯
Trade secrets (e.g., through confidentiality obligations)	◯	◯	◯	◯	◯	◯
Technological measures (e.g., access restrictions, encryption)	◯	◯	◯	◯	◯	◯
Smart contracts	◯	◯	◯	◯	◯	◯
Others	◯	◯	◯	◯	◯	◯

In your view, what could be the benefit of an exclusive right that covers databases containing machine generated data and that you can exercise against anybody, not just people with whom you have a contractual relationship?

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

ℹ Check all that apply
Please choose **all** that apply:

- ☐ It offers the opportunity to better regulate the relationship with clients, including licences
- ☐ It offers protection against third-party infringements (i.e. unauthorised use of machine generated data) which is highly needed
- ☐ It offers protection without triggering unnecessary costs
- ☐ It would provide an incentive to license or share the data from my database
- ☐ In my view there is no considerable benefit
- ☐ Don't know
- ☐ Other:

Which means do you use or plan to use to protect the databases containing machine generated data against unauthorised use?
*

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please choose the appropriate response for each item:

	Not using and no plans to use	Already using	Plan to use in the next 12 months	Not aware of	Don't know
Database sui generis right	◯	◯	◯	◯	◯
Copyright	◯	◯	◯	◯	◯
Competition law	◯	◯	◯	◯	◯
Contracts (e.g., through detailed licensing agreements)	◯	◯	◯	◯	◯
Trade secrets (e.g., through confidentiality obligations)	◯	◯	◯	◯	◯
Technological measures (e.g., access restrictions, encryption)	◯	◯	◯	◯	◯
Smart contracts	◯	◯	◯	◯	◯
Other, please specify below	◯	◯	◯	◯	◯

Please specify:

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?)

Please write your answer here:

To what extent is sui generis database right useful for the purposes of your data protection compliance, e.g. to protect personal data from unauthorized access, etc.?

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* **I** have enforced the database right myself' *or* '**I** have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

ⓘ Choose one of the following answers
Please choose **only one** of the following:

- ◯ Useful
- ◯ Potentially useful, but we have not used it in practice
- ◯ Not useful at all
- ◯ Not applicable / Don't know

VI. POLICY OPTIONS

What is your view on the following options for a potential change in the Database Directive specific to machine generated data ?
*

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* **I** have enforced the database right myself' *or* '**I** have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please choose the appropriate response for each item:

	Strongly agree	Agree	Disagree	Strongly disagree	Don't know
Machine-generated data would be explicitly excluded from the scope of application of the sui generis right	◯	◯	◯	◯	◯
Replacing the sui generis right, particularly as applied to MGD, with an alternative protective mechanism against certain unauthorized uses that could be categorized as abusive or unfair from competition perspective	◯	◯	◯	◯	◯
Machine-generated data would be included in the scope of application of the sui generis right together with a new data access regime (e.g. new exception or compulsory licences)	◯	◯	◯	◯	◯
The sui generis right would only apply to machine-generated data if the maker of the database takes an additional step, namely taking certain proactive protection measures (such as technical measures)	◯	◯	◯	◯	◯

Technical measurements could be: (i) clearly expressed intention to apply the protection of the sui generis right, e.g. in machine-readable format, (ii) application of reasonable technical protection measures to mark the protected nature of the machine-generated database

What is your view on the following options for remaining with the status quo or a potential general change in the Database Directive? *

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please choose the appropriate response for each item:

	Strongly agree	Agree	Disagree	Strongly disagree	Don't know
The Database Directive would remain in its current form	◯	◯	◯	◯	◯
The sui generis right would be repealed	◯	◯	◯	◯	◯
Public bodies would be excluded from the sui generis right	◯	◯	◯	◯	◯
Extension of lawful user rights	◯	◯	◯	◯	◯
Exceptions to sui generis right would be expanded in line with broader general copyright exceptions like private copying and research	◯	◯	◯	◯	◯
Broaden the sui generis right to include insubstantial parts (with potential extension to exception or the lawful user right).	◯	◯	◯	◯	◯

Compared to the current situation, how would you assess the following options in terms of changes in costs and benefits for your company? Taking 0 as the current situation, please assign a score from -3 to +3 to the costs and the benefits of each option: *

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

	Cost (0 – neutral; -3: very low, +3 very high)	Benefit (0 – neutral; -3: very low, +3 very high)
The sui generis right would be repealed.		
Excluding machine-generated data from the scope of application of the sui generis database right.		
Replacing the sui generis right, particularly as applied to MGD, with an alternative protective mechanism against certain unauthorized uses that could be categorized as abusive or unfair from competition perspective.		
Machine-generated data would be included explicitly in the scope of application of the sui generis right together with a new data access regime for the benefit of the database users (e.g. new exception or compulsory licences) for the benefit of database users .		
The sui generis right would only apply to machine-generated data if the maker of the database takes an additional step, namely taking certain proactive protection measures (such as technical measures).		

Please feel free to elaborate on potential costs and benefits:

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please write your answer here:

Compared to the current situation, how would you assess the following supplementary options for changing aspects of the sui generis right? Taking 0 as the current situation, please assign a score from -3 to +3 to the costs and the benefits of each option: *

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

	Cost (0 – neutral; -3: very low, +3 very high)	Benefit (0 – neutral; -3: very low, +3 very high)
Public bodies would be excluded from the sui generis right		
Extension of lawful user rights		
Exceptions to sui generis right would be expanded in line with broader general copyright exceptions like private copying and research		
Broaden the sui generis right to include insubstantial parts (with potential extension to exception or the lawful user right).		

Please feel free to elaborate on potential costs and benefits:

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please write your answer here:

How would you assess the consequences for your company of excluding databases containing machine-generated data from the scope of application of the sui generis database right ? (Please select on a 1-5 scale from 1 – Very negatively to 5 – Very positively) *

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please choose the appropriate response for each item:

	1 - Very negatively	2- Negatively	3 –Not affected	4 -Positively	5- Very positively	Don't know
Legal protection from ex-employees and other third parties	◯	◯	◯	◯	◯	◯
Obtaining legal certainty	◯	◯	◯	◯	◯	◯
Cost of protection	◯	◯	◯	◯	◯	◯
Access to data	◯	◯	◯	◯	◯	◯
Cost of data access	◯	◯	◯	◯	◯	◯
Enter new markets and compete against other firms	◯	◯	◯	◯	◯	◯
Develop new/value-added products	◯	◯	◯	◯	◯	◯
Investment in database creation	◯	◯	◯	◯	◯	◯
Revenues generated from the production and/or exploitation of databases	◯	◯	◯	◯	◯	◯
Innovation and research activity	◯	◯	◯	◯	◯	◯
Other, please specify below	◯	◯	◯	◯	◯	◯

Please specify:

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please write your answer here:

How would you assess the consequences of including databases containing machine generated data in the scope of application of the sui generis database right together with a new data access regime (e.g. new exception or compulsory licences) for your company? (Please select on a 1-5 scale from 1 – Very negatively to 5 – Very positively) *

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please choose the appropriate response for each item:

	1 - Very negatively	2- Negatively	3 –Not affected	4 -Positively	5- Very positively	Don't know
Legal protection from ex-employees and other third parties	◯	◯	◯	◯	◯	◯
Obtaining legal certainty	◯	◯	◯	◯	◯	◯
Cost of protection	◯	◯	◯	◯	◯	◯
Access to data	◯	◯	◯	◯	◯	◯
Cost of data access	◯	◯	◯	◯	◯	◯
Enter new markets and compete against other firms	◯	◯	◯	◯	◯	◯
Develop new/value-added products	◯	◯	◯	◯	◯	◯
Investment in database creation	◯	◯	◯	◯	◯	◯
Revenues generated from the production and/or exploitation of databases	◯	◯	◯	◯	◯	◯
Innovation and research activity	◯	◯	◯	◯	◯	◯
Other, please specify below	◯	◯	◯	◯	◯	◯

Please specify:

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please write your answer here:

What is your view on the following statements:

An exclusion of machine generated data from the database right increases the risk of… *

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please choose the appropriate response for each item:

	1 - Disagree strongly	2 - Disagree	3 - Neutral	4 – Agree	5 – Agree strongly	Don't know
Contractually based restrictions imposed by the database holders	◯	◯	◯	◯	◯	◯
Decrease of available data due to the decrease of trust in sharing/providing access to databases (Data lock-ins)	◯	◯	◯	◯	◯	◯
Users to lose access and use regime included in the Directive	◯	◯	◯	◯	◯	◯
Other, please specify below	◯	◯	◯	◯	◯	◯

Please specify:

Only answer this question if the following conditions are met:
Answer was 'Legal expert' *or* 'Senior Legal Officer/ Counsellor ' *or* 'Other' at question '4 [A4]' (Which of the following best describes your position within your organisation?) *and* Answer was 'I am familiar with the Database Directive' *or* 'I have enforced the database right myself' *or* 'I have been subject to a legal claim due to a database right' at question '33 [E1]' (How familiar are you with the Database Directive?)

Please write your answer here:

CLOSURE

Many thanks for your responses!

Would you be willing to participate in an in-depth interview on the subject with experts of this study? If so, please provide your email address:

Please write your answer here:

Do you have any general comments?

Please write your answer here:

Many thanks for your participation!

Submit your survey.
Thank you for completing this survey.

C.2 Survey Results

The survey was targeted to specific countries and to specific professionals. One of the core questions the review of the Database Directive addresses is, whether or not (or under which conditions) machine generated data (MGD) should be included or excluded for the scope of the *sui generis* right. Thus, the survey addressed companies in a range of sectors that are either using already substantially IoT or are likely to do so in the near future. Furthermore, countries were selected where IoT plays a key role (such as Germany, Spain, Sweden) but the selection also considered the variety of small and large as well as the geographic spread of EU member states and followed closely the choice of countries as used in the ATI survey[293].

Since potential changes are of key legal relevance, we also addressed legal professional services. To answer the questionnaire, it was sent to IT experts and legal experts in the identified companies.

A few sections of the survey addressed all respondents, however, once the respondent identified him- or herself as either IT or legal expert, the respondent was then only seeing the relevant questions since the questionnaire was branched into two areas: on the one hand a more technical part addressed questions on use cases, the relevance of MGD, access, or costs. These questions were open to IT experts. On the other hand, several questions addressed potential legal changes. These questions were open to the legal experts.

It was however possible to share the survey, therefore, a third category of respondents 'other' was provided. This group of people saw the whole survey but was informed that it was possible to skip questions.

The Commission's steering committee also suggested to provide the survey to research actors. While the design of the survey was not amended, the survey was diffused to research organisations (in the wider sense). These respondents chose very often 'other'. Since some of the opinions from this group is markedly different from the industry perspective, we highlight relevant aspects.

Beside a larger number of respondents from the research sector, there were a few coming from a particular industry association and their national members. Since most of their open responses were coordinated, - in an open public consultation this is considered a campaign – it needs to be treated separately. Therefore, insights from these responses are highlighted.

Overall, in the following analysis it is important to consider that the total number of responses varies per question.

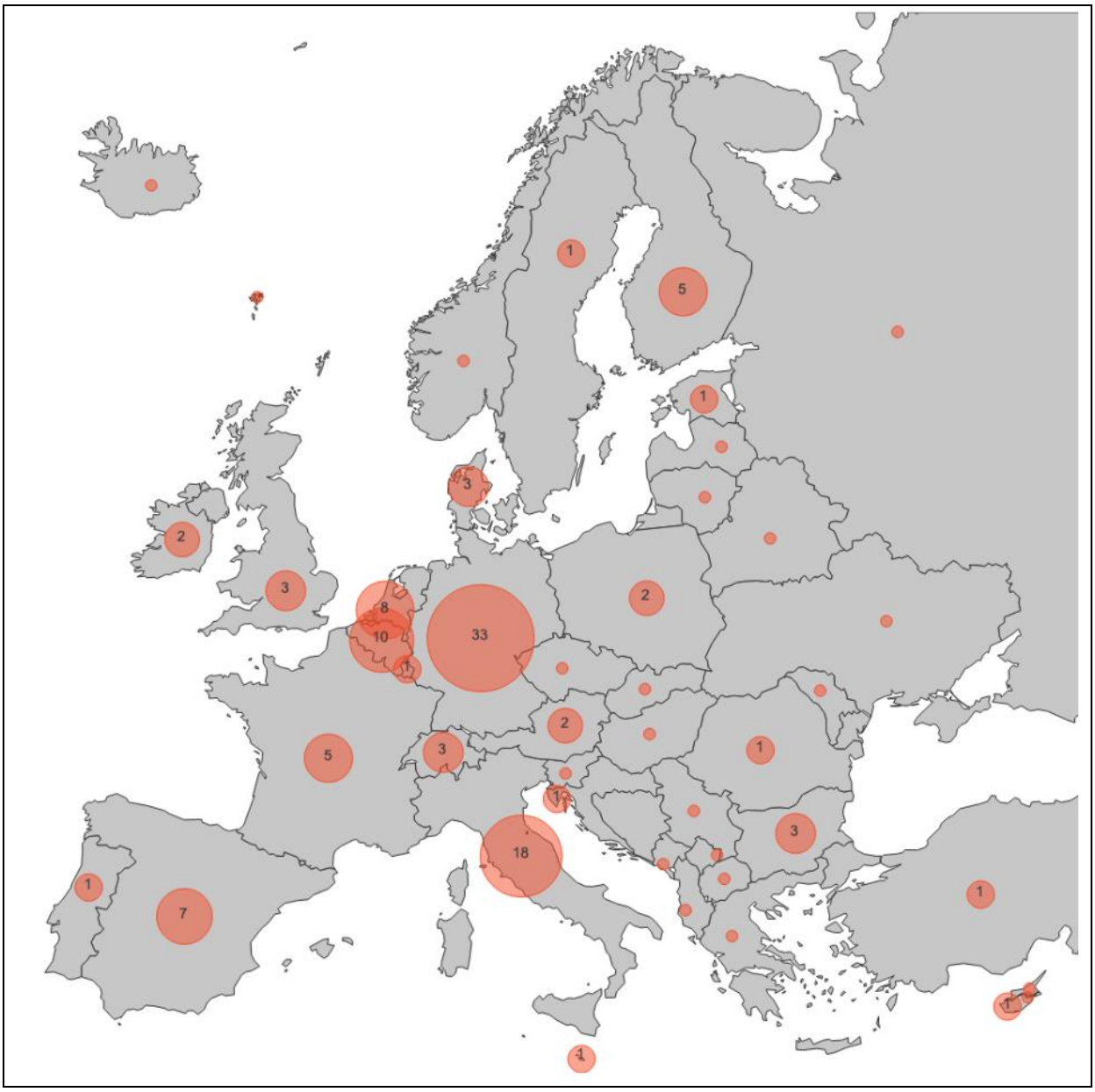
C.2.1 Profile of the respondents

As indicated, there were nine targeted countries. In a late stage, the EC, DG RTD sent out survey links to its stakeholders in order to integrate also the view from the research community. Therefore, responses from other EU countries were obtained as well.

Out of the total of 114 valid responses, 29% (33) came from Germany and 16% (18) from Italy. Responses also come from eighteen other EU Member States, as well as three from non-EU countries (UK, Switzerland, Turkey).

[293] See : https://ati.ec.europa.eu/

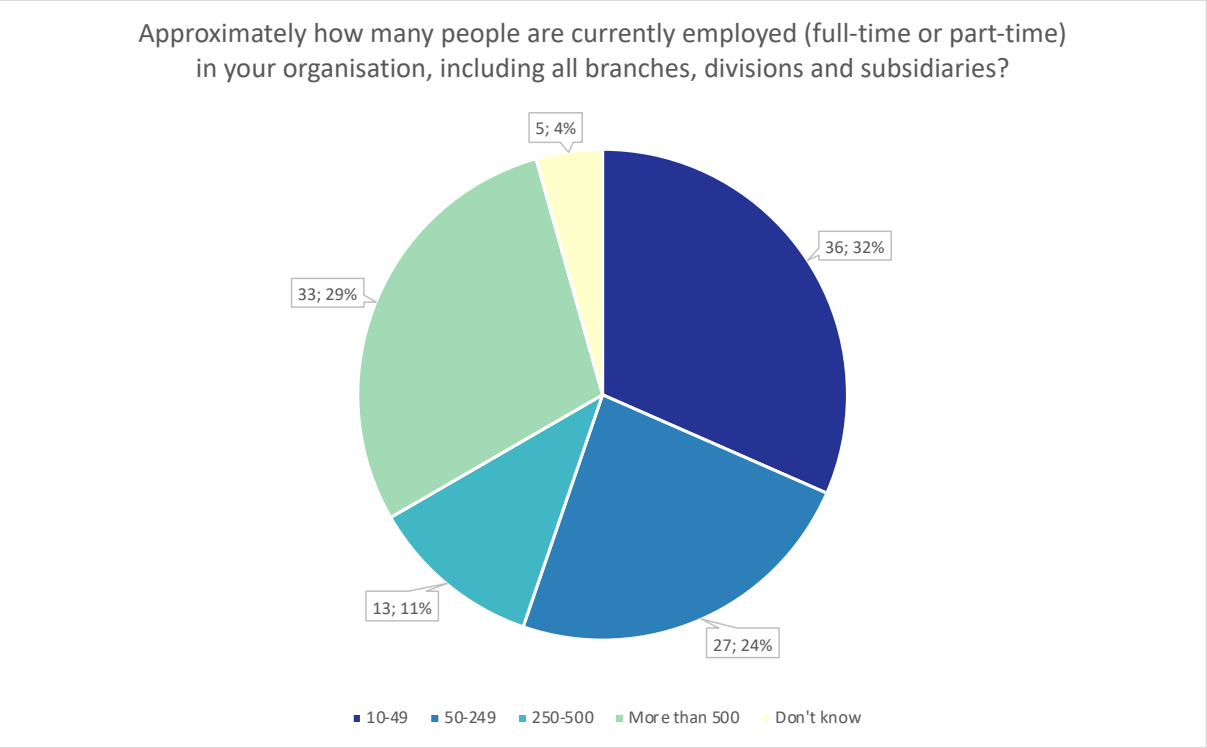
Figure 19: Country of respondents' organisations



Source: Survey for the Database Directive Review
Note: based on A1 n=114

In terms of size of the organisations of the respondents, one third (33%) work at a small organisation/business with up to 49 employees, while another third (33%) work in much larger organisations/businesses with more than 500 employees. Five respondents were not able to provide this information.

Figure 20: Size of organisation

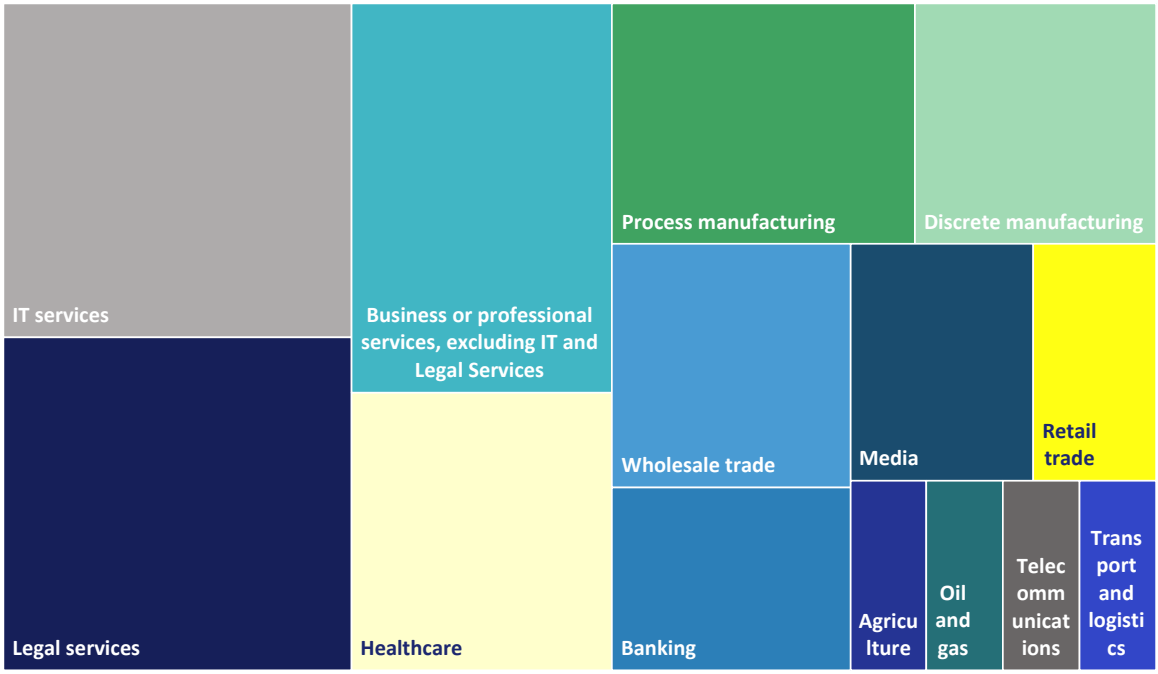


Source: Survey for the Database Directive Review
Note: based on A2: n=109

Out of the 114 valid answers, 53 work in the targeted industries, while 60 respondents chose 'Other'. This includes several (wider) research stakeholders including respondents active in the field of 'RDI', 'civil society' organisations or 'governmental agency'.'

From the available targeted sectors, professionals from legal and IT services were strongly represented.

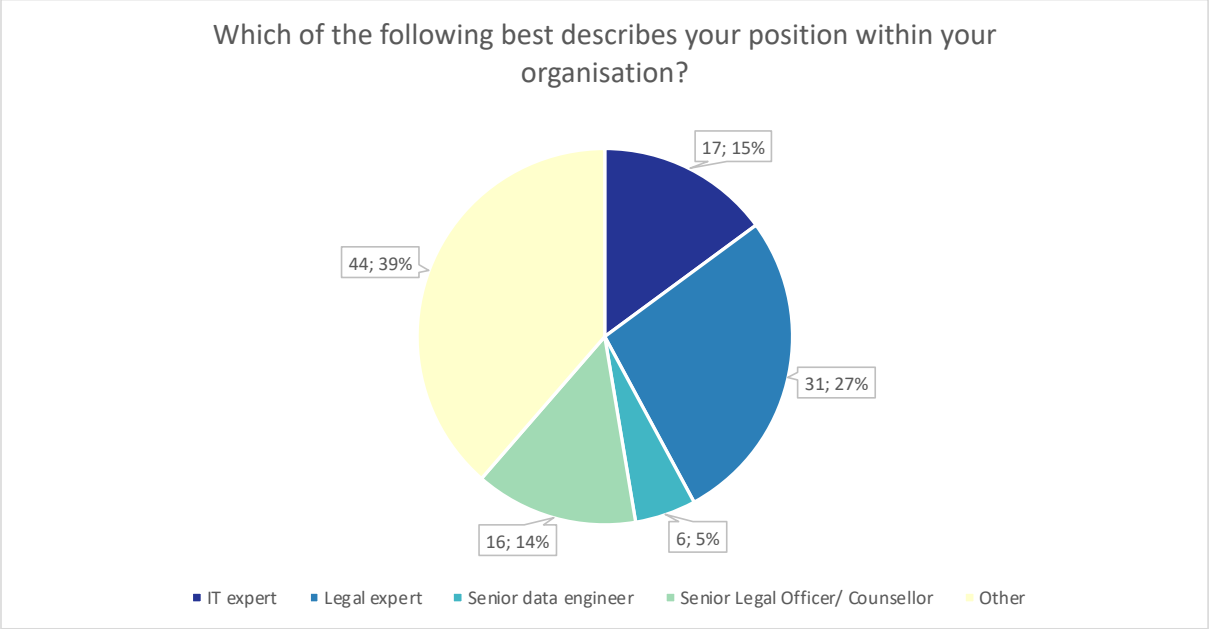
Figure 21: Industry of organisation's primary business



Source: Survey for the Database Directive Review
Note: based on A3 n = 53

Since the survey was sent to people who identified themselves either as a legal or IT experts, 47 identified themselves as legal experts/senior legal officer and 23 as IT experts/senior data engineer. 44 other respondents did not consider falling within any of the two categories.

Figure 22: Position within their organisation

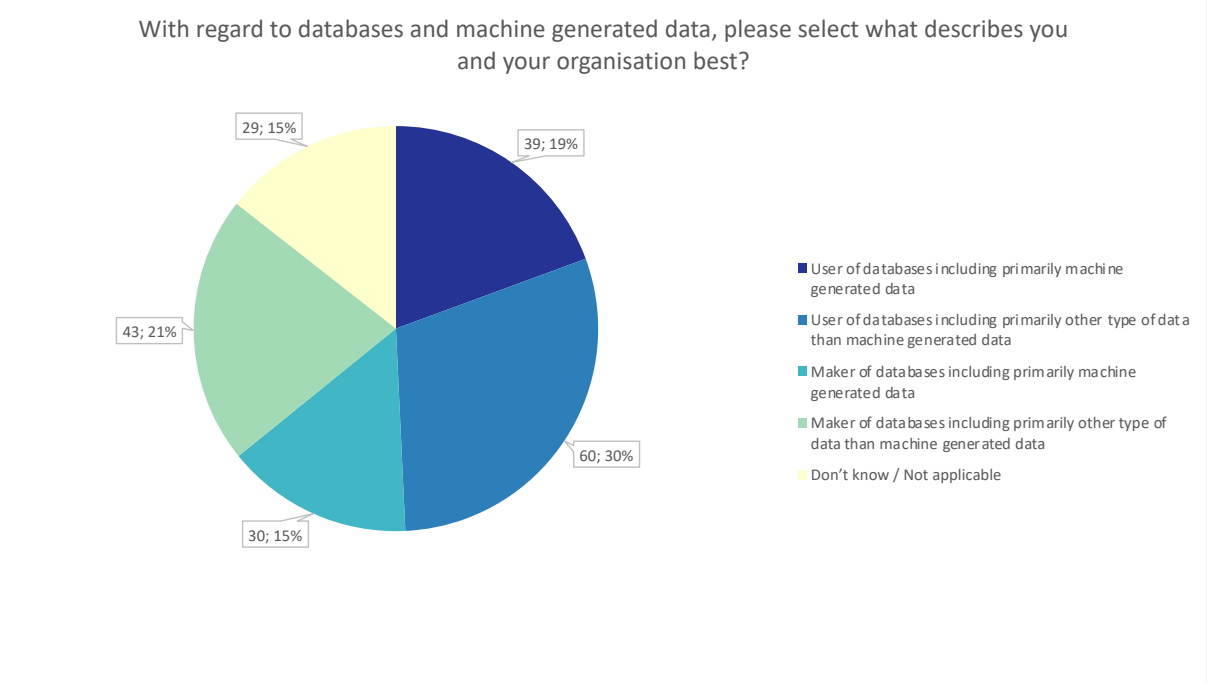


Source: Survey for the Database Directive Review
Note: based on A4 n = 114

When asked about the characteristic of identifying their organisation either as a *maker* or a *user* of databases, almost 50% indicated that they were predominantly *users*, while 36% indicated *maker.* 29 other chose “don’t know/not applicable” (among which, the six industry associations).

Since multiple answers were possible, many respondents selected more than one option. For example, two thirds of the *users* of databases containing machine generated data also indicated the use of other databases. Two-thirds of the users also indicated to be *maker* of machine generated and half of them produced other kinds of databases.

Figure 23: Characterisation of organisations by database maker or user

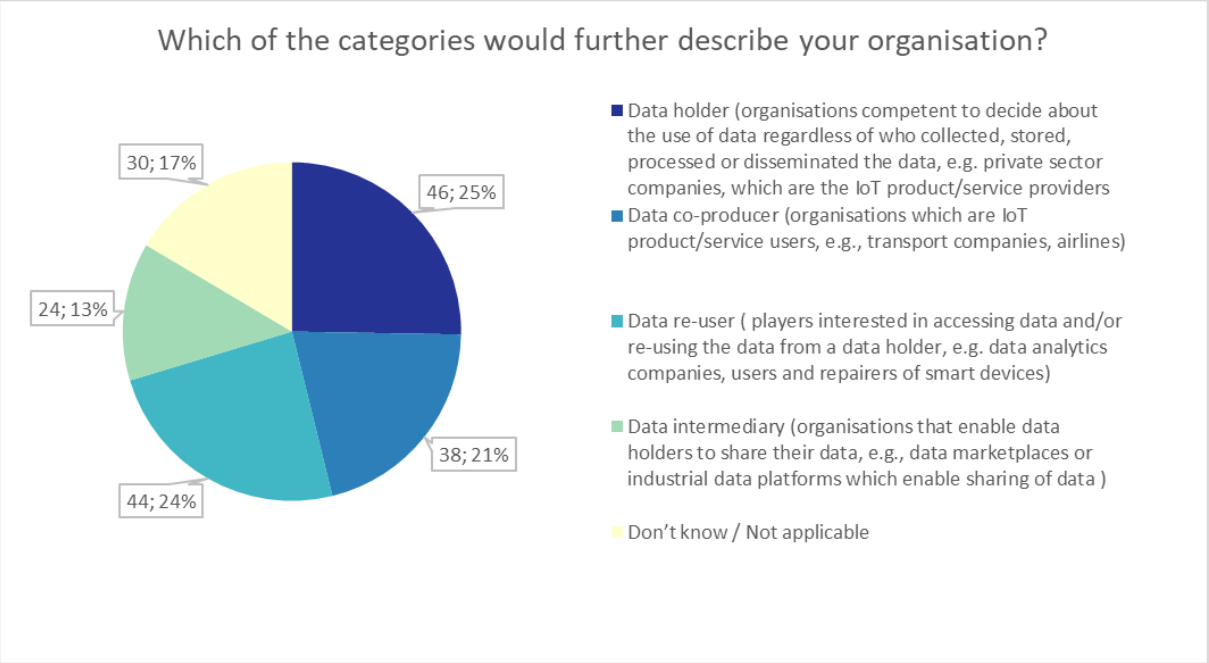


Source: Survey for the Database Directive Review
Note: Based on A5 n = 114

Following a distinction used by Deloitte to describe four key data characteristics, the following question asked whether the respondent's organisation could be described as either data holder, co-producer, re-user, or intermediary.

If one leaves out the responses from the six industry associations, it is interesting to note that one third of the respondents (40) chose more than one category. This suggests that **a clear-cut categorisation of companies as holder, re-user etc. is – at least for more than one third of companies – not realistic.**

Figure 24: Data categories that further describe the organisation

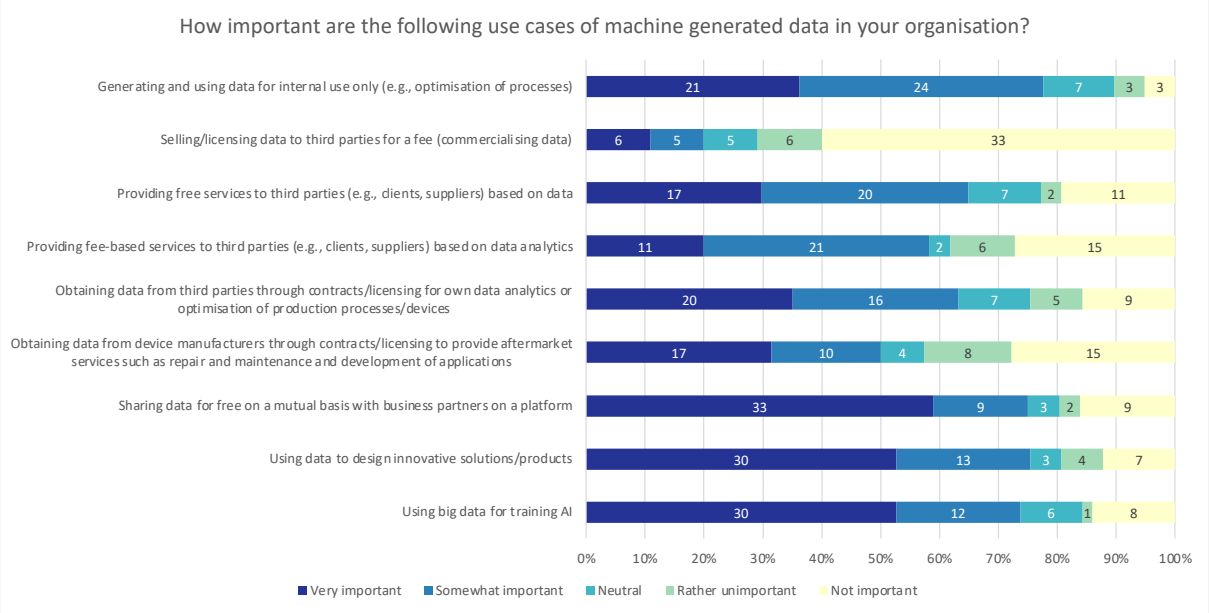


Source: Survey for the Database Directive Review
Note: Based on A6 n = 114

C.2.2 Questions on the technological-commercial context of data and databases

The following question was only available to IT and 'other' respondents. A total of 67 responses were thus considered. Overall, one can summarise that **the generation, using and sharing of data is very importance to the respondents' businesses**, while the commercialisation of data (i.e. selling/licensing of data) is not a popular option (with almost 70% of respondents classifying it as 'not important' or 'rather unimportant').

Figure 25: Importance of use cases of machine generated data to the organisation

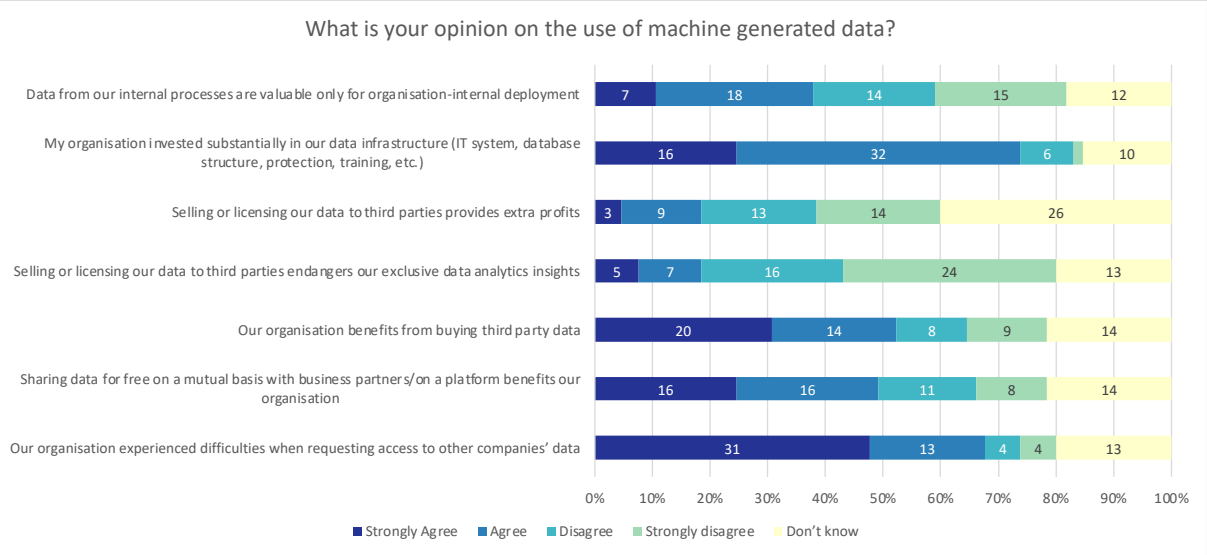


Source: Survey for the Database Directive Review
Note: Based on B1 n = 67

Respondents share the view that within a data economy, data access and sharing are important.

In terms of experience with the use of machine generated data, more than 70% of respondents claim that their organisation invested substantially in data infrastructure (IT system, database structure, protection, training, etc.) and 68% has experienced difficulties when requesting access to other companies' data. In addition, more than half of the respondents (52%) also report that their organisation benefits from buying third-party data and 49% have been sharing data for free on a mutual basis with business partners.

Figure 26: Opinion on use of machine generated data



Source: Survey for the Database Directive Review
Note: Based on B2 n = 66

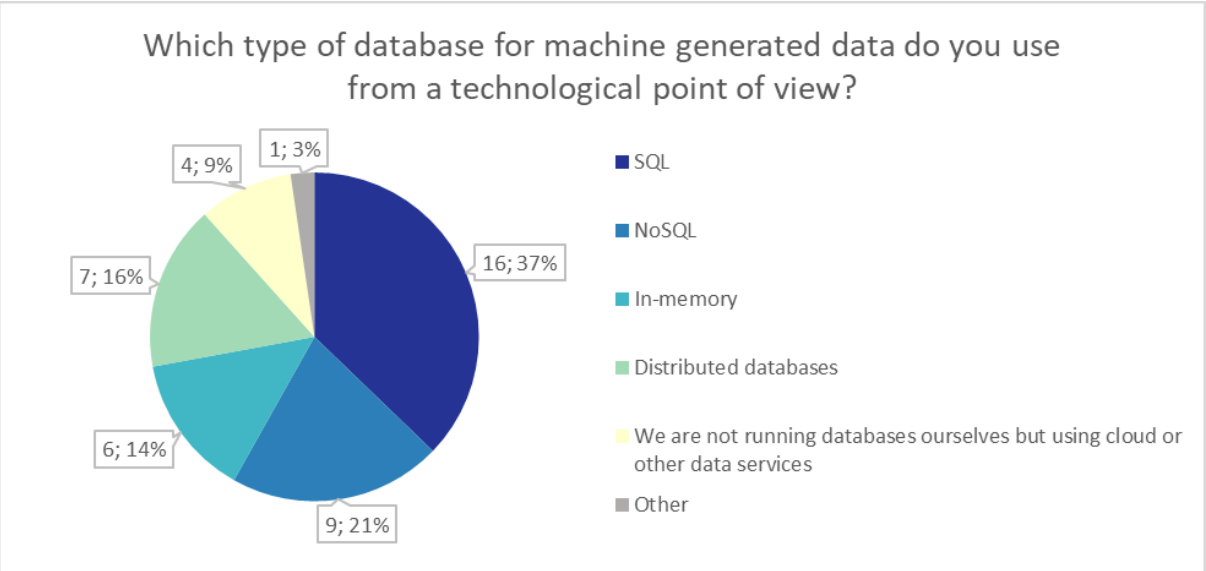
C.2.3 Generation and collection of data and development of databases

This section addressed only IT experts, but also 'other' respondents answered occasionally. The analysis is based on the 23 **IT experts** which responded.

The first question asked about the technical infrastructure of the databases: *Which type of database for machine generated data do you use from a technological point of view*?

While multiple answers were possible, the highest use was the data format SQL, followed by NoSQL, in-memory solutions, cloud services, and distributed databases.

Figure 27: Type of databases used for machine generated data

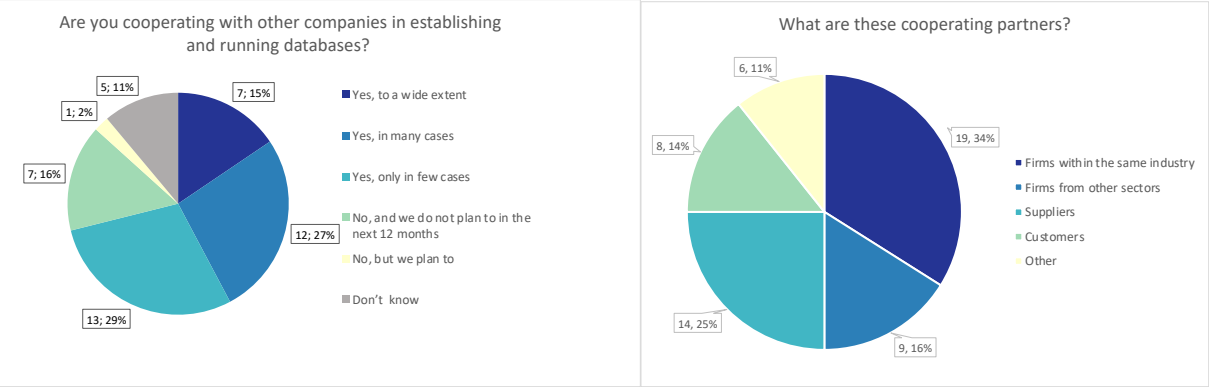


Source: Survey for the Database Directive Review
Note: Based on C1 n = 23

Cooperation with other companies in establishing and running databases is common. 32 out of 45 respondents (71%; IT experts and 'other') have collaborated (Question C2: Are you cooperating with other companies in establishing and running databases?).

While multiple answers were possible, collaboration within the same industry was the most common for 60% of the 32 respondents with collaboration experience, followed by suppliers (44%).

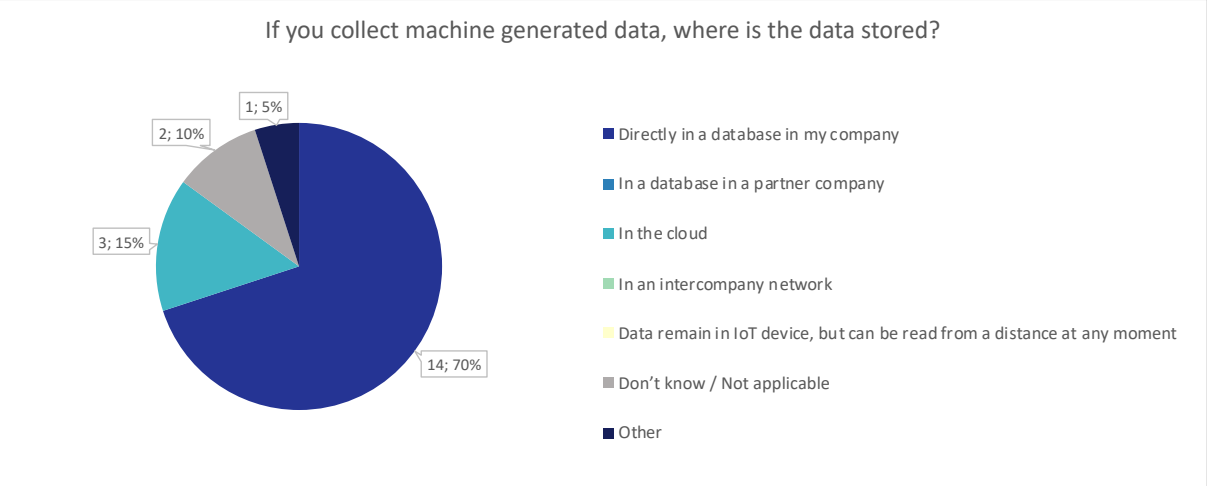
Figure 28: Cooperation and relationship with cooperating partners



Source: Survey for the Database Directive Review
Note: based on C2 n = 45 (left) and 32 (right) graph

Another question asked, where machine generated data was stored. For 70% of the 20 responding IT experts, machine generate data is most often stored in the database of their company. Storage of data in a cloud seems to be a minority practice with 15%.

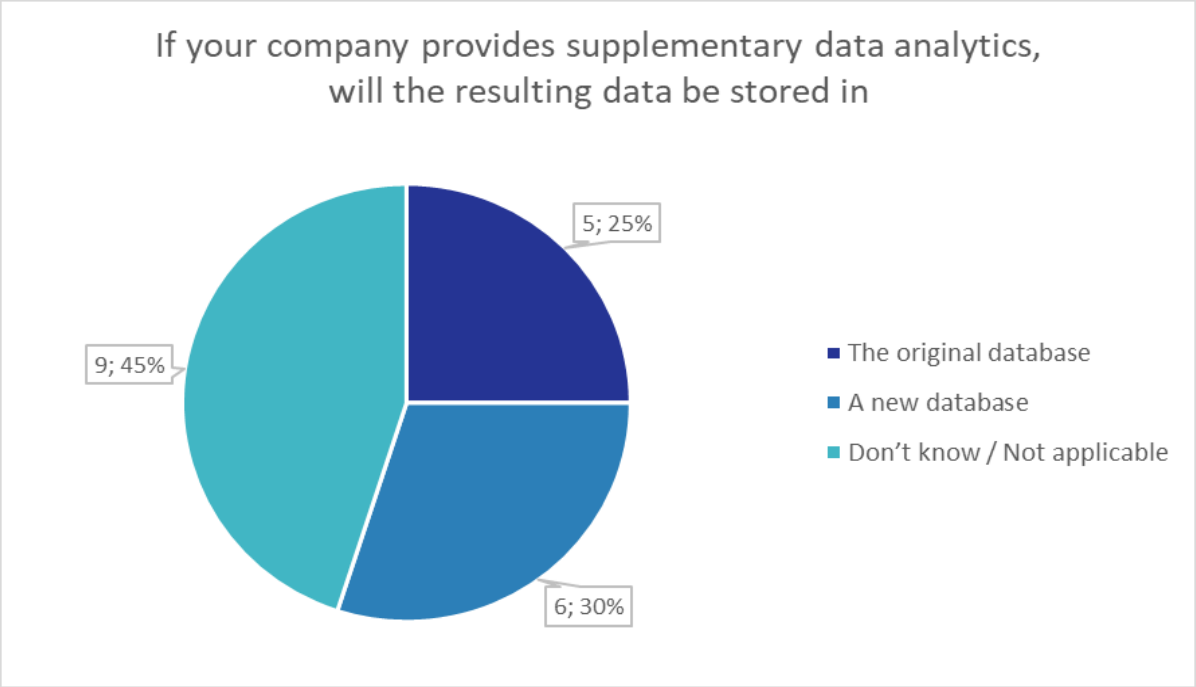
Figure 29: Storing of machine generated data



Source: Survey for the Database Directive Review
Note: based on C3 n = 20

Only half of the respondents knew where supplementary data analytics was stored. The remaining 11 respondents who knew provided an almost equal view: while for half the resulting data from additional data analytics were stored in a new database, the other half indicated that it would be included in the original database.

Figure 30: Storing of supplementary data analytics

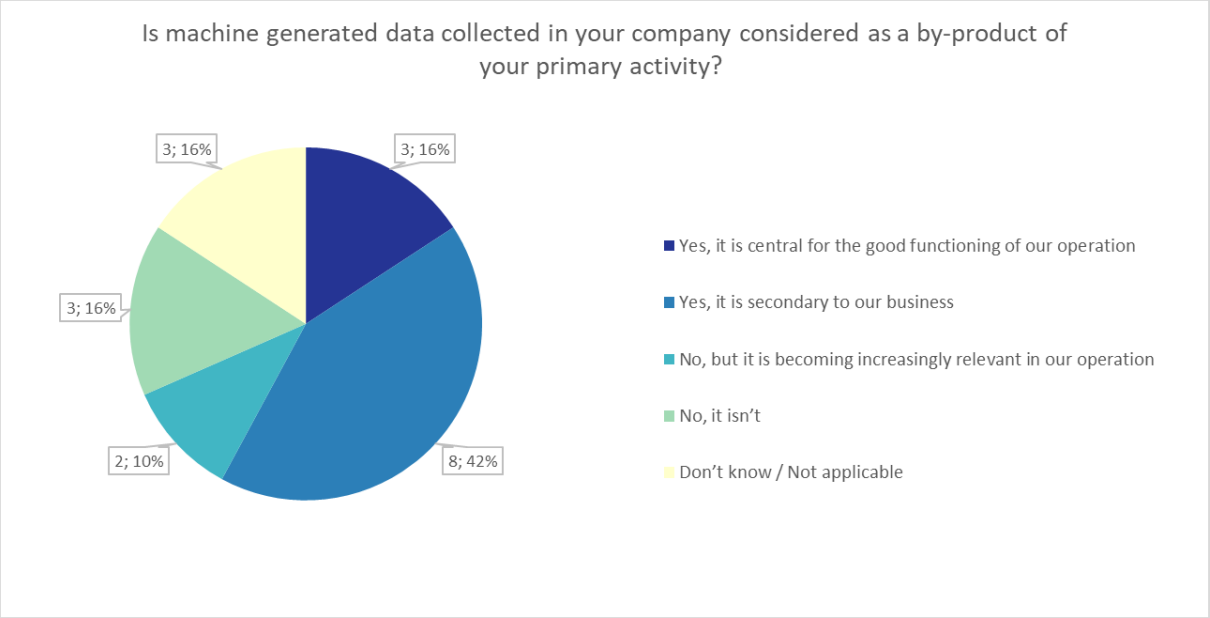


Source: Survey for the Database Directive Review
Note: Based on C4 n = 20

The next question asked if machine generated data was considered to be a by-product of the company's primary activity. The majority answered positively: the machine generated data is relevant for the business of almost 60% of the respondents' companies with 42% claiming "it is secondary to our business".

Three IT experts chose "it is central for the good functioning of our operations", while another three indicated that it isn't "but is becoming increasingly relevant" for the company.

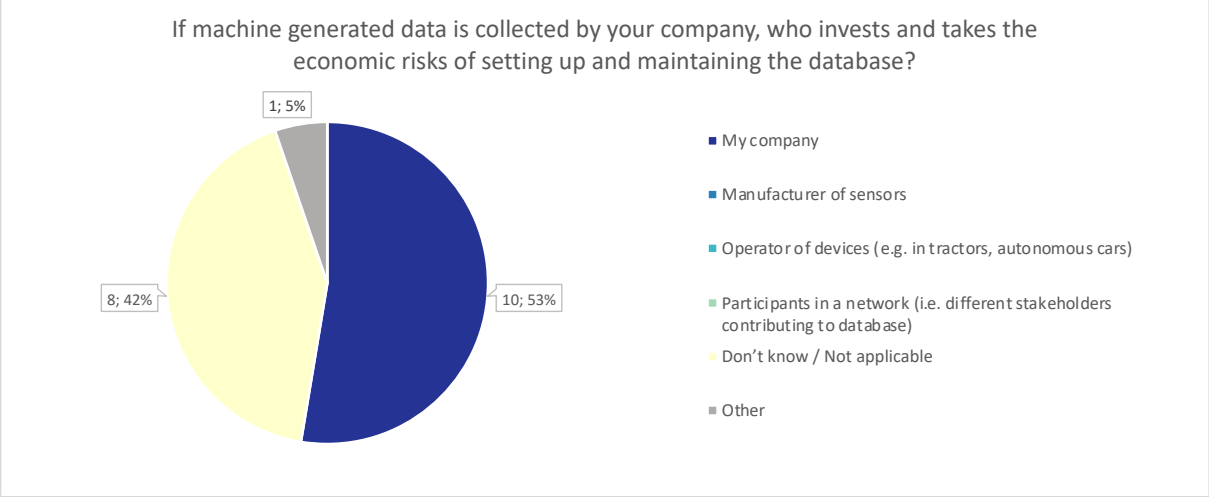
Figure 31: Is the machine generated data considered a by-product?



Source: Survey for the Database Directive Review
Note: Based on C5 n = 20

When it comes to the question of who takes the economic risks of setting up and maintaining a database of machine generated data, 42% of the respondents (8 out of 19) do not know while 53% (10) indicate that the risk will be taken by their own company.

Figure 32: Economic risks of setting up and maintaining databases



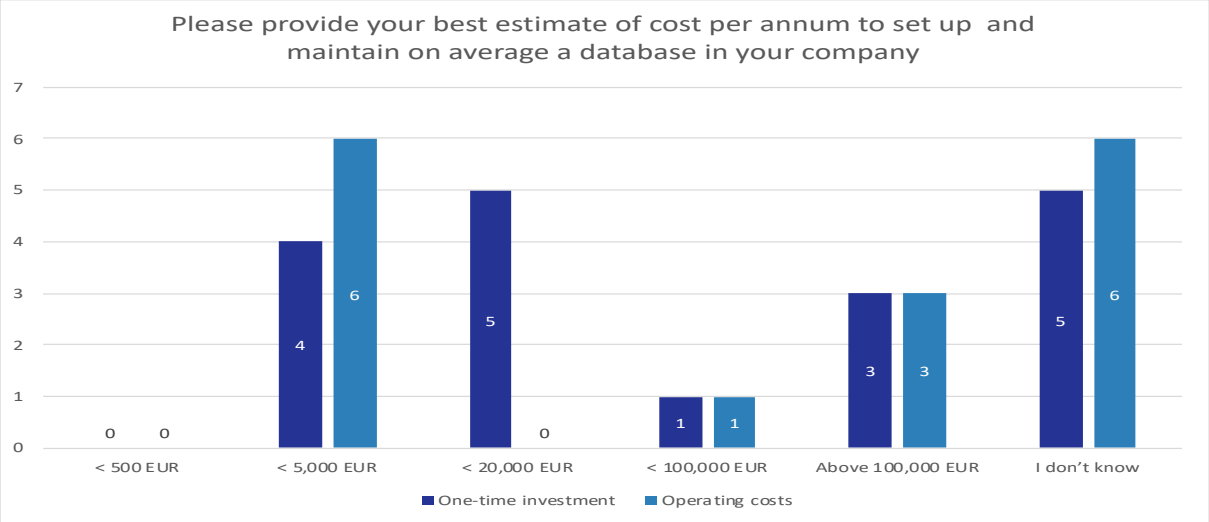
Source: Survey for the Database Directive Review
Note: Based on C6 n = 19

An important factor for the *sui generis* right is on costs for setting up and maintaining a database. More than fifteen respondents provided their best estimates on one-off investment costs and operating costs. In both cases, the lowest investment estimate is less than €5.000, while few respondents report spending more than €100.000.

It is worth cross-checking the size of the organisations of the respondents: the three respondents indicating more than €100.000 are rather large firms (with more than 250 employees). Likewise, in the case of operating costs, the three respondents indicating costs above €100.000 are from large firms (one above 250 and two above 500 employees).

One respondent, who indicated a high cost, remarked that the investment costs depend on the licence model, named user, CPU, etc. and therefore, can vary considerably. In the respondent's view, the maintenance costs are about one fifth of the initial investment costs.

Figure 33: Average cost per annum of set up and maintaining a database



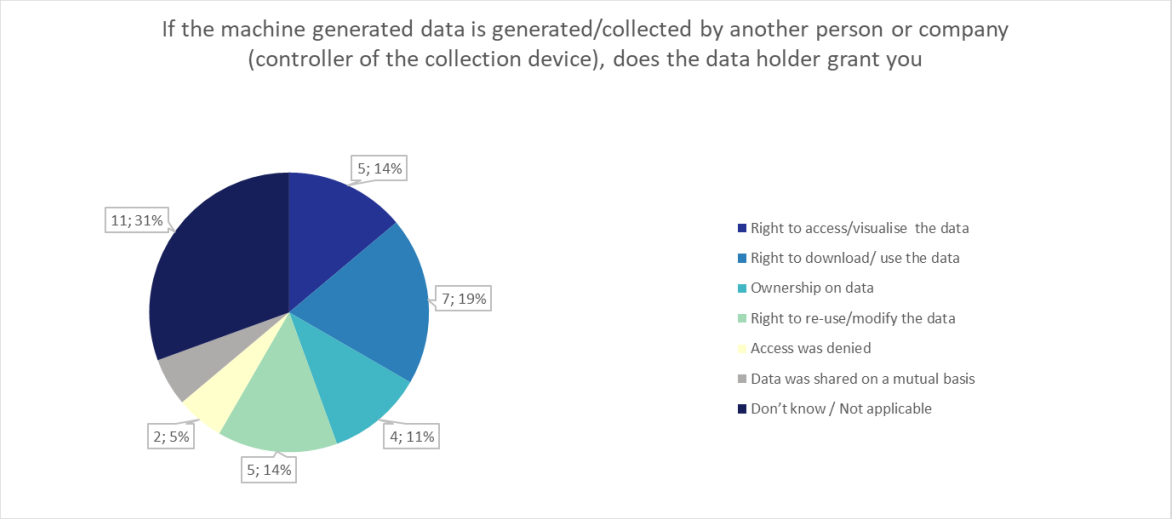
Source: Survey for the Database Directive Review
Note: Based on C7 n = 18 IT experts only (Answer item: One-time investment) and 16 (Answer item: Operating costs)

When it comes to accessing machine generated data that is controlled by another person or company, half of the respondents were not aware whether or not the data holder grant them any right on the data being generate/collected.

More than 30% of respondents (7 out of 23) indicate that they can download and use the data, while more than 20% (5) indicate being able to re-use or modify or to visualize the data. Only two IT experts indicated that access was denied.

However, among the 'other' respondents, in particular the industry associations and three research stakeholders mentioned that access was denied.

Figure 34: Access to machine generated data generated/collected granted to respondent



Source: Survey for the Database Directive Review
Note: Based on C8 n = 23

The questions on estimated costs to establish a contract for accessing data was answered by seven persons only (three indicating less than €1000, one indicating between €1.000 and €5.000, one between €5000 and €10.000, one between €10.000 and €20.000, and another indicating more than €20.000).

C.2.4 Sharing and access to data and databases

Questions under this section of the survey were directed to the **IT experts** only but again, a few 'other' respondents provided some information too.

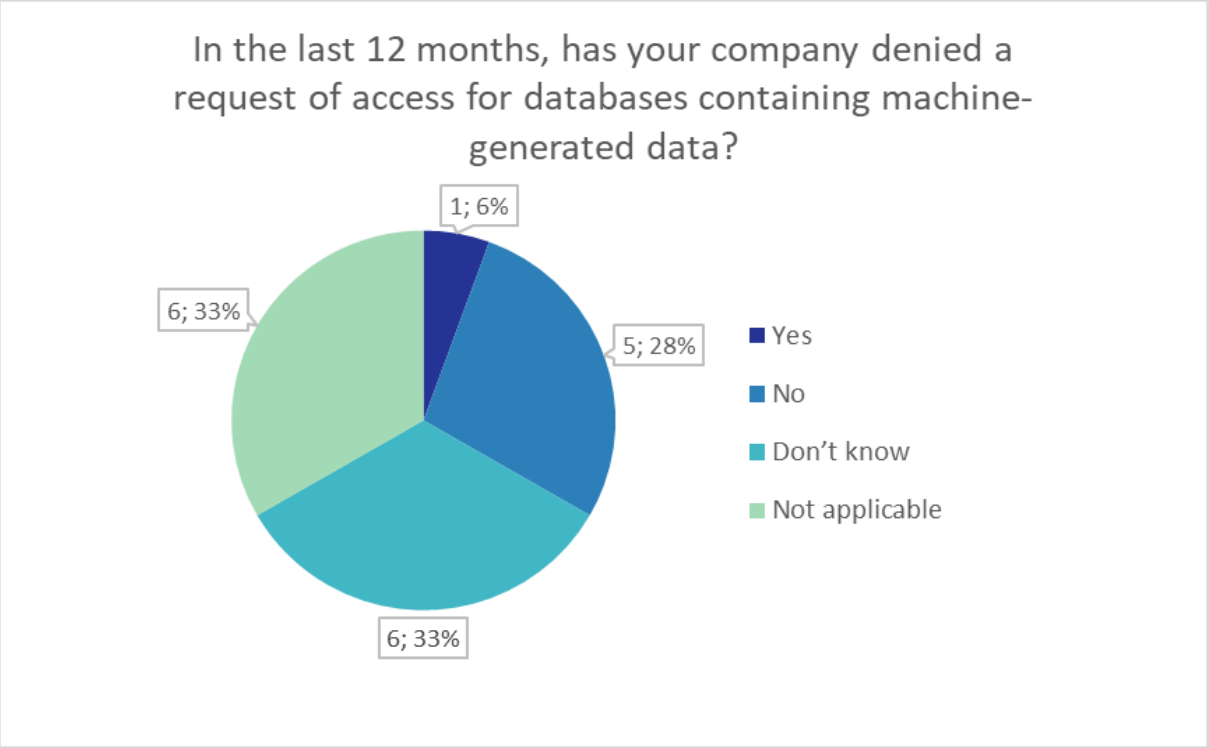
The question D1 asked "In the last 12 months, how often has your company shared its databases containing machine generated data or requested access to/use of another company´s databases?". From a total of 8 respondents answering that they share databases, 3 answer that they share on a daily basis, 3 that they share several times, 2 that they only share a few times. 10 don't know or think that this is not applicable.

The two following questions enquired about the practice of sharing databases containing machine-generated data or requesting access to use other companies' databases of machine-generated data, over the past 12 months (Question D1: In the last 12 months, how often has your company shared its databases containing machine generated data or requested access to/use of another company's databases?).

18 out of 22 IT experts surveyed responded. 7 report that access to data was relevant to their organisation while 8 report that sharing of data was relevant to their organisation ('on a daily basis', 'several times' or 'only a few times').

Asked whether their company has recently denied access to databases of machine-generated data, 11 answered that their company never denied access or that they do not know, while only one small IT company reported having denied access to such databases. Reasons for denying access were both cost factors but also the fear of a potential loss of data exclusivity.

Figure 35: Access to machine-generated data produced by respondents' companies

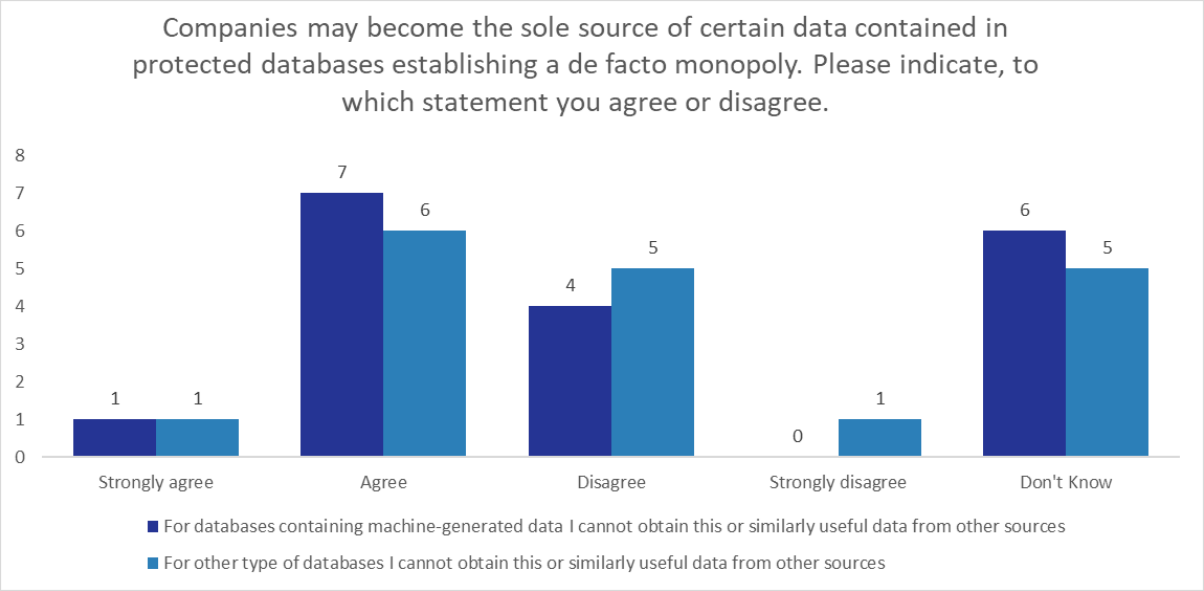


Source: Survey for the Database Directive Review
Note: Based on D2 n = 18

De facto data monopolies are an issue – but not only in terms of databases containing machine generated data, but also other types of databases.

Among the respondents who strongly agree are the six industry associations.

Figure 36: Views on de facto monopolies for sole sources of data



Source: Survey for the Database Directive Review
Note: Based on D3 n = 18

When asked if respondents encounter any problem when they tried to obtain access to databases containing machine generated data (Question D4: When you tried to obtain access to databases containing machine generated data, did you encounter any problems?), 5 out of 15 respondents indicated yes, while 3 did not have problems and 7 other did not know.

Further asked about the encountered problems, the respondents, were able to indicate the relevance of six pre-identified problems.

Four indicated that the database was legally protected and there was no licensing available, three indicated the lack of interoperability, and two respondents indicated that the data in the database was kept secret, while one that the costs to obtain access to the database (licensing costs, infrastructure developments, etc.) were too high.

None indicated the fact that the database was protected with technical measures and that there was no market for the type of database needed being among the faced problems.

The responses from the industry associations provided an example of problems faced: "*Vehicle manufacturers (OEMs) refuse to grant access to live in-vehicle data (from sensors etc.). When access is granted, oppressive entry restrictions are put in place (high fees, complex contracts, delays, no access to live data from moving vehicles, long latency, need to divulge information about repairer and the individual vehicle involved etc.).*" Another respondent indicated that "*this is a not a simple question: problems with semantics, provenance, formats, access, and legal status assuming access" are impeding factors*."

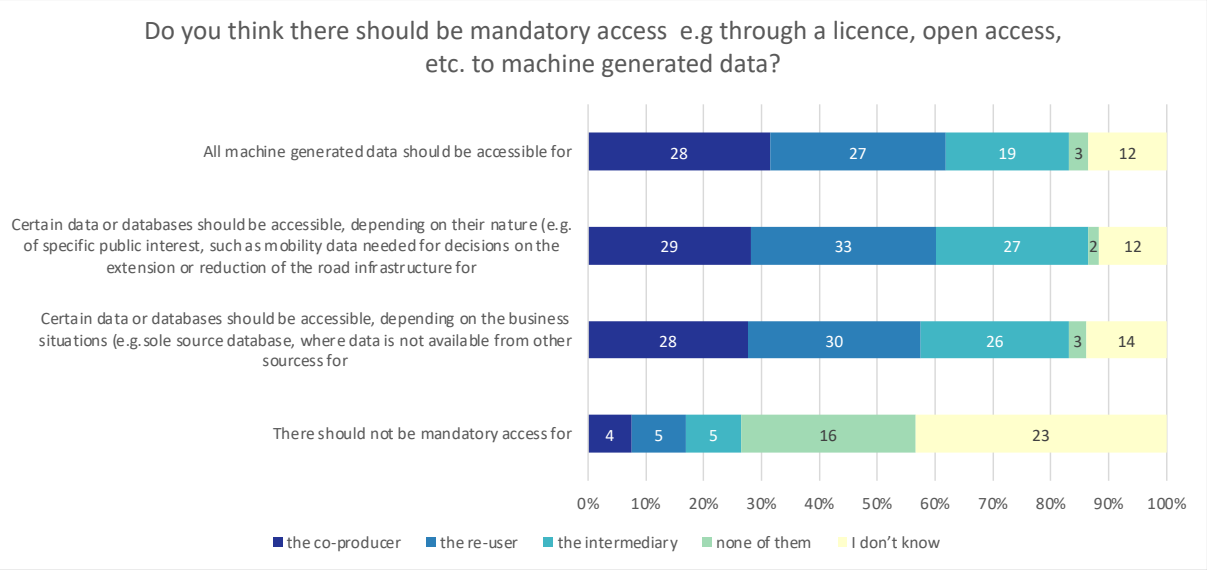
Asked about mandatory access for various types of data producers or users, there was a clear view that certain machine generated data should be accessible to co-producers and to re-users, and (with a minimal less high preference) also to intermediaries.

Respondents favour mandatory access to all three categories (co-producers, re-users and intermediaries) of all machine-generated data or certain data or datasets (depending on their

nature and business situation). Overall, the remaining responses favoured mandatory access over non-mandatory one which may partly be explained that most respondents had identified themselves earlier as users or re-users and not only as database makers.

The industry associations voiced for mandatory access regardless of the situation.

Figure 37: Views on mandatory access to machine generated data



Source: Survey for the Database Directive Review
Note: Based on D5 n = 32

Three respondents indicated why mandatory access should *not* be granted. All indicated that it affects the freedom to contract too much.

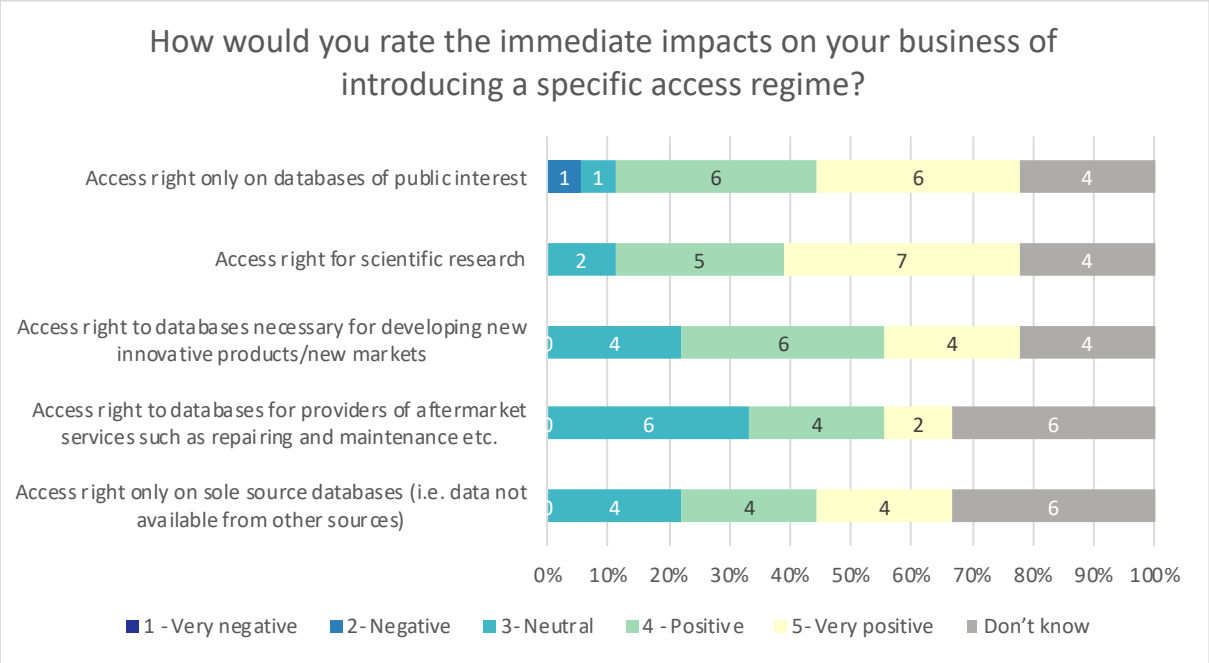
When taking the answers to the next question on how should access be granted (Question D5b: If you think there should be access to databases containing machine generated data, how should access be granted?), there is a clear view of the coordinated industry "*as a free access*" (22 or 69%). Four other respondents provided a choice with "*as a compulsory license under FRAND terms*".

Asked about potential gains or losses if databases containing of machine generated data would be mandatorily accessible (Question D6: Suppose databases containing machine generated data would be mandatorily accessible. Could you provide an estimate of the potential gains and losses in revenues for the data holder?), almost three-quarter of the respondents were not able to provide an estimate, while other respondents indicated that revenues are not affected, neither positively or negatively (six in both options).

Concerning the question on access regimes, the industry associations had a very clear view on the usefulness of an access right to databases for developing innovations and for aftermarket services. Likewise, the other respondents had a positive opinion regarding the introduction of a specific access in terms of developing innovative products/new markets, for the scientific research and for the public interest.

Results from the total sample provide similar results as the sub sample only containing responses of IT experts. When asked about the immediate impacts of introducing a specific access right only on sole-source databases, the majority of IT experts responding to the survey question (23 out of 48) consider it "positively" or "highly positive". For 6 respondents, the impact would be neutral, other 6 expressed it would be "negative" or "very negative", and 14 respondents do not know.

Figure 38: Immediate impacts of introducing a specific access regime



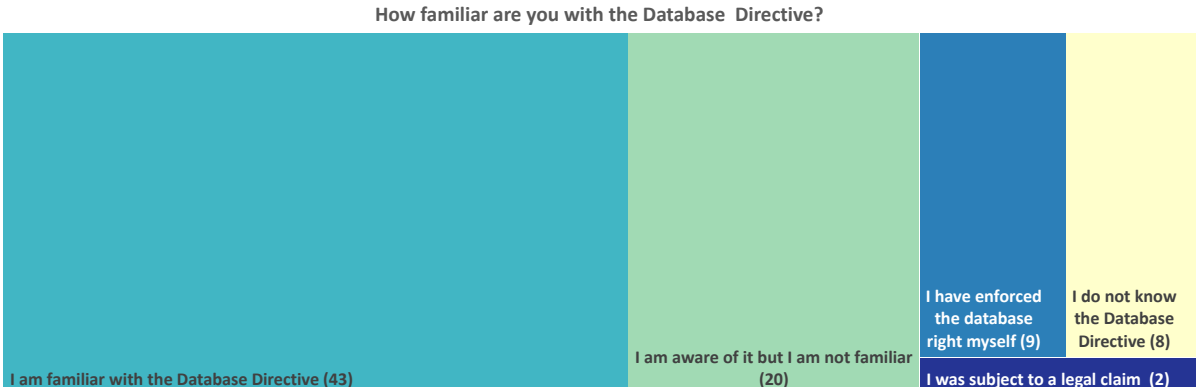
Source: Survey for the Database Directive Review
Note: Based on D7 n = 18

C.2.5 *Sui generis* and other types of database protection means

Questions under this section of the survey were directed to **all respondents**. However, not all 114 participants provided answers to this set of questions. Therefore, the number of valid answers falls in some questions below 70 responses.

In terms of familiarity with the Database Directive, less than 50% of all respondents confirm being familiar with the Database Directive (42), while slightly more than 20% are aware but not familiar (19) and another 9% has enforced the database right themselves (9, of all respondents with legal expertise). Eight of the respondents state not being aware of the Database Directive.

Figure 39: Familiarity with the Database Directive



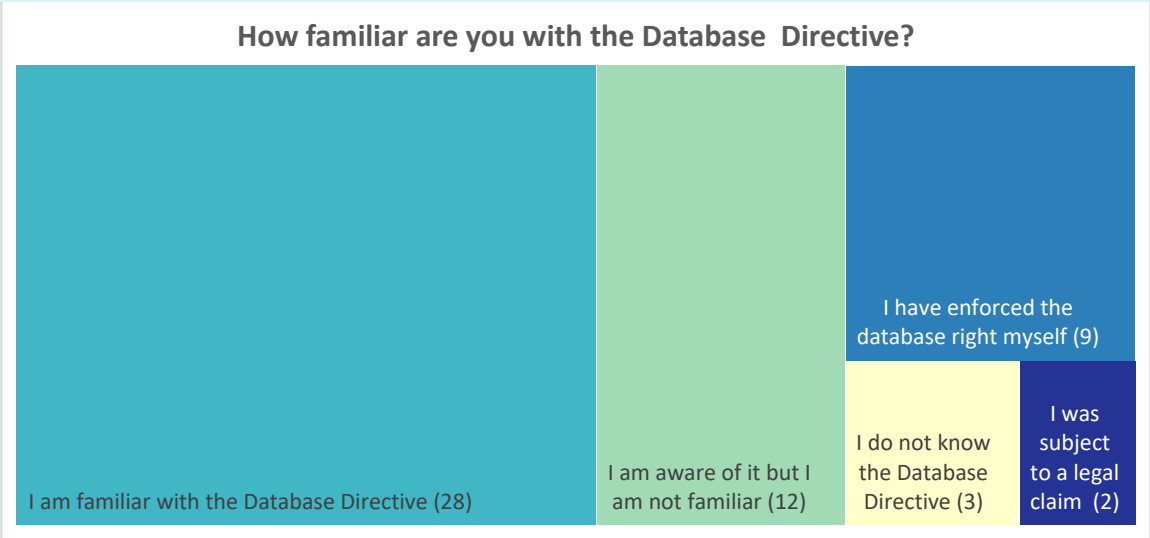
Source: Survey for the Database Directive Review
Note: Based on E1 n =82

Legal experts

In terms of familiarity with the Database Directive, 60% of responding legal experts (28 out of 47 respondents) claimed being familiar with the Database Directive, while 10% were aware but not familiar (12) and another 10% has enforced the database right themselves (9). Only two were subject to a legal claim. Finally, three legal experts did not know the Database Directive.

The R&D stakeholders with legal background declared being familiar with the Database Directive (10 out of 12 respondents). Two of them declared having enforced the Database Directive themselves.

Figure 40: Familiarity with the Database Directive, legal experts only



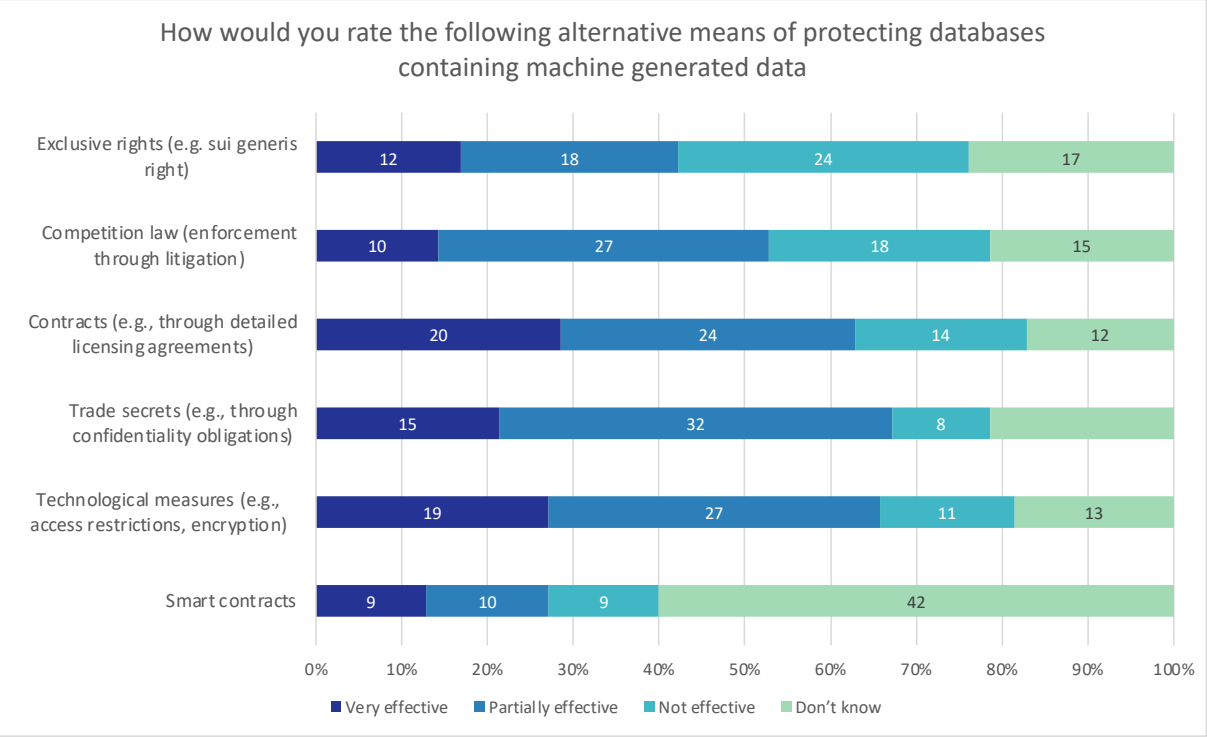
Source: Survey for the Database Directive Review
Note: Based on E1 n = 47

In the following question, all survey participants were asked to provide their views on a set of alternative means of database protection.

Among the preferences of the respondents, there are four different protection means which are preferred to the *sui generis* right as an effective means to protect databases containing machine-generated data. In order of priority, these protection means are:

1. trade secrets ('verry effective' and 'partially effective' above 65% of responses),
2. technological measures (approx. 65%),
3. contractual obligations (above 60%) and
4. competition law measures (above 50%).

Figure 41: Rating of alternative means of protecting databases containing machine generated data



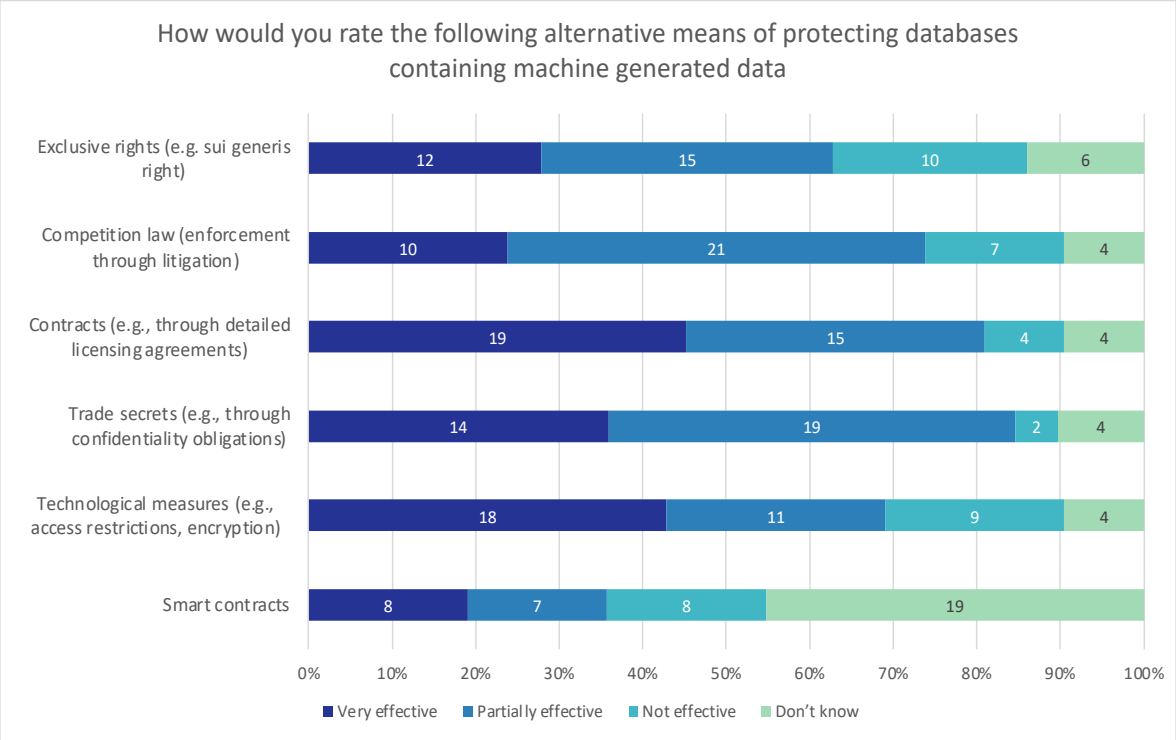
Source: Survey for the Database Directive Review
Note: Based on E2 n = 70

Legal experts

For legal experts the most effective means to protect databases containing machine-generated data are trade secrets and contract obligations. Both options are at or above 80% of agreement. By contrast, smart contracts are not widely known and are the least valued alternative for legal protection.

The responses of the R&D experts with legal background follow the same patterns of the other legal experts, except in the opinion regarding the relevance of the *sui generis* right to protect machine-generated datasets. This subgroup seems to be considering the exclusive right as rather 'not effective' (7 out of 11 respondents).

Figure 42: Rating of alternative means of protecting databases containing machine generated data (legal experts only)



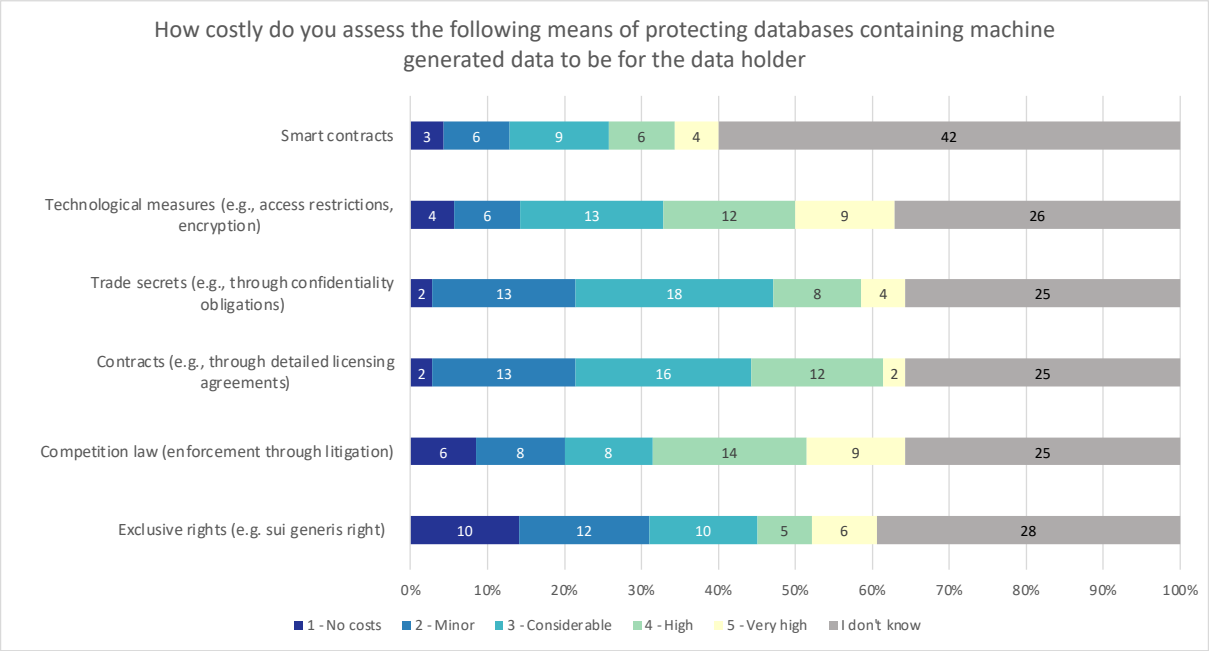
Source: Survey for the Database Directive Review

Note: Based on E2 n = 43

Asked about the estimated costs of the different protection means, most respondents provided an opinion about the seven suggested protection means, except for the smart contracts, for which most did mention 'do not know'.

Results suggest that to almost one third of the respondents the *sui generis* right has mostly 'no cost' or 'minor' costs while trade secrets, contracts and technological measures were deemed being significantly more costly.

Figure 43: Estimated costs of different protection means for the data holder



Source: Survey for the Database Directive Review
Note: Based on E3 n = 71

Legal experts

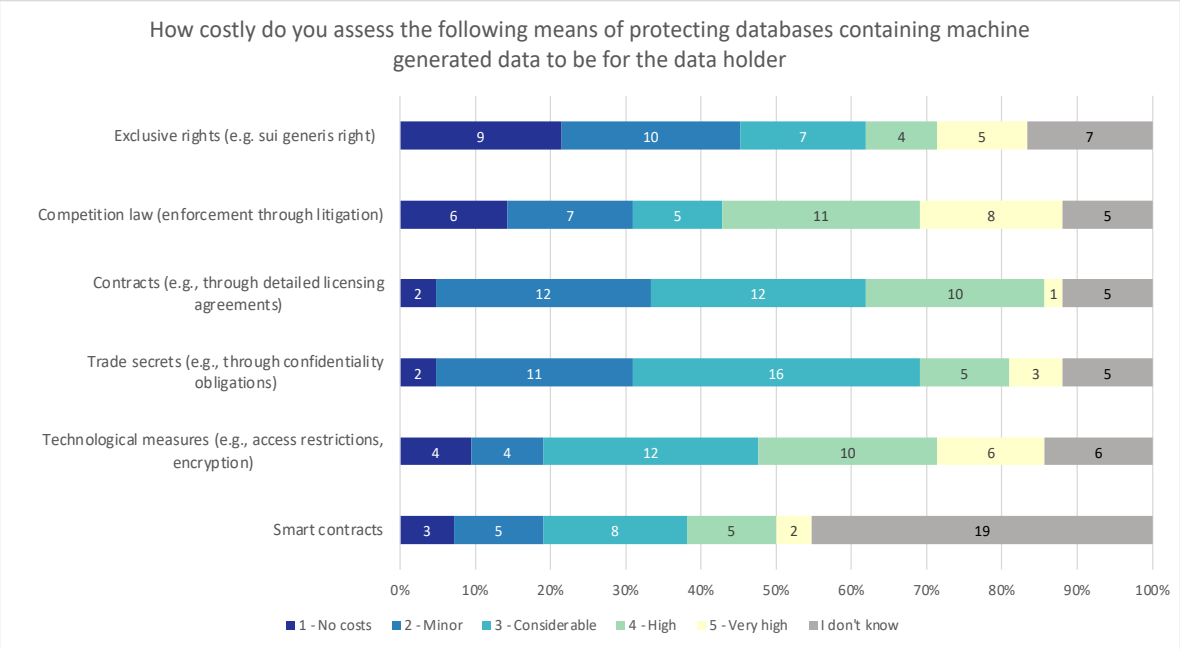
Most legal experts expressed an opinion about the seven suggested protection means.

The legal experts found the *sui generis* right as having mostly 'no cost' or 'minor' costs. This was also shared by legal practitioners with prior enforcement knowledge.

Trade secrets, contracts and – particularly – competition law and technological measures were deemed as having 'considerable', 'high' or 'very high' costs.

The views of the R&D experts with legal background follow overall the same pattern of the broader sample of legal experts.

Figure 44: Estimated costs of different protection means for the data holder – view of legal experts only

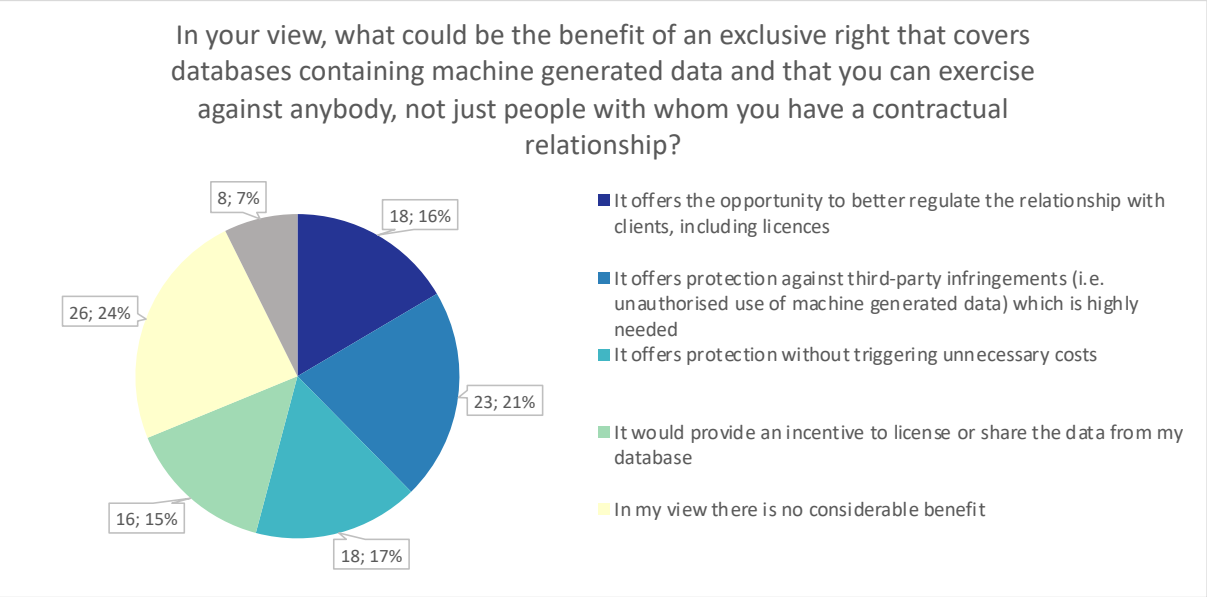


Source: Survey for the Database Directive Review

Note: Based on E3 n = 42

When it comes to the question on benefits of an exclusive right, respondents do not provide consistent responses. In fact, on the statement 'In my view there is no considerable benefit' three-quarter do not agree. But when it comes to individual statements of benefits, such as incentives or protection, around three-quarter do not see those as potential benefits of the *sui generis* right.

Figure 45: Benefits of an exclusive right



Source: Survey for the Database Directive Review
Note: Based on E4 n = 109

Legal experts

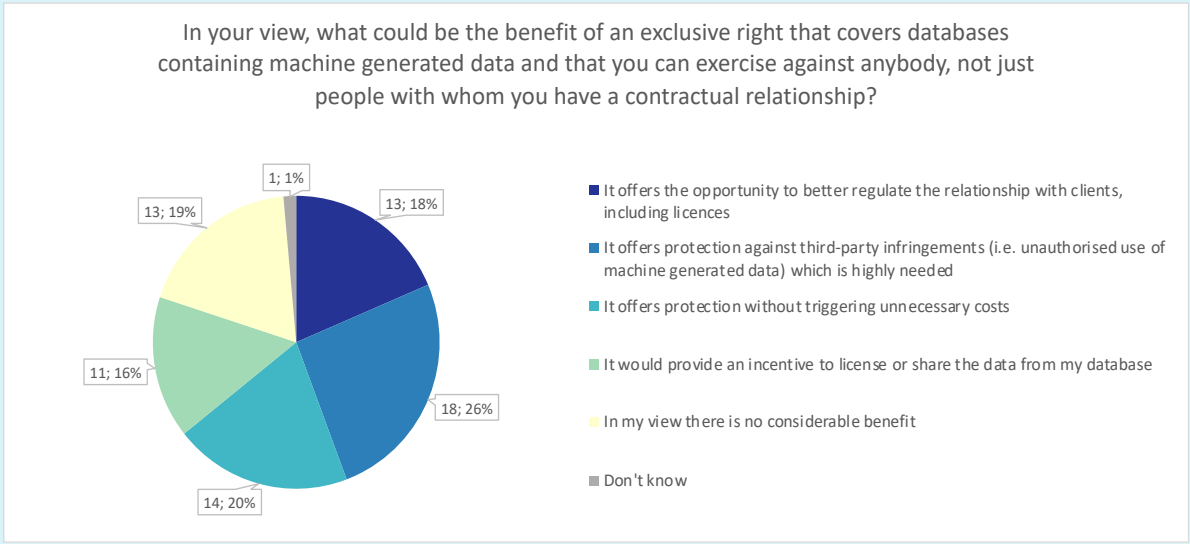
Slightly more than one third of the legal experts find that extending the exclusive right to machine-generated data would 'offer needed protection against third-party infringements', one-third think that it 'offers protection without triggering unnecessary costs', while almost one-third see that 'It would help in contractual relations with clients'.

Almost one fourth of the responding legal experts state that in their opinion, the introduction of the exclusive right for machine-generated datasets would not bring any particular benefit.

Interestingly, the R&D experts with legal background solidly think that an exclusive right that covers databases containing machine generated data would not bring considerable benefits (more than 70% of the respondents belonging to this subgroup, namely 8 out of 11).

Thus overall, only one third of the legal experts see benefits to include databases of machine generated data in the *sui generis* right.

Figure 46: Benefits of an exclusive right – views of legal experts only

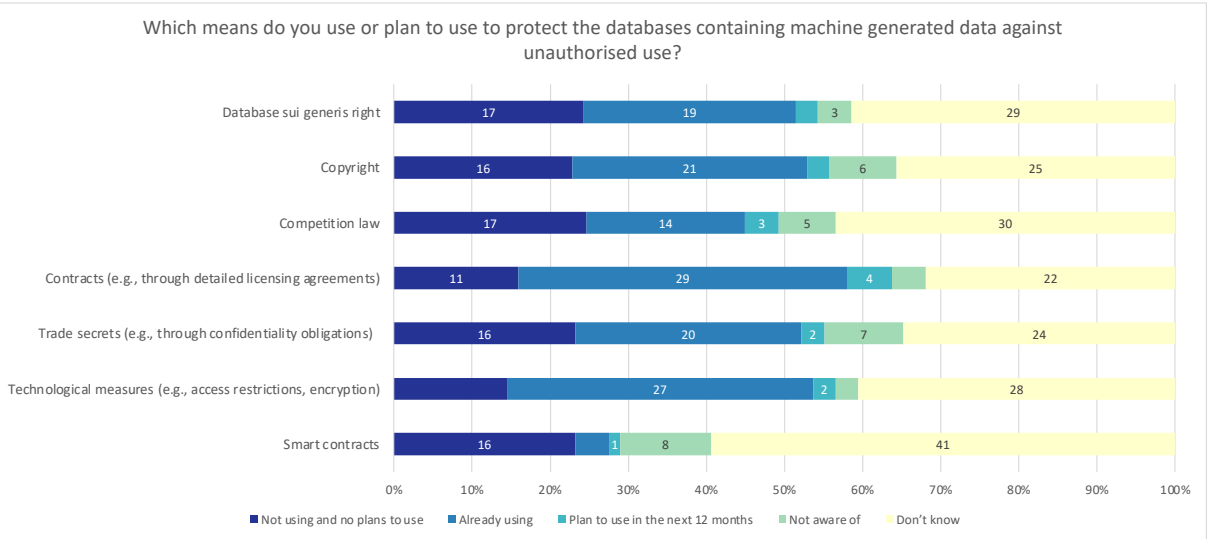


Source: Survey for the Database Directive Review

Note: Based on E4 n = 47

In terms of the means to protect databases containing machine generated data, the *sui generis* right is not the most common protection means. Technological measures and in particular contracts are the two means that are already used more often and are more likely to be used in the near future.

Figure 47: Plans to protect the databases containing machine generated data against unauthorised use



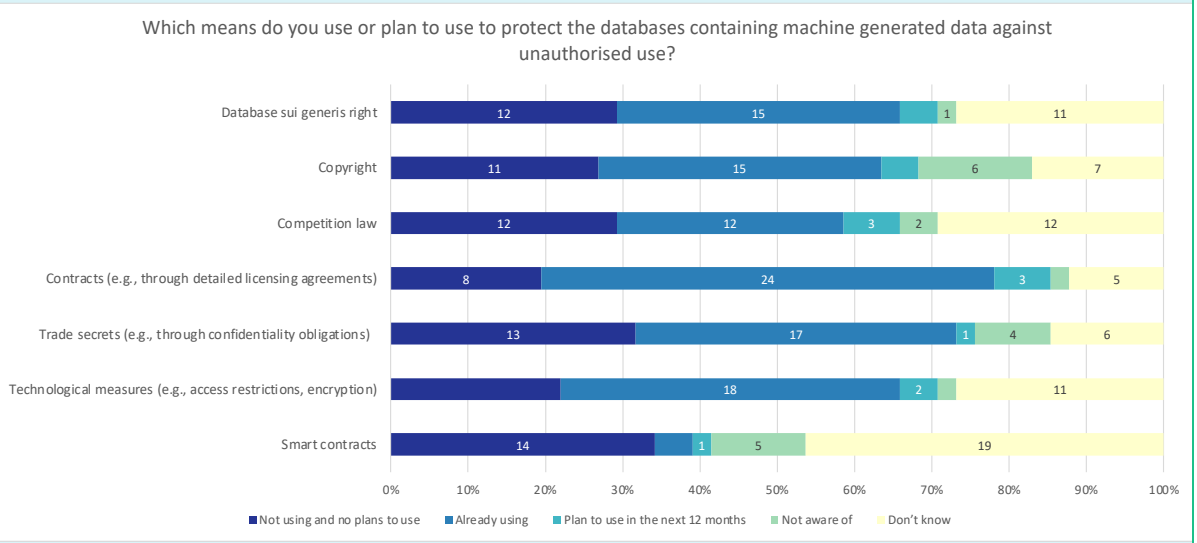
Source: Survey for the Database Directive Review
Note: Based on E5 n = 70

Legal experts

Despite being used by more than 30% of the legal experts, the *sui generis* right seems to be not their preferred means of protection. In fact, trade secrets, technological measures, and in particular contracts are means that are more used and more likely to be used.

One respondent added "*We do use the database sui generis right, even if the scope of protection concerning databases containing machine generated data remains unclear and, therefore, it is not an effective means of protection*."

Figure 48: Plans to protect the databases containing machine generated data against unauthorised use



Source: Survey for the Database Directive Review

Note: Based on E5 n = 41

The question E6 asked "to what extent is *sui generis* database right useful for the purposes of your data protection compliance, e.g., to protect personal data from unauthorized access, etc.?". From a total sample of 39 respondents, 14 respondents found it not useful at all, 6 respondents thought it is potentially useful, but they have not used it in practice, and only 5 found it useful.

Legal experts

To the question on the level of perceived usefulness of the *sui generis* right for the purposes data protection compliance, 24 legal expets provided an answer. Half of them found that the *sui generis* right would 'not (be) useful at all'. Only eight found it 'useful' or 'potentially useful'.

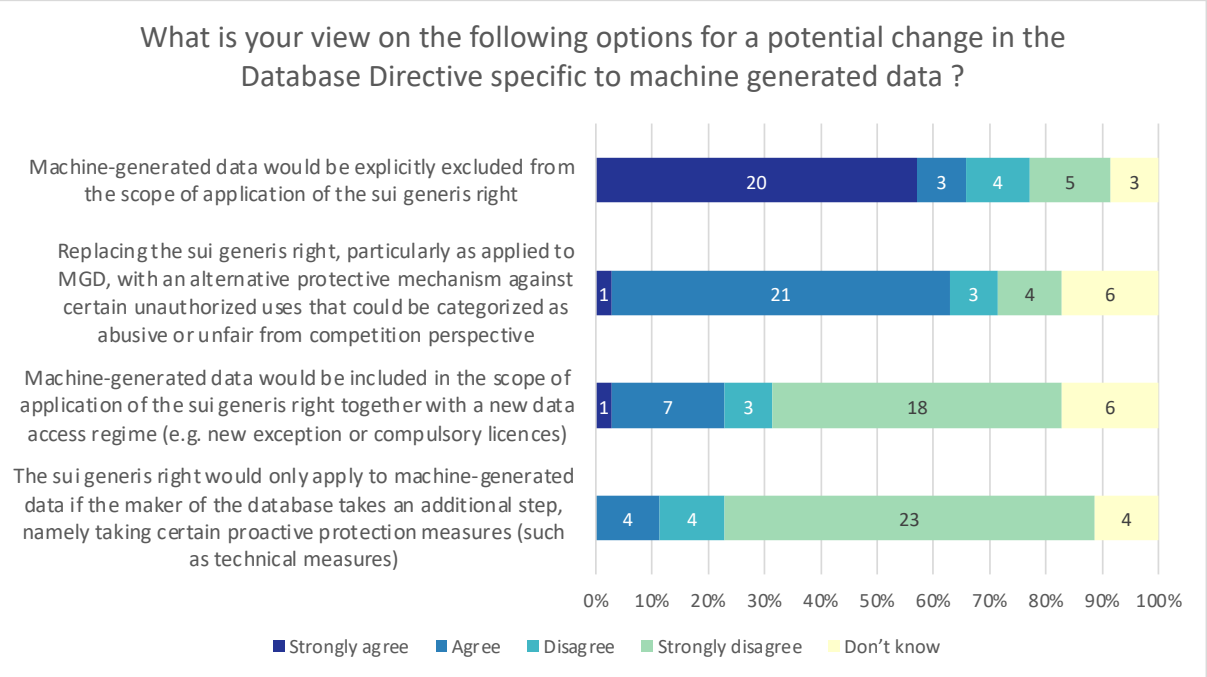
C.2.6 Policy options

Questions under this section of the survey were directed to **the legal experts** but also a few 'other' respondents answered,

Thirty-five respondents provided their views on the various options and sub-options which are in the discussion of the review of the Database Directive. **The option to 'exclude machine generated data explicitly' is welcomed** by more than 60% of the respondents – about 50% 'strongly agree'. Replacing the *sui generis* right with an alternative protection mechanism is also agreed by roughly 60%, however, this option is not strongly agreed to. The inclusion of MGD – also only under specific circumstances – is however strongly agreed or agreed by 15-20%.

The coordinated responses from the industry associations show a strong agreement to the exclusion of machine generated data from the *sui generis* right, an agreement to replace the right with an alternative protection mechanism, and a strong disagreement with the two remaining options.

Figure 49: Views on the options on a potential change in the Database Directive specific to machine generated data



Source: Survey for the Database Directive Review
Note: Based on F1 n = 35

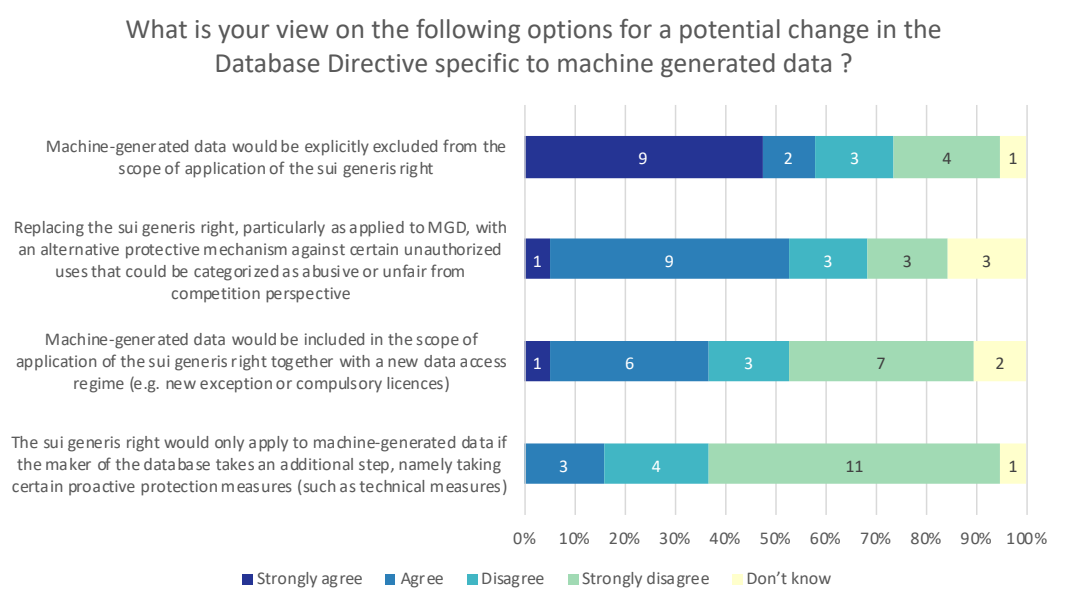
Legal experts

19 respondents with (some) legal expertise provided their views on the various options. The option to exclude machine generated data explicitly is welcomed by almost 60% of respondents ('agree' or 'strongly agree' answers), while the application of the *sui generis* right to machine generated data under conditions (such as proactive protection measured

by the maker) is rejected by approx. 80% of respondents ('disagree' or 'strongly disagree' answers).

The R&D stakeholders with legal background agree to exclude machine-generated data from the protection of the Database Directive and they also disagree to extend the *sui generis* right to machine-generated data under certain conditions.

Figure 50: Views on the options on a potential change in the Database Directive specific to machine generated data (views of legal experts only)



Source: Survey for the Database Directive Review

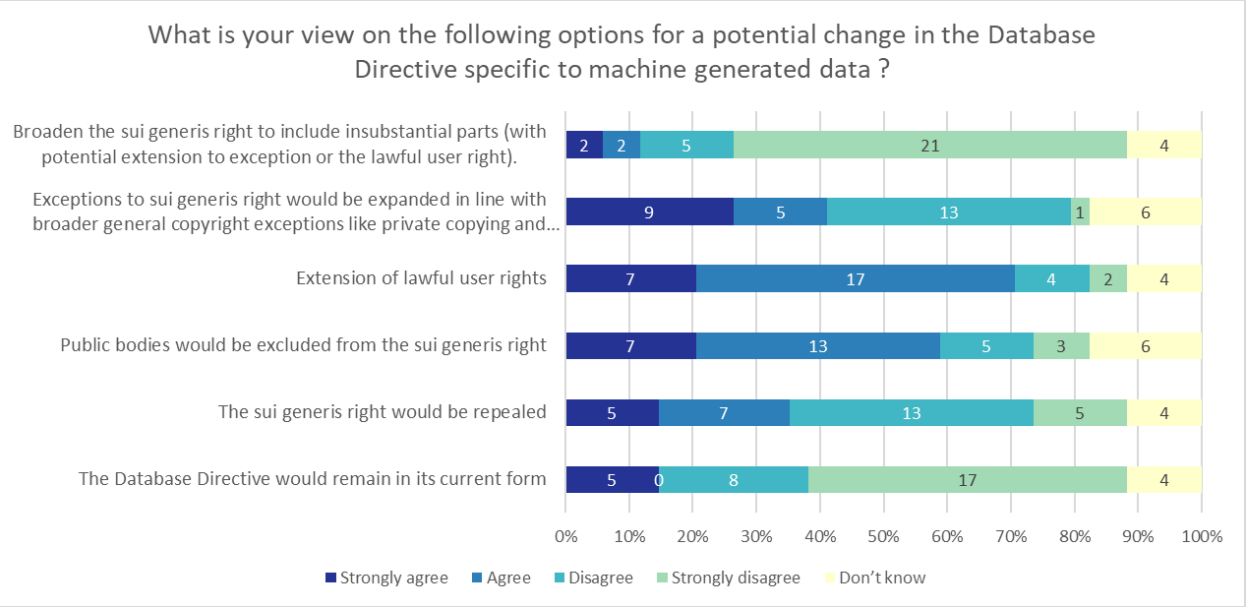
Note: Based on F1 n = 19

Asked about the options of 'maintaining the *status quo'* or of 'introducing general changes' to the Database Directive, there is clearly a favourite option: 70% agree or strongly agree to extend rights of lawful users. The other options receiving favourable agreement by half of the respondents is the exclusion of public bodies from the *sui generis* right.

About 70% of the respondents disagree to maintain the Database Directive in its current form.

The industrial associations' views about these options are also clear: they strongly disagree to keep the Database Directive in its current form, to broaden it to include insubstantial parts, to repeal the Database Directive and to expand exceptions. They agree on excluding public bodies and extending lawful user rights.

Figure 51: Views on the options for remaining with the status quo or a potential general change in the Database Directive



Source: Survey for the Database Directive Review
Note: Based on F2 n = 34

Legal experts

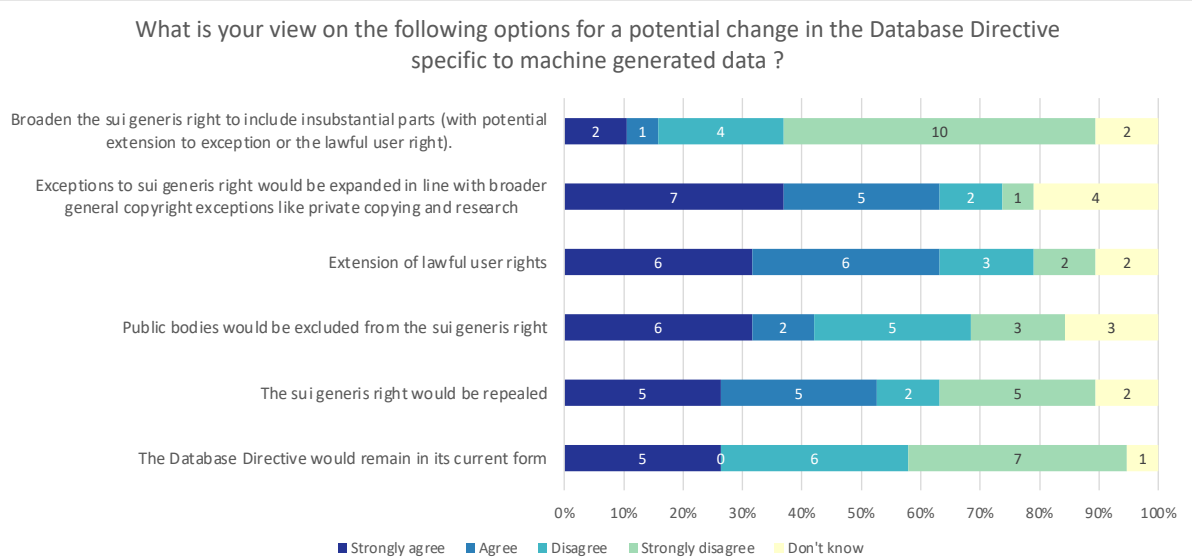
Only looking at the legal experts' answers, no uniform agreement can be found on the options. In fact, for some options, their views were rather split while in some, one can see a common direction.

The option to introduce exceptions to the *sui generis* right coupled with exceptions to copyright and extending rights of lawful users obtained mostly favourable opinions (more than 60% 'strongly agree' or 'agree'). About 50% agree or strongly agree to repeal the *sui generis* right altogether. The option of 'broadening the *sui generis* right to include insubstantial parts (with a potential extension to exception or the lawful user right)' is agreed or strongly agreed by three respondents while all other disagree.

Five (25%) of the legal experts are in favour to keep the Database Directive as it is while about 70% disagree to maintain the Database Directive in its current form.

R&D stakeholders with legal expertise are in favour of extending rights of lawful users and of introducing exceptions to the *sui generis* right (in both cases, 7 out of 8 advocate for these two options).

Figure 52: Views of legal experts on the options for remaining with the status quo or a potential general change in the Database Directive



Source: Survey for the Database Directive Review

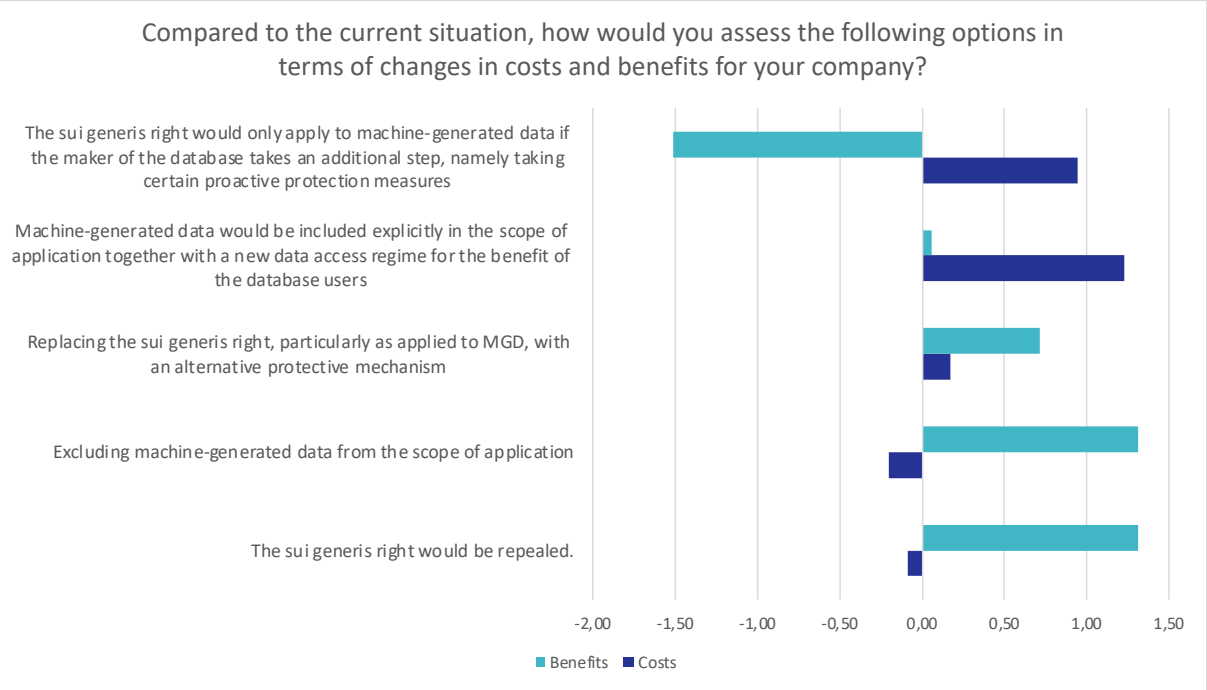
Note: Based on F2 n = 19

Asked to provide estimates on costs and benefits for each of the five options by indicating values in the range of -3 (very low), 0 (neutral) to 3 (very high), the option of applying the *sui generis* right to MGD in combination of proactive steps has the least benefits and the highest costs among all options. On the other end, repealing the *sui generis* right would have very limited costs but high benefits. The option to exclude MGD is also seen as an option with high benefits and limited costs.

For the respondents overall, the exclusion of MGD has higher benefits than the inclusion.

The industry associations indicated high costs for the option of explicitly including machine generated data with a new access regime while in terms of benefits, they estimated the highest benefits if the *sui generis* right would be repealed and if machine generated data would be excluded. Least beneficial would be if the *sui generis* would only apply to machine-generated data if the maker of the database takes additional steps.

Figure 53: Assessment of different options in terms of costs and benefits



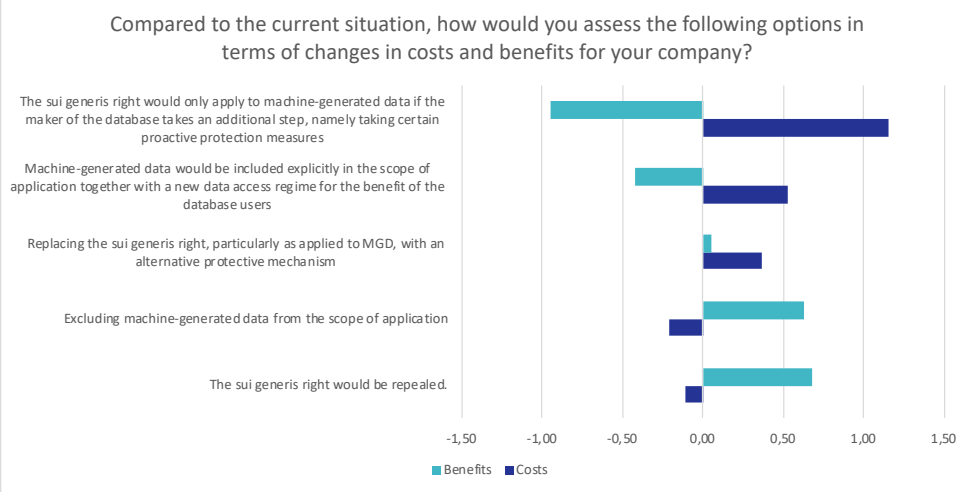
Source: Survey for the Database Directive Review
Note: Based on F3 n = 35

Legal experts

The 19 legal experts among the respondents have rather similar views but in terms of most benefits, repealing the *sui generis* right has slightly the highest benefits and very limited costs. In terms of costs, the option to include machine generated data in combination with additional steps is the most costly option and it also offers the least benefits. One respondent remarked: "*Finding rightholders of databases, reviewing the scope of protection and finding a license creates significant costs. The creation of databases is mostly based on pre-existing data(bases) and the sui generis* right *is complicating things significantly. The benefit is low. Reuse of data and any data driven business model is complicated*."

R&D legal experts expect low costs and high benefits should the *sui generis* right be repealed.

Figure 54: Assessment of different options in terms of costs and benefits -views of legal experts only



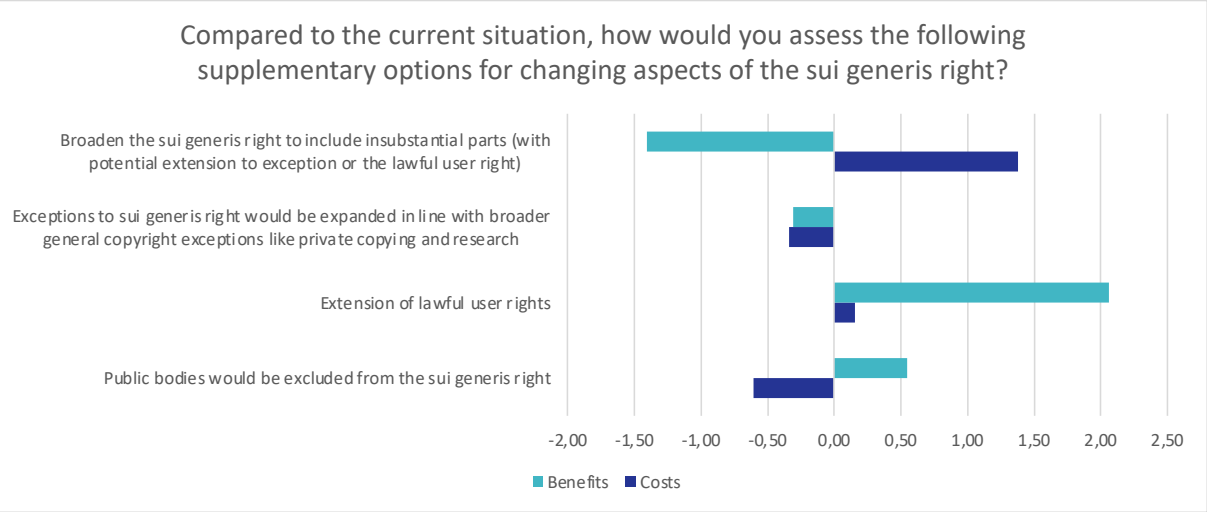
Source: Survey for the Database Directive Review

Note: Based on F3 n = 19

Similarly, supplementary options were tested on their expected costs and benefits. To the 34 respondents, the least beneficial and most costly supplementary option would be to broaden the *sui generis* right to include insubstantial parts. Also, the expansion of exceptions is not seen as beneficial, but also not as costly. The highest benefit and lowest cost is assessed for the supplementary option to extent lawful user rights.

For the industry associations, the highest costs would come with the option to broaden the *sui generis* right to include insubstantial parts (with a potential extension to exception or the lawful user right). The associations see the main benefit in the extension of the lawful user and the least benefits in broadening the *sui generis* right to include insubstantial parts and expanding exceptions to the *sui generis* right.

Figure 55: Assessment of different supplementary options in terms of costs and benefits

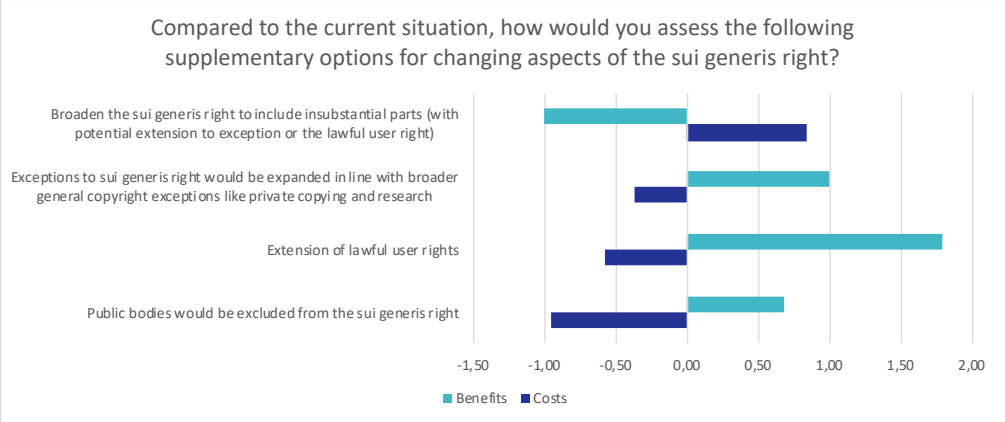


Source: Survey for the Database Directive Review
Note: Based on F4 n = 34

Legal experts

Similarly for legal experts, the least beneficial and rather costly option would be to broaden the *sui generis* right and include insubstantial parts. This option would also have the highest costs among the supplementary options. In terms of highest benefits, the option to extend lawful user rights was seen as most beneficial while its costs are relatively low. The least costs but considerably benefits are seen with the option to expand exceptions. In this option, there are thus marked differences between legal experts and other respondents. Legal experts also see even less costs and a similar level of benefits for the option to exclude public bodies from the *sui generis* right.

Figure 56: Assessment of different supplementary options in terms of costs and benefits – views of legal experts only

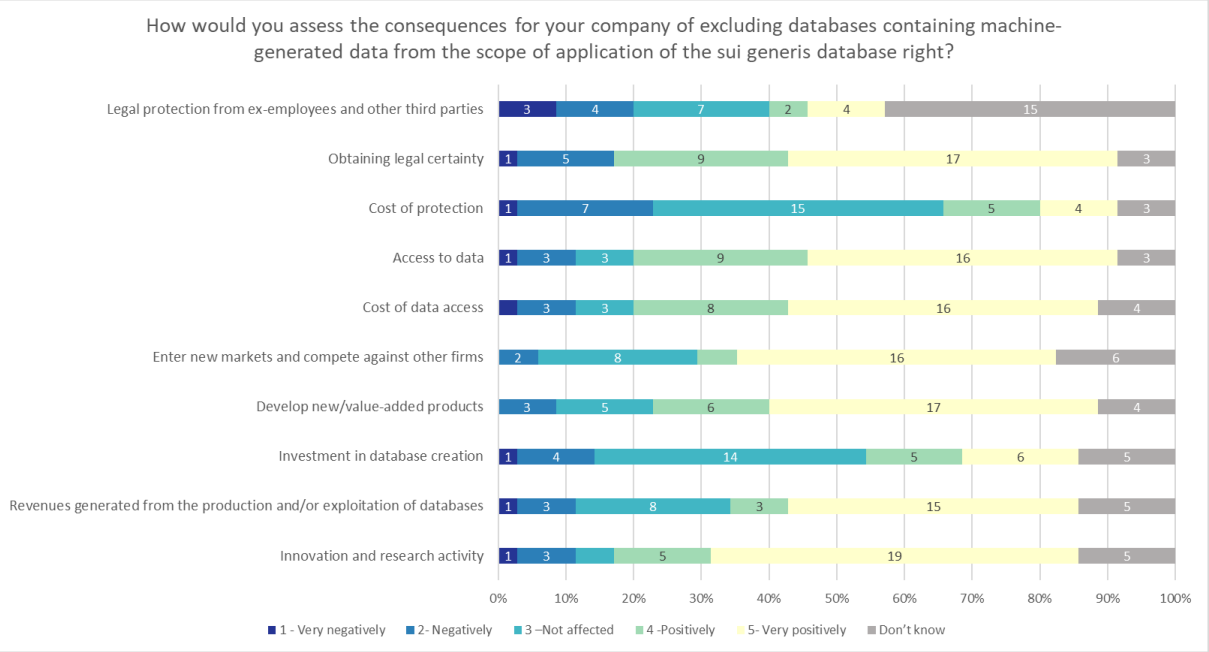


Source: Survey for the Database Directive Review
Note: Based on F4 n = 19

The exclusion of databases containing machine-generated data from the *sui generis* right would bring negative consequences to 10-20% of the businesses and operations. Only in terms of cost of protection about 25% indicated negative consequences. Exclusion would mainly not affect investments in databases. In seven out of the ten consequences, the respondents consider the exclusion as positive, for example, in terms of obtaining legal certainty or to develop new, or value-added products.

An industry representative further pointed out that "*Excluding machine generated data from the sui generis right and easy access to such data would foster innovation and competition with regard to data driven business models*".

Figure 57: Consequences of excluding databases containing machine generated data from the scope of application of the *sui generis* right



Source: Survey for the Database Directive Review
Note: Based on F5, n = 35

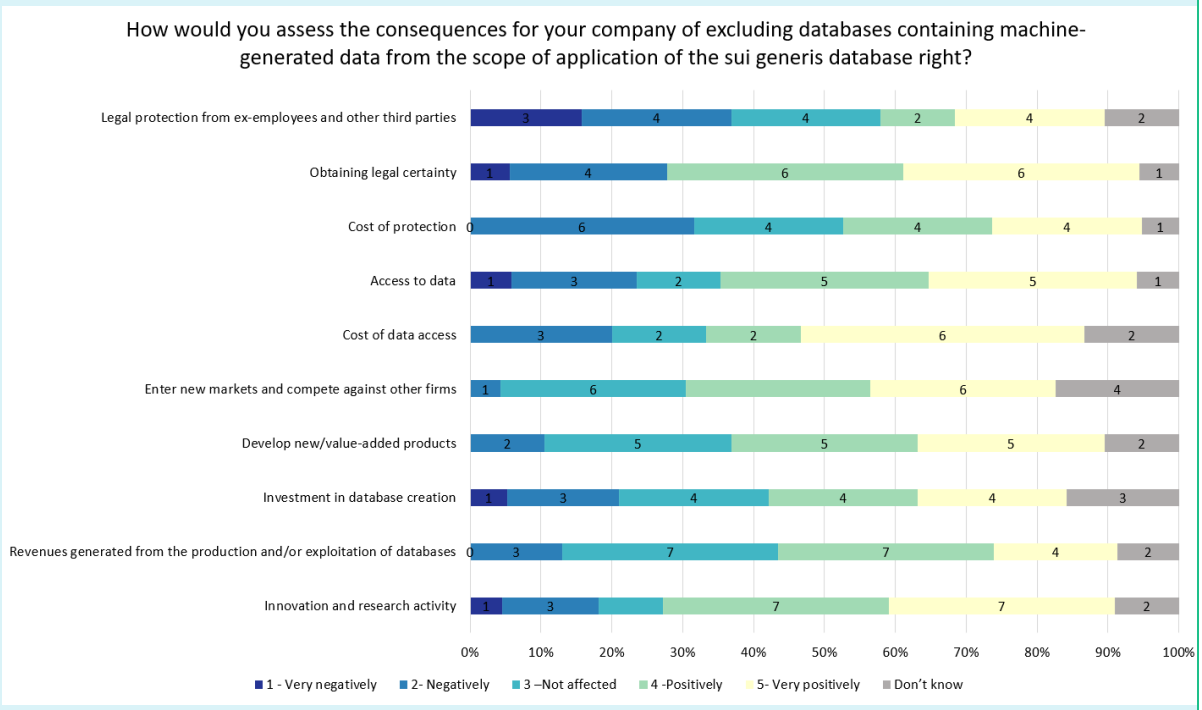
Legal experts

When it comes to the consequences of including databases containing of machine generated data, more than 50% of the legal experts consider that this would probably have negative or very negative effects on access to data for the business of their company/organisation. 45% think that such change would have a negative or very negative effects on the costs of data access, costs of protection, or on obtaining legal certainty. Positive effects were seen in obtaining legal certainty as well as research and innovation.

One legal expert remarked that: "*The problem is that [it] is difficult [to] assess whether or not data is based on a substantial investment and/or if the data [is] machine-generated or not. Accordingly, the protection is unclear. Unclear IP Rights a[re] negative to innovation and rightful users. Illegal users do not care anyway. Hence, it is best to repeal the sui generis* right*. it does not work properly*."

Several R&D legal experts also advocate for the repealing of the *sui generis* right and oppose the creation of new rights to protect machine-generated data. In their view, both options would undermine open science, research and innovation public policy developments, and research. 90% fear that the inclusion of machine-generate data in the *sui generis* right would have a negative or very negative effect particularly on the access to data, the costs of protection, and the costs of data access.

Figure 58: Consequences of excluding databases containing machine generated data from the scope of application of the *sui generis* right – views of legal experts only



Source: Survey for the Database Directive Review

Note: Based on F5, n = 19

The consequences of including databases containing machine generated data into the scope of application of the *sui generis* right together with a new data access regime were also surveyed. 35 respondents provided their views on a set of ten areas of potential change.

21 respondents (60%) think that such change would affect the cost of data access 'negatively' or 'very negatively'. 18 respondents, or more than 50% point out that it would bring more legal certainty and for 15 respondents, or more than 40% it could produce additional revenues from the production and/or exploitation of databases.

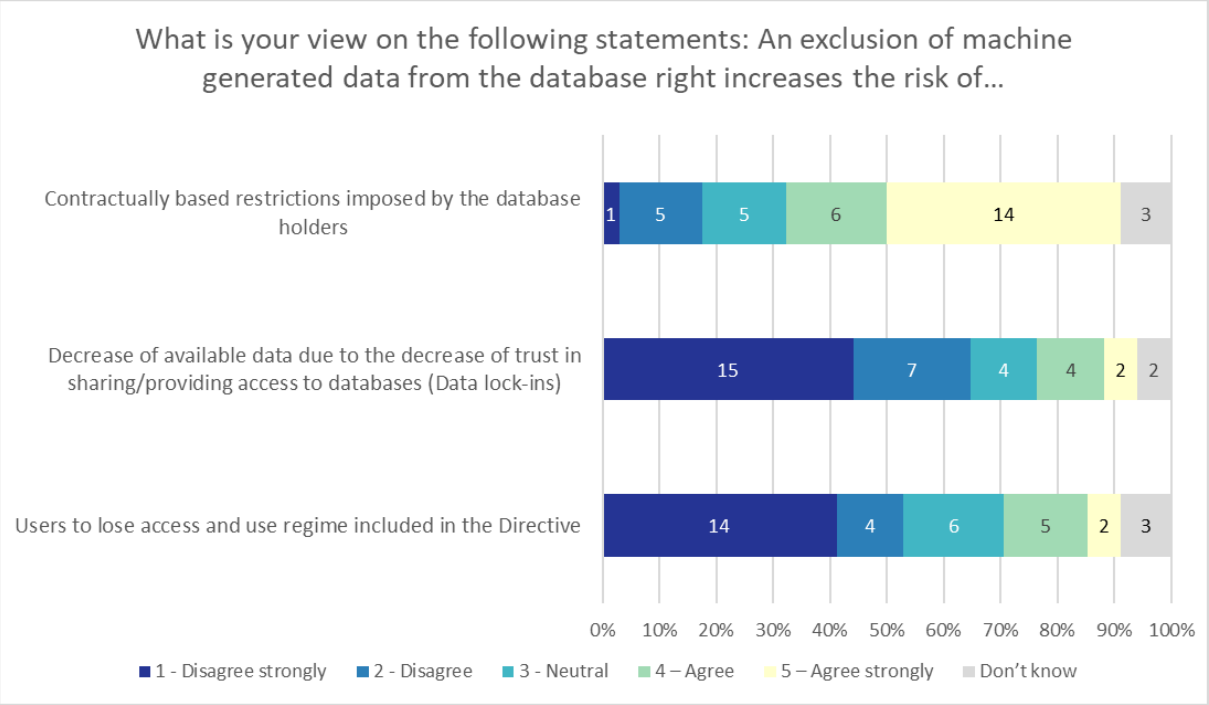
The industry associations voiced a negative effect on the cost of data access.

Finally, respondents were asked about potential risks due to the exclusion of machine generated data from the *sui generis* right.

Almost 40% of the respondents considered that this would increase the risk of contractually based restrictions imposed by the database holders.

They disagree or disagree strongly to the two other risks, namely that this would limit available data and lead to data lock-ins or that users would lose the access and use regime included in the Database Directive.

Figure 59: View on risks due to an exclusion of machine generated data from the database



Source: Survey for the Database Directive Review
Note: Based on F7 n = 34

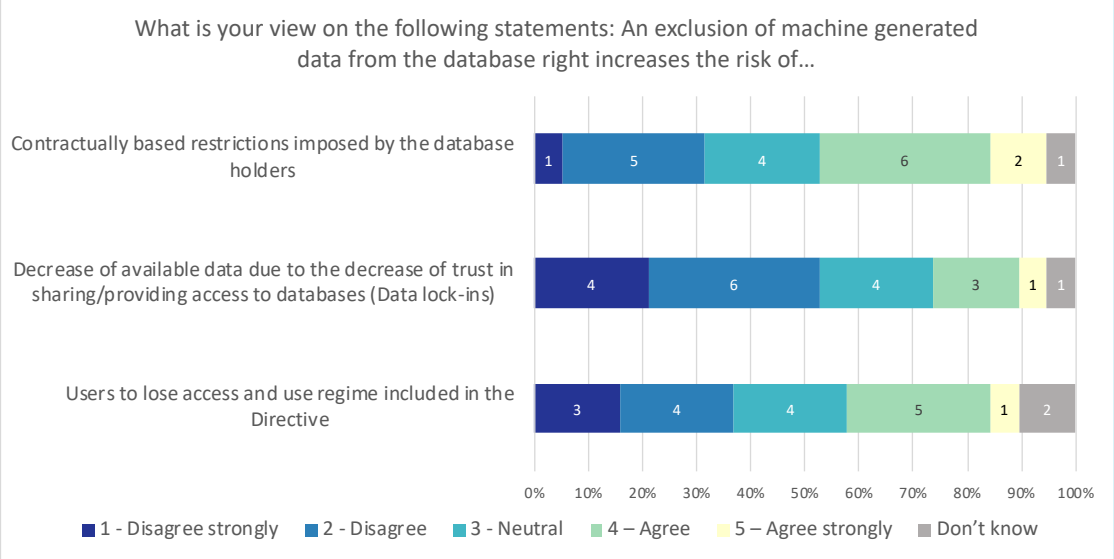
Legal experts

More than half of the respondent legal experts consider that this would translate into data lock-ins (i.e. decrease of available data due to the decrease of trust in sharing/providing access to databases).

Neither of the two other potential risks (i.e. 'Contractually based restrictions imposed by the database holders' and 'Users to lose access and use regime included in the Database Directive') are clearly seen as likely in case of exclusion of machine generated data from the *sui generis* right: the sum of 'disagree strongly' and 'disagree' equals the sum of 'agree' and 'agree strongly' responses.

R&D stakeholders with legal expertise do not think that an exclusion of the machine-generated data from the *sui generis* right would decrease available data due to the decrease of trust in sharing/providing access to databases, thus producing data lock-ins (only one thinks so). They rather think (4 out of 8) that it would bring to contractually based restrictions imposed by the database holders.

Figure 60: View on risks due to an exclusion of machine generated data from the database – views of legal experts



Source: Survey for the Database Directive Review

Note: Based on F7 n = 19

D Annex D: In-depth interviews with legal experts and industry sections

D.1 In-depth interview approach and limits

Two interview programmes with selected industry stakeholders and a brainstorming workshop with legal experts we carried out to complement the online survey. As such, in-depth interviews allow to provide valuable insights on specific topics while providing a more interactive setting necessary to legal forward thinking as well as understanding concerns and needs of the different organisations consulted.

The first strand of in-depth interviews consisted in a group discussion with legal experts held online on the 01/07/2021. In order to foster the discussions, an interactive whiteboard was prepared on Mural which displayed the rules of participation as well as the support documentation, a visual of the options to be discussed and the main topics of investigation to be covered during the workshop. The discussions were moderated by a legal expert of the study team and facilitated by a senior consultant trained for this kind of exercise in order to increase the contribution of each participant and balance the different point of views. In addition, each participant was able to interact with the online whiteboard providing written inputs, comments, and reactions without interrupting the intervention of the different speakers. Following the group interview, a summary transcript was shared with the EC. It was also completed by an ad-hoc interview with a legal expert specialised in competition law and familiar with the problematic of the Database Directive.

In total**, 7 legal experts** have directly contributed to the study, and 6 participated to the online brainstorming workshop.

The second strand of the interview programme consisted in in-depth interviews with representatives from different industry sectors. Similar sources were used to design the interview guidelines and tailored them to the specificity of each interviewee (legal background, IT, executive positions). A more flexible format was used to allow for an efficient and effective interaction with the respondents, implementing mostly grouped interviews (2 to 6 participants) to cover the different functions within an organisation.

The selection of organisations paid a specific attention to find the adequate balanced set of interviewees to ensure that all the relevant industries and stakeholder groups were covered and reached out, along the same lines as presented in the case of the survey in the previous section. Given the specificities of the subject, a combination of desk research, existing networks of contacts within the consortium, previous studies and the B2B provider Leadiro plc were used to identify key contacts within the targeted sectors. The provider allowed to retrieve direct emails from LinkedIn. Search by function/role/title of the job were limited to “legal” and “IT” and “executive” positions to increase the potential relevance.

A list of **45 organisations** was contacted with the following profile:

Sectors	Executive positions	IT positions	Legal positions	Total
Aerospace and Aviation			1	1
Agriculture	2			2
Automotive			3	3
Banking/Accounting/Financial		5	3	8
Connected devices	3		1	4
Construction	1			1
Creative industry	1		2	3
Energy		1		1
IT industry			1	1
Machinery/IoT/Sensors			1	1
Manufacturing and automotive		1	5	6
Pharmaceutical	3	1	2	6
Repair & maintenance	1		1	2
Retail and wholesale		1	1	2
Sports			2	2
Transportation and Shipping		2		2
Grand Total	**11**	**11**	**23**	**45**

Following the initial identification, the recruitment phase started on the 29/06/2021 with the 20 first organisations identified, followed by a second batch of organisation identified on the 20/07/2021. In order to increase the interest of the potential organisations, the invitation emails stated the objectives of the study, the potential interest that interviewees might have in participating (tailored to the specific companies) to an interview, the cover letter to increase the legitimacy as well as a clear define timeframe.

The team also contacted a number of law firms from its network in order to promote the study and recruit potentially interested corporate clients to interview.

Once the first contact was successful, interviewees were provided with the overview of the considered policy option, a background note including the main topic of investigation in view of preparation of the interview as well as a Teams invitation to join the online meeting.

Despite 2 to 3 reminders, as well as attempts to reach out interviewees by phone (when contacts were available) – the study team was able to conduct **only 5 in-depth interviews accounting for 14 participants** and have clear answers from 7 organisations (2 being either not interested or not available). Only the findings of 4 in depth-interviews were used for the report. Overall, a series of limitations must be considered:

- **Technical limitation**: most of the contacts identified and relevant organisation do not have their updated contact information available online which resulted in a certain amount of rebounces and created delays in identifying the correct domains/addresses

- **Awareness limitation**: identified contacts were either not directly involved or unaware of the Database Directive and its relevance for their activity which led to difficulties for them to identify within their organisation the relevant contacts that might be willing to contribute.
- **Timing limitation**: The summertime period led to an important number of Out of Offices and difficulties to schedule interviews before the end of August.

D.2 Interview guidelines

Interview partners received a short background information. Overal, the study team used a number of questions to guide the interview. In the following, the key points from the different questions posed are summarised.

Asked about the usefulness of the Database Directive's *sui generis* right in the data economy, interviewees argued that they find the database directive to be ineffective and/or an impedement and gave as an example that the length of protection is 15 years. In a sectorial context such as automobile, after 15 years, the car is not valuable thus, too long protection period. Others were worried about personal data protections laws. Only one interviewee stated that its company had been able to add value through the Database Directive.

When it comes to alternative mechanisms such as contractual agreements, smart contracts, technological solutions, or trade secrets, interviewees were asked if those were of more use. Two different arguments were put forward, one being that they would rather like to see more access rights as a more useful. Among the other existing tools, others were argueing for contracts. This instrument gives a precise legal framework and keeps them from battling in court.

Rather diverse views were voiced on the question if the Database Directive should be kept as it is, repealed, or adapted. In fact, there was no general consensus on this question. Some interviewees argued the Database Directive should be adapted to include machine generated data (MGD) while others wanted the latter to be excluded due to the oblique nature of the database directive. Only one interviewee wanted it to be kept as it is.

Given the rise of machine generated data (interviewees were asked if this kind of data should be explicitly included in or explicitly excluded from the Database Directive. The majority of the interviewees agreed that MGD should be included in the Directive, however, their opinions differ slightly on whether they should be granted a *sui generis* right for the database based on the investment. The interviewees whose organisations deal predominently in data collection wanted it to be included. They argued about the costs to set up the database and create the algorithms while those who use data predominently prefer it to be excluded as they believe it is to be a by-product and therefore should be excluded.

In terms of 'mixed datasets' which contain MGD and other data, interviewees thought that this is becoming more and more common practice, but also that it is costly to separate the data now. Due to technical advancement, this task could be cheaper in the future. Also, some claim that the collection of data from machines is becoming more common practice and could be the main source of data collection in the future.

Effects of the exclusion of MGD are predominantly of a financial sort. Interviewees believe that it would be costly to exclude MGD. The Directive functions as an incentive to create and add value to the data, and protects their investment. However, other are in favour to exclude MGD. They argue that data holding companies charge too high prices for access. This hinders business growth, is costly, and disincentiveses its use.

Asked about an appropriate protection period, interviewees do not provide one consistent answer. In one particular industry, the need for real time access was stated and thus no protection period of MGD suggested while for manurfacturing industries, a period of ten years was suggested.

With respect for protection to be extended it has been suggested that the IP could last for 20 years which includes 10 years protection and then 10 years renewal - if applied for. However, the general consensus seems that MGD should not be included as it would be inefficient to introduce/keep a 'timestamp' for each single data point generated, in particular in case of continuous sensored data, it would also be more difficult to define when it starts and stops and could be costly for the producer.

Interviewees were also asked about introducing conditions, should MGD be included. The interviewees were not sure how useful a mechanism could be. To have the *sui generis* protection is a prefered option. They also favour clarity and transparency not only around the law but also about the data itself to understand how useful it is. Interviewees agree that the machine manufacturer should be paid to be incentivised.

In terms of a new right that could be introduced instead of the inclusion of MGD, the general thought here was that it could complicate things unnecessairly. Some already have trade secrets and other forms of protections in place, however, from a consumer persepective it would be important to have options.

Interviewees were asked about benefits of other potential changes which could be thought of in case of an exclusion of MGD. Asked about the introduction of general copyright limitations, German interviewees pointed out that this would already exist in the German Urhebergesetz. It may disincentive data sharing which may lead to problems; even big companies cannot get all the data they need. Although, as long as the company is protected by either a database right or copyright law, the feeling is that it would still function as an incentive.

When it comes to the inclusion of a set of exceptions (such as permission of the use of search engines, conditional web-scraping), the inclusion of options such as web-scraping could have negative impacts as it could infringe on the database protection. However, exceptions could be made for areas such as news reporting and advertising, as conditions could be built upon for these exceptions.

A moot aspect is the term 'lawful user'. Interviewees indicated that clarification would be welcome, however, it should not be used for the whole database as it could disincentivise data producers. Contractual agreements for certain parts of the database would be optimal. Access of lawful users would include a form of portability through which users can grant access to certain data of their machine, for example to repairers and other providers.

Interviewees also pointed out that many manufacturers have their own interpretation of GDPR. This could cause friction between manufacturers and consumers (and data intermediaries) but it is imperative that consumers give their consent to data use. Some interviewees believe that the consumer should be entirely in control of their own data.

On another suggestion, namely on the use of exisiting copyright exceptions, interviewees find it unclear how existing copyright exceptions would be used. They are in favour of a threshold or instrument to be used to calculate what is deemed to be 'substantial'. One interviewee was in favour of keeping things the way they are.

Interviewees welcomed the introduction of a compulsory licensing regime under FRAND terms for sole-source databases containing MGD and/or any sole-source databases. In general,

FRAND terms would have a positive impact across all businesses. However, it was also stressed that 'fair price' and 'compulsorary' would have to be very well defined.

When it comes to the situation of public databases that enjoy the protection of the *sui generis* right and impacts on own businesses, interviewees mentioned that public databases could add value to their own business. They stated that everything has a business model even open source data. Exclusion or inclusion – for some this is more of a political question.

When asked if the interviewees own company databases contain also personal data, the general consensus was that this is out of the scope of the *sui generis* right but falls under GDPR. Some data such as on payrolls is kept in a database but GDPR guidelines apply to this.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

KK-09-22-006-EN-N

doi: 10.2759/647387
ISBN 978-92-76-46550-8