

Evaluation of the Impact of Framework Programme supported Social Sciences and Humanities Research

A bibliometric approach

Interested in European research?

Research*eu is our monthly magazine keeping you in touch with main developments (results, programmes, events, etc.).

It is available in English, French, German and Spanish. A free sample copy or free subscription can be obtained from:

European Commission
Directorate-General for Research
Communication Unit
B-1049 Brussels
Fax (32-2) 29-58220
E-mail: research-eu@ec.europa.eu
Internet: <http://ec.europa.eu/research/research-eu>

Pictures © Shutterstock, 2010

EUROPEAN COMMISSION

Directorate-General for Research
Directorate L - Science, Economy and Society
Unit L1 - Horizontal Aspects and Coordination

Contact: Anne-Sophie Malherbe

*European Commission
Office SDME 07/51
B-1049 Brussels*

E-mail: Anne-Sophie.Malherbe@ec.europa.eu

Evaluation of the Impact of Framework Programme supported Social Sciences and Humanities Research

A bibliometric approach

Authors

Dr. Viola Peter (Project leader)

Lorena Rivera Leon

Dr. Yann Cadiou

Mathieu Doussineau

***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

(*) Certain mobile telephone operators do not allow access to 00 800 numbers
or these calls may be billed

LEGAL NOTICE:

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information.

The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (<http://europa.eu>).

Cataloguing data can be found at the end of this publication.

Luxembourg: Publications Office of the European Union, 2010

ISBN 978-92-79-15370-9

ISSN 1018-5593

doi 10.2777/18202

© European Union, 2010

Reproduction is authorised provided the source is acknowledged.

Table of Contents

Executive summary	2
Assessment of the overall impact of the Key Action	3
<hr/>	
1. Introduction	4
1.1 Objectives and aims	4
1.2 Definition of SSH	5
1.3 FP-funded SSH research background	5
1.4 Bibliometric analysis in socio-economic research	6
1.5 Structure of the report	9
<hr/>	
2. IHP participants – descriptive statistics	10
2.1 SSH Database establishment	10
2.2 Descriptive analysis	10
<hr/>	
3. The impact of FP5 projects’ scientific publications	16
3.1 The sample	16
3.2 Sample descriptive analysis: type of publications	20
3.3 Sample descriptive analysis: authors	25
3.4 Preparing citation analysis of FP-published articles	29
3.5 Benchmarking FP-publications with other publications of the authors	31
3.6 Book publications	33
3.7 Research networks	35
<hr/>	
4. Benchmarking the “status” of the Key Action in the international scientific scene	42
4.1 Identification of journals – defining the pool of documents	42
4.2 Retrieval of all bibliographic information	43
4.3 Limitation to journal articles	43
4.4 Attribution of total number of citations per article to each author	44
4.5 Establishing lists of the top 100 European authors	44
4.6 Matching the top 100 individuals with the IHP-database	44
<hr/>	
5. Results and recommendations	46
5.1 Technical results	46
5.2 Limitations of the use of bibliometric analysis	47
5.3 Recommendations for further refinements of the approach	48
<hr/>	
Bibliography	50

List of Exhibits

Exhibit 1:	Number of projects by project call.....	10
Exhibit 2:	Number of participating institutions (participants) per project (all FP5-IHP projects)	11
Exhibit 3:	Number of participating institutions (participants) by call	11
Exhibit 4:	Number of project coordinators by country (all FP5-IHP projects).....	12
Exhibit 5:	Number of coordinators by call and country	13
Exhibit 6:	Origin and number of participants by country (all FP5-IHP projects).....	13
Exhibit 7:	Origin of participants by call and country.....	14
Exhibit 8:	Total number of information source by type and call.....	17
Exhibit 9:	Sample statistics: Share of participants by country.....	20
Exhibit 10:	Sample statistics: error margins per type of publications	22
Exhibit 11	Productive FP-authors: FP publications vs WoS publications and citation rates	28
Exhibit 12:	Links between authors in FP-resulting books	37
Exhibit 13:	Links between authors in book chapters.....	38
Exhibit 14:	Links between authors in books and chapters.....	39
Exhibit 15:	Links between authors in FP journal articles publications	40
Exhibit 16	Age and gender profiles by call.....	47

List of Tables

Table 1: Coverage of the database	10
Table 2: Project type by call	14
Table 3: Average financial contributions by call and project type	15
Table 4: The SSH sample (1) (contin.).....	18
Table 5: Sample statistics: Project numbers, average financial contribution and average number of participants by project type and call year	20
Table 6: Standard deviation, confidence levels and standard errors.....	22
Table 7: Overview of the sample by project (contin.)	23
Table 8: Number of authors per project (total, inside and outside consortium)	25
Table 9: Sample statistics: Publications per author (totals, on shares, by in- and outside consortium authors).....	27
Table 10: Publication numbers by type and year (1).....	29
Table 11: Sample descriptive statistics of authors contributing to journal articles....	30
Table 12: Total number of authors contributing to a given number of articles per call year	30
Table 13: Sample statistics: Overall citation results comparing FP publications to overall oeuvre of authors	32
Table 14: h-index by number of publications (1)	33
Table 15: Citation rates for FP-resulting book publications using Google Scholar ..	34
Table 16: Average number of citations per FP-book and per FP journal article by publication year	35
Table 17: Average number of authors per book	36
Table 18: Number of authors per chapter.....	38
Table 19: Number of authors per FP journal article publication	40
Table 20: Number of documents by field (2002-2006, top impact journals)	43
Table 21: Number and share of articles and authors by field	43
Table 22: Number and share of articles and authors by field	44
Table 23: Number and share of articles and authors by field.....	44
Table 24: Number of participants in the socio-economic Key Action among the 100 Europeans with most citations in high impact journals	45
Table 25: Number of institutions/departments participating in the Key Action amongst the affiliations of the 100 Europeans with most citations in high impact journals	45

Executive summary

The evaluation of the scientific impact of the programme for the social sciences and humanities (SSH) – The Key Action Socio-Economic Research of the 'Improving the Human Potential' (IHP) programme of the 5th Framework Programme for Research and Development (FP5), was commissioned by DG-RTD in 2008 and performed by Technopolis Consulting Group in 2009. It aimed at measuring and assessing the scientific impact of the Key Action by means of bibliometrics.

In this respect, the evaluation had an explorative character as bibliometric analysis, although being fairly well established, has not been widely used neither for the evaluation of research programmes, nor for analyses of social sciences and humanities.

The objectives of this study were:

1. To establish the profile of scientific publications from European cooperative research supported by the Key Action, and to perform a bibliometric evaluation of this profile by using citation analysis.
2. To build benchmarks to establish the "status" of the Key Action in the international scientific scene.

Key findings

- The Key Action funded 203 projects (147 Research Projects, 38 Thematic Networks and 18 Accompanying Measures) and the average financial contribution for all three project types was € 761.000. On average, there were ten project partners which teamed up for an average of 32 months. Participants from 39 countries were identified.
- Based on the overall structure, a stratified sample of 40 research projects and thematic networks was constructed. The list of publications reported in the final reports and the projects' websites was included in a publication database. The data base showed:
 - 981 authors which authored 1.449 publications;
 - The majority of the publications (almost 45%) were 'grey literature' – working papers, reports, etc., followed by scientific articles (28%), individual book chapters (19%) and books (8%)¹.
- On this basis, we estimate the publications output of the Key Action to be in the range between 6.300 and 7.100 publications, among those about 1.600-2.200 journal articles and 340-730 books.² These resulted from 185 research projects and thematic networks that brought together 2.250 participants from 39 countries consuming €150 million over eight years.
- Published articles in peer-reviewed international journals resulting from Key Action research are more cited than the average publication of the same author(s). The average citation score for the total oeuvre of the same authors was 2.9 citations

¹ The shares for books and book chapters was lower when this publication type was analysed further in an additional step at the end of the project.

² The publication output may be somewhat underestimated given that publications resulting from the projects may also be started only after the end of the projects – thus they would not necessarily be included in the final reports. On the other hand, a number of publications were mentioned in the reports which dated from before the project even started, or in their inception year – which seems highly unlikely for book publications for example.

compared to 4.2 for the FP5-resulting articles (three year citation window). Also, for the same authors, on average 31% of articles remain non-cited, while only 11% of the FP-resulting articles remain non-cited.

- Within the same population, articles by female authors obtain significantly higher citation rates than male authors – this holds for their average article but in particular for their FP5-resulting article.
- Co-publications are very common in the sample. On average, publications from research projects have 1.5 authors, and those from thematic networks 1.9. The co-authors are not necessarily co-operating partners within a given research project - most likely about half of the co-authors are not associated with the Key Action's project.
- Journal articles form a much more important part of the total oeuvre resulting from the Key Action than was originally expected based on existing literature (e.g. Hicks 1999). Also for the articles in our sample a citation window of 3 or 4 years seemed appropriate for evaluation purposes. Thus, we conclude that bibliometric analysis is much more suitable for the evaluation of programmes like the Socio-Economic Key Action than previously thought.

Structural impacts of the Key Action

Given that SSH funding via the European Commission had just started with FP4, the results for FP5 show that the Key Action has attracted a large number of researchers. Key Action projects have allowed several researchers to establish their researchers careers' and others to strengthen, establish or widen their research networks.

The analysis suggests that the structural effects of the Key Action are important and are mutually enforcing scientific excellence. These are:

- The Key Action offered an opportunity for younger researchers to pursue research of international interest that is accepted in refereed journals. These articles were often the starting point of their research career.
- The chance to pursue international research collaboration and obtain important results was more successfully taken and achieved by a large share of female researchers. Previous results showing female researchers to be more selective in terms of output and more successful in terms of citation and impact was confirmed in our sample and thus holds also with the community funded research projects in SSH.

Benchmarks and conclusions

- Our attempt to benchmarking showed that very few of Europe's top highly cited social scientists were identifiable among the researchers who participated in the programme. However, the applicability of the concept of highly cited authors in SSH is highly debatable especially in the context of lack of comprehensive coverage of databases.
- On the basis of this study we conclude that:
 - It is feasible to establish a quantitative view of the scientific impact of the programme, and the results of this study can be used as a benchmark for an evaluation of SSH in following Framework Programmes with a time-lag of one or two years.
 - Establishing an external benchmark with reference to which one could assess the impact of the programme remains a challenge. Such benchmarks require:
 - Comprehensive coverage in bibliographical databases;
 - Comprehensive coverage in the data-bases covering the researchers that worked in the programme

Evaluation of the Impact of Framework Programme supported Social Sciences and Humanities Research: A bibliometric approach

1. Introduction

Scientific and technological development is recognised as essential for the functioning and development of industrialised countries like the Member States of the European Union (EU). Its contribution to growth, employment, the protection of consumers and of the environment, to greater competitiveness and to resolving the most important issues facing society makes consensus today.

This report is documenting the efforts and approaches made in order to evaluate scientific excellence in the specific Key Action for the social sciences and humanities (SSH) – ‘Improving the Human Potential’ (IHP) of the 5th Framework Programme (FP5). This task was commissioned by DG-RTD Directorate L – Science, Economy and Society in 2008, and executed in 2009 by Technopolis Consulting Group³.

1.1 Objectives and aims

The objectives of this contract were as follows:

1. Using citation analysis in order to address the impact of scientific publications resulting from teams that have participated in FP5 projects,
2. To build benchmarks to establish the “status” of the Key Action in the international scientific scene.

The terms of reference of the study also mentions the use for the analysis:

“It will provide indications as to how well the FP supported SSH projects are performing in relation to their ex ante assessments, and will also provide suitable baselines for future ex post evaluations.”

The objectives are broken down into three tasks:

1. Establishment of a profile of scientific publications from European cooperative research projects, and a bibliometric evaluation of this profile.
2. Development of appropriate benchmarks to assess the overall status of the Key Action in the international scientific scene.
3. Reporting on the impact of European research on advancement of knowledge in SSH, as well as on the quality and performance of European research.

In the offer, the contractor pointed out to several problems arising using bibliometric analysis, in particular concerning the social sciences and humanities, and in relation to measuring excellence. Even if various approaches were tested and some very interesting results were obtained, the study can confirm several of the shortcomings and warnings that were mentioned in previous academic articles. The following thus recapitulates the most relevant problems.

³ Evaluation of the Impact of Framework Programme Supported Research on EU excellence in Social Sciences and Humanities (SSH).

Objectives:

- Bibliographic profile
- Bibliometric assessment
- Benchmarking the status of the programme

Definition of SSH

There is no unique and uniform used definition of the social sciences and humanities. Widely used is for example the *UNESCO nomenclature for fields of science and technology*, which provides 13 fields at 2-digit level for the socio-economic sciences and humanities, and which includes fields from Anthropology to Philosophy.

The most widely used classification for statistical purposes is possibly ISCED, the *International Standard Classification of Education*, another UNESCO scheme in its revision of 1997. The scheme lists 25 fields, among those also the humanities and social sciences, business and law. These fields cover the following:

- 22 Humanities
 - Religion and theology; Foreign languages and cultures: living or ‘dead’ languages and their literature, area studies;
 - Native languages: current or vernacular language and its literature;
 - Other humanities: interpretation and translation, linguistics, comparative literature, history, archaeology, philosophy, ethics.
- 31 Social and behavioural science
 - Economics, economic history, political science, sociology, demography, anthropology (except physical anthropology), ethnology, futurology, psychology, geography (except physical geography), peace and conflict studies, human rights.
- 32 Journalism and information
- 34 Business and administration
- 38 Law

Is this compatible to what was funded under the IHP Key Action? The Key Action does not provide a definition, but the titles of the calls and the funded topics are quite telling. Compared to the UNESCO definitions, the IHP themes are almost exclusively in the field of social sciences (e.g., Social and Economic Challenges of Changing Family Structures, Employment and unemployment, or The relationship between employment and growth). The themes covered socio-economic aspects such as internationalisation, technology, (un-)employment, welfare, education and training, but also a number of socio-political themes including European integration, enlargement and identity, governance issues, or societal change. The humanities were only included in terms of capacity building (The development of European infrastructures for comparative research in the social sciences and humanities, Support activities to stimulate the development of social sciences and humanities in the European research Area).

For its bibliometric assessment – in particular when it comes to specific fields, one is more or less bound to the fields as defined by the Social Science Citation Index and its producer, Thomson Reuters. Here, fields are defined by a set of journals that can be removed or added depending on their usefulness and other unknown criteria by the producing company.

1.2 FP-funded SSH research background

Funding of socio-economic research in the Framework Programme dates back to the 4th programme (1994-1998). Under FP4, socio-economic research was targeted under the so-called TSER programme (**T**argeted **s**ocio-**e**conomic **r**esearch). The programme received 147 Mecu, (1.1% of the total budget), which enabled funding for a total of 162 projects. Under FP5 (1998-2002), SSH research was integrated in the so-called Fourth Activity, equivalent to the line for funding of socio-economic research *and* the training and mobility of researchers, which, under FP4 obtained 6%

Programme focussed almost exclusively on social science themes

(792MECU) and now under Fp5, 1.28 bn Euro, or roughly 10%. SSH received funding under the Key Action Socio-economic Research which is the basis for the analysis here. And which was part of the IHP programme (Improving the human research potential and the socio-economic knowledge base),.

Within the Key Action, three calls for proposals were made which addressed socio-economic research topics. The Key Action offered funding via three types of projects: research projects (RP), thematic networks (TN), and - only available for the last call-, accompanying measures (AM).

1.3 Bibliometric analysis in socio-economic research

Measuring the impact and performance of research programmes is challenging due to the very nature of research such as uncertainties, effects of transferability, the cumulative character of research, and long-term impacts. Therefore different tools are used to measure the output of research programmes and projects: peer review, reputation assessment, as well as bibliometric analysis.

Bibliometrics can basically be defined as the analysis of the scientific outcome published for example by a researcher, a research team, an institution or country. It can further be specified by scientific discipline. The scientific outcome can be in a number of forms such as books, book chapters, a journal article, a contribution in a newspaper etc. One can label this strand of analysis the counting of publications. This provides information on quantity, but neither provides it a relative use nor does it tell anything about the scientific use or impact.

As it is rather difficult to obtain a picture of all scientific outcome, bibliometricians tend to use a dedicated database. For more than 40 years, the main database for analysing the scientific performance of researchers was the Science Citation Index (SCI), starting in 1963, and the Social Science Citation Index (SSCI), starting in 1973. The Arts & Humanities Citation Index followed in 1978. All indexes were developed and maintained by the Institute for Scientific Information (ISI), which is now owned by Thomson Reuters. The main idea behind such an index is to collect citations of *scientific articles*. The database thus contains several thousand journals, and provides citation counts between these journals. The more a journal is cited, the higher its scientific impact.

This business is very powerful these days and bears several risks. Even if the usefulness of bibliometric analysis as a tool to measure scientific production is not very much questioned, there are strong arguments against the use of pure numbers of output, citations, impact factors and the frequency of non-cited papers (Evidence 2007).

1.3.1 Bibliometric analysis in the social sciences - limitations

There are a number of important concerns that affect the application and applicability of bibliometric analysis to evaluation, and need to be taken into account when drawing conclusions from bibliometric findings.

Publication habits differ largely between scientific fields. While in the biomedical sciences hardly any researcher publishes a book, historians publish up to 60% of their research in books rather than in a journal (Hicks 1999). A good biomedical researcher will be able to publish around five articles from a given research project while an engineer will hardly be able to publish one.

Language plays an important role. There are fields, for example economics or mathematics, where the use of the English language dominates. In most disciplines, however, national languages are equally significant and a number of important publications are in national language journals. The aspect of national literature is a very influential factor: as social sciences analyse society, they are bound to their social context and are inherently more national. Furthermore, research agendas are

For bibliometric analysis
SSH differs from other
fields of science

influenced by national policy making and last but not least, social science research is often published in national languages. The producers as well as the readers of the social science literature are nationally oriented. Social science researchers from a given country cite fewer international articles and are cited more often by their fellow compatriots, even if the main body of literature in a given discipline is produced elsewhere (Hicks 1999). However, due to the internationalisation trend, which can also be seen to some extent in the social sciences, one may see some more homogeneous international social sciences. Hicks (1999) mentions among the factors that "...in recent years, the EU began to fund social science research, and this should increase the international orientation of European social scientists and hence the share of their work written in English and indexed in the SSCI." (Hicks 1999, p. 207).

While bibliometricians have learned to deal with the various biases in the Science Citation Index (SCI), this is less the case for the Social Science Citation Index (SSCI), which is much less often used for bibliometric analysis than the SCI. The SSCI covered at the time of the analysis for this project (summer 2009) 2.350 journals. The journals are predominantly written in English (82%), followed by Dutch (8.2%) (Archambault et al, 2006).

Another important consideration is that theoretical publications and literature reviews tend to be more cited than publications that present empirical research results. Theoretical work tends to be transferable for many countries and thus it gets incorporated into the knowledge base of researchers of a given discipline in many countries. Empirical work on the other hand is often bound by national geographic or language-wise boundaries. A journal which gets covered in the database such as the Web of Science (WoS), becomes quasi automatically a 'high impact' journal: articles in this journal will obtain citations and thus, visibility. The journal becomes part of the so-called 'core literature' and a publication outside of this 'core literature' will be less well known.⁴ As the number of citations are the base for the impact of a journal, the chances for a theoretically oriented journal to become a leading impact journal are by far higher than an empirically-oriented one. As in addition to the empirical orientation, SSH research is focussing on geographically or language-wise 'national' issues, it is mainly published in national journals. Thus the 'core body of literature' for SSH researchers is not what is covered in the journal databases such as WoS or Scopus, but it is often treated as such when bibliometric analysis based on these databases is used in evaluations.

Hicks (1999) listed a number of factors, which need to be taken into account for a bibliometric analysis of the social sciences and humanities as there are distinctive features which differ from the natural sciences:

- fragmented literature and book publishing
- citations to books
- national literature
- scholarliness of journals

"The social science literature is fragmented because social scientists develop less consensus and adhere to more competing paradigms than do natural scientists. Lack of consensus within a field has been associated with a higher proportion of books in that field's literature because journal publishing has often been seen both as a signal of greater consensus and as a unifying force in itself" (Pierce 1987). According to Bourke et al (1996), natural scientists publish about 85% of their research findings in scientific journals, while the figure was about 61% for the social sciences and

There is no core, established body of literature in SSH;
SSH tends to vary by country and discipline

⁴ A famous example for a seminal work which got largely ignored by its peers was Gregor Mendel's work on genetic selection in 1865 (Versuche über Pflanzen-Hybriden). It was not published in a major journal and got rediscovered only 16 years after his death.

humanities. Within the social sciences and humanities, this share varies largely between disciplines: while it is relatively high in economics, it is very low in history, indicating that the latter is highly fragmented while for the former, a 'core literature' exists. Whatever discipline one analyses, a journal-based bibliometric analysis will include a smaller fraction of the total research output in the social sciences than in the natural sciences.

The second aspect concerns citations to books. They account for about 40% of all citations (Hicks 1999). While the SSCI includes the citations to books which are referenced in the journal articles covered, books are not included in the database.

While SSH research is often written in national language, the Social Science and the Arts & Humanities indices are dominated by English-speaking journals. A recent analysis estimated a 20-25% overrepresentation of English-language journals in the ISI database for the SSH fields (Archambault 2006). A somewhat older analysis suggested that 60% of all SSHI articles have a US address, and 20% have UK addresses (Andersen 2000). Godin (2002) similarly estimated the most productive countries in SSH related research using the ISI databases, the US, UK, Canada, and Australia – followed by Germany, the Netherlands, and France. Given the size of the Canadian and Australian research systems, it seems unlikely that they produce more than Germany or France.

Finally, the scholarliness of journal articles is mentioned. Again, due to the absence of an international 'core literature' for most social sciences disciplines, there is not much consensus on the scholarliness of journals. While a researcher publishing in journal x finds the journal scholarly, other experts may do not share this view.

1.3.2 Any alternatives around?

While ISI and its products seem to hold a monopoly for bibliometric analyses, this monopoly is slightly trembling. Due to rising critique of the use of impact factors, the database-internal biases and due to technical progress, competing systems are being built. The two new systems are 'Scopus' and 'Google Scholar'. Scopus is another commercial venture linked to Elsevier B.V. while Google Scholar is available for free on the web. While several bibliometricians and users welcome the new competitors, the tools and options for analysis differ. A systematic comparison concerning scope, quality, and analytics between the different databases is ongoing (Norris/Oppenheim 2007).

Scopus was launched in 2004 and covers now 16.500 peer-reviewed journals (including 1200 Open Access journals), 600 trade publications, 350 book series and about 3,6 million conference papers. There are more than 4.100 journals and book series in SSH. 42% of these journals are US based, followed by 26% from the UK, 9% from the Netherlands, and 5% from Germany. The shares of Canada and Australia, which have a higher share than the Netherlands and Germany in the ISI databases, obtain in Scopus 1.8% and 1.3% respectively. European countries make up 47%. Comparing the SSH journals in Scopus and ISI by country, favours Scopus: not only is the absolute number higher than in ISI with currently 2.350 journals, but most EU countries are better covered in Scopus than in ISI: A comparison between the countries' shares covered in Archambault et al (2006) and the list of journals covered in Scopus shows that for example Belgium is covered in SSH with five journals in ISI, but 61⁵ in Scopus. For France the numbers are 56 (ISI) and 206 respectively. Countries, which are not at all or almost not at all covered in ISI, are included in Scopus with sometimes quite high numbers (e.g., Hungary 26 journals, Czech Republic 20, Denmark 16, Spain 72).

⁵ Several journals are classified in more than one field.

Beside the coverage, another question concerns the type of analysis that can be performed. What can be done in Scopus is citation analysis: number of citations by an individual document, the number of citations of an author (including and excluding self-citations), which implies also analysis of the number and type of documents. Furthermore, Scopus provides the so-called h-index (Hirsch-Index)⁶ as standard information per author.

1.3.3 Citation analysis of socio-economic research

Another important aspect for bibliometric analysis relates to timing. A publication needs time: the actual time to be produced, the time spent under review, and finally the time to be discovered by others in order to get cited. While the actual production period may be similar for most scientific disciplines, the time under review can already vary considerably, and the time to find a place in the scientific community varies even more between fields. While there are differences between social sciences and humanities disciplines, they share a relatively slow uptake in the scientific world. While an important article in the medical sciences can obtain 100 and more citations within a couple of months after its publications, these figures are rather unseen for any SSH article. There are several reasons such as the diversity within almost any SSH discipline which hampers the development of a dominating 'scientific trajectory' and a – scientifically – more homogenous scientific community that absorbs new findings quickly. Whatever the reasons, it needs to be taken into account should a useful citation analysis be undertaken.

1.4 Structure of the report

The structure of the report reflects mostly the step-wise approach as planned. Following the introduction in chapter 1, chapter 2 provides a description of the database establishment and the main descriptive statistical results, e.g., who participated in which call and by which project type. This chapter should provide the reader with a simple overview of the Key Action Socio-economic Research. The database served not only for obtaining an overview, but provided the basis for the sample that was constructed for conducting the bibliometric analysis.

Chapter 3 provides the approach and results for setting up a scientific profile of the Key Action while chapter 4 provides the same for the benchmarking exercise.

Chapter 5 provides conclusions and suggestions for an improvement of the data basis

⁶ The Hirsch index h is defined as the number of papers with citation number higher or equal to h . Proposed by J.E. Hirsch (2005), the index characterises the scientific output of a researcher. The index itself is a subject of various research papers, e.g., Bormann, L., Daniel, H.-D. (2009): The state of h index research. Is the h index the ideal way to measure research performance?, *EMBO Reports*, 10 (1); Kosmulski, M. (2009): New seniority-independent Hirsch-type index, *Journal of Informetrics* 3 (4)

2. IHP participants – descriptive statistics

2.1 SSH Database establishment

A database with information about the Key Action was constructed through information supplied by the Key Action management, which combined general data on the project with publications that were attributed to the projects in their final reports as well as on their web-sites. The data-base covers all the projects selected in the three calls for proposals of the Socio-economic Research Key Action, including Research Projects, Thematic Networks and Accompanying Measures.

Table 1: Coverage of the database

FP5-IHP Key Action	1999	2001	2002	IHP Key Action total/average
Total sum of projects	43	71	89	203
Average number of participants	11	12	9	10
Average financial contribution	740.891	776.063	759.524	761.362

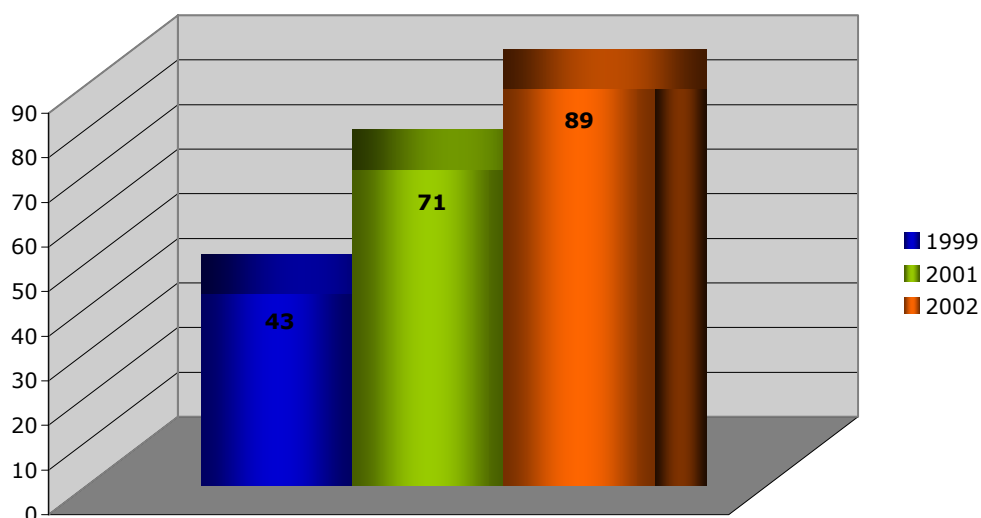
Source: Technopolis-Group
Data: FP5-IHP database

2.2 Descriptive analysis

The descriptive statistical analysis of the information about the three calls and thus the FP5 funded projects are shown in the following exhibits.

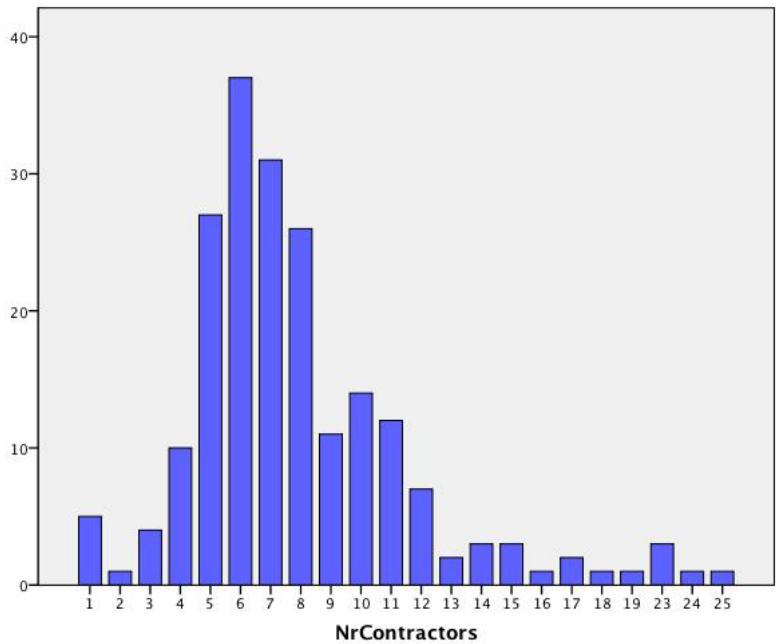
Following Exhibit 1 one can see that there are almost double the figure of projects in the third call with 89 than in the first one with 43. In total, 203 projects were funded by the Key Action.

Exhibit 1: Number of projects by project call



Source: Technopolis-Group
Data: FP5-IHP database

Exhibit 2: Number of participating institutions (participants) per project (all FP5-IHP projects)

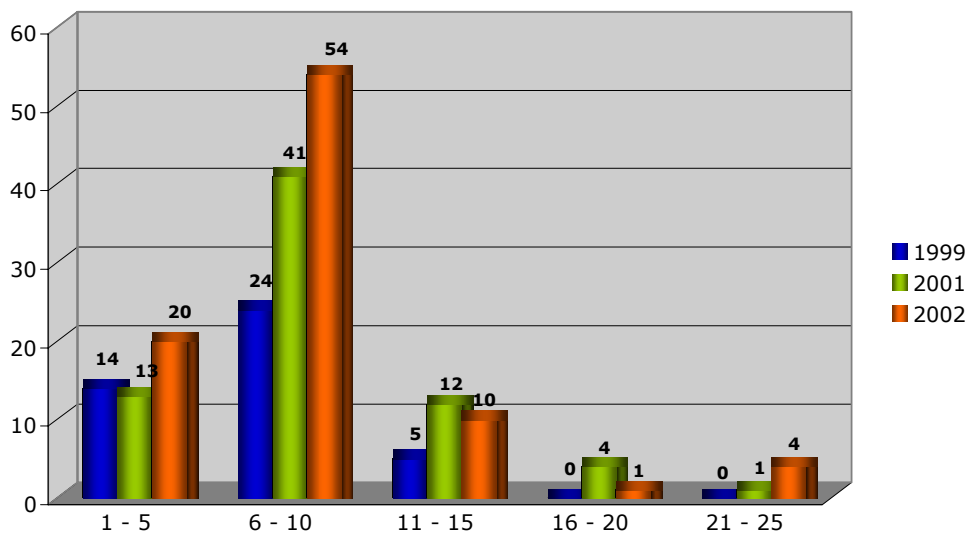


Source: Technopolis-Group
Data: FP5-IHP database

Exhibit 2 shows that for the IHP Key Action, 119 projects (almost 60%) had between 6 and 10 participants (i.e. coordinator(s) and partners), another 23% were in the range with up to five participants. The remaining 18% had numbers up to 25. The IHP average was 8 participants per project.

On average 8 partners collaborated per project

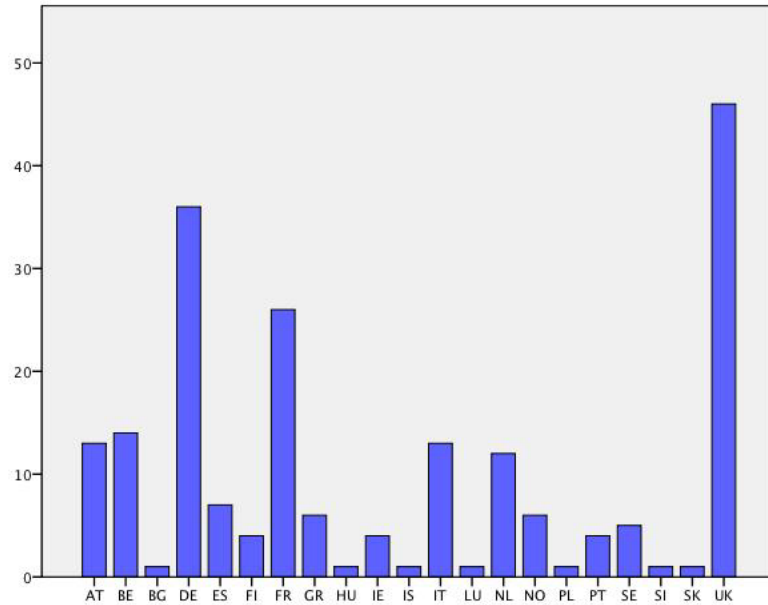
Exhibit 3: Number of participating institutions (participants) by call



Source: Technopolis-Group
Data: FP5-IHP database

While there are differences for all three calls, overall, the majority of projects had between six and ten participants.

Exhibit 4: Number of project coordinators by country (all FP5-IHP projects)

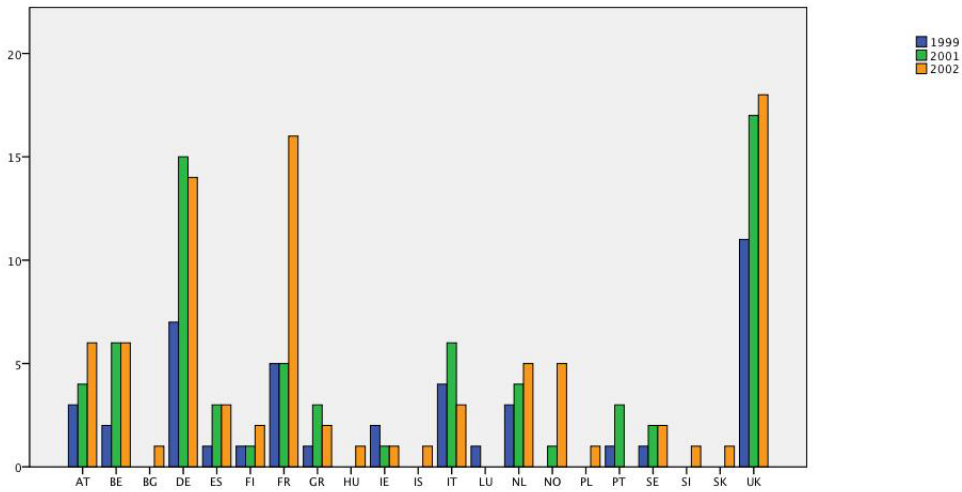


Source: Technopolis-Group
Data: FP5-IHP database

Dominance of UK coordinators

If one then starts looking at the country of coordinators and participants (Exhibits 5- 7), one notices for the EU15 countries that only Denmark has no project coordinator. From the then Candidate countries, Bulgaria, Hungary, Poland and Slovakia all coordinated one project each. From the EFTA countries, Norway (6) and Iceland (1) were the countries also seen as coordinators, while Switzerland did not coordinate one project. As can be clearly seen from Exhibit 4, the only country sticking out visibly is the UK with 46 project coordinators. It is followed by Germany (36), France (26), Belgium (14), Austria and Italy (both 13 each).

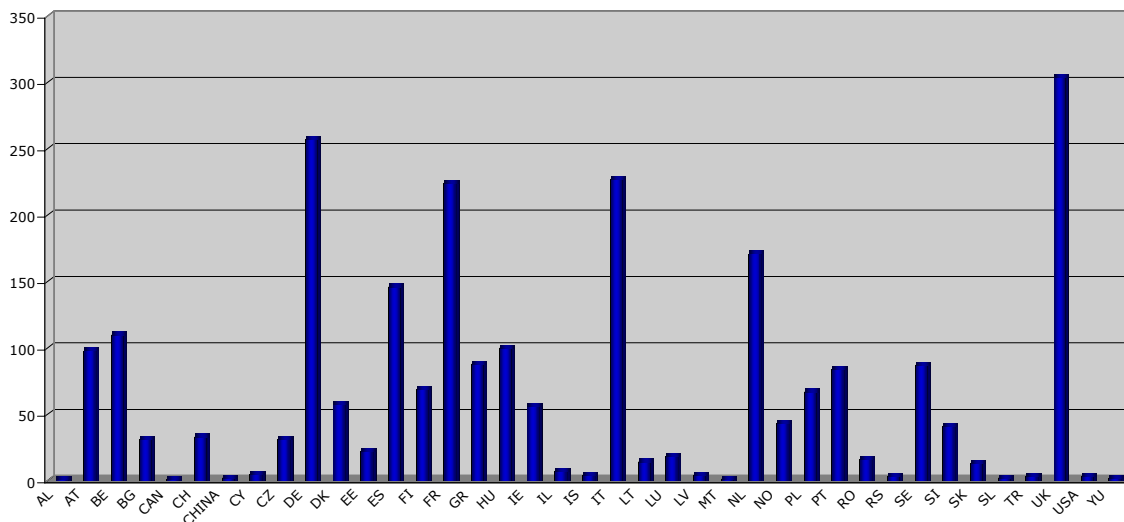
Exhibit 5: Number of coordinators by call and country



Source: Technopolis-Group
Data: FP5-IHP database

The non-EU15 countries' coordinators were in particular present in the 3rd call (2002) – of which a number of projects from EFTA and then Candidate Countries originate. In terms of shares by call of coordinators, the dominating UK maintained shares between 20-26% for the three calls, Germany between 16-21% and France between 7-18%. French coordinators were in particular rare for the projects in the second call (5 out of 71), but there were 16 projects coordinated by French participants in the third call – almost as many as coordinated by UK participants (18).

Exhibit 6: Origin and number of participants by country (all FP5-IHP projects)

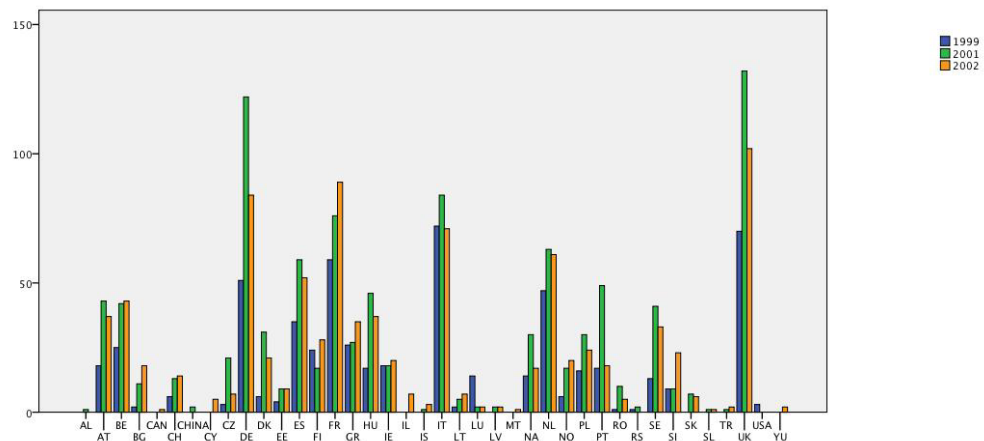


Source: Technopolis-Group
Data: FP5-IHP database

Majority of partners are from the UK

When it comes to the country origin of participants (including coordinators and partners), the number of countries is with 39 higher than the number of countries that were associated to FP5. Participants from all EU15 Member States, all ten new MS, the EFTA countries with the exception of Liechtenstein, several of the presently associated countries to the Framework Programme such as Albania, Turkey, Israel etc. as well as a few participants from Canada, USA, and China were identified. The majority here as well came from the UK (12.1%), followed by Germany (10.2%), Italy (9%), France (8.9%), and the Netherlands (6.8%). Again, due to the fact that several participants are counted more than once, the numbers should not be overstated; the shares provide a better picture of the reality.

Exhibit 7: Origin of participants by call and country



Source: Technopolis-Group
Data: FP5-IHP database

When looking at the absolute numbers, some countries seem to have been more successful in a particular call, but this view cannot be maintained when looking at the shares. There are ups and down for almost all countries in terms of absolute number as well as in terms of shares. For the UK for example, the shares were relatively stable with 12.1%, 12.9% and 11.2% respectively for the three consecutive calls, while for Finland, the shares were 4.1%, 1.7%, and 3.1% respectively.

Finally, analysing the three calls by type of project, the following information can be retrieved (Tables 2 and 3). First of all, the majority of projects with 72% were Research Projects (RP), followed by 19% of Thematic Networks (TN) and – only available from the third call – 9% of Accompanying Measures (AM).

Table 2: Project type by call

	1999	2001	2002	FP5
RP	37	58	52	147
TN	6	13	19	38
AM	0	0	18	18
Total	43	71	89	203

Source: Technopolis-Group
Data: FP5-IHP database

The projects differed in terms of total financial contribution by type of project as well as between the three calls (Table 3) – this can be seen looking at the calls as well as at the project types.

Table 3: Average financial contributions by call and project type

	1999	2001	2002	FP5 average
RP	800.424	843.839	1 045.443	904.227
TN	373.770	473.680	472.765	457.447
AM			236.226	236.226
Average	740.891	776.063	759.524	761.362

Source: Technopolis-Group

Data: FP5-IHP database

Key Findings on the Key Action

- The IHP-Key Action **evolved in terms of scale** in a number of characteristics such as the number of participating countries, and the financial contribution per funded project.
- The Key Action equally **evolved in terms of structure**: the available projects types were extended to Accompanying Measures in the 3rd call.
- In terms of distribution of the participants, **the action reflects very much the countries' overall funding of and performance in SSH** based on higher education expenditures on R&D (HERD) on SSH as well as the overall share of SSH publications in the Web of Science (WoS) by Thomson Reuters.
- The countries **most often participating are the largest three EU Member States**: the UK, Germany, and France, followed by Italy and the Netherlands.

3. The impact of FP5 projects' scientific publications

In order to analyse the impact of scientific publications that resulted of FP5 research, a citation analysis was conducted. As mentioned in part 1, citation analysis is largely limited to articles (and a limited number of other types of documents such as letters, editorials etc.), of journals, which are either covered in the Science Citation Indices of Thomson Reuters or Scopus. This does exclude the bibliometric analysis of a large share of scientific outcome namely books and book chapters which are in particular prominent in the SSH fields. However, one can hypothesise that publications resulting from FP5 that have made it into scientific journals covered in the SSCI or Scopus, have passed a very high quality barrier already.⁷

We constructed a representative sample in order to assess the impact of FP5-IHP funded projects by using impact analysis of publications. The study team provided the suggestion for obtaining a sample of publications via the list of publications that can be constructed based on the final reports of the projects. Factors that were considered when forming the sample included:

- Timing – the publication date and its impact on citation analysis,
- Type – the Key Action has financed different types of projects,
- Is a project best represented by its best publications or by a different sample?

Another requirement was that "the resulting sample of publications can then be assessed bibliographically by comparing their impact with the impact of publications of the same authors from other research projects, as well as with the average impact of publications in the same journals."

The study team opted for the construction of a sample of all publications based on a sample of projects rather than a sample of publications for all projects. The following section 3.1 provides thus the sampling approach as well as information on the sampled projects.

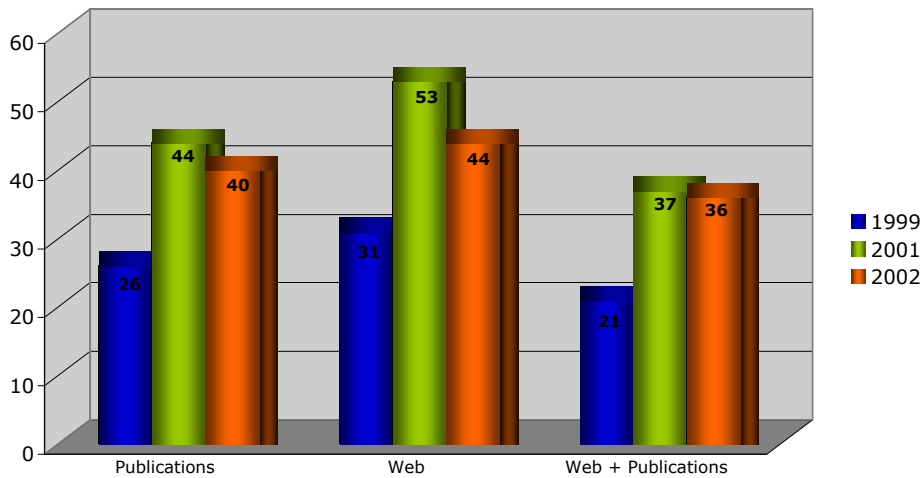
3.1 The sample

3.1.1 Considerations for sampling

For the development of a sampling strategy, one aspect may be the availability of websites and/or publication lists included in the final reports. Exhibit 8 shows the number of projects where this information is available. The advantage of an available publication list is that it may be used as a starting point as a number of authors and indicated titles are mentioned. The availability does however not mean, that these lists are comprehensive. Very often, only a few indicated publications are mentioned whereas in other cases, much more is mentioned. Concerning the availability of publication lists and websites, the relative shares for the first and the second call are similar around 49%. Availability for both is much lower for the 3rd call with only 40%, i.e. only 40% of the final project reports provide lists.

⁷ The journals covered in SSCI are peer-reviewed journals only. Similar as to Scopus, journals can be suggested to be included. There is an established review process in both systems that include and exclude new journals. However, the criteria for this process are not transparent.

Exhibit 8: Total number of information source by type and call



Source: Technopolis-Group
Data: FP5-IHP database

By inserting the data on participants in the FP5-IHP database, we also noted that there were **differences between the names of participating researchers mentioned in the synopsis reports (which are based on project proposals) and the final reports**. In a few cases, there were even more participants mentioned in the final report than initially mentioned. In a limited number of projects we noted mobility of researchers (change of affiliation between project start (project proposal) and ending (final reports), and in other cases, originally included persons were not any more included in the final reports. Given that several projects lasted for three to four years, one would expect that also due to the mobility of researchers, not all projects were executed by exactly the same people that started them. This adds another layer to the complexity for the identification of involved researchers and their publications resulting of the IHP projects.

Another aspect for a possible over- or under representation of projects by country of origin of the participants was tested. We compared the share of participants by country of origin with some proxy indicators. Given that the majority of participants are from academia, we took as relevant indicators the amount spent by country for the higher education sector (HERD in mio pps), the share of SSH funding in academia (SSH by HERD in mio pps), and the number of SSH publications. The analysis of **the overall economic and publication data of countries does by and large match the FP5-IHP distribution of projects funded by country of origin** with some small differences such as an overrepresentation of Italian and an under-representation of Spanish participants. The problem of under- representation is also in particular a question concerning small countries. However, our analysis showed that their distribution within the IHP Key Action matches largely the funding structures of SSH. This is an important finding as it provides a reason against a possible criterion of "full MS coverage" in the sample. To choose 'country' as a criterion and to include at least a project in order to have all MS represented is not an optimal criterion for the following reason: the distribution of project participants by country within the IHP Key Action is similar to the overall economic and scientific share of the countries. The decision to have them included in the sample nonetheless would inevitably lead to an overrepresentation of these countries.

The basic descriptive analysis of the Key Action (see chapter 2) showed that the three calls differed in terms of number of funded projects and *project types*. Accompanying measures (AM) were only included in 2002. However, the information available on this type of projects is limited compared to the other types. There were also only 18 AM

and their average financial contribution was much lower than the two other types (Table 3). Furthermore, accompanying measures have more structuring effects as objectives than research projects and thematic networks; therefore there was reason to leave this type of project out of the sample.

Not all 'Areas' of the work Key Action were covered equally by each call. Call areas 8 and 11 were not represented in 1999 projects for example, but the remaining areas are represented in every call. There is also not systematically a TN by area while one finds RPs in all areas and calls. In almost all areas, there are projects that provided publication lists in their final reports, only for the reports in two areas and in two calls where there was only one project, no publication list was mentioned.

3.1.2 Sampling

The study team obtained a stratified sample which took into account the following criteria for sampling 32 research projects and eight thematic networks (21.6% of all RP and TN funded via the FP5-IHP Key Action):

- Call year
- Area by call
- Area by Task
- Availability of publication list

Given the coverage by 'area', 34 projects would be sampled according to that criteria (and availability of publication lists). Another six research projects (two by area) were then added based on the criteria number of projects per area, publication list and, if necessary, financial contribution. One thematic network was included which did not provide a publication list. The reason for that was the fact that it was the only project in that call covering one of the particular areas and tasks.

This results in a stratified sample per call of 40 projects, as included (anonymised) in Table 4. The sampled projects were the basis for the publication database.

Table 4: The SSH sample (1) (contin.)

Project (coded)	Nr. of participants	Call	Area	Task	Project Type
99-1	4	4	10	10	RP
99-2	9	4	7	3	TN
99-3	2	4	2	6	RP
99-4	9	4	9	1	TN
99-5	6	4	5	5	RP
99-6	13	4	1	7	TN
99-7	10	4	3	2	RP
99-8	6	4	10	11	RP
99-10	12	4	12	1	RP
99-11	14	4	4	9	RP
99-12	7	4	4	2	RP
01-1	8	5	1	3	RP
01-2	8	5	7	4	RP
01-3	12	5	11	7	RP

Project (coded)	Nr. of participants	Call	Area	Task	Project Type
01-4	8	5	6	2	RP
01-5	4	5	5	5	RP
01-6	8	5	10	6	RP
01-7	7	5	4	3	RP
01-8	6	5	9	2	RP
01-9	6	5	12	2	RP
01-10	6	5	10	4	RP
01-11	5	5	8	2	RP
01-12	23	5	11	7	TN
01-13	11	5	3	6	TN
02-1	9	5	2	3	TN
02-2	7	6	9	2	RP
02-3	7	6	5	1	RP
02-4	7	6	2	3	RP
02-5	8	6	4	2	RP
02-6	10	6	8	2	RP
02-7	6	6	7	1	RP
02-8	7	6	11	3	RP
02-9	10	6	3	1	RP
02-10	7	6	8	2	RP
02-11	10	6	10	3	RP
02-12	7	6	7	2	TN
02-13	5	6	6	4	TN
02-14	12	6	1	5	TN
02-15	4	6	12	4	TN

Source: Technopolis-Group

Data: FP5-IHP database

(1) Number of participants: individual institutions; Call, Area, Task: the figures were provided by the Commission Service.

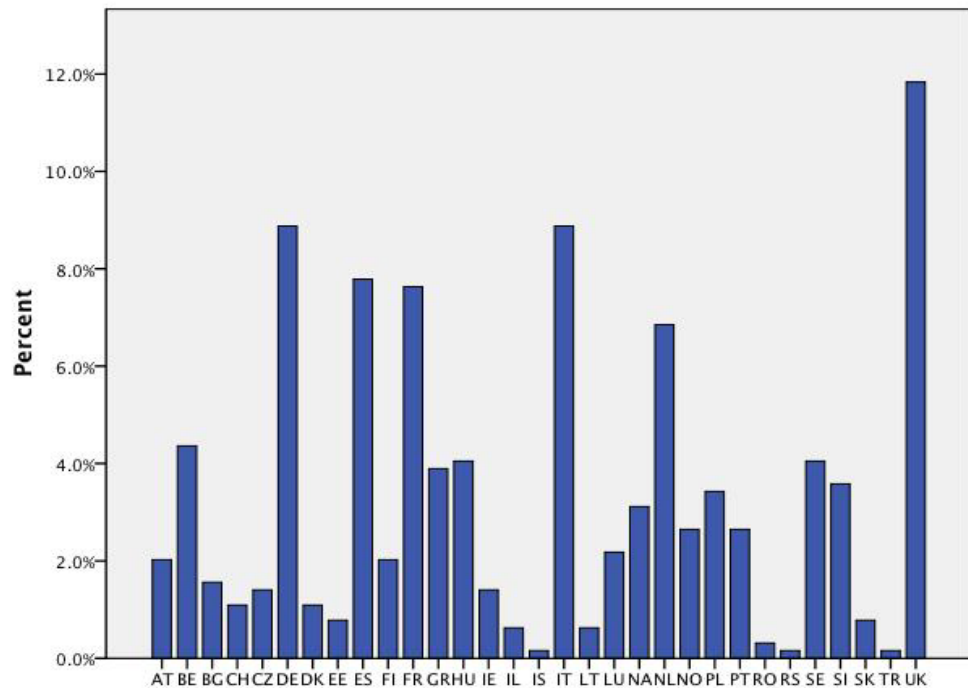
Analysed by call, these 40 sampled projects have averages comparable to the overall Key Action averages (Table 3). Concerning the country coverage for example, the distribution of participants is matched well. Both results presented in Table 5 and in Exhibit 9 are in line with the full database. The stratification resulted equally in a representative sample.

Table 5: Sample statistics: Project numbers, average financial contribution and average number of participants by project type and call year

		1999	2001	2002
RP	Number of projects	9	11	10
	Average number of participants	8	6	7
	Average financial contribution	884.644	868.426	1.104.826
TN	Number of projects	3	3	4
	Average number of participants	10	14	7
	Average financial contribution	405.000	580.923	469.380

Source: Technopolis-Group
Data: FP5-IHP database

Exhibit 9: Sample statistics: Share of participants by country



Source: Technopolis-Group
Data: FP5-IHP publication database

3.2 Sample descriptive analysis: type of publications

The analysis of the final reports, the synthesis report published by the Commission Service, as well as still existing project websites obtained for these 40 projects a total of 1.449 publications. These publications were all included in the publication database, which included information about the authors, the type of publication, and all available bibliographic information such as title, journal, publication date, and language of the publication.

There are basically four types of publications:

- scientific articles, published (or submitted) in national or international journals,
- edited (planned or published) books,
- individual book chapters, and
- grey literature (working papers, series of working papers established often in the course of the project, or the project deliverables (work package results)).

The **'grey literature' category dominated** quantitatively. About 45% of the publications were in this category, followed by 28% of scientific articles, 19% book chapters, and 8% books (Table 6). The two latter categories overlap to some extent: while some projects only mentioned one edited book (collective volume) once as main output, other projects mentioned the book and/or book chapters individually as output per project partner. Given the voiced interest of the Commission Service about options to use books as basis for a impact analysis, the study team looked into this aspect in more detail (chapter 3.6).

Highest share of publication output is 'grey literature'

Language wise, it is not really a surprise that **English language publications dominate** with 80% followed by almost 10% of German-language (Table 6). What is more surprising is the relative large share of French publications with 3% given that the total number of publications from France in SSH is rather low (in absolute numbers, France comes only as the 12th country out of the EU27, see Annex A 3.1, Figure 2). It may partly be explained that a share of Belgian publications is published in French but it may also show that French SSH researchers are more productive if they participate in a FP project compared to their average publication habits. For all other countries beside Germany/Austria, there is quite a noticeable propensity to publish in English.

The **variance in output** by sampled project is very high: the lowest output is zero. This can be found in three projects. In all three cases, the coordinator as well as the project partners were addressed and asked to provide information but in two cases, there was no response. Beside these 'outliers' the lowest publication number is four, the highest 189. On average 36 publications can be noted. Again, the vast differences can partly be explained with differing 'publication habits': while some mention each and every work package output, others do not include them in their list of publications, and a collective volume can be separated in addition into individual chapters.

An important question concerns the overall degree of representativeness of the sample for the overall Key Action in terms of publications by type. In order to see if our sample results hold true for the whole Key Action, standard deviation and error margins were calculated based on the results of the 40 individual projects.

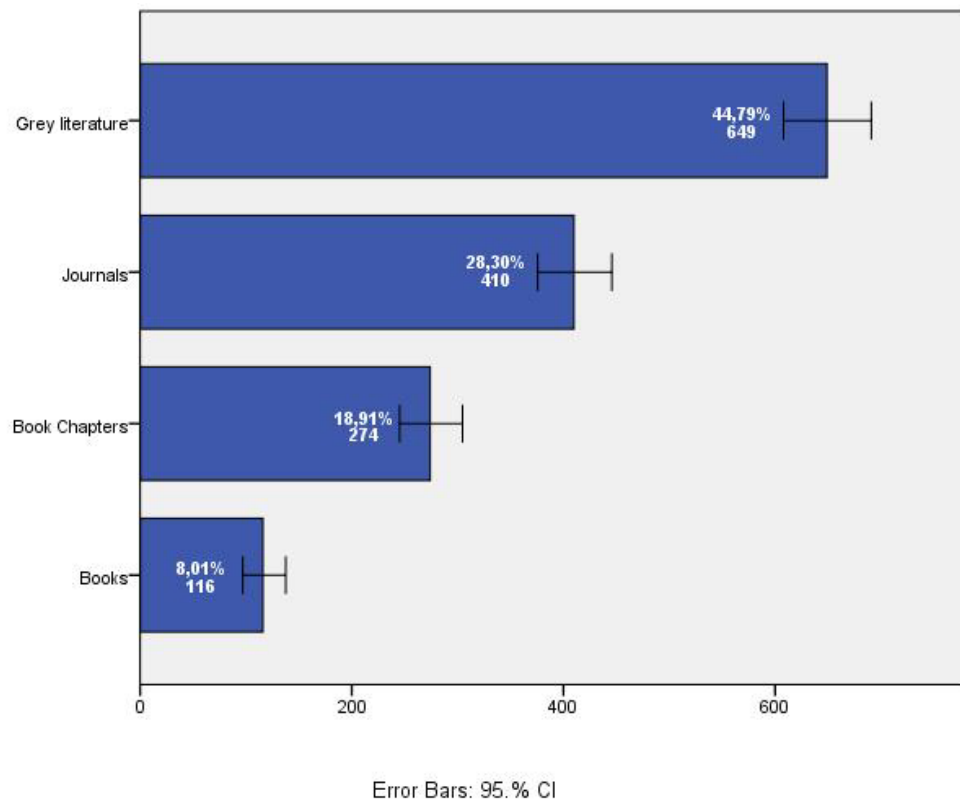
At 95% confidence interval, we can conclude from Table 6 that all projects have between 2 and 4 book publications, 5-8 book chapters, 9-12 journal articles and 15-18 other publications of the grey literature type. Standard deviation from this is the lowest for books with 5.7 and the highest for the grey literature with 20. The deviation is however rather small. In terms of error level, again the lowest error is calculated for books and book chapters.

Table 6: Standard deviation, confidence levels and standard errors

	Books	Book chapters	Journal articles	Other (Grey literature)
Low confidence interval (in absolute numbers)	1.8544	5.4612	8.5153	14.6488
Upper confidence interval (in absolute numbers)	3.9456	8.2388	11.9847	17.8012
Standard deviation	5.7459	11.7289	17.9211	20.4870
Error bars				
Low confidence interval	104.7	251.0	374.9	608.8
Upper confidence interval	127.2	297.0	445.1	689.2

Source: Technopolis-Group
Data: FP5-IHP publication database

Exhibit 10: Sample statistics: error margins per type of publications



Source: Technopolis-Group
Data: FP5-IHP publication database

Table 7: Overview of the sample by project (contin.)

Project Project (coded)	Project Type	Number of publications by type						Publication languages									
		Books	Journals	Book Chapters	Others	Totals	EN	DE	FR	IT	ES	PT	Other				
99-1	RP	6	24	24	11	65	44	20	0	0	0	0	0	0	1		
99-2	TN	0	0	0	32	32	32	0	0	0	0	0	0	0	0		
99-3	RP	0	3	0	9	12	12	0	0	0	0	0	0	0	0		
99-4	TN	14	17	27	36	94	81	2	1	0	0	0	0	0	10		
99-5	RP	3	13	8	3	27	18	0	0	1	1	1	0	0	6		
99-6	TN	2	8	0	1	11	11	0	0	0	0	0	0	0	0		
99-7	RP	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
99-8	RP	4	0	1	0	5	5	0	0	0	0	0	0	0	0		
99-9	RP	1	11	4	8	24	19	3	1	0	0	0	0	0	1		
99-10	RP	0	0	0	16	16	16	0	0	0	0	0	0	0	0		
99-11	RP	0	7	1	2	10	9	0	1	0	0	0	0	0	0		
99-12	RP	0	0	2	5	7	7	0	0	0	0	0	0	0	0		
01-1	RP	1	12	6	51	70	69	0	1	0	0	0	0	0	0		
01-2	RP	1	3	10	60	74	41	18	2	3	0	0	0	8	2		
01-3	RP	0	34	16	32	82	61	5	0	0	0	0	0	0	16		
01-4	RP	0	7	2	3	12	8	0	1	1	1	0	0	0	1		
01-5	RP	1	3	2	0	6	5	0	0	0	0	0	0	0	1		
01-6	RP	1	3	0	0	4	4	0	0	0	0	0	0	0	0		
01-7	RP	0	6	0	0	6	5	0	1	0	0	0	0	0	0		
01-8	RP	6	6	0	0	12	4	8	0	0	0	0	0	0	0		
01-9	RP	0	1	3	2	6	6	0	0	0	0	0	0	0	0		
01-10	RP	1	7	9	0	17	10	6	0	0	0	0	0	0	1		

Project	Project (coded)	Project Type	Number of publications by type					Publication languages									
			Books	Journals	Book Chapters	Others	Totals	EN	DE	FR	IT	ES	PT	Other			
01-11	RP		7	20	16	58	101	80	0	14	1	0	0	0	6		
01-12	TN		0	0	0	0	0	0	0	0	0	0	0	0	0		
01-13	TN		2	0	0	49	51	47	0	4	0	0	0	0	0		
02-1	TN		0	0	0	29	29	29	0	0	0	0	0	0	0		
02-2	RP		1	6	4	49	60	34	26	0	0	0	0	0	0		
02-3	RP		4	5	0	2	11	10	1	0	0	0	0	0	0		
02-4	RP		1	5	0	0	6	6	0	0	0	0	0	0	0		
02-5	RP		1	1	11	4	17	13	1	0	0	2	0	1	1		
02-6	RP		3	4	3	15	25	21	0	1	0	0	0	3	3		
02-7	RP		7	10	21	6	44	34	3	0	0	0	0	7	7		
02-8	RP		2	9	3	0	14	14	0	0	0	0	0	0	0		
02-9	RP		0	0	0	17	17	17	0	0	0	0	0	0	0		
02-10	RP		1	20	2	12	35	33	0	0	0	0	0	2	2		
02-11	RP		15	66	55	53	189	160	6	5	1	12	0	5	5		
02-12	TN		0	0	0	61	61	58	0	3	0	0	0	0	0		
02-13	TN		31	92	38	23	184	130	36	12	0	0	0	6	6		
02-14	TN		0	0	0	0	0	0	0	0	0	0	0	0	0		
02-15	TN		0	7	6	0	13	11	1	0	1	0	0	0	0		
Total			116	410	274	649	1449	1164	136	47	8	16	9	69	69		
Average			8.0	28.3	18.9	44.8	100.0	80.3	9.4	3.2	0.6	1.1	0.6	4.8	4.8		

Source: Technopolis-Group
Data: FP5-IHP publication database

3.3 Sample descriptive analysis: authors

When establishing the literature database it got obvious, that not all authors mentioned in the publications were equally participating in the FP. As many names were not included as coordinators or principal researcher, we extended the initial Key Action database (which is based on the reports) for these 40 projects and searched the project websites in order to obtain a better view who were really the associated researchers with these projects (see also chapter 3.4).

Table 8 provides the numbers of authors per project – differing between those which were also mentioned in the synopsis or final reports as consortium partners, thus included in the database as ‘principal investigators’ and in the following labelled as authors in consortium and those, which appeared in the listed publications but were not among the identified principal investigators and were thus labelled as authors outside the consortium.

A total of 981 authors were thus identified. For the 40 projects there were on average 8.1 consortium partners and another 16 from outside the consortium authoring the output.

981 authors: only one third were principal investigators.

Table 8: Number of authors per project (total, inside and outside consortium) (contin.)

Project acronym	Project Type	Total number of authors	Number of authors in consortium	Number of authors outside consortium
99-1	RP	12	8	4
99-2	TN	61	4	57
99-3	RP	18	4	14
99-4	TN	43	14	29
99-5	RP	29	23	6
99-6	TN	12	7	5
99-7	RP	NA	NA	NA
99-8	RP	9	5	4
99-9	RP	21	11	10
99-10	RP	29	27	2
99-11	RP	8	8	0
99-12	RP	10	3	7
01-1	RP	36	23	13
01-2	RP	28	24	4
01-3	RP	63	27	36
01-4	RP	19	7	12
01-5	RP	13	12	1
01-6	RP	7	6	1
01-7	RP	8	4	4
01-8	RP	9	2	7
01-9	RP	7	1	6
01-10	RP	11	1	10

Project acronym	Project Type	Total number of authors	Number of authors in consortium	Number of authors outside consortium
01-11	RP	38	5	33
01-12	TN	NA	NA	NA
01-13	TN	46	5	41
02-1	TN	37	2	35
02-2	RP	12	4	8
02-3	RP	14	4	10
02-4	RP	7	5	2
02-5	RP	19	17	2
02-6	RP	28	16	12
02-7	RP	19	5	14
02-8	RP	9	2	7
02-9	RP	30	11	19
02-10	RP	21	4	17
02-11	RP	50	8	42
02-12	TN	40	9	31
02-13	TN	147	2	145
02-14	TN	NA	NA	NA
02-15	TN	11	6	5
		981	326	655

Source: Technopolis-Group
Data: FP5-IHP publication database

When comparing the authors 'in the consortium', i.e. those clearly identified as participants in the projects, and the authors outside the consortium, we can see from Table 9, that the share of outside authors is very high for authors that are mentioned only once. One may hypothesize that the external authors are part of the professional networks of the FP participating authors being frequently made co-author. The importance of the own professional network which goes beyond the FP-project is especially pronounced for participants with only a few FP-publications. If one looks at the absolute number of authors with up to seven publications resulting from the Key Action, the total number of authors identified as being outside the consortium is by far higher than the authors from within. There is an opposite effect for the authors with large publication numbers (i.e., more than seven publications resulting of the Key Action). Although being only a few authors in absolute numbers – (see Table 9), one can note that their co-authors are more often from within the consortium than with outsiders. To put it differently, the more co-publications occur for a given author, the more likely it is that these publications are authored with project partners.

The analysis using 'inside' and 'outside' consortia members may however be biased due to the fact that a number of co-workers from a given institution may not have been mentioned as principal investigators or project members and thus were simply not captured. We analysed this possibility in a later step for a limited number of articles (see Chapter 3.4).

Table 9: Sample statistics: Publications per author
(totals, on shares, by in- and outside consortium authors)

Nr. of publications per author	Total number of authors	Nr. of authors outside consortium		Nr of authors in consortium	
		Absolute	In shares	Absolute	In shares
1	598	465	71.1	133	40.8
2	138	81	12.4	57	17.5
3	75	36	5.5	39	12.0
4	38	21	3.2	17	5.2
5	24	14	2.1	10	3.1
6	19	11	1.7	8	2.5
7	14	4	0.6	10	3.1
8	13	5	0.8	8	2.5
9	8	3	0.5	5	1.5
10	7	4	0.6	3	0.9
11	5	2	0.3	3	0.9
12	3	1	0.2	2	0.6
13	7	0	0.0	7	2.1
14	6	0	0.0	6	1.8
15	3	1	0.2	2	0.6
16	3	1	0.2	2	0.6
17	4	1	0.2	3	0.9
18	2	0	0.0	2	0.6
19	1	0	0.0	1	0.3
21	2	1	0.2	1	0.3
22	3	0	0.0	3	0.9
23	1	0	0.0	1	0.3
25	1	0	0.0	1	0.3
27	1	1	0.2	0	0.0
28	2	2	0.3	0	0.0
32	1	0	0.0	1	0.3
55	1	0	0.0	1	0.3
Total	980	654	66.7 (100)	326	33.3 (100)

Source: Technopolis-Group

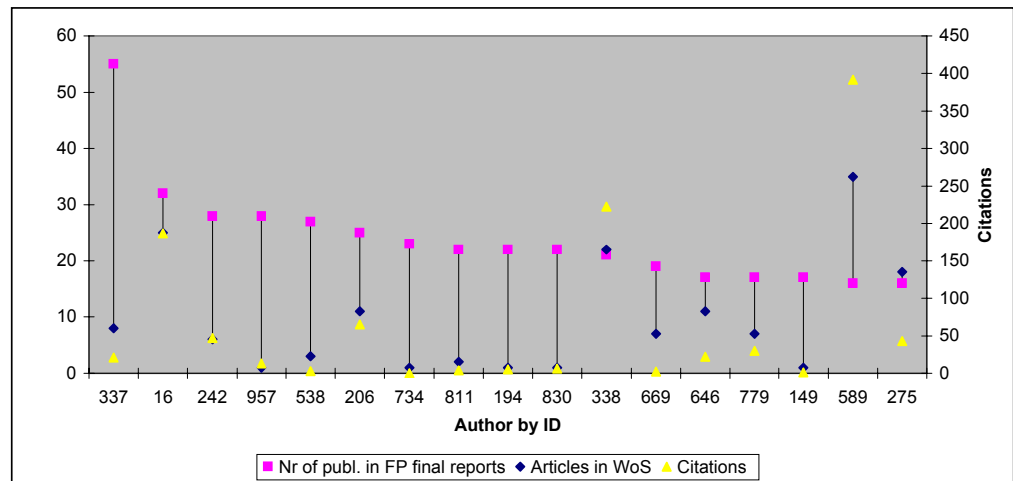
Data: FP5-IHP publication database

Note: One author participated in 2 projects: 99-00036 (RP) and 02-50028 (TN). The author is outside the consortium in both projects.

If one looks at the most productive authors in the sample, i.e. those with at least 15 publications mentioned, the numbers get very small. In the sample only 22 people have that many publications.

Are these productive people also outstanding in terms of scientific impact? We searched for them in the Web of Science, were 17 were identified. Exhibit 11 shows their WoS performance. The authors are sorted according their decreasing FP-mentioned publications, so author 337 has 55 FP publications, author 16 32 etc.⁸ there is quite a discrepancy between the numbers. While author 337 faced a huge leap behind, the authors 338 and 589 have more articles covered in the WoS and both obtained equally the highest citation scores.

Exhibit 11 Productive FP-authors: FP publications vs WoS publications and citation rates



Source: Technopolis-Group
Data: FP5-IHP publication database, WoS

The benchmarking exercise based on scientific profiles using Scopus as the database (see chapter 3.8) showed a similar discrepancy. Comparing the citation impact of the two most prolific authors overall (Nr. 337 and Nr. 16) and taking their output from 2000-2007 and a two year citation window, author Nr. 16 obtained 35 cites and 337 six. In terms of total citation score for all their peer-reviewed articles, author Nr 16 is number six of all authors with a total citation score of 384 whereas Nr. 337 is coming at 31st with a total score of 65. Here entry into the publication market shows: while author number 16 publishes since 1984, number 337 started only in 1997.

The study team examined the FP-resulting body of literature carefully and concluded that not all mentioned publications were the result of the FP participation: While for 47% of the total publication output no year was mentioned, a small number of supposedly resulting publications were published even prior to the start of the projects (Table 10).

⁸ Given that the first names are not included in full but only with initials, there is room for errors in identifying the right person – in particular if the last name is a rather common name.

Table 10: Publication numbers by type and year (1)

	Books	Book chapters	Journal articles	Other	Total
1986	1	0	0	0	1
1990	1	0	0	1	2
1997	1	0	0	1	2
1999	0	0	0	1	1
2000	0	0	9	8	17
2001	0	0	20	32	52
2002	2	1	48	41	91
2003	1	4	94	95	194
2004	0	5	111	71	187
2005	0	1	71	97	169
2005	0	0	1	0	1
2006	0	1	20	7	28
2007	0	0	7	2	9
2008	0	0	3	1	4
Planned (1)/	0	0	8	2	10
no year	110	262	18	290	680
Total	116	274	410	649	1449

Source: Technopolis-Group

Data: FP5-IHP publication database

(1) this includes forthcoming publications and those under review

Almost no book or book chapter was provided with a publishing year – which is not really surprising given that the publishing procedure for books can take up to several years. What one can however see is the precise provision of the publishing date for more than 95% of the journal articles published. The peak for this type of publication is 2003-2004.

3.4 Preparing citation analysis of FP-published articles

As mentioned before, a citation or impact analysis can only be performed for articles in database covered journals such as SSCI or Scopus. The study team therefore limited the following analysis to the 410 articles (based on journal names) identified in the sample.

Top-down matching journals to SSH high impact journals

The first approach tested was a matching of the journals of these articles with the top journals according to impact factor in 15 sub-disciplines. None of the articles was published in any of the 150 journals mentioned. Given that the "top journals" and their impact factors are heavily influenced by the US publication and citation behaviour, the team opted for choosing a more bottom-up approach and analysed the journals in which these articles appeared.

Bottom-up identification of articles in SSCI/Scopus covered journals

We identified 115 articles which were published in 66 SSCI covered journals. Another 50 articles were identified in 32 journals which were in addition covered in Scopus. Thus, for the 66 SSCI journals, impact factors were available via the Journal Citation

Index, and citation analysis for a total of 165 articles seemed possible. For the 32 journals only covered in Scopus, impact factors were not available.

These 165 articles were then further analysed in terms of authors. A total of 210 authors were identified. Of those almost half seemed to come from outside the project consortium. After investigating these authors in detail it was concluded that 20% of them were in fact people who worked in the project but did not appear as such in the project reports.

The split between consortium members and non-consortium members was almost equal with 102 to 98. Now, given that this small sample is highly relevant for the further analysis, the team wished to check if these 98 non-consortium members⁹ were really outside the consortium, or other consortium members, which were just not mentioned in the reports. All of the still existing project websites were thus searched for the list of participants. 20 authors were identified in addition as consortium members. The list of the 120 authors identified as consortia members and having at least one FP project related article in a peer reviewed journal as well as the list of institutions hosting more than one identified author were provided separately.

Cleaning articles for impact analysis

The exercise so far aimed at identifying FP-participants as authors of peer-reviewed articles. The aim now moved towards analysing the impact of this output. The next step therefore took the number of articles in the SSCI and Scopus identified journals as the basis. A search in Scopus by either/or the titles of the articles, the author(s), and the list of publications by journal identified only 157 of the 165 articles that appeared to have a sound bibliographic reference (Tables 11 and 12).

Table 11: Sample descriptive statistics of authors contributing to journal articles

Number of authors contributing to journal articles in sample, FP5	196
Total number of journal articles in sample (FP5)	157
Average number of articles per author	1.56
Minimum number of articles per author	1
Maximum number of articles per author	7

Source: Technopolis-Group
Data: FP5-IHP publication database

Table 12: Total number of authors contributing to a given number of articles per call year

Call year	Total number of authors contributing to journal articles	Total number of journal articles in sample
1999	35	29
2001	48	40
2002	113	88

Source: Technopolis-Group
Data: FP5-IHP publication database

⁹ Almost 40 non-consortia authors are mentioned in one Thematic Network which lists its many core members.

These 157 articles and the identified authors were then the basis for the benchmarking exercise described in section 3.5. below. The analysis took into account details on the publication: whether or not it was a co-publication, whether or not it was among the most cited publications of the oeuvre, the cites per year etc., and it was the basis for the establishment of individual profiles for those authors, identified as FP-participants.

3.5 Benchmarking FP-publications with other publications of the authors

The following task – the analysis whether or not FP-publications do relatively better or not compared to the overall oeuvre of the researchers (stemming from other projects), was analysed by first identifying all articles of the identified authors in Scopus with the relevant citation scores (i.e., without self-citations of the author). 1.315 journal articles were identified for 110 authors.

As mentioned previously, the length of the citation window is an important aspect. Therefore three different calculations were made:

- (1) the longest possible citation window (i.e., an article published in 2002 obtains all citations from 2002 to 2009, an article published in 2006 obtains those 2006-2009 etc.,
- (2) the standard two year citation window (i.e., for the article published in 2002, all citations obtained in 2002, 2003, and 2004 are counted),
- (3) a three year citation window. This one can only be used for publications until 2006.

Table 13 shows the results comparing 2 years and 3 years citation windows for a total of 1.313 publications (i.e., the journal article oeuvre of all authors identified with a peer-reviewed article) compared to their 157 FP publications. It shows several results that are in line with the literature, for example that the number of citations increases with a longer citation window and the share of non-cited literature decreases. However, on average, the longest possible citation window which the team used as well, is not the best option: there is something like a half-life period: for many publications citation rates are increasing – for a couple of years, but then they become less cited and given the denominator 'number of years' the citation rate for a very long period of six to eight years is lower than say for a medium-long one of 3-4 years. This is somewhat contradictory to the belief that publications in the SSH need longer to get cited. The analysis of the FP journal article publications shows that they are cited rather quickly after publication, but their peaks are also quickly reached. It would need some further analysis in order to analyse this finding in more detail.

When benchmarking FP publications in comparison to all publications of the sampled participants, there is clear evidence that the **FP publications are on average more cited than the rest of the oeuvre**, and their proportion of non-cited articles is with 10.9% substantially lower than the average proportion of non-cited publications of the same authors (32.7% for 2007)

FP articles are on average more cited than other articles of the participants

Table 13: Sample statistics: Overall citation results comparing FP publications to overall oeuvre of authors

Publication Year	2 years				3 years			
	Min	Max	Average	Share of non cited publications	Min	Max	Average	Share of non cited publications
ALL PUBLICATIONS (1.313 articles)								
2000	0	14	1.36	53.4%	0	27	2.36	42.6%
2001	0	14	1.68	40.0%	0	24	2.95	23.8%
2002	0	11	1.62	41.5%	0	18	2.71	26.8%
2003	0	20	1.54	38.9%	0	39	3.13	26.7%
2004	0	11	1.30	55.4%	0	29	2.87	38.6%
2005	0	12	1.80	41.8%	0	40	3.55	29.6%
2006	0	11	1.99	38.4%	0	12	2.59	29.3%
2007	0	14	1.35	40.7%				
FP PUBLICATIONS (156 articles)								
2000	0	1	0.50	50.0%	1	3	2.00	0.0%
2001	0	7	3.40	20.0%	0	12	5.20	20.0%
2002	0	8	2.55	27.3%	1	13	3.91	0.0%
2003	0	15	2.07	30.0%	0	37	4.03	13.3%
2004	0	11	1.40	53.3%	0	29	3.51	35.6%
2005	0	12	2.59	14.8%	0	40	6.15	7.4%
2006	0	8	4.30	5.0%	0	8	4.30	0.0%
2007	0	8	1.56	33.3%				

Source: Technopolis-Group
Data: FP5-IHP publication database

Gender

As the study team had also identified the gender of the consortium member authors, gender was used in a further analysis as differential. The first check concerned the citation scores according to gender. Again, adding evidence to more recent bibliometric analysis using gender, **journal articles of women are more cited than those of male authors**. This holds for all publications with average citation scores for women of an average of 5.2 compared to 4.6 for men, as well as looking only at the FP publications. Here the average citation scores for women were 9.2 compared to 6.5 for men.

In the pool of authors of peer-reviewed articles, the study team located more female than male researchers. The analyses of the articles according to call shows that between 63% and 73% female authors had participated in the Key Action.

h-index: a measure to address the impact of an author

An interesting indicator associated with 'impact' of a researcher is called the h-index (Hirsch index). The h-index indicates the number of articles A, which have at least been cited A times. An author with an index of 5 has at least five publications that were at least cited five times. The study team checked if the FP publication of an author belonged to those articles of his or her oeuvre, which is included in the calculation of his or her h-index.

Female authors obtain higher citation scores than men

While this index is not independent of size (i.e., the age, impact and overall productivity of an author), the indication whether or not an FP article is among the h-indexed covered articles, can be indicative concerning the overall average productivity of a researcher. In order to make some calculations, each article of an author was assigned the h-index of the author. For co-publications, the article was assigned to each author with his or her relevant h-index. The results in Table 14 show that if all publications of an author are taken into account (the first column 'all publications') an h-index between 12-15 is the highest obtained, and this concerns only 0.3% of the publications. The majority of articles is in the range of four to seven (55.5%) indicating that at least one of the authors has an h-index between four and seven. Concerning the FP publications one can see that the share of an h-index of one to three is much higher compared to the one for all publications. It suggests that many of the FP articles were among the first journal article publications of the authors or put it otherwise: the FP publication is a starting point for the academic publishing of many younger researchers.

Table 14: h-index by number of publications (1)

h-Index	All publications		FP publications	
	# of publications	%	# of publications	%
1 to 3	300	22.8%	56	35.9%
4 to 7	729	55.5%	70	44.9%
8 to 11	220	16.7%	28	17.9%
12 to 15	61	4.6%	2	1.3%
No h-index (1)	4	0.3%	0	0.0%
total	1314	100.0%	156	100.0%

Source: Technopolis-Group

Data: FP5-IHP publication database

(1) The h-index is calculated by Scopus for publications from 1995 onwards. There were however two authors with a minimum number of articles which have not received a citation – therefore, they do not have an h-index.

3.6 Book publications

Given that books are estimated to be the most important publication type for several SSH disciplines, but also knowing the limits using standard citation analysis, the manageable output classified as 'book' was further analysed using Google Scholar.¹⁰

This involved a further cleaning: out of the 116 books that were identified and classified in the database as such, the individual search revealed that five books were only book chapters, two were special issues of a journal and 13 were labelled as 'other', as they were available as pdf documents (suggesting they were in the end not edited), some within a series of a university or research institute, but they were not published by a publishing house. Four of the books were published before the projects had started and another seven were published during the starting year. One may also question 12 books published in 2003 as being the result of the third call – given that the projects started only in 2002 - to be the outcome of the Framework Programme funding. Due to the long preparation period needed for book publishing, the study team decided to ignore books dating either from the first year of the project or previous years

¹⁰ This analysis was executed at the end of the project on the demand of the DG. The results of the analysis and the changes they would have brought to the overall publication statistics of the sample, are not reflected in the descriptive analysis of the sample. This supplementary work would have decreased book and book chapter publication shares considerably.

The individual analysis also revealed one publication being a policy report to a ministry, another one a Master's thesis, and 19 could not be found in Google Scholar¹¹. Citation analysis using Google Scholar is a tedious process. The system does provide the number of citations of a collective volume type of book – and sometimes the citations of individual chapters, but not always the total number of citations of the book. It is equally almost impossible to define citation windows other than going manually through the citing publications and hoping that a publication date is mentioned (which is not always the case). Therefore, book citation analysis using Google Scholar is largely bound to the 'longest possible citation window', namely the time since a book got published until the day of the citation analysis. From a quality point of view, the possibly most serious problem is the number of received citations of a given publication. No matter what type of document is counted equally for this measure. A book which gets 20 cites can be cited 20 times in a presentation. Of course it can also include cites in peer reviewed journals or other books, but also any other grey literature – working papers, thesis, conference, etc.

Given these methodological constraints, the results for the book citation analysis are only illustrative.

Table 15: Citation rates for FP-resulting book publications using Google Scholar

	Number of books considered	Min-max number of citations	Average number of citations per book
1999	27	0-247	41
2001	18	1-230	36
2002	59	0-387	27
Total	104		35

Source: Technopolis-Group

Data: FP5-IHP publication database

Erratic results using book publications

Interestingly, in each of the three calls, one can find one outlier book publication that received by far the highest number of citations. The publication with the highest citation score was investigated further: it is a publication mentioned in the list of publications of a final report as: "Les réseaux sociaux, 2ème édition actualisée et augmentée, Paris, Armand Colin, 2004" by A. Degenne and M. Forsé. In Google Scholar the book is included together with two similar items of the authors appeared: "Introducing social networks", a book from 1999 with its second edition in 2004, or a book called: Les réseaux sociaux, dating of 2003 – which is equally available as pdf. It received about 400 citations until January 2010. When opening the pdf, it is an article, published 2004 in a French journal called Mathematics and Social Sciences. This suggests that Google Scholar is not yet a reliable tool for bibliometric analysis. Interestingly, on the project website, this book is not mentioned as the researcher's output of the project.

Table 16 compares the average number of Google Scholar hits for books resulting from the Key Action with the average number of citations for the refereed journal articles. Given the reliability issues of Google Scholar, the comparison underscores the relative quality of the journal publications that resulted from the Key Action. Furthermore, the exercise showed that the tendency to publish books based on research projects is somewhat exaggerated, as only 60% of the publications declared as books could be identified as such.

¹¹ This experience raises questions as to the reliability of information found in project reports.

Table 16: Average number of citations per FP-book and per FP journal article by publication year

	Average number of Google Scholar reported citations per FP-book	Average number of citations per FP journal article in Scopus
2001	18.3	20.0
2002	52.2	9.1
2003	39.4	10.4
2004	39.4	7.8
2005	12.3	9.3
2006	7.8	5.5

Source: Technopolis-Group

Data: FP5-IHP publication database

3.7 Research networks

Given the large material in terms of publication types and the complete information on authors, a social network analysis (SNA) was performed using the information of all publication output of the sample – including books and book chapters, which could not (or only with non satisfactory results) be analysed in terms of scientific impact so far. A co-authored/co-edited book as any other co-publication is a place where authors meet. SNA can provide a clearer view and mathematical analysis of the relationships among the FP authors, and given that the full scientific output can be considered, structuring effects may be observed. This will help in assessing the impact of FP5 on shaping the European Research Area measured through co-publications. With social networks analysis, collaboration patterns can be analysed and key players (here: authors), as well as core groups around which other groups are structured, can be identified. Again, due to ongoing cleaning efforts, the initial list of 1.450 publications (i.e. the total output from the 40 projects) was reduced to 1.313, forming the basis of the following analysis.

The following graphs show author IDs. They are equivalent to network nodes. The relationships or flows between these nodes are the ties (links). The position of the nodes in the graph – whether they are in the centre or the periphery, provides information on the involvement of authors in the most prominent publications by the number of authors involved. There are some aspects worth noting: for example, the degree of centrality of a network. Highly centralised networks depend on their core nodes – these are the ones via which the less central nodes are connected. If they are removed, the network loosens or falls apart. In a less centralized network, the network is more resistant; several nodes can be removed without preventing the remaining nodes to still reach each other. Large networks can also be made of several small networks. The following aspects are important to bear in mind:

- The degree of centrality is measured by the number of connections of nodes. The more connections, the more central a node. The most central nodes can be called “Hub(s)”, given that they often connect very peripheral nodes, nodes with only one or a few connections. There are equally connectors, mavens, leaders, bridges, or isolates that may be detected.
- The betweenness centrality measure reveals the number of direct connections with other nodes, the number of direct links can be small, but these can be crucial for bridging two parts of the network. This parameter reveals the influence over information flows. This measure reveals what we can call the “broker(s)”
- The closeness centrality is obtained regarding the geodesic distance (shortest direct or indirect paths) between a large number of nodes. This provides a good position to monitor the information flows in the network. Due to the shortest ways, they have the best visibility into what is happening in the network.
- The proximity between nodes indicates a collaboration frequency between nodes. Hence, core groups can easily be distinguished within the graphs.

Exhibit 11 shows the **links between the identified authors of books** in the sample.¹² The links between authors indicate collaborations. The non-connected nodes on the left side of the graph indicate that these authors did not collaborate with other FP participants to produce their books. On the other hand, there are several authors collaborating in the production of two or more books. This is especially the case for authors 340, 682, 735, 607, 398, and 571. The size of the nodes is directly related to the betweenness centrality which reveals the direct and indirect connections with other authors and the role of a 'broker'. There is an indirect connection when an author is on the path between two other authors. This is graphically displayed by the size of the node. This centrality indicator seems to be the most appropriate to indicate key authors.

Table 17: Average number of authors per book

	Nr. of Publications	Share (in %)
More than 6 authors	1	0,9%
5-6 authors	2	1,8%
3-4 authors	30	26,3%
2 authors	45	39,5%
1 author	36	31,6%
Total	114	100 %

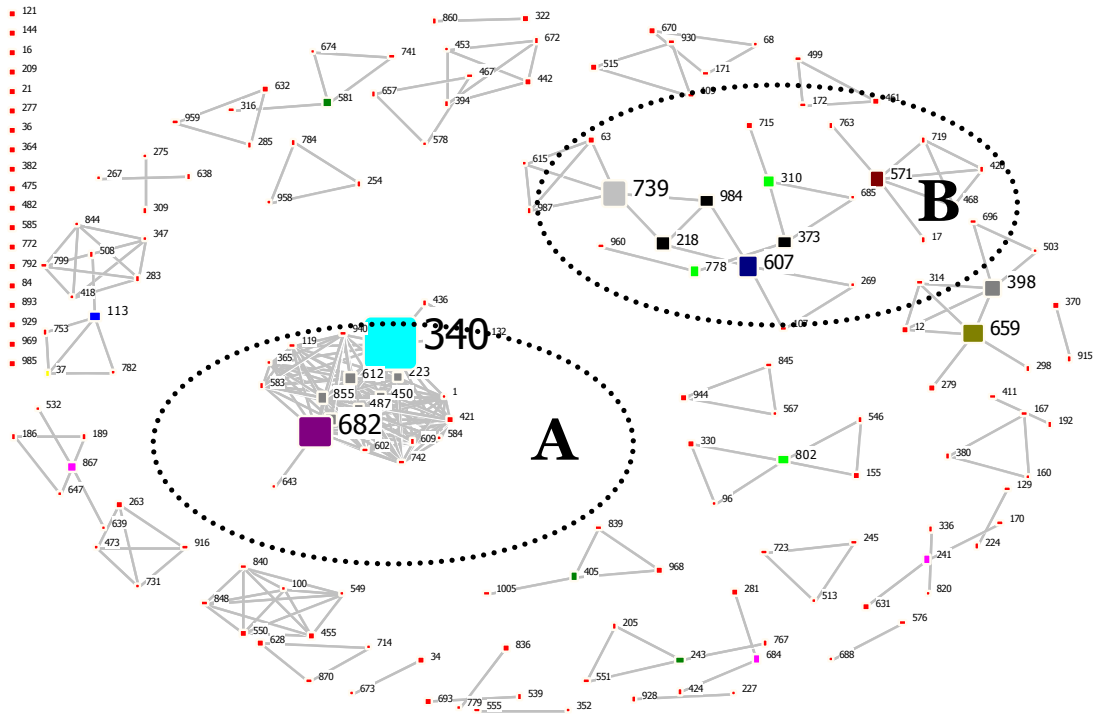
Source: Technopolis-Group
Data: FP5-IHP publication database

The average number of authors per book is 2.26. Table 17 shows the distribution of the numbers of authors involved in a book by five groups. The majority of books is co-written while 31.6% of the books are published by one author only.

It appears that only one book has involved more than six authors: the collective volume resulting of 99-4 has involved 26 authors. This group appears in the area A of Exhibit 12, and two authors are in a particular exposed position: author 340 and author 682. The area B shows the collaboration between authors in four distinct books where authors 739 and 607 appear to be the most connected.

¹² Here the analysis used the originally identified books – prior to the cleaning efforts made as described in chapter 3.6. This implies that a smaller number of books were not published as such but appeared as any other grey co-authored literature and

Exhibit 12: Links between authors in FP-resulting books



Source: Technopolis-Group
 Data: FP5-IHP publication database

The study team then looked at the publication type “book chapter”. While the work on a book is often in form of a collective volume output of the research project, a FP-book chapter can also be a contribution to an external book publication. Many authors listed individual book chapters, Exhibit 13 analyses this network.

The average number of authors per chapter was 1.7. As presented in Table 18, more than 50% were written by a single author, 27% by two authors and 17% by three or four authors.

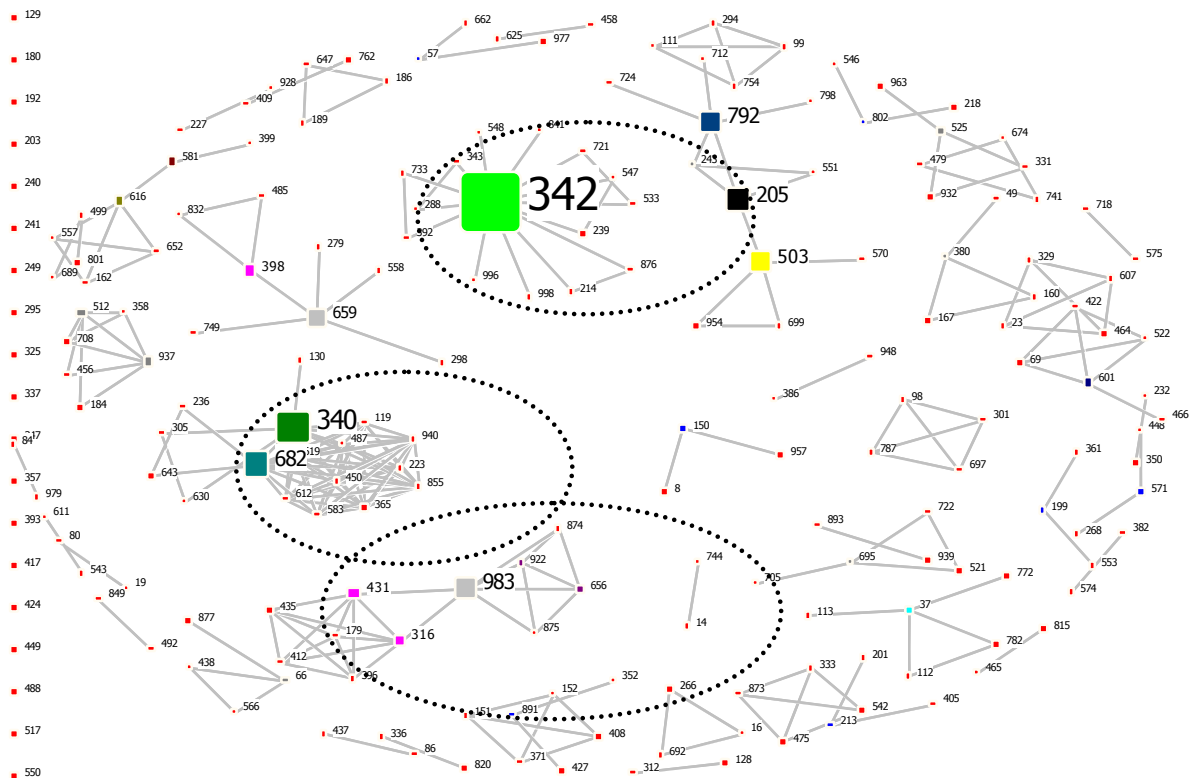
Table 18: Number of authors per chapter

	Number of Publications	Share (in %)
More than 6 authors	1	0,4
5-6 authors	1	0,4
3-4 authors	46	16,8
2 authors	74	27,1
1 author	151	55,3
Total	273	100,0

Source: Technopolis-Group
Data: FP5-IHP publication database

The upper oval of Exhibit 13 indicates that author 342 is involved in the writing of 13 different chapters. This author is the focal point of a network made by collaboration between two or three authors. The network of the 99-4 project is built around four collaborative chapters where authors 340 and 682 appear as focal points – similar to the 'book' analysis above, which identified both authors as central to the network.

Exhibit 13: Links between authors in book chapters

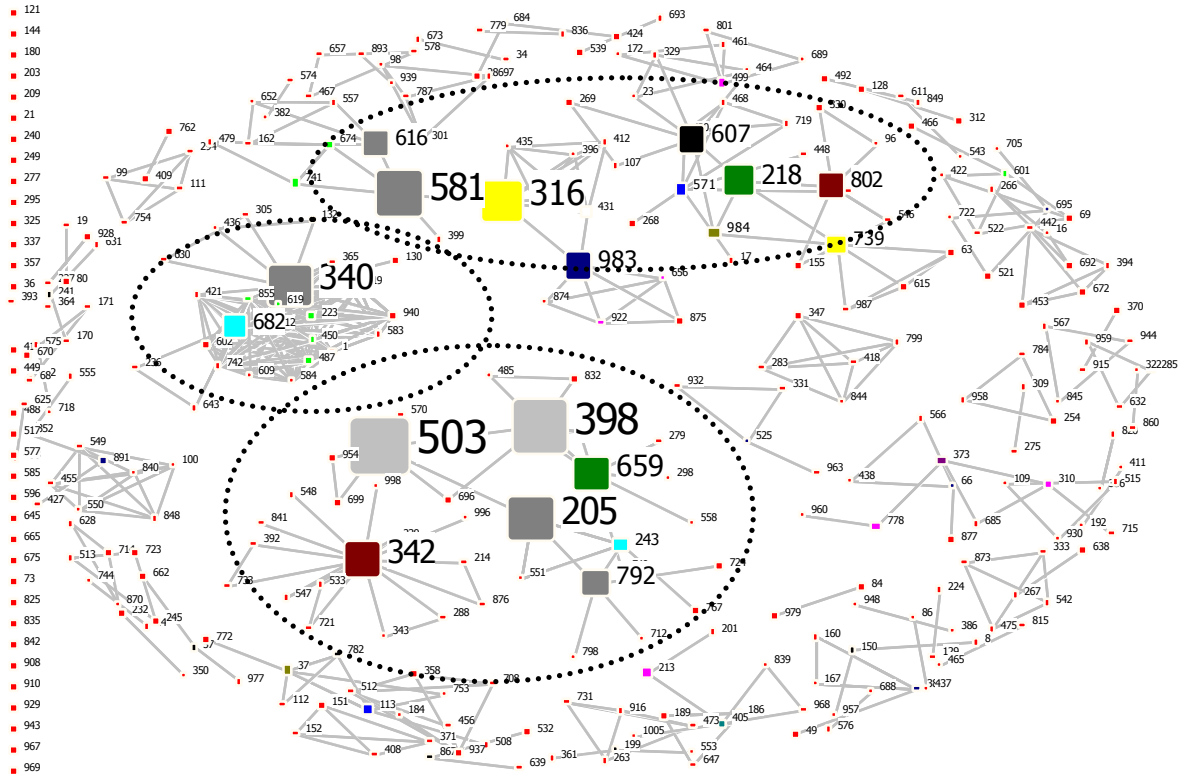


Source: Technopolis-Group
Data: FP5-IHP publication database

Exhibit 14 shows the collaboration between authors in books and chapters writing. From this graph, several distinct author networks can be seen. It appears that less than 50 authors have a central role and only 15 are key players and have an influence

on the network. This structure suggests that there are in fact only one or two main authors per project that drive the editing of a collective volume.

Exhibit 14: Links between authors in books and chapters



Source: Technopolis-Group
Data: FP5-IHP publication database

An important aspect concerns the question about the status of the participating researchers. Status is often connected to age and impact, while the latter comes with a longer publication history and thus more options for being cited. Given DG-RTD's interest to see if the FP attracts excellent researchers, the study team noticed through the various steps of the project, that the FP participation served a number of young researchers – doctoral students as well as post-docs - as a means to publish articles in peer reviewed journals and as a stepping stone for establishing a career. As the provided data on projects does not include the age of the researchers, proxies based on publication histories were constructed. This was possible by retrieving the biographic information of all journal articles for those 157 researchers with a peer-reviewed article, which were in our sample. The researchers were classified as following:

- The *senior researchers* who started to publish at least 4 years before the FP publication. They are indicated by a triangle in Exhibit 14;
- The *experienced researchers* who started to publish at least 3 years before their FP publication. They are indicated by a square;
- The *junior researchers*, which are represented by circles. They have started to publish less than 3 years before the FP publication.

Key findings

What did the network analysis reveal? Note that the fact that cross project publications are not seen could well be the result of the construction of the sample which aimed at including the projects according to different areas and topics. There is no standard or average project construct responsible for the joint scientific publications of books or collective volumes and their individual chapters. There is a bit of everything: there are projects where senior and experienced researchers dominate the book publications, but there are equally projects where junior researchers are involved. Concerning the journal articles, one finds a rather surprisingly high number of articles by junior researchers. Given that the publication process in a peer-reviewed high-impact journal is more difficult than publishing in a non-leading journal, one may conclude that a number of young researchers were able to obtain important research findings worth to be included in the leading journals. The network analysis also suggested that young researchers opt for scientific journal articles as publication means.

This outcome of the Key Action, namely the aspect of benefiting young researchers, seems to be a key structural aspect which – if extrapolated – has potentially a very wide positive effect.

The limited network analysis and the results on the benchmarking analysis do however **not suggest to change the structures dramatically – but to take more advantage of the positive effects**. It is important to have senior and experienced researchers involved as they are in general the initiators of these international projects and bring in their professional networks.

It is equally benefiting in terms of scientific impact, to have a large share of **female researchers** participating. Given that their output is often smaller in terms of quantities but is more cited, their FP resulting contributions is either of a higher quality or more useful for the research system – or a combination of both. Therefore it does make a difference if female researchers are participating in a project.

The **development of scientific careers** is another issue which is in particular relevant. The Framework Programme has an impact on initiating or aggravating the careers of young researchers. The share of peer-reviewed journal articles single authored by young researchers indicated that there are not only the usual network externalities of establishing own professional networks, but the actual research with its international and comparative aspects seems to be of high relevance that it even allows young and non-established authors to publish in established journals. To our knowledge, this aspect has not yet been analysed in greater detail.

4. Benchmarking the “status” of the Key Action in the international scientific scene

Status benchmarking limited on four fields

The second task of this study concerned the analysis of the attractiveness of the Key Action and the provision of an international benchmark. The study should therefore identify highly cited authors in four fields:

- economics,
- sociology,
- political sciences,
- innovation (policy).

The choice of fields was based on a study of the disciplinary composition of FP5 consortia which showed that sociology, political science and economics dominated¹³. Innovation policy was a narrower field of public policy which was traditionally important in European SSH research including the Key Action. In order to set a benchmark for the four fields, a limited set of high-impact journals and the articles in these journals were defined to set the benchmark. The analysis was then to compare the pool of authors identified via the high-impact journals and the FP-IHP Key Action participants.

The basic approach for all four fields was the following:

- (1) Identification of a limited number of high impact journals in the Social Science Citation Index.
- (2) Retrieval of all bibliographic information for all publications between 2002 and 2006.
- (3) Analysis of the full set of publications per field.
- (4) Limitation to journal articles.
- (5) Attribution of total number of citations per article to each author.
- (6) Aggregation of citations per author.
- (7) Establishing lists of the top 100 European authors
- (8) Comparative, manual check of the top 100 list names with the IHP-database (2.510 participants),
- (9) Comparative, manual check of the top 100 list names and the cleaned list of FP authors (981).
- (10) Manual analysis of institutional affiliation match.

4.1 Identification of journals – defining the pool of documents

The identification of the journals in the four fields was based on the Social Sciences Journal Citation Index (2007). Those top ten journals were chosen which did have a European editor (i.e., the publishing house). The decisive indicators were the provided one-year impact factor (including self citations) and the 5-year average impact factors.

¹³ A. Benavot, S. Erbes-Seguin and S. Gross (2005) Interdisciplinarity in EU-Funded Social Science Projects, in M. Kuhn and S. Remoe (eds) Building the European Research Area: socio-economic research in practice, New York, Peter Lang

Not for all journals the latter is available – some journals were not covered or did not exist that long.

The approach for the identification of journals differed for the field ‘innovation’. Here, the identification of the journals was based on a keyword search and then combined with a check whether or not this journal was an SCI covered journal and thus allowing impact analysis of the journal. Details of the identification of the journals as well as the various output tables on the four fields are provided separately.

4.2 Retrieval of all bibliographic information

After the identification of the journals, the information on all types of documents published between 2002-2006 were retrieved and stored in a database. The information contained all necessary bibliographic information as well as the number of citations per article.

For more than 10.000 documents, bibliographic information was retrieved and included in a database. Table 20 provides the figures by field. Innovation policy was by far the largest field

Bibliographic information for 10.000 journal articles for 2002-2006

Table 20: Number of documents by field (2002-2006, top impact journals)

	Number of documents retrieved (2002-2006) and share of total (in brackets)
Economics	2.224 (24.7%)
Sociology	1.453 (14.2%)
Political Sciences	2.305 (22.5%)
Innovation Policy	3.949 (38.6%)
Total	10.231 (100%)

Source: Technopolis-Group

Data: FP5-IHP publication database

4.3 Limitation to journal articles

The bibliographic information retrieved for the published documents contained six types: articles, reviews, notes, conference papers, editorials and errata. Given that ‘reviews’ tend to be much more cited than articles, but are not presenting any novelty as such, it is common in bibliometric analyses, to take out this category. All other categories but articles are also most often left out for citation analysis. The study team followed this common practise and limited the analysis to articles.

The whole ‘pool’ of documents served for determining the average and the highest number of citations per journal in order to benchmark the individual authors.

The following figures include the results from the first set of economic documents retrieved. For this first set, a more detailed analysis was done than for the second one.

Almost **8.000 articles and more than 13.000 authors were the basis for the benchmarking exercise**. Again by far the largest field in terms of articles and authors was innovation policy (Table 21).

Basis of analysis: 8.000 articles and 13.000 authors

Table 21: Number and share of articles and authors by field

	Number of articles (2002-2006)	Total number of authors
Economics	2.238 (28.8%)	3.370 (25.2%)
Sociology	905 (11.6.%)	1.206 (9.0%)
Political Sciences	1.288 (16.6%)	1.704 (12.7%)
Innovation Policy	3.344 (43.0%)	7.091 (53.0%)
Total	7.775 (100%)	13.371 (100%)

Source: Technopolis-Group

Data: FP5-IHP publication database

4.4 Attribution of total number of citations per article to each author

Considering that the 2-year citation window, which is often used for citation analysis, may not be the optimal in SSH, the total number of citations obtained by a publication was used. Thus articles in 2002 have a citation window 2002-2009, articles from 2003 from 2003-2009, etc. each total citation score was then attributed to each author of the article (i.e., for co-publications).

In the next step, the citation scores were aggregated at author level: if there were two or more articles by one author, the aggregated citation score provided the sum for all citations obtained.

The comparison of the total number of citations per field as well as the average number of citations per author may be the result of different citation behaviour between scientific disciplines or simply due to differences in the degree of coverage of the different literature by the databases.

Table 22: Number and share of articles and authors by field

	Total number of citations (1)	Average number of citations per author
Economics	87.119 (57.0%)	25.9
Sociology	11.916 (7.8%)	13.2
Political Sciences	16.203 (10.6%)	12.6
Innovation Policy	37.527 (25.6%)	11.3
Total / Average	144.614 (100%)	15.8

Source: Technopolis-Group

Data: FP5-IHP publication database

(1) Representing an 'inflated' figure due to double-counting at author's level

4.5 Establishing lists of the top 100 European authors

Looking at the Top 10% of the authors in each field as covered by Scopus, one finds 23 individuals in economics compared to 12 in sociology, 13 in political sciences and 16 in innovation. Each economist in the Top 10% share has a cumulated citation score of 257 and a maximum of 743.

Table 23: Number and share of articles and authors by field

	Range of top 10% citations per author	Nr of authors in Top 10%
Economics	743-257	23
Sociology	125-26	12
Political Sciences	208-29	13
Innovation Policy	347-61	16
Average	355-93	16

Source: Technopolis-Group

Data: FP5-IHP publication database

4.6 Matching the top 100 individuals with the IHP-database

The last step in the benchmarking of individual authors consisted of a manual check, comparing the lists of the identified 100 European most cited researchers in each field (based on the citation counts of articles in top SSCI journals) with the IHP-database. The four separate lists of people were matched against the IHP-database and its included 2.510 participants.

There was an extremely low overlap between the lists (see Table 24). In a next step, the organisations that employed the highly cited authors were compared with the affiliations of the principal investigators in the consortia. The reason for this was the perception that the database of participants does not record the complete involvement of researchers in the Key Action. As can be seen in Table 25, the overlap was a bit more pronounced although the data sets both on the institutional affiliation of authors and on the participating organisations in consortia presented important difficulties.

Table 24: Number of participants in the Socio-economic Key Action among the 100 Europeans with most citations in high impact journals

	Number of Top 100 individuals participating in the FP-IHP Key Action
Economics	3
Sociology	6
Political Sciences	3
Innovation Policy	6

Source: Technopolis-Group

Data: FP5-IHP publication database

Table 24 probably underestimates somewhat the involvement of researchers amongst the 100 most cited in SSCI journals, and Table 25 clearly overstates it. It is obvious that very few of the 100 European researchers who are most cited in SSCI journals in economics, sociology, political science as well as in the journals covering innovation policy have worked in the Socio-economic Key Action.

Table 25: Number of institutions/departments participating in the Key Action amongst the affiliations of the 100 Europeans with most citations in high impact journals

	Number of institutions/departments identified among Top 100-affiliations participating in the FP-IHP Key Action
Economics	40
Sociology	54
Political Sciences	44
Innovation Policy	34

Source: Technopolis-Group

Data: FP5-IHP publication database

One reason for this is that the sets of journals chosen differ from the publication means of FP-participants and that the journals with the higher SSCI impact do most likely not match the research topics of the Key Action. This is certainly partly the case, as special features within a discipline lead to different publication and citation habits. In the case of economics for example, the most apparent difference is the strict theoretical/econometric focus of the top journals versus a rather applied, empirical focus of the economics research projects. Indeed, if one compares the set of journals used in this analysis and the –equally peer-reviewed, international - journals analysed in the sample, there is a marked difference.

However, an important consideration is that the problem of measuring citation impact is not resolved by restricting analysis to a set of journals, as lack of coverage results in a double bias. There is a loss of articles competing for citations and also a loss of citations. This is why the lack of coverage cannot be compensated by segmenting the field. In our effort we found that people with a specific disciplinary background are not bound to a set of disciplinary journals. We found a number of political scientists among the top sociologists, geographers among the top economist and so on. While representing fields/disciplines by a set of journals may be useful for measuring output in such and such a field, it is not useful for benchmarking purposes, which require truly comprehensive data sources.

5. Results and recommendations

Key programme data

- 2.510 individual participants
- 203 projects

5.1 Technical results

Concerning the first objective of the study - the establishment of a **profile of scientific publications** from European cooperative research projects, and a bibliometric evaluation of this profile, the following key data for the Key Action can be listed:

- IHP Key Action statistics:
 - There were 2.510 participations (coordinators and principal investigators) mentioned for 203 projects;
 - The average number of participants per project was 10;
 - The average project length was 32 months;
 - Coordinators came from 21 countries while
 - Partners came from 39 countries;
 - The average financial contribution was € 761.362.

Key sample data:

- 40 projects
- 981 authors
- 1.449 publications

The descriptive analysis provided the basis for establishing the sample. Starting with collecting the information on publications of 40 projects, 981 authors were identified. A match with the overall database (and the 2.510 participations mentioned above), 326 (33%) were identified as consortium members and 655 (67%) were considered external authors.

These 981 authors published 1.449 publications. The largest share of this is so-called labelled 'grey literature', i.e., working papers, reports etc. with 45%, followed by journal articles with 28.3%, book chapters (19%), and books (8%).

- 166 peer-reviewed journal articles

- 210 authors
- 52% participants in the IHP programme.

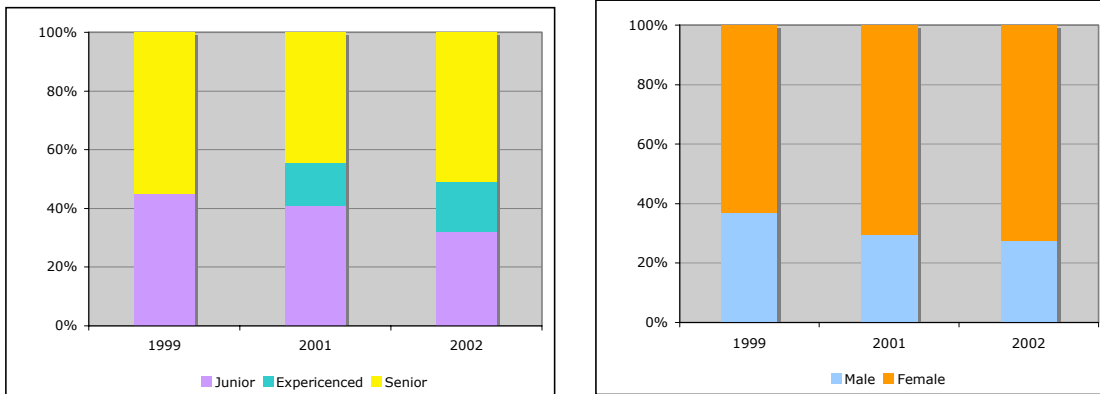
Given that the task was on providing an impact analysis, the sample was further limited to peer-reviewed articles. Among the total number of 410 journal articles, 166 peer-reviewed articles were identified via the list of journals covered in the Web of Science and Scopus. These 166 articles were authored by 210 individuals. A check of the IHP-database and project websites revealed only 110 authors as having participated in the FP5-IHP Key Action, and 100 as external authors.

For the 110 researchers of the journal articles that were identified as participants of the Key Action, scientific profiles were established. These profiles contained all journal articles by a researcher as included in Scopus. The profiles were analysed in terms of scientific impact but they also allowed to categorise the researcher as senior, experienced, or junior, and by gender (see Exhibit 15). The following results are important:

- The analysis of the profiles revealed that the FP5-IHP articles do better than the average publication, and are often among the best publications for a number of authors.
- FP resulting articles perform better than the average article.
- There are several young researchers managing a single authored publication with the FP5-IHP publication often as the starting one of their career.
- Female authors outperform male authors in terms of citation impact.
- In terms of impact, female authors obtained in general higher citation scores than male authors. This holds for their average publication and is very marked for their FP5-IHP publications. While there is a difference between the male and female authors of FP publications as well as the average publication, the biggest benefit and publish in peer reviewed journals.

difference is between the FP5-IHP publications and all other publications of women.

Exhibit 16 Age and gender profiles by call



Concerning the second task of **establishing benchmarks**, the analysis was focussed on four fields: economics, sociology, political sciences and innovation policy.

For each field, ten top journals ('high impact journals') served as the basis. For the period 2002-2006, bibliographic information for 10.000 documents included in these journals was retrieved and analysed. A total of 13.000 authors and 8.000 journal articles were identified which served as the basis for the impact analysis. The first outputs of this analysis were lists of top researchers in terms of citation rates. In a second step, these were matched to the IHP Key Action participants. The matching rate was 3.4%, meaning that **3.4% of the 100 European, most cited individuals in high-impact SSCI journals in the fields selected were certainly also participating in the IHP Key Action**. While this is probably an underestimate it is nonetheless low.

Low participation of highly cited authors

5.2 Limitations of the use of bibliometric analysis

Bibliometric analysis is certainly an important tool in the evaluation of scientific performance. However, there are clearly limits which show in the evaluation of SSH research.

Concerning the establishment of scientific profiles, the use of bibliometric methods leads to valid results and interesting information which can be used for the comparison of previous ex-ante evaluation results.

The analysis revealed important findings concerning the publication behaviour in SSH: book publications are less dominant for SSH than anticipated. According to our findings, books and book chapters are not the dominant publication type for FP scientific output. Even if some books may not have been included in the project final reports, the book type may be estimated to account for 25% of all publications. This is much less than the share of 61% of book publications, Bourke et al¹⁴ (1996) have mentioned for the social sciences and humanities. This may be the result of rapidly changing funding mechanisms and promotion decisions which are more and more based on 'scientific impact' – available for journal articles but not the other publication means.

No dominance of book publications

¹⁴ Bourke, P, L. Butler, B. Biglia (1996): Monitoring research in the periphery: Australia and the ISI indices, research Evaluation and Policy Project, Monograph Series No. 3, 72p

Another finding concerns the issue of the citation window. While the literature suggests that SSH research needs more time¹⁵ to be absorbed by peers and to get cited, the citation analysis of this study suggests that this ‘ripening’ period is not too different from other fields of science. A three to four year citation window will impact the overall citation score of an article – but a longer period can negatively impact the scores per article as well as per author. In this respect it seems that a number of publications do not obtain more citations over a longer period but instead tend to be forgotten. Of course, a small number of eminent publications that belong to the core of a field¹⁶ will remain cited, but for the majority of articles this is not the case. The longest possible citation window is therefore not the best option. In this study, the longest possible window was seven years and it was applied to the publications from 2002.

5.3 Recommendations for further refinements of the approach

While the general approach yielded interesting insights, important issues remain concerning the IHP database and the information which is collected at EU level:

Improvements:

- standardisation of institutional names
- required acknowledgement of funding in publications
- use of EU CVs

- As long as there is no standard in providing the institution name and the department/faculty or other sub-level affiliation, the analysis has shortcomings. The introduction of the unique identifier PIC may resolve this problem for FP7 and onwards.
- The second shortcoming is at individual level: as long as only the coordinator and principal investigators are listed, the bibliometric analysis is limited and biased¹⁷, and overall structural insights are limited as well. In this respect, a full sample analysis as used in this study provides an array of options for further insights that are beyond the pure scientific impact analysis.
- Bibliometric analysis could also be improved if the FP research output was identified more clearly. In order to identify journal articles (as well as other publication means) more systematically, it would be useful to contractually bind the authors to acknowledge the specific Key Action funding in their publications. Database mining allows for identifying publications with such a tag rather easily.
- Finally, the analysis could be improved if curricula vitae were to be taken into account. CVs do not only contain most likely a list of publications, but provide also information on age and affiliation(s). With this sort of input, a more telling and accurate analysis will be possible. Given that the EU CV is getting more and more standard, and that they are included in the proposals (and often available on the web as well), a bibliometric approach may be enhanced by personal information from the CVs.

¹⁵ A precise length of citation window is not recommended in the scientific literature yet, but a seven year period can be found as a suggestion.

¹⁶ This can be observed for example for G. Dosi’s article “Technological paradigms and technological trajectories...” of 1982. It is a standard reference article that is still growing with an annual average rate of 10%, having obtained only since 1996 until now 860 citations.

¹⁷ A survey of the identified researchers as done in the mid-term assessment of FP6, where the surveyed researchers provided their lists of FP originating publications obtains more likely the lists of successful participants than less successful ones (in terms of publication output). Therefore, the analysis of such a sample may be more inclusive of all articles, but biased.

Bibliography

- Andersen, H. (2000): Influence and reputation in the social sciences – how much do researchers agree?, *Journal of Documentation*, 6(6), pp. 674-692.
- Archambault, E., et al (2006): Benchmarking scientific output in the social sciences and humanities: The limits of existing databases, *Scientometrics*, 68/3, pp 329-342.
- Benavot, A.; Erbes-Seguin, S.; Gross, S. (2005) *Interdisciplinarity in EU-Funded Social Science Projects*, in M. Kuhn and S. Remoe (eds) Building the European Research Area: socio-economic research in practice, New York, Peter Lang
- Bornmann, L.; Daniel, H.-D. (2009): *The state of h index research. Is the h index the ideal way to measure research performance?* EMBO Reports, 10 (1);
- Bourke, P.,; Butler, L.; B. Biglia (1996): *Monitoring research in the periphery: Australia and the ISI indices*, research Evaluation and Policy Project, Monograph Series No. 3.
- Evidence Ltd for Universities UK, 2007, *The use of bibliometrics to measure research quality in UK higher education institutions*
- Glänzel, W.; Schubert, A. (2005): Domesticity and internationality in co-authorship, references and citations, *Scientometrics* 65 (3), pp. 323-342
- Hicks, D. (1999): The difficulty of achieving full coverage of international social science literature and the bibliometric consequences, in: *Scientometrics*, 44, 2, pp. 193-215.
- Kosmulski, M. (2009): New seniority-independent Hirsch-type index, *Journal of Informetrics* 3 (4)
- Norris, M.; Oppenheim, C. (2007): Comparing alternatives to the Web of Science for coverage of the social sciences' literature, *Journal of Informetrics* 1 (2), pp. 161-169;
- Pierce, S. (1987): *Characteristics of professional knowledge structures: Some theoretical implications of citation studies*. LISR, 9, 143-171.
- Schubert, A.; Glänzel, W. (2004) *Cross-national citation patterns of the EU15 and the Candidate Countries*. 8th Science and Technology Indicators Conference 2004

European Commission

EUR 24311 – Evaluation of the Impact of Framework Programme supported Social Sciences and Humanities Research - *A bibliometric approach*

Luxembourg: Publications Office of the European Union

2010 — 56 pp. — 17,6 x 25,0 cm

ISBN 978-92-79-15370-9

doi 10.2777/18202



How to obtain EU publications

Publications for sale:

- via EU Bookshop (<http://bookshop.europa.eu>);
- from your bookseller by quoting the title, publisher and/or ISBN number;
- by contacting one of our sales agents directly. You can obtain their contact details on the Internet (<http://bookshop.europa.eu>) or by sending a fax to +352 2929-42758.

Free publications:

- via EU Bookshop (<http://bookshop.europa.eu>);
- at the European Commission's representations or delegations. You can obtain their contact details on the Internet (<http://ec.europa.eu>) or by sending a fax to +352 2929-42758.

The report presents the results of an evaluation study of the impact of Framework Programme supported Social Sciences and Humanities research by means of bibliometric analysis. It covers research carried out under the Socio-economic Key Action of the 5th Framework Programme, and aims to provide a benchmark for the evaluation of the scientific impact of subsequent programmes.

