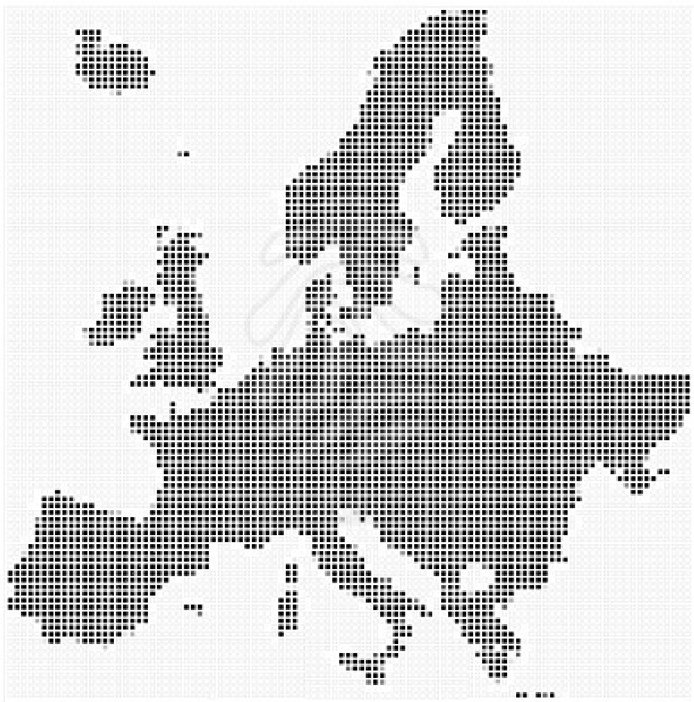
USING PERFORMANCE INDICATORS IN MONITORING THE IMPLEMENTATION OF ICT RESEARCH IN FP6 AND FP7



D5-Final Assessment Report

Tender: – SMART 2008/0032 / OJ 2008/S 174-231772

KITeS-CESPRI - BOCCONI UNIVERSITY

Stefano Breschi is the author of this report. Gianluca Tarasconi has managed the construction of the database of scientific publications and patents. Federica Uccella, Marinella Tortora, Stefano Spagnoli, Maurizio Polenghi and Daniele Quattrocchi have provided invaluable research assistance. This report is based on the results of our research and analysis and does not necessarily reflect the views of the European Commission. The author is the sole responsible for any errors or omissions.

Milan, June 2010.

Summary

List of tables

List of figures

Executive Summary

This report contains the results of a pilot study aiming to assess the scientific and technological achievements of collaborative projects funded by the Information Society and Media Directorate General (DG INFSO) of the European Commission in the context of FP6 and FP7. To this purpose, the study has examined a data set on the scientific publications and patents produced by the funded projects. Raw data have been collected by the DG INFSO itself through a survey sent to the project coordinators. KITeS-Bocconi has then undertaken a thorough work of data cleaning and standardisation. Moreover, it has also carried out the work of matching data on scientific publications with the Thomson-ISI Web of Science database and data on patents with the Espacenet and the EPO-Patstat databases. The resulting data set contains for each published article and each patent reported by project coordinators a full set of bibliographical information that allow assessing several aspects of the scientific and technological activity of DG INFSO projects. It has to be remarked that the work of raw data cleaning and standardisation took a large amount of the time and resources devoted to the study due to the very poor quality of data received. In this respect, the experience gained in carrying out the study suggests a few possible ways to improve the process of data collection. In particular, it suggests the adoption of a web-based platform to collect data from project coordinators and the reduction in the redundancy of information requested from project coordinators through the use of the Digital Object Identifier (DOI) system. This would save time for project coordinators to enter relevant data and it would also reduce the amount of noise and errors contained in the filled questionnaires.

The analytical part of the study has investigated three important aspects of the scientific and technological activities carried out by DG INFSO projects. First, it investigated the scientific and technological *productivity* of projects. Second, it evaluated the *quality* of the scientific output. Third, it assessed the extent of *international*, *inter-regional* and *inter-organizational* collaboration. Most of the analytical part of the study has focused on FP6 projects.

As far as the first aspect is concerned, results suggest that the distribution of scientific and technological output is concentrated in relatively few projects. Around fifty percent of all FP6 projects did not report any scientific articles, while the fraction increases up to eighty five percent in the case of patents. This result has to be qualified in two ways. First, a correct assessment of the scientific productivity should include also conference and proceedings paper, which have been left out in the present study due to the data quality problems mentioned above. Second, the scientific and technological productivity differs across projects funded through different instruments. In particular, our results suggest that Networks of Excellence (NoE) outperform other projects in terms of both scientific and technological performance. Integrated Projects (IP) are on average more productive than Strategic Targeted Research Projects (STReP) in terms of publications and patents. However, this larger productivity is almost entirely explained by the larger size in terms of participants and financial resources. In this regard, the preliminary results of the study seem to indicate the existence of diminishing returns with respect to the size of projects. In other words, it is possible that the FP6 has funded "too large" projects whose larger size has not been matched by a proportionally larger amount of scientific and technological output.

As far as the *quality* of output is concerned, we carried out a benchmarking analysis of the citations received by scientific articles produced by DG INFSO projects with respect to the population of articles in the same scientific area. Results show that on average the number of citations received by project articles is larger than for the average article in the same journal and scientific field. This result suggests that EU DG INFSO projects have been able to attract top quality researchers and research teams.

The analysis of co-authorship patterns shows that DG INFSO articles are significantly more likely to involve intra-European collaboration than the average European article in the same field, thereby suggesting that FP projects have been highly effective in reinforcing the extent of intra-European collaboration in research. However, this result has to be qualified. On the one hand, DG INFSO articles are also relatively less propense to show extra-European collaboration than other European articles. On the other hand, the regional concentration of the scientific output produced by DG INFSO projects looks significantly larger than the regional concentration in the benchmarking population of articles. Finally, analysis of co-patenting reveals that the extent of joint patent ownership in the context of DG INFSO projects is significantly higher than in the benchmarking population of patents. This finding probably signals the need for organisations participating in EU funded projects to devise schemes for sharing the intellectual property rights over the results of joint research. Moreover, even though company-company ownership is the most prevalent type of joint ownership, a significant amount of co-patents involves PROs and universities.

Extended Summary

The report illustrates the results and findings of a pilot study aiming to assess the scientific and technological achievements of collaborative projects funded by the Information Society and Media Directorate General (DG INFSO) of the European Commission in the context of FP6 and FP7. To this purpose, the study has examined a data set on the scientific publications and patents produced by the funded projects. Raw data have been collected by the DG INFSO itself through a MS-Excel survey sent to the project coordinators. KITeS-Bocconi has then undertaken a thorough work of data cleaning and standardisation. Moreover, it has also carried out the work of matching data on scientific publications with the Thomson-ISI Web of Science database and data on patents with the Espacenet and the EPO-Patstat databases. The resulting data set contains for each published article and each patent reported by project coordinators a full set of bibliographical information that allow assessing several aspects of the scientific and technological activity of DG INFSO projects.

Data quality issues

A first set of results emerging from the study concerns the methodological issues related to the process of data collection and data quality. In this respect, it has to be remarked that the work of raw data cleaning and standardisation took a large amount of the time and the resources allocated to the study due to the very poor quality of the data received from project co-ordinators. The experience gained in carrying out this work suggests a few possible ways to improve the process of data collection and the quality of the data. In particular, it suggests the adoption of a web-based platform to collect data from project coordinators and the reduction in the redundancy of information requested from project coordinators through the use of the Digital Object Identifier (DOI) system. The DOI is a unique label that allows identifying in a univocal and persistent way a computer readable object that can be found on the internet. Given that virtually all publishers are nowadays are adopting the DOI system and therefore label the articles and publications they issue with this system, collecting this single information from project coordinators would significantly speed up the process and would also improve the precision and reliability of the data collected. Appendix 1, which accompanies this report, elaborates a concrete and detailed proposal along these lines.

The analytical part of the study has investigated three important aspects of the scientific and technological activities performed by the DG INFSO projects. First, it examined the scientific and technological *productivity* of projects. Second, it evaluated the *quality* of the scientific output. Third, it assessed the extent of *international*, *inter-regional* and *inter-organizational* collaboration. Most of the analytical part of the study has focused on FP6 projects given the relatively low number of FP7 projects, articles and patents in the database. The major results emerging from our analysis are summarised below.

Assessment of productivity

- The distribution of scientific productivity looks rather skewed. Around 50% of all FP6 projects have not reported the production of any scientific articles. This fraction is even higher for FP5 and FP7 projects, though this result is less significant given the relatively low number of FP5 and FP7 projects in the database. The picture changes if one examines the publication of papers in *conference proceedings*. In this case, the fraction of FP6 projects with no publications decreases to around 18%. The two figures are not fully comparable, given the different scientific importance of articles published in peer-reviewed journals and papers published in conference proceedings. Extending the assessment analysis to including conference papers represents a natural and fruitful avenue for further research. At the same time, such an extension would probably require improving the process of data collection from project coordinators.

- The degree of skewness in the distribution of patents among projects is even larger than the one observed for the scientific output. Only 41 out 199 FP5 projects (21%) have applied for at least one new patent and the fraction is even lower for FP6 projects (109/927=12%). Yet, even though the fraction of patenting projects looks lower for FP6 than for FP5 projects, the average number of patent applications per patenting project is larger for the former than for the latter: 2.5 vs. 1.8. Please note that these figures must be interpreted with some caution. Due to the poor quality of data sent back by project coordinators we could not trace information on all patent applications claimed to be produced by projects.

- An important result emerging from the analysis is that the scientific and technological productivity differs across projects funded through different instruments. In particular, our results suggest that Networks of Excellence (NoE) outperform other types of projects in terms of both scientific and technological performance. NoEs have generated on average 29.7 scientific articles per project as compared to 8.7 for Integrated Projects (IP) and 4.6 for Strategic Targeted Research Projects (STReP). Moreover, even though IPs look apparently more productive than STRePs, this higher productivity is almost entirely explained by their larger size in terms of participants and financial resources. In this regard, the preliminary results of the study seem to indicate the existence of diminishing returns with respect to the size of projects. In other words, it is possible that the FP6 has funded "too large" (IP) projects whose larger size has not been matched by a proportionally larger amount of scientific and technological output.

- Concerning the filing of new patent applications, IPs and NoEs seem to exhibit a higher propensity to patent than STRePs: the fraction of all projects reporting at least one new patent application is equal, respectively, to 23%, 29% and 10%. Moreover, considering only projects with new patent applications, the average number of patents per project goes from 3.05 for IPs, to 2.07 for NoEs and 2.16 for STRePs. Once again, these productivity differences arise at least partly from the different amount of resources allocated to projects. In particular, once the number of participants and the amount of financial resources are taken into account, no significant difference remains between IPs and STRePs either in the propensity to patent or in the number of new patent applications. At the same time, results confirm that NoEs tend to exhibit a higher propensity to patent than IPs and STRePs.

- These results are to some extent surprising given the focus of NoEs on scientific research activities and the low participation of industry in this type of projects. An analysis of the profile of patent applicants, however, reveals that while companies account for the bulk of patents in the case of IPs and STRePs, universities and public research organisations play a major role in the case of NoEs with 50% of all patents applications. In addition to this, the fraction of jointly owned patents in the case of NoEs is around twice as large as the one observed for IPs and STRePs. These findings seem to suggest that a higher degree of research integration and a lower level of conflicts among partners around IPR issues have emerged in the case of NoEs. Testing this hypothesis represents, in our view, an interesting avenue for further research.

Assessment of quality

- The assessment of the *quality* of scientific articles produced by DG INFSO projects has been based upon an analysis of the number of citations received in the scientific literature. The quality of the articles resulting from European projects has been benchmarked by comparing it with the quality of the population of all other articles published in the same journals and scientific areas. Specifically, we have examined the three scientific fields more relevant for the DG INFSO projects, i.e. computer science, materials science, and optics.

- Concerning the *scientific impact* of articles produced by DG INFSO projects, results show that the probability they receive forward citations from other articles is generally no lower and often significantly higher than the average article published in the same journal and in the same scientific field. More specifically, the fraction of uncited articles produced by DG INFSO projects is significantly lower than the corresponding fraction in the benchmarking population.

- In addition to this, the mean observed citation rate (MOCR) is systematically larger than the mean expected citation rate (MECR), thereby suggesting that DG INFSO articles are not only more likely to receive at least one citation, but that on average the number of citations received is larger than for the average article in the same journal and scientific field. In particular, DG INFSO articles in the field of optics received on average 1.3 times more citations per article than the average article in this field. Similar results hold for the scientific field materials science, but not in the field of computer science. In this latter case, the omission of conference papers (a major vehicle of knowledge dissemination in this field) is likely to be responsible for the result.

- Overall, the results confirm that DG INFSO projects have been highly effective in attracting top quality researchers and research teams from the research fields relevant for the ICT area.

Assessment of research collaboration

- The propensity of DG INFSO projects to engage in international and inter-regional scientific research collaboration has been tested by examining the patterns of co-authorship and by benchmarking it with the pattern emerging in the population of all European authored articles in the same journals and scientific areas.

- The propensity to engage in international research collaboration is significantly larger for the authors of articles produced in the context of DG INFSO projects than for the average European article in the same scientific fields. In particular, this study finds that DG INFSO articles are more likely to involve *intra*-European collaboration than the average European article. The share of DG INFSO articles involving at least two affiliations from different ERA countries is around twice as large as in the benchmarking population. This result suggests that the FPs have been effective in reinforcing the extent of intra-European collaboration in research. At the same time, this study finds that DG INFSO articles are also characterised by a lower propensity to engage in *extra*-European collaboration compared to other European articles in the same fields.

- The regional (NUTS2 and NUTS3 level) concentration of the scientific output produced by DG INFSO projects looks significantly larger than the spatial level of concentration in the benchmarking population of articles. The number of European regions with a positive number of articles represented within DG INFSO projects is generally much lower than the number of European regions capable of a scientific production in the benchmarking population of articles. As a consequence, the spatial concentration of the scientific output (Herfindahl index) in the case of DG INFSO projects is around twice as large as in the benchmarking population of articles for most of the scientific fields examined here. Despite the larger concentration of scientific output, however, the spatial patterns of scientific activity look rather similar. A linear and a rank correlation analysis of the number of scientific articles produced by NUTS2 European regions within DG INFSO projects and in the benchmarking population confirms the existence of a large (though far from perfect) and positive correlation between the two variables.

- The majority of patent applications produced by DG INFSO projects have been extended to the European Patent Office (EPO). Focusing on this subset of patent applications, the most important technology areas, covering 90% of all EPO patents reported by DG INFSO projects, are electrical engineering (which comprises telecommunications, computer, audiovisual technology and semiconductors) and scientific instruments (which comprise medical technology, control technology and optics). The analysis of research collaboration has been carried out by benchmarking patents produced by DG INFSO projects with the population of other European invented patents in these two technological fields.

- Analysis by type of applicant shows that the share of DG INFSO patents applied by universities and public research organisations (PRO) significantly exceeds the share observed in the benchmarking population of patents in the same technological fields. This suggests that the propensity to patent and/or to engage in research leading to patented inventions by universities and PROs is significantly larger in the context of EU funded projects than in the corresponding population of European institutions.

- Analysis of co-patents, i.e. patents jointly applied by two or more independent organisations, also reveals that the extent of joint patent ownership in the context of DG INFSO projects is significantly higher than in the benchmarking population of patents. This finding probably signals the need for organisations participating in EU funded projects to devise schemes for sharing the intellectual property rights over the results of joint research. Moreover, even though company-company ownership is the most prevalent type of joint ownership, a significant amount of co-patents involves PROs and universities.

- The extent of spatial concentration of patenting activity is comparable to the one observed for the scientific publication activity. The top ten NUTS3 regions in terms of patents account for about 33% of all EPO patent applications produced by DG INFSO projects. This finding can be probably interpreted as evidence of complementarity between patenting and publishing activities.

- Finally, an analysis at the level of individual researchers show that around 27% of all inventors of EPO patents applied by DG INFSO projects are also authors of one or more scientific articles reported by the same projects. These individuals that we can label as author-inventors play a crucial role as social connectors between the domain of open science and the world of private technology. Our analysis also shows that around 14% of such individuals have a corporate affiliation. Even though benchmarking these figures (i.e. assessing the extent to which the observed values are larger or lower than expected) is extremely difficult, we can quite confidently conclude that the degree of interplay between scientific research and technological development achieved in the context of DG INFSO projects looks quite remarkable.

Limitations of the study

It is important to point out that the results reported in this study have to be interpreted with some caution due to the limitations of the empirical methodology adopted.

- The study has explored only one thematic area, i.e. ICT, of the FP and the results can be hardly generalised to other fields.

- The data on scientific and technological output of EU funded projects come directly from project co-ordinators, they are hardly verifiable in an independent way and therefore are subject to possible reporting bias, even though *a priori* it is difficult to say whether the bias should go in the direction of over-estimating or under-estimating the true output. For example, we noted that several projects reported patents whose priority date is prior with respect to the starting date of the project. We argued that this phenomenon can be probably related to the IPR rules that oblige project participants to grant access rights to their knowledge (or pre-existing know-how) to another participant if the latter needs such access rights in order to carry out its own work under the project or to use its own knowledge resulting from the project. However, this is no more than a conjecture that needs to be verified. More generally, we believe that testing the accuracy of what reported by project coordinators is imperative in order to derive robust conclusions from studies like this one.

- The assessment of scientific performance of DG INFSO funded projects has been mostly based on articles published in mainstream scientific journals indexed in the ISI-WoS SCI and SSCI databases. As repeatedly argued in this report, a more correct assessment of the scientific productivity of research projects in the ICT area should take into account that conference proceedings and papers play a particularly important role as a vehicle of knowledge dissemination in this field.

- Even though this report has used state of the art bibliometric and scientometric techniques, one should be cautious before drawing too simple conclusions and policy implications based on the results reported here. For example, one of the conclusions of this study is that the scientific and technological performance differ across projects in terms of funding instruments. In particular, Networks of Excellence outperform IP and STRePs both in terms of production of scientific articles and propensity to introduce new patent applications. Yet, concluding from these results that the superior performance of NoE has to be imputed to the funding instrument would ignore the possibility that the best scientists and research teams are selected into these projects and therefore generate a higher number of articles and are more likely to produced patented inventions. Testing the existence of these potential endogeneity problems is necessary before being able to draw robust conclusions.

- Finally, one should point out that, no matter how important, scientific articles and patents represent only one of the several and multifaceted socio-economic benefits that are expected from the FPs.

Avenues for further research

The limitations of the study briefly discussed above suggest as many potential avenues for further research.

- First, we believe that broadening the set of publications considered, in particular taking into account conference proceedings, is essential in order to get a correct assessment of the overall scientific productivity of research projects. More generally, we would recommend extending the bibliometric analysis to other sources of bibliographic information, such as the *Scopus* database, produced by Elsevier, and the *Google Scholar* database.

- Along similar lines, for future evaluations it would be useful to complement data on articles and patents with other indicators of performance, such the formation of start-up companies, the mobility and career promoting effect of FP participation and so on. Yet, two problems have to be solved before being able to use effectively such indicators. First, the data collection process should be improved to ensure that comparable, standardised and consistent data are collected. Second, the data should be as far as possible verifiable in an independent way in order to avoid any possible reporting bias.

- Despite the importance of collecting other indicators of performance, we believe that further quantitative analyses of data on publications and patents are needed in order to better understand the organization of research activities in the funded projects.

- Analysis of forward citations received by patents generated by the projects might help assessing the importance of inventions resulting from the funded research, whereas analysis of backward citations could be used to assess the inter-sectoral and international flows of knowledge, e.g. the countries of origin of knowledge "used" by patents. Similarly, further analyses of patents should include an analysis of the so-called "non-patent references" (i.e. scientific publications and other published material) reported on the front pages of the patents. The analysis could help gauge the extent to which European scientific research proves to be relevant in the development of ICT patents resulting from DG INFSO patents.

- Although a key result of this study is that articles resulting from projects have a higher than average scientific impact, the interpretation of these results is not entirely clear due to endogeneity problems. Is the production of better articles an outcome of project participation or is it simply the consequence of the fact that the best scientists and research teams participate in FP projects? In a similar way, the study finds that the scientific and technological productivity markedly differ across projects funded through different instruments. In particular, NoEs tend to outperform IPs and STRePs in terms of number of articles produced and propensity to generate new patent applications. Yet, concluding from these results that the superior performance of NoEs has to be imputed to the funding instrument would ignore the possibility that the best scientists and research teams are selected into these projects and therefore generate a higher number of articles and are more likely to produce patented inventions. Solving issues like these requires further efforts of data collection and econometric analysis.

- Finally, we would recommend that subsequent evaluation studies include a validation stage of quantitative results. In particular, a key result of this study is that the scientific and technological productivity markedly differs across projects and funding instruments. Understanding the reasons for such differences calls for the use of qualitative and case-based studies, where selected research projects- both highly successful projects and less successful ones according to their output performance- are sampled for additional data collection. This more qualitative study should assess the impact of project-specific and R&D environmental factors that have impinged on each project's performance.

I. Introduction

This document is the Final Assessment Report "Using Performance Indicators in Monitoring the Implementation of ICT Research in FP6 and FP7". The report provides the results and findings of a pilot study aimed to assess the amount and quality of scientific and technological output produced by EU funded projects in the ICT area. In particular, the report illustrates the findings that emerge from the bibliometric analysis of the scientific and technological research output produced by projects funded by the Information Society and Media DG of the European Commission (from now on DG INFSO).

The report is organised into two main parts. The first part deals with the bibliometric analysis of the scientific publications produced by DG INFSO funded projects. After a short description of the database used in section II.1, section II.2 provides a thorough analysis of the scientific productivity by project and FP, by funding instrument and by strategic objective of projects. In this context, particular attention is paid to the role played by the new funding instruments introduced in the FP6. This is followed in sections II.3 and II.4 by an analysis of the publication activity by scientific field. Two major types of analyses will be carried out in this context. First, we will examine what is the average impact, in terms of citations received, of scientific publications produced by EU funded projects. Second, we will analyse to what extent scientific publications produced by EU funded projects involve authors from different countries and regions. In both cases, a benchmarking analysis will be carried out to assess the extent to which the patterns observed for the sample of EU projects deviates from the general pattern observed for the population of European articles. The second part of this report (section III) regards the analysis of the patenting activity of DG INFSO funded projects.

The analysis carried out in this study is based upon a data set of scientific articles and patents resulting from DG INFSO funded collaborative projects. Appendix 1 to this report provides a technical and detailed description of the methodology adopted to collect and process the data in order to construct the bibliometric database.

II. Bibliometric analysis

II.1 Database description

The bibliometric analysis presented in this report is based upon a data set of scientific articles produced by a sample of EU projects funded by the DG INFSO. Raw data on the scientific output of these projects have been collected through a wave of monitoring questionnaires sent out by the DG INFSO itself in the years 2005, 2006, 2007 and 2008 to all running projects.

As far as this study is concerned, the raw data on the scientific output of projects have been provided by the DG INFSO in two batches. The first batch has been delivered at the beginning of the study in January 2009 and it refers to data collected through the questionnaires sent out in the years 2005, 2006 and 2007. The second batch has been delivered at the beginning of January 2010 and it refers to data collected through the questionnaire sent out in the year 2008.

The raw data have been preliminarily processed in order to clean and standardise the information provided by project coordinators. The details of this process are fully discussed in the technical appendix to this report. Since the scientific output reported by project coordinators contains very heterogeneous items, ranging from scientific articles published in peer-reviewed journals to talks given to invited seminars, press releases and websites, a specific procedure has been adopted to *separate* scientific output amenable to bibliometric analysis from other items. The procedure has been based upon the use of *Zotero*, an open source extension of the Firefox browser that is able to detect books and articles on the web, to store the reference and to manage it through reference management tools, such as Bibtex (for more details see http://www.zotero.org). The principle behind *Zotero* is simple. On many research websites, such as the websites of major academic publishers, digital libraries, Pubmed, Google Scholar and so on, *Zotero* is able to detect when an article or a book is being viewed. The information concerning the article or book, i.e. author names, title of the article, title of the journal, publication year etc. can then be saved through a mouse click to a local database for use with reference management software.

All items identified through *Zotero* have been matched to the Thomson-ISI Web of Science (WoS) database, which provides information from more than 5,000 peer-reviewed international journals covering a wide spectrum of scientific disciplines (or subject categories). Fuller details on the process of identification and matching of scientific articles, as well as data quality assessments are discussed in the technical appendix. Here, it is important to note a few points:

i) The process of cleaning and standardising the documents provided by project coordinators through the use of *Zotero* has been performed both for the first and the second batch of data received from the DG INFSO (see above). As far as the matching of the identified items with Thomson-ISI is concerned, however, we proceeded as follows. Articles identified in the first batch were matched to the original source article in Thomson-ISI in order to retrieve full bibliographical information, i.e. name of authors, affiliations, citations received and so on. Articles identified in the second batch were matched to Thomson-ISI only by journal name, in order to separate articles published in journals indexed in Thomson-ISI from articles published in other sources. In other words, for articles contained in the second batch of data (and not already reported in the first batch), we are able to identify those articles that are published in journals indexed in the Thomson-ISI Web of Science (WoS) database, but we have no bibliographical information on aspects such as the number of authors, the affiliation of these authors, the impact of the article in terms of citations received and so on. Since the benchmarking analysis carried out in sections II.3 and II.4 crucially depends upon this information, this type of analysis will be carried out only for the first batch of scientific articles.

ii) The Thomson-ISI WoS database actually consists of several different databases, the most important of which are the Science Citation Index Expanded (SCI-Expanded), the Social Sciences Citation Index (SSCI), and the Conference & Proceedings Citation Index Science (CPCI-S). While the SCI and the SSCI contain information on articles published in peer-reviewed scientific journals that meet specific criteria established by Thomson-ISI (and thereby exclude other journals that do not qualify for inclusion), the CPCI-S provides information on published literature from the most significant conferences, symposia, seminars, colloquia, workshops worldwide.

iii) For the purposes of the present study, scientific output reported by project coordinators has been matched only to the SCI and SSCI databases, and not to the CPCI-S database. As a significant portion of the scientific output produced by projects in the ICT field is likely to finds its way to publication as conference proceedings papers and articles, *this represents a major limitation of the present study*. There are two major reasons explaining this choice. First, the CPCI-S has become available as part of the ISI-WoS only at the end of October 2008 and most libraries do not yet include CPCI-S in their subscription to the ISI-WoS. However, the second and most important reason, as we discuss in detail in the technical appendix, is related to the very poor quality of the raw data on conference papers provided by project coordinators. For example, in many cases only the acronym of the conference is used or the year of the conference is missing and so on. In these circumstances, even the adoption of sophisticated matching algorithms would still require a significant amount of manual work to ensure the quality of data. Given the very large number of conference papers to clean and match, this work would have far exceeded the amount of resources devoted to this study. Nonetheless, we firmly believe that *including conference papers in the bibliometric analysis could represent a natural and fruitful extension of this study*. At the same time, our experience in carrying out this work also suggests that any future evaluation study of FP projects based on bibliometric indicators would strongly benefit from *improving the methodology to collect scientific output data from project coordinators*. Among other things, this would allow to shifting resources devoted to evaluation studies away from the trivial, but very time consuming tasks of standardising and cleaning information to the more productive tasks of data analysis. In the technical appendix, *we elaborate a few concrete proposals on how the data collection procedures could be improved*.

Table II.1 provides a few summary statistics of the bibliometric database of scientific output produced by DG INFSO projects. Overall, the database contains information on 1417 projects. The vast majority of projects in the database (927 or 65%) have been funded under the 6th Framework Programme. The 1417 projects have been responsible for 6786 scientific articles published, 4359 (64%) collected in the first batch and 2427 (36%) collected in the second batch[1].

Table II.1 – Summary statistics of the bibliometric database

Number of DG INFSO projects	1417 (100%)
of which funded under	
FP5	199 (14.0%)
FP6	927 (65.4%)
FP7	291 (20.6%)
Number of scientific articles – ISI (WoS) [a)]	6786
of which	
1st batch (collected in period 2005-07)	4359
2nd batch (collected in period 2008-09)	2427
Other documents	43914

[a)] Articles published in journals indexed in ISI-WoS (SCI-Expanded and SSCI)

In addition to this, the 1417 project coordinators have also reported other types of scientific output, which have been classified according to the type of document. In par-

[1] The 4359 articles collected in the first batch have been published in 807 distinct journals indexed in the ISI-WoS (SCI and SSCI) database. It must be pointed out that the 4359 documents collected in the first batch are not all original research articles. The frequency distribution according to the field DT (document types) includes 3389 articles, 830 proceedings papers, 75 reviews, 41 editorial materials, 20 letters, 2 meeting abstracts, and 2 corrections. It is important to note that the editorials, letters, news items, and meeting abstracts are usually not included in *article counts* in the Journal of Citation Reports (JCR) because they are not generally cited. The number of articles given for journals listed in JCR includes primarily original research (i.e. articles and proceedings papers) and review articles. At the same time, it has to be also noted that *citation counts* in JCR do not distinguish between citations to letters, reviews, or original research articles, even though only original research and review articles are used in the denominator of the impact factor calculation. Unfortunately, we are unable to provide similar statistics for the second batch of articles collected, even though it is quite reasonable to assume that the distribution in terms of article types is quite similar to the one observed for the first batch. In what follows, we will examine the whole set of 6786 documents reported by DG INFSO projects, without making a distinction between original research articles, letters or editorial materials. Given that almost 99% of all items reported by projects and collected in the first batch of data are either articles or reviews, we believe that this choice should have no impact on results. For more details on how the number of journal articles and the impact factors are calculated in the JCR see http://admin.isiknowledge.com/JCR/help/h_using.htm.

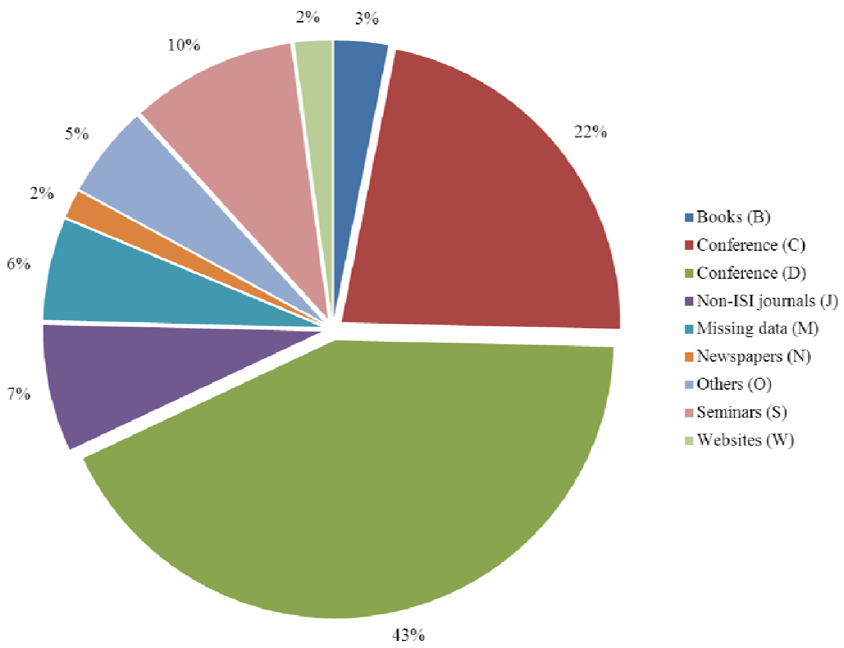
ticular, we have attempted to classify all documents not matched with the ISI-WoS (SCI and SSCI) database into nine distinct categories[2]:

a) Books, book series, book chapters (B)
b) Scientific articles published in journals not indexed by ISI-WoS (K)
c) Articles published in proceedings of conferences (indexed through *Zotero*) (C)
d) Articles presented at conferences (not indexed through *Zotero*) (D)
e) Websites of projects, articles in websites (W)
f) Invited talks, seminars, workshops not published (S)
g) Newspapers, press releases, tv news (N)
h) Other documents, e.g. MSc and PhD theses, posters, etc. (O)
i) Missing data, i.e. not enough information to classify item (M)

Overall, the 1417 projects have reported 43914 items that do not match with the ISI-WoS (SCI and SSCI). The distribution of these documents by category type is illustrated in Figure II.1. The vast majority of them is represented by papers and articles presented to conferences. They account for around 65% of all other documents produced by DG INFSO projects. For around one third of them (9772 documents corresponding to category C) we have been able to find information through Zotero, i.e. Zotero found a website of a publisher or of a library indexing the article and thus reporting information on the authors, title of the article, title of the conference, publication year and so on. It is thus likely that at least a fraction of such conference articles could be successfully matched with the ISI-WoS CPCI-S database. For the other conference papers (18717 articles corresponding to category D), we have not implemented any check using Zotero. However, also in this case, we cannot exclude that a certain fraction of these documents could be found in conference and proceedings indexed by ISI-WoS CPCI-S database. Given the weight of conference papers in the overall scientific output reported by coordinators of DG INFSO projects, we believe that extending the bibliometric analysis to such documents could represent a fruitful avenue for further research in order to confirm and generalise the results found in this study.

[2] Once again, the technical appendix provides a more detailed discussion of how the classification of other documents has been performed.

Figure II.1 – Distribution of other documents by category



The analysis in the context of the present study will focus on the subset of scientific output represented by the 6786 scientific articles published in scientific journals indexed in the ISI-WoS (SCI and SSCI). Despite a few limitations widely recognised in the literature (e.g. the poor coverage of social sciences, law and humanities, and the bias in favour of English-language journals), there is also some consensus that articles reported in the ISI-WoS SCI and SSCI databases provide a satisfactory representation of high-quality, internationally accepted basic research. Moreover, these scientific publications are also likely to represent a more homogenous (and thus *countable*) set of documents, as compared to conference proceedings, books and book chapters. Finally, for these articles we have access to the full bibliographic information (i.e. publication year, affiliations of authors, number of citations received, title of journal) that allows us to carry out a thorough bibliometric assessment of the performance of EU funded projects and to benchmark it with respect to the general pattern for the whole population of European articles[3].

[3] The specific advantages and limitations of different bibliometric data sources are widely acknowledged and discussed in the literature. The advent of new databases and tools, such as Scopus and Google Scholar, has raised question about the use of ISI-WoS as the only or even the main source of information (Meho and Yang, 2007). Similarly, the potential pitfalls of bibliometric indicators as measures of knowledge creation and diffusion have been widely discussed in the literature. Providing an overview of

However, given that other types of documents, most notably conference papers, books and articles published in other scientific journals, account for the bulk of the scientific output reported by project coordinators, we also use the available information on these documents to assess some aspects of the performance of DG INFSO projects. In particular, in the next section, we provide a broad evaluation of the scientific productivity by project, funding instrument and strategic objective using information both on scientific articles and on other types of scientific documents.

II.2 An assessment of the scientific productivity

II.2.1 The issue of double counting

In this section, we aim to provide an assessment of the scientific productivity by FP, project, funding instrument and strategic objective. To this purpose, however, we have to deal preliminarily with the issue of double-counting. The issue arises from the fact that the same scientific article may have been reported as output by different projects[4]. Table II.1 reports the distribution of articles in the data set according to the number of different projects in which the same article has been listed as an output. The vast of majority of articles have been reported in only one project. However, there are 402 articles, around 6% of all articles that have been reported by more than one project. In particular, there are 225 articles reported by two projects, 13 articles reported by three projects, 162 articles reported by twelve different projects and 2 articles reported by thirteen different projects.

---

this literature goes beyond the purpose of this study. A useful summary of the main methodological issues may be found in the appendix to the Second European Report on *Science and Technology Indicators* (http://www.ucm.es/BUCM/be/docs/i+d/REPORT/Appendix.pdf) and in the OECD *Report Bibliometric Indicators and Analysis of Research Systems: Methods and Examples*, also downloadable at http://www.olis.oecd.org/olis/1997doc.nsf/LinkTo/NT00000902/$FILE/05E79150.PDF.

[4] Double-counted articles may represent the output of different projects as long as there are synergies among those projects. However, they could also signal a problem of misreporting output by project coordinators, either because they have incentive to over-report their productivity or because they are unable to allocate articles to specific projects. A possible way of approaching this problem would be that of using the information contained in the fields FU and FX that ISI-WoS provides for each article. The two fields contain information, respectively, on funding acknowledgements and other types of acknowledgement. As long as articles reported by project coordinators are the output of a specific EU project, those two fields should contain reference to it. We performed this check for the articles in the first batch. Unfortunately, only 53 of the 4359 scientific articles report a non-missing value in the fields FU and FX. However, in all such cases acknowledgement of EU funding is reported.

Table II.2 – Distribution of double-counted articles by number of projects

Number of projects	Number of articles	%	Number of articles	%
	1st batch		1st and 2nd batch	
Only one project	4094	93.9	6384	94.1
Two projects	98	2.2	225	3.3
Three projects	3	0.1	13	0.2
Twelve projects	162	3.7	162	2.4
Thirteen projects	2	0.1	2	0.0
Total	4359	100	6786	100

In a similar way, Table II.3 reports the distribution of articles according to the number of different Framework Programmes in which the same article has been reported as an output. One can observe that the vast majority of articles (about 97% of the total) have been reported either as FP5, FP6 or FP7 output. However, there are 43 articles that have been reported both in FP5 and FP6 projects, 24 articles that have been reported both in FP6 and FP7 projects, and 164 articles that have been listed in FP5, FP6 and FP7 projects. The presence of this phenomenon needs to be taken into account when counting the number of articles by Framework Programmes and projects. A possible approach to this issue is to avoid any double counting by taking each article only once. Table II.4 reports the distribution of articles without double-counting. If the same article has been reported in projects belonging to two different FPs, the article is attributed to the earlier of them, e.g. an article reported in two projects, one of which in the FP5 and the other in the FP6, is counted as an article produced within the FP5. Looking at the table we observe that the FP6 projects account for about 81% of all scientific articles in the sample. This fraction is higher than the share of all projects: FP6 projects account for around 65% of all DG INFSO projects for which valid questionnaires were collected. The larger fraction of articles compared to the share of projects observed for FP6 is largely due to the fact FP7 projects that represent around 20% of all projects for which valid questionnaires have been collected have started only recently and therefore have had less time to produce published articles compared to FP6 projects. Table II.5 reports the same distribution of articles with double counting, i.e. the same article is counted as many times as the number of FPs in which it has been reported as output. Please note that the total number of articles now exceeds the number of articles in the database as a result of the double counting.

Table II.3 – Distribution of double-counted articles by Framework Programme

Framework Programmes	Number of articles	%	Number of articles	%
	1st batch		1st and 2nd batch	
Only FP5	627	14.4	636	9.4
Only FP6	3541	81.2	5478	80.8
Only FP7	0	0	438	6.5
FP 5 and FP6	27	0.6	43	0.6
FP5 and FP7	0	0	0	0.0
FP6 and FP7	0	0	24	0.3
FP5 and FP6 and FP7	164	3.8	164	2.4
Total	4359	100.0	6783	100.0

Note: for three articles in the 2nd batch no information on the project was available.

Table II.4 – Distribution of ISI-WoS publications by FP (no double counting)

Framework Programmes	Number of articles	%	Number of articles	%
	1st batch		1st and 2nd batch	
FP5	818	18.7	843	12.4
FP6	3541	81.2	5502	81.1
FP7	0	0.0	438	6.5
Total	4359	100.0	6783	100.0

Note: for three articles in the 2nd batch no information on the project was available.

Table II.5 – Distribution of ISI-WoS publications by FP (double counting)

Framework Programmes	Number of articles	%	Number of articles	%
	1st batch		1st and 2nd batch	
FP5	818	17.4	843	11.8
FP6	3717	79.1	5709	79.5
FP7	164	3.5	626	8.7
Total	4699	100.0	7178	100.0

Note: for three articles in the 2nd batch all information on the project were missing.

Given that FP6 projects represent the bulk of projects in the database and account for the largest fraction of all articles, in the rest of this section we will focus our attention on the subset of projects funded under the FP6. Even focusing within the same FP, we need to adopt some rules to count articles by project. Unless otherwise stated, we will follow the rule to attribute an article to a project, even if the same article has been reported by another project within the same FP. The rationale for adopting this criterion is that we cannot exclude the existence of synergies across projects and thus we cannot rule out that the same article is the joint outcome of different projects. At the same time, we will also adopt the rule that if an article has been reported in projects belonging to two different FPs, the article is attributed to the earlier of them. The rationale for this rule is that the same article cannot be assumed to be the outcome of projects that have been funded under different FPs.

II.2.2 Scientific productivity by project and Framework Programme

Our assessment of the scientific productivity of DG INFSO projects starts from a simple analysis of the distribution of scientific output by project and Framework Programme. Table II.6 reports the distribution of FP5, FP6 and FP7 projects according to the number of ISI-WoS (SCI and SSCI) scientific articles they have reported as output. The fraction of projects that have reported no article is fairly different among the FPs: it goes from around 50% for the FP6, to 59% for the FP5 and 68% for the FP7. Similarly, the average number of articles per project goes from 6.1 for the FP6, to 4.3 for the FP5 and 1.4 for the FP7. In the case of FP5, however, one must observe that only a few projects are included in the sample, whereas for the FP7 it should be noted that they have started only recently and therefore they had less time to produced published articles than FP6 projects. We believe that this represents a further reason to focus the analysis on FP6 projects.

Overall, the picture emerging from the data is one suggesting a rather skewed distribution of the scientific productivity, with very few projects producing a relatively large number of scientific articles, and with the majority of projects with few or no articles. Such a conclusion, however, would be misleading as a large fraction of the scientific output produced by DG INFSO projects is represented by documents that are published or made publicly available through other means than peer-reviewed scientific journals. In particular, articles published in the proceedings of conferences and articles published as book chapters or specialised journals not indexed by ISI-WoS seem to constitute very important outlets for the research carried out by EU projects in this field.

Table II.6 – Number of scientific articles per project, FP5, FP6 and FP7

Nr of articles[a)] per project	FP5	%	FP6	%	FP7	%
0	117	58.8	460	49.6	197	67.7
1	22	11.1	102	11.0	37	12.7
2	11	5.5	64	6.9	15	5.2
3	9	4.5	38	4.1	11	3.8
4	4	2.0	28	3.0	7	2.4
5	11	5.5	35	3.8	7	2.4
6	1	0.5	15	1.6	2	0.7
7	1	0.5	18	1.9	2	0.7
8	3	1.5	16	1.7	0	0.0
9	1	0.5	12	1.3	1	0.3
10	3	1.5	14	1.5	2	0.7
>10	16	8.0	125	13.5	10	3.4
Total	199	100.0	927	100.0	291	100.0
Mean	4.3		6.1		1.4	
Std dev	16.8		17.2		4.8	
Min	0		0		0	
Max	204		189		43	

a) Scientific articles published in journals indexed in ISI-WoS (SCI and SSCI).

To take into account this crucial aspect and draw a more correct assessment of the overall scientific productivity of DG INFSO projects, we have included in our analysis also those articles that have been published through other means than ISI-WoS (SCI and SSCI) journals. In particular, from the whole set of other documents (see above Table II.1 and Figure II.1) we have extracted the subset of documents corresponding to conference papers, books and book chapters and articles published in non-ISI journals and for this subset we have computed the distribution of projects by number of documents produced. Results are reported in Table II.7. The picture emerging is in many respects quite different. First, the shares of projects that do not report any output drop to 18% for the FP6, i.e. 82% of all projects have produced at least one paper, to 36% for the FP5 and to 29% for the FP7. Second, the distribution looks also quite skewed but towards the right-tail of it. Around 57% of all FP6 projects have in fact reported ten or more documents and the average number of documents per project is 31.2. However, the table still confirms the higher productivity of FP6 compared to FP5 and FP7 projects.

Table II.7 – Number of other papers per project, FP5, FP6 and FP7

Nr of papers per project	FP5	%	FP6	%	FP7	%
0	72	36.2	165	17.8	84	28.9
1	8	4.0	27	2.9	36	12.4
2	5	2.5	26	2.8	26	8.9
3	9	4.5	28	3.0	19	6.5
4	6	3.0	18	1.9	14	4.8
5	8	4.0	27	2.9	14	4.8
6	1	0.5	23	2.5	14	4.8
7	5	2.5	21	2.3	10	3.4
8	7	3.5	15	1.6	6	2.1
9	5	2.5	31	3.3	3	1.0
10	2	1.0	18	1.9	6	2.1
>10	71	35.7	528	57.0	59	20.3
Total	199	100.0	927	100.0	291	100.0
Mean	12.6		31.2		7.5	
Std dev	24.6		57.7		18.6	
Min	0		0		0	
Max	244		728		202	

Note: other papers include papers published in proceedings of conferences, books and book chapters and articles published in scientific journals not indexed in the ISI-WoS.

In comparing the results reported above in Table II.6 and Table II.7 one should bear in mind a few key differences between scientific articles and other documents. Articles indexed in the ISI-WoS (SCI and SSCI) database are published in peer-reviewed scientific journals. Therefore, they have to pass through a selection process and have to meet with high standards of originality[5]. Given the length of the selection process in peer-reviewed journals, the publication of conference proceedings has become at least in recent years a suitable alternative either to speed up the dissemination of knowledge before results are published in a peer-reviewed journal or even as the final outlet for research re-

[5] Please note that conference proceedings are sometimes published as guest-edited issues in peer-reviewed scientific journals, such as the ones indexed in ISI-WoS (SCI and SSCI).

sults. Conference proceedings may perform different functions, such as gathering feedbacks before submission to a regular journal, disseminating knowledge and stimulating discussion within a field, and transmitting information that would otherwise be difficult for various reasons to include in a scientific article (Drott 1999). It is therefore clear that including conference proceedings and similar papers into a bibliometric assessment, such as the one carried out in this study, will produce a more complete picture of a given discipline's scientific production (Glänzel et al. 2006; Butler e Visser 2006). This is particularly true for engineering and computer sciences, i.e. the scientific fields more relevant for the projects funded by DG INFSO, where proceedings have gained even more prominence than articles in the diffusion of knowledge. While these arguments would suggest the need to extend the coverage of bibliometric assessment to conference proceedings, it is also necessary to point out that the different selection process characterising articles and conference papers is likely to have an impact upon their relative *scientific importance*. In a recent work, Lisée et al. (2008) examine the impact of conference papers by computing the fraction of all citations contained in the ISI-WoS SCI that are directed to papers published in proceedings. Their results are quite striking: only 1.7% of all references made in the natural sciences and engineering refer to conference proceedings and the share is declining over time. A partial exception to this pattern is represented by Computer sciences and Engineering where the share of all references made to proceedings is, respectively, equal to 20% and 10%. In addition to this, they also show that the average age of cited proceedings is significantly lower than the average age of cited literature, which suggests that papers in conference proceedings tend to become obsolete at a faster rate than the scientific literature in general (Lisée, Larivière, and Archambault 2008).

On the basis of the discussion above, as the bibliometric assessment carried out in this study is mostly based upon scientific articles produced by DG INFSO projects, any findings emerging from it must be interpreted as referring to a subset of the whole scientific production of EU funded projects. At the same time, the subset examined is likely to represent the most important portion of such production, at least in terms of scientific impact[6]. The inclusion of conference proceedings papers in the bibliometric

[6] Please note that scientific articles are also likely to represent a more homogenous (and thus *countable*) set of documents, as compared to conference proceedings, books and book chapters.

analysis would allow one to achieve a more complete and exhaustive assessment of the scientific *production* of DG INFSO projects. However, as already argued above, this objective requires either to allocate more resources to clean and standardise the information provided by project coordinators or improving the data collection process. With these caveats in mind, in the next section we turn to examining the extent of scientific production by funding instrument.

II.2.3 Scientific productivity by funding instrument

In order to assess whether there are differences in scientific productivity according to the instrument used to fund research, all projects in the database have been classified according to this dimension. Since funding instruments have been radically reformed passing from the 5^{th} to the 6^{th} Framework Programmes, and given that FP6 projects represent around 65% of all projects and 81% of all articles in the database, our analysis is limited to this subset of projects. Table II.8 reports the distribution of all scientific articles produced by FP6 projects according to the funding instrument, while Table II.9 reports the average number of articles by funding instrument as well as the share of projects funded with a specific instrument that reported at least one article. The evidence shows that the share of articles exceeds the share of funded projects for three instruments: networks of excellence (NoE), integrated projects (IP) and integrated infrastructure initiatives (I3). NoEs stand out in this respect: for this funding instrument, the average number of articles per project and the share of projects with a non-null scientific production are significantly larger than the average value of these variables for all instruments (see Table II.9). Specific targeted research projects (STREP) represent an interesting case. Whereas they account for a low share of all articles (lower than their share on all projects) and they produce a lower than average number of articles per project, the share of projects with a non-null scientific production is relatively large, thereby signalling a lower dispersion across projects in terms of scientific capabilities compared to other types of instrument.

Table II.8– Distribution of scientific articles by funding instrument, FP6

Funding instrument	Number of DG INFSO projects	%	Number of scientific articles	%
Co-ordination actions (CA)	47	5.1	110	1.9
Integrated infrastructure initiatives (I3)	8	0.9	152	2.7
Integrated projects (IP)	189	20.4	1637	28.8
Networks of excellence (NoE)	49	5.3	1455	25.6
Specific support actions (SSA)	137	14.8	41	0.7
Specific targeted research projects (STREP)	497	53.6	2286	40.2
Total	927	100.0	5681	100.0

Scientific articles published in journals indexed in ISI-WoS (SCI and SSCI).

Table II.9– Average number of scientific articles per project by funding instrument, FP6

Funding instrument	Mean	Std dev	Min	Max	Projects with articles	% of all projects
Co-ordination actions (CA)	2.3	8.6	0	53	8	17.0
Integrated infrastructure initiatives (I3)	19.0	49.7	0	142	4	50.0
Integrated projects (IP)	8.7	19.2	0	151	122	64.6
Networks of excellence (NoE)	29.7	44.4	0	189	37	75.5
Specific support actions (SSA)	0.3	1.1	0	8	14	10.2
Specific targeted research projects (STREP)	4.6	10.2	0	110	282	56.7
All instruments	6.1	17.2	0	189	467	50.3

Scientific articles published in journals indexed in ISI-WoS (SCI and SSCI).

Table II.10– Distribution of other papers by funding instrument, FP6

Funding instrument	Number of DG INFSO projects	%	Number of papers	%
Co-ordination actions (CA)	47	5.1	703	2.4
Integrated infrastructure initiatives (I3)	8	0.9	281	1.0
Integrated projects (IP)	189	20.4	10250	35.4
Networks of excellence (NoE)	49	5.3	6224	21.5
Specific support actions (SSA)	137	14.8	413	1.4
Specific targeted research projects (STREP)	497	53.6	11062	38.2
Total	927	100.0	28933	100.0

Other papers include papers published in proceedings of conferences, books and book chapters and articles published in scientific journals not indexed in ISI-WoS (SCI and SSCI).

Table II.11 – Average number of other papers per project by funding instrument, FP6

Funding instrument	Mean	Std dev	Min	Max	Projects with papers	% of all projects
Co-ordination actions (CA)	15.0	28.7	0	130	32	68.1
Integrated infrastructure initiatives (I3)	35.1	48.5	0	141	7	87.5
Integrated projects (IP)	54.2	66.7	0	404	174	92.1
Networks of excellence (NoE)	127.0	155.6	0	728	38	77.6
Specific support actions (SSA)	3.0	5.5	0	26	57	41.6
Specific targeted research projects (STREP)	22.3	25.5	0	244	454	91.3
All instruments	31.2	57.7	0	728	762	82.2

Other papers include papers published in proceedings of conferences, books and book chapters and articles published in scientific journals not indexed in ISI-WoS (SCI and SSCI).

Given that a significant portion of the whole scientific output of DG INFSO projects is represented by other documents than scientific articles, we have repeated the same analysis for the subset of papers corresponding to conference papers, books and book chapters, and non-ISI articles (see previous section).

Results are reported in Table II.10 and Table II.11. By and large, the picture that emerges is similar. IPs and NoEs account for a larger share of papers than projects and they have a higher than average scientific productivity, while the opposite pattern is observed for STREPs. However, it is also interesting to note that while the fraction of all projects with a non-null scientific production increases significantly for all funding instruments once we take into account other papers, this is not true for NoEs (see last columns of Table II.9 and Table II.11). In particular, as far as IPs are concerned, the fraction of projects reporting a non-null scientific production increases from 65% when considering only scientific articles to 92% when examining other papers. This implies that including conference proceedings in order to achieve a more complete bibliometric assessment seems to be particularly important for IPs. On the other hand, the share of projects with a positive scientific production increases only from 76% to 78% in the case of NoEs, thereby suggesting that articles and papers have a more complementary role for this type of projects than for IPs.

II.2.4 Scientific productivity by strategic objective

The analysis carried out in the previous section has been replicated by classifying the projects according to the *strategic objective* of the EU calls for proposals under which they have been funded, rather than according to the type of funding instrument. Strategic objectives and scientific and technological priorities for the IST area are defined in the so-called Work Programme, which defines the objectives and technical content of the calls for proposals, the implementation plan and the criteria to be used for evaluating the proposals responding to these calls. The Work Programme is periodically updated to reflect new objectives and priorities. Since the strategic objectives are defined at a rather detailed level, we have grouped all projects into ten areas that define at a macro level broad objectives and priorities to be achieved. Each of the ten areas represents a sort of umbrella under which more specific objectives are identified[7]. The analysis has been carried out only for FP6 projects for the same reasons discussed above. Results are reported in Table II.12 and Table II.13, for scientific articles, and in

Table II.14 and Table II.15 for the other papers. Starting from the former, we observe that for the two objectives corresponding to *components and micro-systems* and *future and emerging technologies* the share of all scientific articles is significantly larger than the share of all projects, i.e. these areas contribute to scientific productivity more than expected on the basis of their share of projects (see Table II.12). In similar way, the average number of articles per project as well as the share of projects with a non-null scientific production for these two objectives is significantly higher than the average values of these variables (see Table II.13). On the other hand, the contribution to scientific productivity seems to be particularly low when compared to the overall share of projects for the field of *Applied IST research addressing major societal and economic challenges*. The 334 projects funded in this area have produced on average 2.5 articles and only 139 projects (42% of all projects in this area) have reported at least one article.

As long as different objectives correspond to different scientific subfields, part of the differences observed across strategic objectives might be related to different propensities to publish and/or to different means of publication and dissemination of knowledge. Looking at Table II.14 and Table II.15, which report the results of the same kind of analysis related to other papers, the picture changes in a rather significant way. The area corresponding to *Communication, computing and software technologies* becomes one of the most productive in terms of average number of papers per project and fraction of projects with a non-null scientific production. This is probably not surprising in the light of the discussion reported above (see Section II.2.2). *Computer science* is one of the areas in which conference proceedings play a particularly important role as a vehicle of knowledge dissemination and therefore including these papers is probably needed in order to get a correct assessment of the overall scientific productivity.

---

[7] For more details, see http://cordis.europa.eu/ist/activities/activities.htm. The 10 areas identified are Accompanying actions in support of participation in ICT research, Applied IST research addressing major societal and economic challenges, Communication, computing and software technologies, Components and micro-systems, IST Future and emerging technologies, IST Future and emerging technologies, International Co-operation, Knowledge and interface technologies, Nanotechnologies and nanosciences, Policy-orientated research, Research Infrastructures.

Table II.12 – Distribution of scientific articles by strategic objective, FP6

	Number of DG INFSO projects	%	Number of scientific articles	%
Accompanying actions in support of participation in ICT research	43	4.6	0	0.0
Applied IST res. addressing major societal and economic challenges	334	36.0	844	14.9
Communication, computing and software technologies	186	20.1	1010	17.8
Components and micro-systems	121	13.1	1349	23.8
IST Future and emerging technologies	103	11.1	1710	30.1
International Co-operation	24	2.6	6	0.1
Knowledge and interface technologies	51	5.5	216	3.8
Nanotechnologies and nanosciences	37	4.0	376	6.6
Policy-orientated research	5	0.5	2	0.0
Research Infrastructures	23	2.5	168	3.0
Total	927	100.0	5681	100.0

Scientific articles published in journals indexed in ISI-WoS (SCI and SSCI).

Table II.13 – Average number of scientific articles by strategic objective, FP6

Strategic objective	Mean	Std dev	Min	Max	Projects with articles	% of all projects
Accompanying actions in support of participation in ICT research	0.0	0.0	0	0	0	0.0
Applied IST res. addressing major societal and economic challenges	2.5	11.5	0	189	139	41.6
Communication, computing and software technologies	5.4	14.2	0	107	105	56.5
Components and micro-systems	11.1	23.6	0	135	82	67.8
IST Future and emerging technologies	16.6	26.5	0	151	79	76.7
International Co-operation	0.3	1.0	0	5	2	8.3
Knowledge and interface technologies	4.2	8.0	0	43	30	58.8
Nanotechnologies and nanosciences	10.2	15.3	0	80	21	56.8
Policy-orientated research	0.4	0.9	0	2	1	20.0
Research Infrastructures	7.3	29.4	0	142	8	34.8
All strategic objectives	6.1	17.2	0	189	467	50.4

Scientific articles published in journals indexed in ISI-WoS (SCI and SSCI).

Table II.14 – Distribution of other papers by strategic objective, FP6

Strategic objective	Number of DG INFSO projects	%	Number of papers	%
Accompanying actions in support of participation in ICT research	43	4.6	112	0.4
Applied IST res. addressing major societal and economic challenges	334	36.0	7893	27.3
Communication, computing and software technologies	186	20.1	9471	32.7
Components and micro-systems	121	13.1	3727	12.9
IST Future and emerging technologies	103	11.1	4725	16.3
International Co-operation	24	2.6	166	0.6
Knowledge and interface technologies	51	5.5	1447	5.0
Nanotechnologies and nanosciences	37	4.0	1048	3.6
Policy-orientated research	5	0.5	28	0.1
Research Infrastructures	23	2.5	316	1.1
Total	927	100.0	28933	100.0

Other papers include papers published in proceedings of conferences, books and book chapters and articles published in scientific journals not indexed in ISI-WoS (SCI and SSCI).

Table II.15 – Average number of other papers by strategic objective, FP6

Strategic objective	Mean	Std dev	Min	Max	Projects with papers	% of all projects
Accompanying actions in support of participation in ICT research	2.6	5.5	0	22	12	27.9
Applied IST res. addressing major societal and economic challenges	23.6	43.0	0	588	290	86.8
Communication, computing and software technologies	50.9	87.1	0	728	164	88.2
Components and micro-systems	30.8	52.2	0	325	102	84.3
IST Future and emerging technologies	45.9	63.2	0	339	85	82.5
International Co-operation	6.9	10.0	0	32	12	50.0
Knowledge and interface technologies	28.4	48.0	0	282	44	86.3
Nanotechnologies and nanosciences	28.3	26.4	0	98	34	91.9
Policy-orientated research	5.6	5.5	0	11	3	60.0
Research Infrastructures	13.7	31.8	0	141	16	69.6
All strategic objectives	31.2	57.7	0	728	762	82.2

Other papers include papers published in proceedings of conferences, books and book chapters and articles published in scientific journals not indexed in ISI-WoS (SCI and SSCI).

It is also quite interesting to note that the area corresponding to *Applied IST research addressing major societal and economic challenges*, which accounts for almost 36% of all EU funded projects in the database, is the one that shows the highest increase in the fraction of all projects with a non-null scientific production once one shifts from articles to other papers (from 42% to 87%). This suggests that knowledge dissemination through other means than scientific journals seems to be especially important in this field. At the same time, however, we also note that the average number of papers per project in this area is still lower than the average scientific productivity for all projects (23.6 vs. 31.2). In other words, a relatively large fraction of projects in this field are able to produce some kind of scientific output, but the output per project is generally relatively low.

II.2.5 Scientific productivity: a regression analysis

The previous analysis has highlighted a few differences in the scientific productivity of projects according to funding instrument and strategic objectives. However, some of the differences observed might be related to other factors that are correlated with both scientific productivity and with the type of funding instrument or the strategic objective. In order to corroborate some of the findings reported above, we have carried out a simple regression analysis of the possible factors affecting the number of articles and papers produced by DG INFSO projects. In particular, we focused our attention on the FP6 and on the three major funding instruments, i.e. STRePs, IPs and NoEs. Whereas STRePs involve relatively small consortia and narrowly focused research and represent a traditional funding instrument in the context of Framework Programmes, IPs and NoEs have been introduced for the first time in the FP6 with the explicit purpose of facilitating the development of a European Research Area (ERA):

a) IPs are large multi-partner projects involving a wide range of organizations from the research and business communities. They aim at mobilizing a *critical mass of resources* to attain clearly defined objectives in terms of scientific and technological knowledge and/or results applicable to products, processes and services. The ultimate goal of this instrument is to obtain results with a significant and direct impact on European industrial competitiveness or contributing to the solution of important societal/global problems.

b) NoEs are large multi-partner projects aimed at reinforcing European scientific and technological excellence by pooling and networking a *critical mass of resources* and expertise. They are primarily intended to combine and cross-fertilize existing strands of research around a common core issue and are also more likely to involve publicly supported research organizations and to have less centralized or hierarchical structures than IPs. Moreover, they aim at creating a progressive and lasting integration of existing and emerging research activities of the network partners.

The rationale behind the new policy instruments is that world-class centres of excellence already exist in Europe in a wide range of research fields. However, they are often scattered and only loosely connected while their expertise is not always sufficiently well known across Europe, especially by firms that could usefully join forces with them. The integration of these centres into long-term R&D consortia financially supported by the EU and focused on leading-edge research would contribute to enhancing the European position in strategic fields, attracting new resources and expertise, and, most of all, restructuring the way research is carried out in Europe, favouring the development of an overall more collaborative attitude by public and private actors. At the same time, while the creation of these networks is to be supported with European financing, their activities should not become dependent on this support. European funding is in fact meant to complement resources deployed by the participants and should take the form of a fixed grant for integration (for NoEs) and of a percentage grant to the budget (for IPs). In other words, European action is meant to be a stimulus for centres of excellence to 'cluster around' common long-term objectives, network on a permanent basis and self-organize the division of tasks and information flows. Table II.16 summarises the main characteristics of the funding instruments used in the FP6. The new IP and NoE instruments have accounted for about half (48%) of the total EC financial contribution in the FP6, while STRePs have accounted for a further 27% (Vonortas, 2008). As a consequence of the new instruments, the typical size of the projects has increased, whereas the number of funded projects has decreased. The average number of partners per project has doubled, rising from 6.1 in FP5 to 12 in FP6. Moreover, the size of projects has largely varied across instruments: on average, IPs involved 25 participants and received

EC funding of €9.5m over four years, NoEs involved 30 participants with EC contribution of €7.5m over four years, while STRePs involved 9 participants and received €2m in EC contribution for three years (Arnold et al., 2009)[8].

The introduction of the new funding instruments and the associated increase in the size of research consortia have been followed by a number of evaluation studies, which have raised a few concerns about the effectiveness of the new policy tools in achieving their objectives (Marimon Report, 2004; Arnold Report, 2009; Arnold et al., 2005; Vonortas, 2008; EPEC, 2009a). Besides the traditional problems related to the cumbersome level of administration and co-ordination work entailed by the participation in the FP and the associated decline in industrial participation (in particular SMEs), criticisms have been raised in particular regarding the size of projects funded with the new instruments, i.e. IPs and NoEs. The perception is that the emphasis put on the idea of "critical mass" has led to a rather artificial enlargement of partnerships well beyond any optimal level, ultimately reinforcing rather than decreasing fragmentation. According to a survey of 275 FP6 project participants, the group of experts led by Ramon Marimon shows that 61% of respondents felt that the partnership was artificially enlarged to fit work programme (Marimon Report, 2004). A similar survey on more than 1200 project participants shows that only 57% agreed or strongly agreed that most partners in IP projects were involved across the range of activities (EPEC, 2009a). In addition to the inefficiencies of scale and the problems of managing large consortia, it has been pointed out that maintaining excellence may be more difficult as the size of the project increases and that issues related to knowledge sharing and intellectual property become more complex with larger and heterogeneous consortia. This seems to be particularly critical in the case of IPs in which the combination of larger consortia, often including industrial competitors, and compulsory contractual conditions granting a contractor access to another contractor's pre-existing knowledge has been a major factor deterring industrial participation in FP projects.

[8] Other notable trends have been the reduction in the number of funded projects from 12391 in FP5 to 5485 in FP6 (excluding Human Resource and Mobility Actions) and the decrease in the success rate of proposals from 26% in FP5 to 18% in FP6.

Table II.16– Main characteristics of FP6 funding instruments

Instrument	Aim	Average duration (in months)	Optimum size (nr of participants)	Funding (€)
Integrated Project (IP)	Generation, demonstration and validation of new knowledge	36-60	3-20	€ 10 million
Network of Excellence (NoE)	Durable integration of the participants' research activities	48-60	6-12	€ 7 million
Strategic Targeted Project (STP)	Generation, demonstration and validation of new knowledge	18-36	3-15	€ 1,9 million
Co-ordinated Action (CA)	Organization and co-ordination of joint activities	18-36	3-20	€ 1 million
Specific Support action (SSA)	Implementation of FWPs	9-30	1-15	€ 0,5 million

Source: adaptation from European Commission (2004). Full document available at ftp://ftp.cordis.europa.eu/pub/fp6/docs/synoptic.pdf.

Assessing to what extent these concerns are justified and may have affected the achievements of FP6 is not an easy task. Most of the evaluation studies so far have relied upon survey evidence focusing on instruments and policy-related issues rather than the results obtained. A few bibliometric exercises have been carried out recently. For example, a recent study focuses on “lead scientists”, defined through a peer nomination process by asking all FP6 project coordinators to select in their project the person whose name is most likely to appear on published articles resulting directly from the project. Results show that individuals identified in this way tend to have a better publication and citation performance than their peers (EPEC, 2009b). This result is important by showing that FP projects are able to attract high-quality and excellent researchers. Yet, the methodology adopted is unable to identify the scientific output that has been produced in the context of the funded project and therefore to address issues related to the achievements of FP. A notable exception is represented by the study of Vanececk et al. (2010). They analyse scientific articles of Czech research teams resulting from FP5 and FP6 projects, by exploiting the electronic repository of the Czech Academy of Sciences. Their results show that the fraction of all Czech articles accounted for by FP projects has been rapidly increasing over time, representing 2.5% of all Czech papers published in 2006. Although this study goes in the direction of identifying the direct output of FP projects, it refers to a single country and therefore is unable to allow generalisations.

In this respect, the data set collected by the DG INFSO provides a unique opportunity to assess to what extent the instruments used to fund projects are associated to systematic differences in the ability to generate scientific and technological output and, in particular, to test to what extent the new funding instruments, i.e. IPs and NoEs, tend to exhibit a better or worse scientific and technological performance than traditional instruments, i.e. STRePs. To this purpose, we carried out a regression analysis in which the dependent variables are, respectively, the number of scientific articles and the number of other papers produced by DG INFSO projects.

As far as the explanatory variables are concerned, the following factors have been included in the analysis:

i) Since the major purpose of the exercise is to assess to what extent the scientific and technological productivity depends on the instrument used to fund a project, two dummy variables have been included among the covariates, taking value one if project *i* has been funded, respectively, as an Integrated Project (IP) and Network of Excellence (NoE), and zero else (i.e. STReP projects represent the baseline case).

ii) A major problem in assessing the effect of instruments is that other factors that are correlated with both the scientific and technological productivity and the type of funding instrument might confound the relation. This is particularly true for the *size of the project*. It is quite reasonable to expect that *ceteris paribus* larger projects either in terms of number of participating organizations or financial resources allocated to them will also be also to generate a larger volume of scientific output (and patents). Actually, as already argued before, the size of projects is one of the key dimensions considered by the European Commission to design the different types of funding instruments. Likewise, the Work Programme defines which types of funding instruments (and therefore implicitly the project size) are entitled to respond to the calls for proposals in strategic areas of intervention. The obvious implication of these differences is that, for example, the higher productivity of IPs compared to other funding instruments could be explained by the larger average size of these projects rather than by any other effect specific to this type of instrument. In order to control for this confounding factor, we have included in our regression analysis a variable capturing the number of organizations participating in the project. Moreover, to account for possible non-linearities in the relationship between project size and output, a squared term has been added.

iii) A covariate measuring the amount of *funding per participant* has been also introduced. Given the number of participants, projects endowed with a larger amount of financial resources per participant are expected to generate a larger volume of scientific and technological output[9].

iv) A third control variable considered is the *duration* (in months) of the project. As Table II.16 suggests, the average duration is largely different across different funding instruments and therefore strategic objectives. Also in this case, it is expected that all else equal longer projects should be able to produce a larger volume of output.

v) For similar reasons, we have included in the analysis a set of dummy variables capturing the years when projects have started their activities. Given the same duration of projects, we may expect that projects that started earlier had simply a longer period of time to produce and publish articles and papers.

vi) A further set of dummy variables has been included to account for the *strategic objective* of the EU calls for proposals under which projects have been funded. As long as different objectives correspond to different scientific and technical subfields, part of the differences in productivity observed across projects might be related to different propensities to publish and patent and/or to different means of publication and dissemination of knowledge.

vii) Another dimension that discriminates among projects and may explain the differential ability to produce scientific output is the composition of the projects in terms of characteristics of participating organizations. In particular, given the higher propensity of academic researchers to publish the results of their research, it is reasonable to expect that all else equal projects hosting a larger fraction of academic institutions should also display a higher productivity. The share of universities on the total number of par-

[9] The same regression analysis has been carried out by including the (log of) amount of funding in euros rather than the number of participants. However, given the high correlation with the number of participants, we carried out separate regressions.

ticipating organisations has been included in the analysis to account for this effect.

viii) Finally, the last control variable included in our regression analysis refers to the asymmetry in the distribution of funding. In particular, for each project we have computed the share of total funding allocated to each participant organization and then we have calculated the Herfindahl index, by summing up the squared shares. The index ranges from zero to one, where values closer to one indicate a larger degree of asymmetry in the distribution of funds among participants. *A priori*, it is not clear what should be the expected impact of this variable on the scientific productivity of projects. On the one hand, a more even distribution of funds can be taken as a proxy of the extent to which the project was characterised by a real co-operation among participants and therefore can be expected to have a positive impact upon the production of scientific output. On the other hand, as long as there are significant scale effects in research, a more concentrated distribution could also be expected to yield positive effects.

As mentioned above, the dependent variables in our regression model are, respectively, the number of scientific articles and the number of other papers. Given the count nature of our dependent variables, a simple ordinary least-squares regression analysis would yield biased results. For this reason, we have adopted here a negative binomial regression model, which is more appropriate for count data (Greene 2002; Greene 2008). The negative binomial regression model is an extension of the Poisson regression model, which relaxes the assumption of equidispersion. Assuming that Y is a discrete random variable that takes integer values, i.e. y_i=0,1,2,3…, the baseline Poisson regression model assumes that the observed count for observation *i* is drawn from a Poisson distribution with mean μ_i, where μ_i is estimated from observed characteristics. The model is thus defined by the following structural equation:

$$E(y_i|x_i) = \mu_i = e^{\propto + x_i \beta}$$

where β is a vector of parameters to be estimated and x is a vector of independent variables. A distinctive feature of the Poisson model is the equality of the conditional mean and variance, i.e. $E(y_i|x_i)=Var(y_i|x_i)$. Since empirical data often violates this assumption, exhibiting over-dispersion, i.e. $E(y_i|x_i)<Var(y_i|x_i)$, the Poisson model may be inadequate, as it underestimates the amount of dispersion in the outcome. In these circumstances, the negative binomial regression model represents a suitable alternative. The model addresses the problem of over-dispersion by adding a parameter that reflects the *unobserved* heterogeneity among observations and that is uncorrelated with the covariates:

$$E(y_i|x_i,\varepsilon_i) = \tilde{\mu}_i = e^{\alpha + x_i\beta + \varepsilon_i}$$

where $\delta \equiv \exp(\varepsilon_i)$ is assumed to have a one parameter gamma distribution $G(\theta,\theta)$ with mean 1 and variance $1/\theta=\kappa$. With this assumption, it is possible to show that

$$E(\tilde{\mu}_i) = \mu_i E(\delta) = \mu_i = E(y_i|x_i)$$

In other words, the Poisson and the negative binomial models have the same conditional mean. However, in the negative binomial model, the variance exceeds the mean:

$$Var(y_i|x_i) = \mu_i[1+\kappa\mu_i] > \mu_i$$

The model can be estimated via maximum likelihood. All estimates reported here have been obtained using the command `nbreg` implemented in Stata 10. Before reporting results, Table II.17 reports the summary statistics for the variables included in the regression model, while Table II.18 reports the correlation matrix[10]. From the descriptive statistics, we observe that the average project has duration of 37 months, 14 participants, of which 28% are academic institutions. The interpretation of the asymmetry index is less simple. However, taking its inverse, we can say that the observed average asymmetry is the same that we would obtain in a project with seven participants evenly sharing the to-

[10] Please note that for the regression analysis we have considered only STReP, IP and NoE projects. This corresponds to 734 out of 927 projects included in our database. Moreover, for 59 of the 734 projects we had not enough information to compute the distribution of funding among participants.

tal funds. STRePs account for 68% of all projects, followed by IPs (25%); NoE represent a very low share of all projects (around 7%). As far as strategic objectives are concerned, applied IST research addressing major societal and economic challenges account for 38% of all projects, followed by communication, computing and software technologies (23%), components and micro-systems (14%) and IST future and emerging technologies (13%).

At the same time, the table also suggests that at least part of the gap in scientific productivity among projects might be related to the different amount of resources allocated to them: the average number of organizations participating in IPs and NoEs is, respectively, 24.8 and 34.1 compared to an average of 9.3 organizations for STRePs. Similarly, we observe that NoEs and IPs receive a financial contribution of, respectively, €5.2 and €8.7 million, whereas the average financial contribution to STRePs is of €2.2 million. Given the wide disparity among projects in the number of participants, however, the variation in terms of funding per participant is relatively less pronounced: the average funding per participant is equal to around €442 and €189 thousand, respectively for IPs and NoEs, and €271 thousand for STRePs. The descriptive statistics also highlight the different composition of projects in terms of type of participating organizations: while universities account for about 53% of all participants in NoEs, they account on average for only 23% and 28% of all participants, respectively, in IPs and STRePs. Finally, as far duration is concerned, STReP projects are relatively shorter with an average duration of 35 months compared to an average duration of 42 and 43 months, respectively, for IPs and NoEs.

Finally, looking at the correlation matrix (Table II.18), it is interesting to note that there is a positive correlation between the number of scientific articles and the number of papers. However, the correlation is far from perfect confirming that at least for some projects the two publication strategies, i.e. publishing in peer-reviewed journals and publishing in conference proceedings, are not necessarily complementary. We also note that a quite large correlation (0.67) exists between the sizes of projects in terms of number of participants and the (log of) total funding. For this reasons, we carried out separate regressions including only one of the two variables at the time.

Table II.17 – Summary statistics of the variables included in the regression analysis

	All projects	IPs	NoEs	STRePs
Number of scientific articles	7.4	8.6	29.6	4.6
	(18.2)	(19.2)	(44.4)	(10.2)
	[0-189]	[0-151]	[0-189]	[0-110]
Number of other papers	37.4	54.4	127.0	22.2
	(62.8)	(66.8)	(155.6)	(25.6)
	[0-728]	[0-404]	[0-728]	[0-244]
Number of participants	14.4	23.2	33.4	9.0
	(11.6)	(13.4)	(17.4)	(3.4)
	[2-99]	[3-99]	[11-90]	[2-30]
Funding per participant (log)	12.6	13.0	12.0	12.4
	(0.4)	(0.4)	(0.6)	(0.4)
	[10.2-14.2]	[11.8-14.2]	[10.2-13.2]	[11-13.8]
Total project funding (log)	15	16.0	15.4	14.6
	(0.6)	(0.4)	(0.4)	(0.4)
	[13.2-17.2]	[14.6-17.2]	[14.2-16.2]	[13.2-15.6]
Asymmetry funding distribution	14.4	9.2	5.6	17.0
	(7.2)	(4.8)	(2.6)	(6.4)
	[2.4-68.4]	[3.2-46]	[2.4-12]	[7.4-68.4]
Share of universities	28.2	23.0	52.8	27.6
	(23.6)	(20.2)	(24.6)	(23.2)
	[0-100]	[0-100]	[0-91]	[0-100]
Duration (months)	37.4	41.8	43.2	35.2
	(7.8)	(9.2)	(11)	(5.8)
	[17-72]	[24-72]	[18-63]	[17-53]
Starting year				
2003	2.32	3.19	8.16	1.41
2004	30.65	35.64	65.31	25.35
2005	10.22	7.98	0.0	12.07
2006	51.77	50.0	26.53	54.93
2007	5.04	3.19	0.0	6.24
Strategic objective				
Applied IST research	37.6	37.23	30.61	38.43
Comm., computing, software technologies	22.75	24.47	38.78	20.52
Components and micro-systems	14.44	14.36	18.37	14.08
IST Future and emerging technologies	12.81	14.36	2.04	13.28
International co-operation	1.5	0.0	0.0	2.21
Knowledge and interface technologies	6.27	7.98	10.2	5.23
Nanotechnologies and nanosciences	4.63	1.6	0.0	6.24
Number of observations	734	188	49	497

The table reports average values; standard deviations in round brackets; minimum and maximum values in square brackets.

Table II.18 – Correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Number of articles	1.000						
(2) Number of papers	0.672	1.000					
(3) Number of participants	0.193	0.302	1.000				
(4) Funding per participant	-0.027	0.021	-0.234	1.000			
(5) Total funding	0.149	0.300	0.677	0.452	1.000		
(6) Asymmetry funding	-0.141	-0.264	-0.653	0.254	-0.586	1.000	
(7) Share of universities	0.177	0.151	0.074	-0.137	-0.033	-0.039	1.000
(8) Duration	0.273	0.168	0.207	0.297	0.399	-0.092	0.195

The table reports the correlation coefficients for the sample of 675 projects. Correlation coefficients have been reported only for the continuous/count variables included in the regression model.

Results of the regression analysis are reported in Table II.19 and Table II.20, respectively, for the number of scientific articles and for the number of other papers. For each table, we have reported four different specifications of the model.

As far as scientific articles are concerned, we observe that scientific productivity is positively associated to projects with a larger share of academic institutions among participating organisations. The coefficient on the duration of project has a positive sign and it is weakly statistically significant, thus showing that longer projects tend to produce a higher number of articles. Results also suggest that increasing the amount of funding per participant is associated to a larger volume of scientific output (see columns (1) and (2) in Table II.19). Concerning project size, the evidence seems to suggest the existence of a U-inverted relationship between the scale of the project, in terms of number of participants, and productivity. Funding collaborative projects with a larger number of participants brings positive returns in terms of scientific output, but these returns are available only up to a certain point, thereby indicating the existence of decreasing marginal returns to increasing the size of research consortia (see columns (1) and (2) in Table II.19). This result has to be interpreted with some caution. Actually, on the basis of the estimated coefficients, the inflection point seems to occur for a size of around 52 participants, which is substantially higher than the average size of both IPs and NoEs. Taking the sample of projects considered here, there are 9 IP (5%) and 8 NoE (16%) projects with more than 50 participants, while there are no STReP projects above that threshold. However, this finding is consistent with the qualitative evidence emerging from survey studies. In a recent evaluation exercise, the vast majority of participants in IPs reported that the project was too big and that the maximum number of participants had to be around 20 partners or less; in the case of NoEs, around 30% of all participants perceived that 48 should be the maximum number of partners, while another 36% felt that the threshold should be less than 60 partners (EPEC, 2009a). The existence of diminishing returns with respect to project size is confirmed by the estimates reported in column 3 of Table II.19, which show that the value of the coefficient on the (log of) total funding allocated to projects is lower than one and strongly statistically significant.

Regarding the variables of interest here results show that even *after accounting for the impact of project size*, the type of funding instrument seems to affect the scientific productivity of projects. In particular, the coefficient on the dummy variable for NoEs is positive and statistically significant, thus indicating that NoEs are on average more productive than STRePs; on the contrary, the coefficient on the dummy variable for IPs is negative and highly significant, thereby suggesting that the returns in terms of publications from investing in IP projects have been substantially lower than those obtained from STRePs. All else equal, NoE projects produced on average 2.9 times as many articles as STReP projects, whereas taking into account their larger size IPs generated on average around 60% fewer articles than STRePs.

As far as the production of other papers is concerned, results are qualitatively similar to those for scientific articles (see Table II.20). In particular, they tend to confirm the existence of a U-inverted relationship between project size and scientific productivity. At the same time, however, we note a few differences. In particular, we note that IP projects are now neither more nor less productive than STREPs. This finding confirms that for projects funded through this instrument, conference proceedings and other papers represent a more significant indicator of productivity than traditional scientific articles. NoEs are also more productive than STREPs by a factor of 3.1 (or 210%).

Table II.19 – Negative binomial regression of the number of scientific articles, FP6

Variables	(1)	(2)	(3)	(4)
Number of participants	0.0791***	0.0732***		
	(0.0181)	(0.0204)		
Number of participants [squared]	-0.00104***	-0.000916***		
	(0.000313)	(0.000342)		
Funding per participant (log)	0.996***	1.067***		
	(0.222)	(0.238)		
Total project funding (log)			0.595***	1.072***
			(0.118)	(0.211)
Asymmetry in funding distribution		-0.0347**		
		(0.0156)		
Share of universities	0.0118***	0.0121***	0.0213***	0.0114***
	(0.00319)	(0.00317)	(0.00351)	(0.00308)
Duration	0.0202**	0.0192*	0.0192*	0.0180*
	(0.0103)	(0.0110)	(0.0113)	(0.0103)
Funding instrument (baseline=STREP)				
Integrated projects (IP)	-0.863***	-1.224***		-0.885***
	(0.312)	(0.332)		(0.315)
Networks of excellence (NoE)	1.059***	0.766**		1.055***
	(0.310)	(0.325)		(0.263)
Project started in (baseline=2003)				
2004	0.450	0.352	0.185	0.425
	(0.434)	(0.431)	(0.655)	(0.425)
2005	0.527	0.495	0.255	0.529
	(0.506)	(0.521)	(0.733)	(0.500)
2006	0.0489	-0.140	-0.485	0.0134
	(0.439)	(0.437)	(0.665)	(0.431)
2007	0.341	0.574	-0.202	0.306
	(0.691)	(0.676)	(0.865)	(0.678)
Constant	-14.64***	-15.03***	-10.11***	-17.14***
	(2.924)	(3.056)	(1.978)	(3.148)
Over-dispersion parameter (log)	0.837***	0.816***	0.927***	0.837***
	(0.0711)	(0.0731)	(0.0689)	(0.0710)
Log-likelihood constant only	-1919.9	-1771.0	-1919.9	-1919.9
Log-likelihood full model	-1777.1	-1633.5	-1798.9	-1777.4
LR test	285.7***	275.1***	242.2***	285.1***
Observations	734	675	734	734

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Regressions include dummy variables for the strategic objectives of projects. Articles include papers published in scientific journals indexed in the ISI-SCI and SSCI databases.

Overall, the results show that the scientific output increases with the project size following a U-inverted relationship, thereby indicating the existence of decreasing marginal returns to increasing the size of research consortia. In this perspective, one might argue that the emphasis put in the FP6 on the idea of achieving a “critical mass” has been probably excessive leading to the formation of “too large” projects.

Table II.20 – Negative binomial regression of the number of other papers, FP6

Variables	(1)	(2)	(3)	(4)
Number of participants	0.0467***	0.0358***		
	(0.0107)	(0.0123)		
Number of participants [squared]	-0.000654***	-0.000489***		
	(0.000158)	(0.000183)		
Funding per participant (log)	0.573***	0.605***		
	(0.144)	(0.153)		
Total project funding (log)			0.636***	0.662***
			(0.0723)	(0.139)
Asymmetry in funding distribution		-0.0235***		
		(0.00904)		
Share of universities	0.00602***	0.00678***	0.0110***	0.00600***
	(0.00186)	(0.00196)	(0.00195)	(0.00183)
Duration	0.000161	0.000563	0.00489	-0.00193
	(0.00683)	(0.00727)	(0.00801)	(0.00673)
Funding instrument (baseline=STREP)				
Integrated projects (IP)	0.0252	-0.0661		-0.0560
	(0.194)	(0.204)		(0.203)
Networks of excellence (NoE)	1.134***	0.992***		1.087***
	(0.237)	(0.255)		(0.214)
Project started in (baseline=2003)				
2004	0.905**	0.828**	0.419	0.882**
	(0.368)	(0.370)	(0.654)	(0.365)
2005	1.455***	1.384***	0.975	1.453***
	(0.409)	(0.421)	(0.687)	(0.406)
2006	0.604*	0.477	-0.0435	0.567
	(0.367)	(0.370)	(0.658)	(0.362)
2007	0.0808	-0.0378	-0.579	0.0433
	(0.416)	(0.444)	(0.689)	(0.411)
Constant	-5.096***	-4.890**	-6.446***	-7.133***
	(1.863)	(1.940)	(1.247)	(2.061)
Over-dispersion parameter (log)	0.126*	0.133*	0.185***	0.124*
	(0.0685)	(0.0706)	(0.0670)	(0.0691)
Log-likelihood constant only	-3349.0	-3082.7	-3349.0	-3349.0
Log-likelihood full model	-3207.1	-2947.3	-3228.6	-3206.1
LR test	283.8***	270.8***	240.8	285.7
Observations	734	675	734	734

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Regressions include dummy variables for the strategic objectives of projects. Papers identify published literature, such as articles (published in journals *not* indexed in the ISI SCI and SSCI database), books, book chapters, conference proceedings and so on.

A second key result of our analysis is related to the existence of significant differences of performance among funding instruments. In particular, Integrated Projects, one of the new funding instruments introduced in the FP6 with the purpose of achieving a higher degree of research integration and support European competitiveness, did not perform particularly well compared to traditional, smaller STReP projects. After accounting for the larger amount of resources allocated to them, IPs have not been able to deliver a higher number of conference papers on average than STRePs, while their scientific performance has been even lower on average than that reported by STRePs. In this respect, the decision to replace the distinction between IPs and STRePs in the FP7 with the one between large and small *Collaborative Research Projects* is probably going in the right direction by suggesting a reduction in the "optimal" size of consortia[11]. On the other hand, results also show that Networks of Excellence, the other major new instrument introduced in the FP6, have performed much better than IPs and STRePs, standing out from the other types of projects in terms of ability to produce a large number of scientific articles and other papers.

It must be pointed out, however, that the results reported here have to be interpreted with some caution due to the limitations of the empirical methodology adopted. In the first place, the study has explored only one thematic area, i.e. ICT, of the FP and the results can be hardly generalised to other fields. Second, the data on scientific and technological output of EU funded projects come directly from project coordinators, they are hardly verifiable in an independent way and therefore are subject to possible reporting bias, even though *a priori* it is difficult to say whether the bias should go in the direction of over-estimating or under-estimating the true output. Third, it has to be pointed out that, no matter how important, scientific articles and papers represent only one of the several and multifaceted socio-economic benefits that are expected from the FP. Finally, the interpretation of results is not totally clear due to endogeneity problems. As long as better research teams are attracted by NoEs, this might explain the superior performance of this instrument compared to IPs and STRePs.

In the next section (section II.3), we focus our attention on the assessment of the relative impact that the scientific production of DG INFSO projects had in different scientific fields. In section II.4, we will then examine to what extent the scientific output emerging from such projects is the outcome of a cross-country collaboration within European countries and regions, and between European countries and other countries. In doing so, we will restrict our analysis to the subset of 4359 ISI-WoS scientific articles collected in

---

[11] According to the most recent report monitoring the implementation of the FP7 during the first two years of operation (2007-2008), the average number of participants in Collaborative Projects is equal to around 11 compared to 18 for NoEs (http://ec.europa.eu/research/index.cfm?pg=reports).

the first batch of the study (see above section II.1), leaving out conference papers and ISI-WoS scientific articles collected in the second batch. As already discussed above, for these documents we lack the critical bibliographical information (i.e. citations received, affiliation of authors etc.) to carry out such a kind of analysis.

II.3 An assessment of scientific impact

II.3.1 Classification of output by scientific field

For the purpose of assessing the impact of the scientific output generated by DG INFSO projects, it is useful to classify scientific articles according to the science field in which they have been produced. Thomson-ISI classifies each journal included in the WoS (SCI and SSCI) database into one or more *subject categories*. According to the 2008 version of the WoS database, there are 173 subject categories for the SCI database and 56 subject categories for the SSCI database. In addition to the subject categories, Thomson-ISI also assigns each journal to one of 22 broad fields[12]. Journals are uniquely classified to a single broad field, while they can be classified under different subject categories in the SCI and SSCI databases. The two classification systems respond to different objectives: the 22 broad fields are used to build the so-called *Essential Science Indicators* and thus serve statistical purposes of detecting research trends and performance for government policy makers, university and corporate research managers etc.[13]; the subject categories reported in the SCI and SSCI databases instead are intended as tools to facilitate information retrieval for scientific research purposes (Leydesdorff e Rafols 2009).

For our purposes, the classification by subject categories is probably more interesting than the 22 broad fields, as it allows a more finely grained analysis of the publication activity of DG INFSO projects. At the same time, however, the large number of subject categories included in the database requires the adoption of a classification system that aggregates subject categories into a lower number of science fields. Taking the whole set of 4359 articles and 807 journals in the database, there are 151 subject categories represented. The average number of subject categories per journal is 1.9 and the average

[12] The table reporting these attributions is available at http://www.in-cites.com/journal-list/index.html.
[13] For more details, see http://esi-topics.com/ and http://sciencewatch.com/.

number of journals per subject category is 11.0. Moreover, the average number of articles per subject category is 51.3. Given the low average number of articles per subject category, an analysis carried out exclusively at this level of aggregation risks to provide biased results, at least in those categories where the number of articles is particularly low.

For the present analysis, we will adopt the reclassification system of ISI subject categories into 14 scientific disciplines recently proposed by Leydesdorff and Rafols (2009)[14]. The methodology followed by the authors can be briefly summarised as follows. Using the 2006 Journal of Citation Reports (JCR) database, an aggregated citation matrix has been built with rows and columns representing citing and cited subject categories. Each cell in the matrix reports the number of times articles in subject category i have cited articles in subject category j. The cells in the matrix have been normalised by computing the cosine similarity index in both the citing and the cited directions[15]. A high value of the index provides an indication that two subject categories are similar to each other in terms of other subject categories they cite (or from which they are cited). Using the normalised matrix, a factor analysis is then applied which suggests a 14-factor solutions, where each factor corresponds to a distinct scientific discipline, i.e. a set of subject categories that are very similar to each other in terms patterns of citations to and from other subject categories (Leydesdorff e Rafols 2009)[16]. Figure II.2 below illustrates the map of the 14 scientific disciplines, while Figure II.3 shows the citation relations among the 172 subject categories comprised within the 14 fields.

---

[14] The classification may be interactively consulted at http://www.leydesdorff.net/map06/index.htm .

[15] The cosine similarity index is defined as

$$\cos(x, y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

where x_i and y_i are vectors. The cosine index thus measures the cosine of the angle between two vectors of n dimensions. Its value ranges from -1 (minimum similarity) to 1 (maximum similarity).

[16] It is worth noting that out of 172 subject categories, 154 (89%) fall in the same factor (i.e. scientific field) in both the citing and cited dimensions. Of the 18 subject categories that are classified differently in the two dimensions 13 of them belong to areas like biomedical sciences, agriculture and chemistry, which are poorly represented among articles published by DG INFSO projects. Only 5 subject categories belong to areas which are relevant for the purposes of our study. For example, the subject category *Computer science*, *Interdisciplinary applications* is classified as *Computer sciences* in the citing dimension, and as *Engineering* in the cited one. Keeping in mind that, in what follows, we will adopt the classification of 172 subject categories into 14 scientific fields according to the citing dimension.

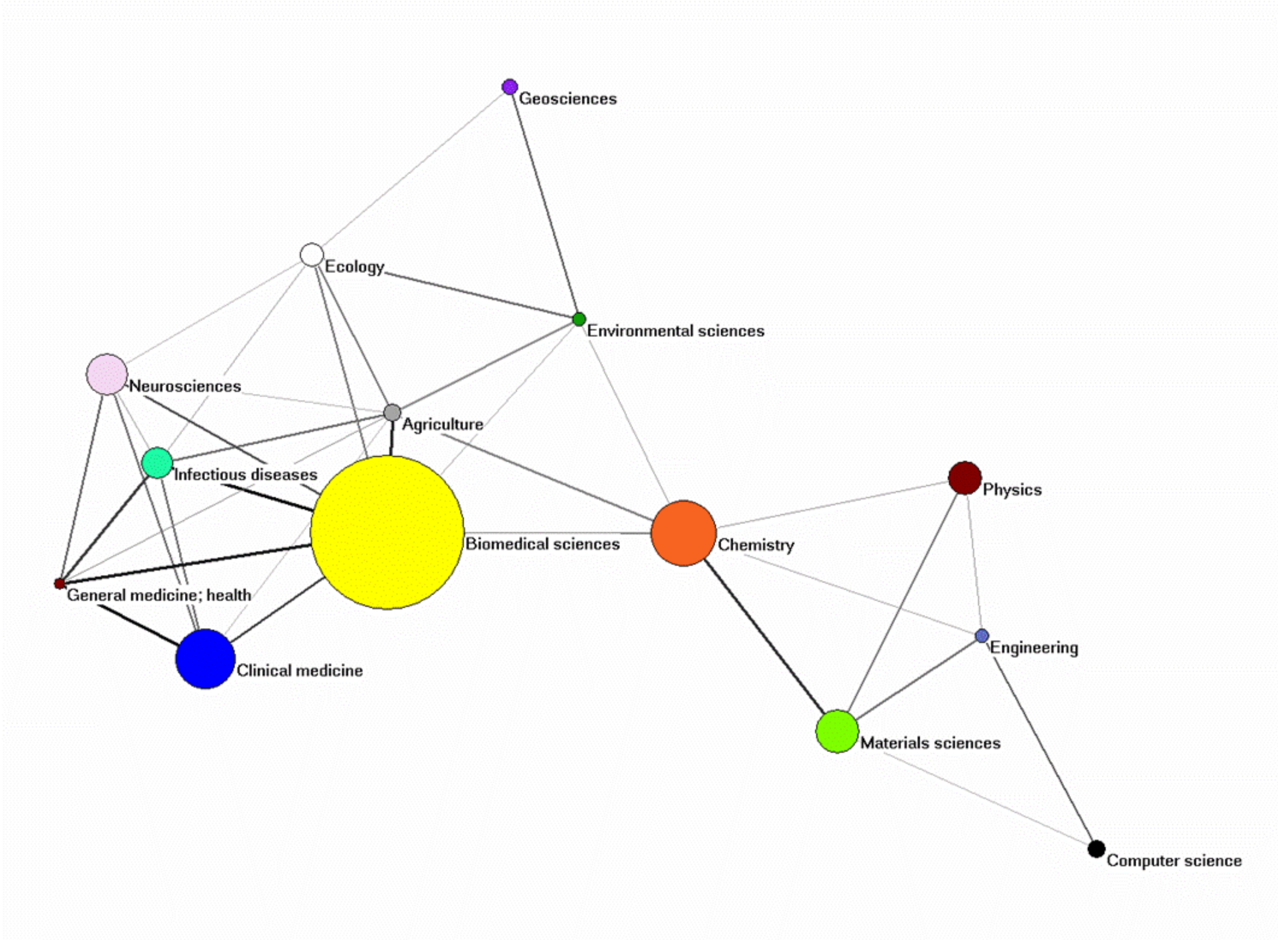
Using the classification of subject categories mentioned above, we have allocated each journal in the database into one of the 14 scientific fields on the basis of the subject category provided by the Journal of Citation Reports. Since a journal may have more than one subject category, we have used the first category reported in the JCR to allocate it in one of the 14 fields[17]. A further field not considered by Leydesdorff and Rafols (2009) has been added to include articles published in the field of social sciences. For each of the resulting 15 scientific fields, we have calculated the following indicators:

(a) The number and the percentage distribution of scientific articles reported by DG INFSO projects and published in the period 1996-2009 among the 15 scientific fields;
(b) The number and the percentage distribution of scientific articles reported by DG INFSO projects and published in the period 2004-2008 among the 15 scientific fields;
(c) The total number and the percentage distribution of scientific articles published in the period 2004-2008 in the same set of ISI-WoS journals in which the DG INFSO articles have been published among the 15 scientific fields.

For the calculation of (c), information derived from the Thomson-ISI Journal of Citation Reports (JCR) has been used. Since the 15 scientific fields largely differ in terms of number of journals and total number of articles, we need to take into account the size of scientific fields when evaluating the distribution of scientific output by DG INFSO projects across areas. The ratio between the fraction of articles published by projects in a given scientific field (point b above) and the fraction of all articles published in ISI-WoS journals in the same field (point c above) provides a useful indicator of the extent to which DG INFSO projects are relatively specialised in a scientific field, given its size in terms of number of journals and articles. In particular, a ratio greater than one for a given scientific field indicates that projects are relatively specialised in publishing in that field; on the other hand, a ratio lower than one indicates a relative de-specialisation.

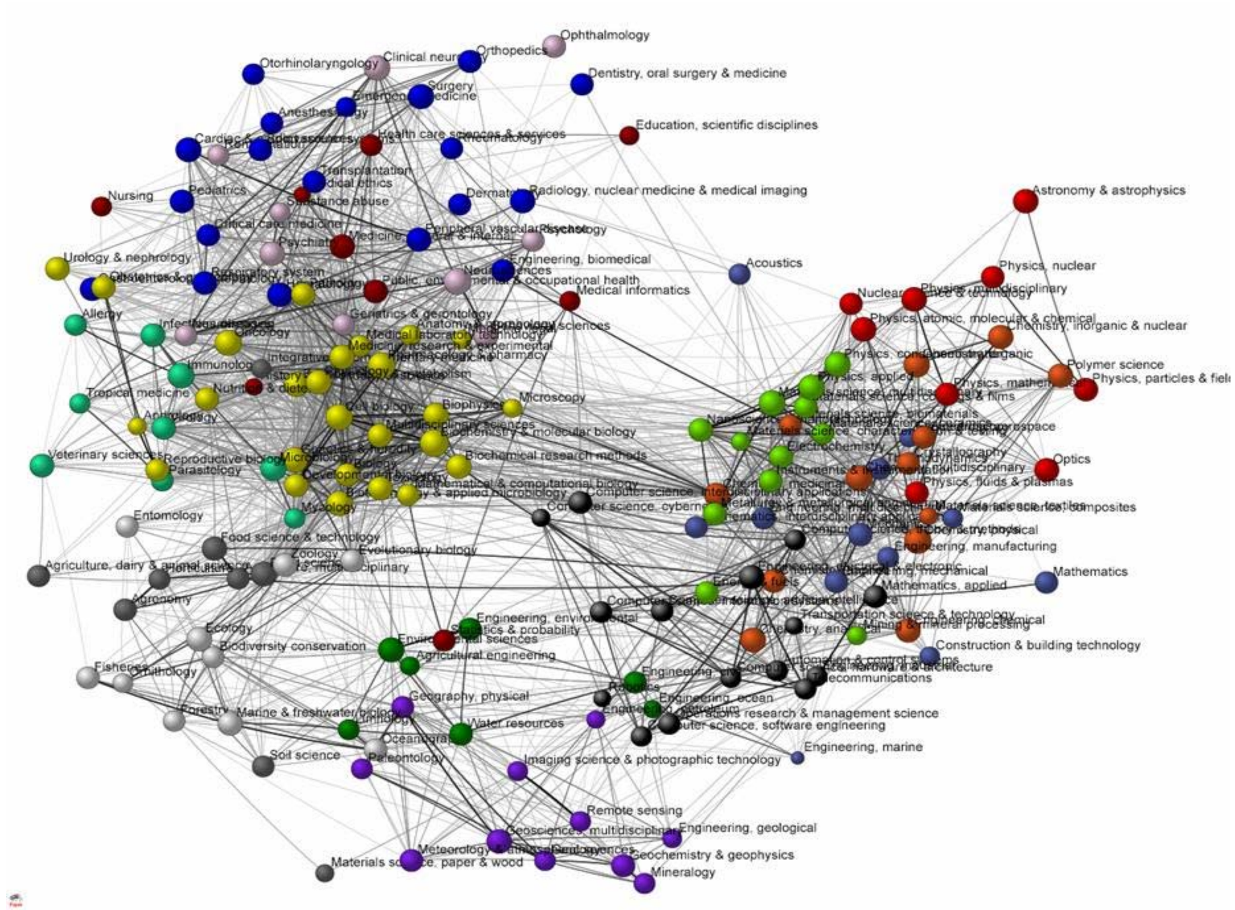
[17] Given that journals are often classified into more than one subject category, a possible objection is that the same journal would be allocated to different science fields when using all subject categories, rather than using just the first one. In order to assess the robustness of our procedure, we have checked this issue. Of all journals, slightly less than 75% are univocally associated to one of the 14 science fields, no matter whether all subject categories or only the first one are used. In view of this, we think that our assumption is acceptable and it should have a negligible impact upon our analysis.

Figure II.2 – Citation relations among the 14 scientific fields identified by Leydesdorff and Rafols (2009)



The figure reports a map of similarity among the 14 scientific fields in the citing dimension (cosine index ≥ 0.2). Node sizes are proportional to the logarithm of the number of citations in each subject category. Source: Leydesdorff and Rafols (2009).

Figure II.3 – Citation relations among the 172 subject categories comprised within the 14 scientific fields



The figure reports a map of similarity in the citing dimension among the 172 subject categories comprised within the 14 scientific fields (cosine index ≥ 0.2). Node sizes are proportional to the logarithm of the number of citations in the respective subject categories. Source: Leydesdorff and Rafols (2009).

Table II.21– Distribution of scientific articles by 15 scientific fields following Leydesdorff and Rafols (2009)

Science field	Articles reported by DG INFSO projects (1996-2009)	% (a)	Articles reported by DG INFSO projects (2004-08)	% (b)	Articles published in ISI-WoS journals (2004-08)	% (c)	(b)/(c)
Computer science	1988	45.6	1893	45.9	256369	21.6	2.1
Physics	791	18.2	731	17.7	143017	12.0	1.5
Materials sciences	608	14.0	584	14.1	186840	15.7	0.9
Chemistry	350	8.0	343	8.3	236619	19.9	0.4
Biomedical sciences	247	5.7	241	5.8	185075	15.6	0.4
Neurosciences	143	3.3	123	3.0	74920	6.3	0.5
Engineering	119	2.7	111	2.7	44133	3.7	0.7
Clinical medicine	38	0.9	35	0.9	20554	1.7	0.5
Social sciences	32	0.7	27	0.7	8622	0.7	0.9
Geosciences	12	0.3	11	0.3	6413	0.5	0.5
Environmental sciences	9	0.2	9	0.2	13070	1.1	0.2
Ecology	8	0.2	8	0.2	3416	0.3	0.7
General medicine	7	0.2	6	0.2	3011	0.3	0.6
Infectious diseases	5	0.1	5	0.1	3934	0.3	0.4
Agriculture	2	0.1	2	0.1	1952	0.2	0.3
Total	4359	100.0	4129	100.0	1187945	100.0	

The table reports the distribution of scientific articles published by DG INFSO projects following the classification into 15 scientific fields proposed by Leydesdorff and Rafols (2009). The frequency distribution has been calculated with reference to articles published in the years 1996-2009 (column a) and in the years 2004-2008 (column b). The table also reports the total number and frequency distribution of articles published in the period 2004-2008 (column c) taking the same set of journals in which project articles have been published and following the same classification into 15 fields. The ratio between columns (b) and (c) is an indicator of the extent to which DG INFSO projects are specialised in the publication of articles in a given field.

Table II.22– Distribution of scientific articles by 22 broad fields

Subject categories	Articles reported by DG INFSO projects (1996-2009)	% (a)	Articles reported by DG INFSO projects (2004-08)	% (b)	Articles published in ISI-WoS journals (2004-08)	% (c)	(b)/(c)
Physics	1589	36.5	1499	36.3	293845	24.7	1.5
Engineering	983	22.6	923	22.4	183229	15.4	1.4
Computer Science	809	18.6	789	19.1	92080	7.8	2.5
Chemistry	259	5.9	254	6.2	218410	18.4	0.3
Materials Science	180	4.1	163	4.0	66971	5.6	0.7
Neuroscience & Behavior	136	3.1	119	2.9	66067	5.6	0.5
Clinical Medicine	115	2.6	109	2.6	79502	6.7	0.4
Biology & Biochemistry	74	1.7	73	1.8	64706	5.5	0.3
Mathematics	59	1.4	55	1.3	19066	1.6	0.8
Multidisciplinary	36	0.8	35	0.9	27836	2.3	0.4
Social Sciences, general	28	0.6	23	0.6	4982	0.4	1.3
Psychiatry/Psychology	24	0.6	22	0.5	7253	0.6	0.9
Molecular Biology & Genetics	16	0.4	16	0.4	10488	0.9	0.4
Economics & Business	12	0.3	11	0.3	3366	0.3	1.0
Environment/Ecology	9	0.2	9	0.2	14257	1.2	0.2
Space Science	7	0.2	7	0.2	17780	1.5	0.1
Agricultural Sciences	6	0.1	6	0.2	4519	0.4	0.4
Pharmacology & Toxicology	6	0.1	6	0.2	4344	0.4	0.4
Geosciences	5	0.1	5	0.1	4183	0.4	0.3
Plant & Animal Science	4	0.1	4	0.1	2992	0.3	0.4
Immunology	1	0.0	1	0.0	1085	0.1	0.2
Microbiology	1	0.0	0	0.0	984	0.1	0.0
Total	4359	100.0	4129	100.0	1187945	100.0	

The table reports the distribution of scientific articles published by DG INFSO projects following the classification of Thomson-ISI journals into 22 broad fields. The frequency distribution has been calculated with reference to articles published in the years 1996-2009 (column a) and in the years 2004-2008 (column b). The table also reports the total number and frequency distribution of articles published in the period 2004-2008 (column c) taking the same set of journals in which project articles have been published and following the same classification into 22 fields. The ratio between columns (b) and (c) is an indicator of the extent to which DG INFSO projects are specialised in the publication of articles in a given field.

Table II.21 reports the value of the above mentioned indicators when articles are re-classified into 15 scientific fields following the classification of subject categories proposed by Leydesdorff and Rafols (2009). For the sake of comparison, Table II.22 reports the same statistics, but using the classification of journals into 22 broad science fields, proposed by Thomson-ISI. Looking at Table II.21, it is not surprising to observe that the fields in which DG INFSO tend to concentrate their research activities are the ones related to computer sciences, physics and materials sciences. These three fields combined account for almost 80% of all research articles published by DG INFSO projects. At the same time, it is also interesting to observe that the share of articles published by EU projects in the fields of computer sciences and physics exceeds the share of all ISI-WoS articles published in these two fields (see last column of Table II.21). In other words, the rate of publication by DG INFSO projects is higher than the expected rate, i.e. they are relatively specialised in these fields.

As far as the field of materials sciences is concerned, the fraction of all articles published by EU projects is similar to the fraction of all ISI articles accounted for by this field, so that the specialisation index takes a value that is only slightly lower than one. Note that, even though the two classifications are hardly comparable, a similar picture emerges from the data reported in Table II.22. Physics, engineering and computer sciences are the fields where DG INFSO projects show a significant degree of specialisation.

II.3.2 Choice of scientific fields for benchmarking

Assessing the impact requires the identification of a suitable sample of *control* articles against which benchmarking the performance of the scientific articles produced by DG INFSO projects. In principle, one would like to select for each of the articles published by projects, all other articles published in the same year and in the same journal, so as to ensure a reasonable degree of comparability. This would imply in our case to download and analyse more than one million of articles (see above Table II.21). Apart from the fact that downloading and parsing information into relevant fields (i.e. citations received, affiliation of authors, countries and regions of authors etc.) is not a trivial task for such a large number of articles, it is also probably not required for benchmarking purposes. As pointed out in the previous section, the volume of scientific production by

DG INFSO projects is rather low in several scientific fields. Therefore, any assessment based upon that low number of articles would probably raise questions of statistical significance. In addition to this, it seems to be quite reasonable to focus the assessment exercise upon those fields in which DG INFSO projects present a relative specialisation.

Following these arguments, we have chosen to perform the impact assessment by selecting articles in three science fields: computer science, physics and materials science (see Table II.21 above). For each of these fields, we have proceeded to identify the most important subject categories within them. More specifically, for each field, we have calculated the share of all DG INFSO articles and the share of all ISI-WoS articles accounted for by the different subject categories comprised within that field. Results are reported in Table II.23, Table II.24 and Table II.25, respectively, for computer science, physics and materials science.

As far as computer science is concerned, we observe several subject categories where the ratio between the two shares is larger than one (see last column of Table II.23). Combining the idea of focusing the assessment on fields of relative specialisation, with the requirement of a sufficiently large number of observations to ensure statistical significance, we have selected four subject categories that will be used in our assessment exercise. The four subject categories (underlined in bold in the table) are: engineering (electrical and electronic), computer science hardware & architecture, computer science information systems, and computer science software engineering. The four categories account for almost 65% of all articles published by DG INFSO articles in the field of computer science.

As far as physics is concerned, we have decided to select only one subject category, i.e. optics. This category accounts for more than 54% of all articles published in this field by DG INFSO projects and it is also the only subject category showing a specialisation index greater than one (see Table II.24).

Finally, for the field of materials science, we have selected the subject category corresponding to applied physics. This category accounts for almost 50% of all articles published in this field by our projects and presents a specialisation index greater than one (see Table II.25).

Table II.23 – Most important subject categories: Computer science

Subject category	INFSO articles (1996-2009)	% (a)	INFSO articles (2004-2008)	% (b)	ISI-WoS articles (2004-2008)	% (c)	(b)/(c)
Engineering, Electrical & Electronic	867	43.6	826	43.6	82221	32.1	1.4
Computer Science, Theory & Methods	243	12.2	235	12.4	41432	16.2	0.8
Computer Science, Artificial Intelligence	181	9.1	173	9.1	49887	19.5	0.5
Computer Science, Hardware & Architecture	160	8.1	152	8.0	12036	4.7	1.7
Computer Science, Information Systems	155	7.8	149	7.9	14255	5.6	1.4
Computer Science, Software Engineering	95	4.8	89	4.7	6608	2.6	1.8
Computer Science, Interdisciplinary Applications	67	3.4	63	3.3	12051	4.7	0.7
Telecommunications	59	3.0	55	2.9	3612	1.4	2.1
Automation & Control Systems	63	3.2	55	2.9	10745	4.2	0.7
Mathematics, Applied	42	2.1	42	2.2	15946	6.2	0.4
Computer Science, Cybernetics	23	1.2	23	1.2	1325	0.5	2.3
Engineering, Industrial	17	0.9	16	0.9	4211	1.6	0.5
Robotics	10	0.5	10	0.5	965	0.4	1.4
Transportation Science & Technology	3	0.2	3	0.2	152	0.1	2.7
Operations Research & Management Science	3	0.2	2	0.1	923	0.4	0.3
Total	1988	100.0	1893	100.0	256369	100.0	

Note: underlined in bold the subject categories selected for the benchmarking. For the list of journals selected for the benchmarking, see Appendix 2.

Table II.24 – Most important subject categories: Physics

Subject category	INFSO articles (1996-2009)	% (a)	INFSO articles (2004-2008)	% (b)	ISI-WoS articles (2004-2008)	% (c)	(b)/(c)
Optics	430	54.4	394	53.9	47357	33.1	1.6
Physics, Multidisciplinary	277	35.0	255	34.9	48200	33.7	1.0
Physics, Atomic, Molecular & Chemical	46	5.8	46	6.3	16510	11.5	0.5
Physics, Fluids & Plasmas	19	2.4	19	2.6	11754	8.2	0.3
Physics, Mathematical	10	1.3	9	1.2	2983	2.1	0.6
Astronomy & Astrophysics	9	1.1	8	1.1	16213	11.3	0.1
Total	791	100.0	731	100.0	143017	100.0	

Note: underlined in bold the subject categories selected for the benchmarking. For the list of journals selected for the benchmarking, see Appendix 2.

Table II.25 – Most important subject categories: Materials science

Subject category	INFSO articles (1996-2009)	% (a)	INFSO articles (2004-2008)	% (b)	ISI-WoS articles (2004-2008)	% (c)	(b)/(c)
Physics, Applied	301	49.5	288	49.3	73957	39.6	1.2
Physics, Condensed Matter	175	28.8	174	29.8	47943	25.7	1.2
Materials Science, Multidisciplinary	89	14.6	80	13.7	33783	18.1	0.8
Instruments & Instrumentation	13	2.1	13	2.2	17571	9.4	0.2
Materials Science, Ceramics	13	2.1	12	2.1	6992	3.7	0.5
Nanoscience & Nanotechnology	11	1.8	11	1.9	2622	1.4	1.3
Materials Science, Coatings & Films	3	0.5	3	0.5	1549	0.8	0.6
Electrochemistry	2	0.3	2	0.3	2088	1.1	0.3
Materials Science, Characterization & Testing	1	0.2	1	0.2	335	0.2	0.9
Total	608	100.0	584	100.0	186840	100.0	

Note: underlined in bold the subject categories selected for the benchmarking. For the list of journals selected for the benchmarking, see Appendix 2.

The last step in the identification of scientific articles for benchmarking has been the selection of relevant journals for each of the science fields and subject categories defined above. The complete list of journals is reported in Appendix 2 to this document. In brief, for each of the science fields and subject categories selected for benchmarking we have examined all articles published in the period 2004-2008 in the journals classified in these science fields and subject categories, with two exceptions. In the case of engineering (electrical & electronic), we have only selected those journals in which DG INFSO projects have published at least two articles. These journals account for 98% of all articles published by projects in this field/category. In the case of applied physics, we have selected only two journals (Applied Physics Letters and Journal of Applied Physics), which together account for almost 88% of all articles published by DG INFSO projects in this field/category. Table II.26 summarises the coverage of the fields, subject categories and journals selected for the benchmarking. The way to read the table is as follows: for example, for the subject category Engineering, electrical & electronic, we have selected 56 journals for the benchmarking exercise; these 56 journals account for 97.7% of all articles published by DG INFSO project in this subject category and the subject category as a whole accounts for 43.6% of all articles published by DG INFSO projects in the field of Computer Science.

II.3.1 Impact analysis

The number of citations received by a scientific article can be considered as an indicator of its impact in the international scientific community. Whether or not citations can also be taken as a signal of quality has been a highly debated question in scientometrics, and a full-fledged review of this debate goes well beyond the scope of this report (Chubin and Garfield 1980; Kaplan 2007). Here, it suffices to say that there are several reasons that motivate researchers to cite previous articles, not all of which related to their quality. Therefore, articles that receive a large number of cites from other papers are usually considered as having a major *impact* for the scientific progress in their field and often (but not always) reflect high-quality cutting-edge research. In addition to this, it is also very important to point out that each scientific field has its own norms and practices governing the citation process, and as a consequence one must be very cautious in comparing citation indicators across fields.

Table II.26 – Coverage of science fields, subject categories and journals used for the benchmarking

	Nr of journals for benchmarking	% of articles by DG INFSO projects in the subject category considered for benchmarking	% of all DG INFSO articles in the science field accounted for by selected subject categories
Science field: Computer science			
Subject cat: Engineering, electrical & electronic	56	97.7	43.6
Subject cat: Computer Science, Hardware & Architecture	34	100.0	8.0
Subject cat: Computer Science, Information Systems	42	100.0	7.9
Subject cat: Computer Science, Software Engineering	21	100.0	4.7
Science field: Physics			
Subject cat: Optics	20	100.0	53.9
Science field: Materials science			
Subject cat: Physics, Applied	2	87.5	49.3

In order to assess the impact of DG INFSO scientific output, we have considered the number of citations received by articles in our database. Information on citations derives from the Thomson-ISI (SCI-S and SSCI) database, which reports for each article the total (i.e. cumulated) number of cites received by other scientific articles up to a certain date. The present release of our database contains citations up to September 2009[18]. As the probability that an article is cited is likely to be a positive function of time, i.e. all else equal older articles are more likely to be cited than younger ones, in the following analysis we will always control for the publication year of articles.

Our assessment starts from an analysis of the number and frequency of uncited articles, i.e. papers that have not received any citation up to September 2009.

Table II. 27 reports the fraction of all articles produced by DG INFSO projects that were uncited by science field and publication year. As expected, the frequency of uncited articles tends to be larger for articles published in more recent years. Given the differences in citation styles across science fields and the fact that the absolute number of articles is very low in several fields (see also Table II.21 above), any comparison across fields is likely to be not meaningful.

For the purposes of the present assessment, it is much more interesting to compare the share of uncited articles produced by DG INFSO projects with the similar share in the benchmarking population of articles, i.e. taking all articles published in the same journals and science fields. This type of analysis is reported in Table II.28 and Table II.2, respectively, for the three scientific fields and for the six subject categories chosen as benchmark. Each table reports the number and the share of articles produced by DG INFSO projects that received no citations up to (September) 2009 and compares them to the number and the share of all articles in the population that were uncited up to (September) 2009. In order to assess the extent to which the observed difference is statistically significant, a z-score is computed and reported together with the p-value in the last two columns of each table. The z-score is an often used statistics to test the difference between two proportions and it is defined as follows:

[18] In the Thomson-ISI database, the number of cites is reported in the field "TC".

Table II.27– Percentage of uncited articles by scientific field and publication year, 15 science fields

Science field	2004	2005	2006	2007	2008
Agriculture	-	-	0.00	0.00	-
Biomedical sciences	0.00	03.03	9.33	11.58	34.78
Chemistry	13.64	3.92	5.93	16.26	36.67
Clinical medicine	0.00	28.57	8.33	8.33	0.00
Computer science	**14.20**	**19.32**	**20.64**	**32.29**	**68.25**
Ecology	0.00	0.00	0.00	0.00	100.00
Engineering	0.00	0.00	17.50	35.29	73.33
Environmental sciences	-	-	0.00	50.00	100.00
General medicine	-	0.00	0.00	0.00	100.00
Geosciences	0.00	0.00	0.00	100.00	50.00
Infectious diseases	-	0.00	0.00	33.33	-
Materials sciences	**7.32**	**0.99**	**9.63**	**15.70**	**38.24**
Neurosciences	0.00	13.33	0.00	6.00	36.36
Physics	**2.50**	**2.59**	**6.52**	**12.15**	**18.18**
Social sciences	33.33	0.00	16.67	0.00	71.43

The table reports the share of articles produced by DG INFSO projects that received zero citations, i.e. uncited, by 2009. Articles are classified by 15 science fields and publication year. A (-) sign indicates that no article was published in that year/field. The three scientific fields selected for benchmarking are highlighted in bold.

Table II.28– Share of uncited articles, DG INFSO projects vs. benchmarking population, three science fields

Science field	Publication year	Number of un-cited articles by DG INFSO projects	share of all DG INFSO articles	Number of un-cited articles in the popula-tion	share of all articles in the population	z-score	p-value
Computer Science	2004	6	0.06	3380	0.17	-2.695	0.007
	2005	41	0.17	4203	0.18	-0.347	0.729
	2006	64	0.18	5651	0.24	-2.798	0.005
	2007	104	0.28	8336	0.34	-2.376	0.017
	2008	90	0.66	16219	0.65	0.377	0.706
Physics	2004	1	0.02	835	0.10	-1.969	0.049
	2005	2	0.02	1216	0.13	-3.567	0.000
	2006	8	0.07	1623	0.16	-2.536	0.011
	2007	14	0.12	2536	0.23	-2.616	0.009
	2008	0	0.00	5934	0.51	-2.694	0.007
Materials science	2004	1	0.06	250	0.04	0.419	0.675
	2005	0	0.00	562	0.07	-1.948	0.051
	2006	0	0.00	963	0.11	-3.410	0.001
	2007	7	0.09	2066	0.22	-2.736	0.006
	2008	4	0.33	6042	0.64	-2.210	0.027

Table II.29– Share of uncited articles, DG INFSO projects vs. benchmarking population, six subject categories

Science field	Publication year	Number of uncited articles by DG INFSO projects	share of all DG INFSO articles	Number of uncited articles in the population	share of all articles in the population	z-score	p-value
Computer Science, Hardware & Architecture	2004	1	0.07	477	0.21	-1.285	0.199
	2005	4	0.14	600	0.24	-1.169	0.242
	2006	13	0.28	824	0.30	-0.255	0.799
	2007	24	0.52	1146	0.41	1.519	0.129
	2008	13	0.72	2069	0.72	0.057	0.955
Computer Science, Information Systems	2004	2	0.18	507	0.21	-0.208	0.835
	2005	6	0.16	625	0.22	-0.910	0.363
	2006	9	0.21	984	0.30	-1.207	0.227
	2007	15	0.34	1243	0.39	-0.639	0.523
	2008	13	0.87	2325	0.68	1.514	0.130
Computer Science, Software Engineering	2004	0	0.00	200	0.16	-1.232	0.218
	2005	2	0.18	254	0.18	0.008	0.994
	2006	5	0.17	364	0.23	-0.732	0.464
	2007	9	0.26	585	0.36	-1.135	0.257
	2008	4	0.57	1327	0.69	-0.668	0.504
Engineering, Electrical & Electronic	2004	3	0.05	2196	0.15	-2.278	0.023
	2005	29	0.18	2724	0.17	0.484	0.628
	2006	37	0.15	3479	0.22	-2.534	0.011
	2007	56	0.23	5362	0.32	-3.004	0.003
	2008	60	0.63	10498	0.62	0.067	0.947
Optics	2004	1	0.02	835	0.10	-1.969	0.049
	2005	2	0.02	1216	0.13	-3.567	0.000
	2006	8	0.07	1623	0.16	-2.536	0.011
	2007	14	0.12	2536	0.23	-2.616	0.009
	2008	0	0.00	5934	0.51	-2.694	0.007
Physics, Applied	2004	1	0.06	250	0.04	0.419	0.675
	2005	0	0.00	562	0.07	-1.948	0.051
	2006	0	0.00	963	0.11	-3.410	0.001
	2007	7	0.09	2066	0.22	-2.736	0.006
	2008	4	0.33	6042	0.64	-2.210	0.027

$$z = \frac{p_s - p_b}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_s}+\frac{1}{n_b}\right)}}$$

where the suffixes s and b indicate, respectively, the sample of DG INFSO articles and the sample of articles published in the benchmarking journals, p_s and p_b indicate the fractions of articles that are were uncited, n_s and n_b are the total number of articles, and

$$\hat{p} = \frac{m_s + m_b}{n_s + n_b}$$

where m_s and m_b are the number of articles that were uncited. The two-tail test of our interest calculates the probability that the two proportions are equal (H0: p_s=p_b) versus the alternative hypothesis that they differ significantly (H1: $p_s \neq p_b$) from a standard normal distribution.

Looking first at Table II.28, we observe that the data tend to reject in most cases the null hypothesis that the fraction of uncited articles is equal. In particular, the evidence seems to suggest that the fraction of uncited articles produced by DG INFSO projects is significantly lower than the corresponding fraction in the benchmarking population. Thus, for example, of all articles produced by DG INFSO projects in the field of Computer science and published in the year 2006, only 64 were uncited by (September) 2009 or 18% of the total. The corresponding number of the benchmarking population is equal to 5651 or 24% of all articles published in this field in the year 2006. The difference between the two fractions is statistically significant at the 99% level.

Once we turn our attention to more detailed subject categories, results reported in Table II.29 show a more articulated picture. On the one hand, no significant difference between DG INFSO articles and the benchmarking sample seems to emerge with respect to the fields of Computer Science, Hardware & Architecture, Computer Science, Information Systems, and Computer Science, Software Engineering. On the other hand, some evidence that project articles have a higher likelihood to be cited than the benchmarking articles can be found for the fields of Engineering, Electrical & Electronic, Optics and Applied Physics.

Overall, the evidence presented above suggests that, at least in the scientific fields considered here, the probability that DG INFSO articles are cited is no lower and often significantly higher than the average article published in the same journals and scientific field. In order to corroborate this result, we have calculated the average number of cites received by articles produced by DG INFSO projects. This information is reported in Table II.30 for the 15 scientific fields used to classify articles. As expected, in most fields the average number of citations tends to decrease for the articles published in most recent years. Likewise, it is also worth noting how the average number of citations per paper differs across scientific fields, thereby justifying a separate analysis by scientific domain.

Rather than commenting in detail the results reported in Table II.30, it is more useful for assessment purposes to compare the citation rates of DG INFSO articles with the corresponding rates for benchmarking articles, i.e. articles published in the same years and journals. In particular, for each of three scientific fields selected for benchmarking and for each publication year we have calculated the following quantities:

1) *Mean observed citation rate* (MOCR): this is defined as the ratio between the total number of citations received up to 2009 by articles published in year *t* by DG INFSO projects and the total number of articles published in year *t* by those projects. This index measures the average impact of the project articles and offers an indication of their visibility, reputation and quality in the international scientific community.

2) *Mean expected citation rate* (MECR): this is the expected citation rate of a single paper defined as the average citation rate of all papers published in the same journals in the same year. More specifically, it is defined as the ratio between the total number of citations received up to 2009 by articles published in year *t* in a given scientific field and the total number of articles published in year *t* in the same field.

Table II.30– Average number of citations received per article, 15 scientific fields

Field	Publication year	Articles	Min	Max	Average	Std. dev
Agriculture	2006	1	5	5	5.0	-
	2007	1	1	1	1.0	-
Biomedical sciences	2004	19	1	218	33.8	60.0
	2005	33	0	209	27.9	43.6
	2006	75	0	238	15.2	29.8
	2007	95	0	175	7.2	18.9
	2008	23	0	8	1.8	2.0
Chemistry	2004	22	0	102	28.8	30.9
	2005	51	0	105	15.1	21.2
	2006	118	0	95	9.4	12.0
	2007	123	0	57	6.1	8.7
	2008	30	0	13	2.5	3.5
Clinical medicine	2004	2	7	14	10.5	4.9
	2005	7	0	53	14.1	18.5
	2006	12	0	12	6.3	3.2
	2007	12	0	15	4.5	5.0
	2008	2	1	3	2.0	1.4
Computer science	2004	162	0	228	9.5	19.1
	2005	383	0	158	7.5	13.5
	2006	562	0	42	4.4	5.6
	2007	575	0	30	2.3	3.4
	2008	211	0	11	0.7	1.5
Ecology	2004	1	17	17	17.0	-
	2005	1	13	13	13.0	-
	2006	2	3	25	14.0	15.6
	2007	3	1	8	3.3	4.0
	2008	1	0	0	0.0	-
Engineering	2004	8	4	28	11.0	8.0
	2005	14	2	15	8.2	3.6
	2006	40	0	40	6.5	8.4
	2007	34	0	11	2.0	2.7
	2008	15	0	4	0.7	1.3
Environmental sciences	2006	3	2	9	5.3	3.5
	2007	2	0	1	0.5	0.7
	2008	4	0	0	0.0	0.0
General medicine	2005	2	4	10	7.0	4.2
	2006	2	4	40	22.0	25.5
	2007	1	2	2	2.0	-
	2008	1	0	0	0.0	-
Geosciences	2004	2	18	26	22.0	5.7
	2005	2	5	22	13.5	12.0
	2006	3	2	27	14.7	12.5
	2007	2	0	0	0.0	0.0
	2008	2	0	1	0.5	0.7
Infectious diseases	2005	1	1	1	1.0	-
	2006	1	3	3	3.0	-
	2007	3	0	6	2.3	3.2
Materials sciences	2004	41	0	82	18.3	20.3
	2005	101	0	100	12.3	13.9
	2006	187	0	62	9.1	9.0
	2007	223	0	28	4.2	5.3
	2008	34	0	8	2.1	2.3
Neurosciences	2004	12	3	114	24.9	29.6
	2005	15	0	46	14.3	14.6
	2006	35	1	53	11.0	11.8
	2007	50	0	36	6.4	7.3
	2008	11	0	12	2.1	3.5
Physics	2004	120	0	198	24.5	31.5
	2005	192	0	172	19.5	25.3
	2006	184	0	122	13.2	15.3
	2007	214	0	101	7.8	11.0
	2008	22	0	17	4.7	4.5
Social sciences	2004	3	0	3	2.0	1.7
	2005	2	5	17	11.0	8.5
	2006	11	0	4	1.5	1.1
	2007	3	1	6	3.3	2.5
	2008	7	0	3	0.6	1.1

The ratio between MOCR and MECR defines the relative citation rate, i.e. it measures whether DG INFSO articles attract more or less citations than expected on the basis of the average citation rates of the journals in which they appeared. The indicator ranges between 0 and infinity, its neutral value is 1. In particular, $RCR < 1$ ($RCR > 1$) means "citation attractivity" is below (above) expectation.

Results are reported in Table II.31 for the three fields used here for benchmarking purposes, i.e. Computer science, Physics and Materials sciences. In addition to MECR and MOCR, the table also reports the minimum and maximum number of cites received by articles, as well as the value of the standard deviation. In order to assess to what extent DG INFSO articles receive a larger number of citations per paper than the benchmarking articles, rather than computing the fraction between MOCR and MECR, i.e. RCR, we have performed a simple t-test comparing the two means.

Results reported in Table II.31 show that MOCR is systematically larger than MECR in all the three fields and for all publication years, thereby suggesting that DG INFSO articles tend to display a higher than expected attractivity or impact. At the same time, however, the difference between the two indicators seems to be statistically significant only in Materials science and Physics, but not for Computer sciences.

Similar findings emerge from Table II.32, which compares MECR and MOCR for the six subject categories used for the benchmarking[19]. The average citation rate of DG INFSO articles tends to be larger than the average citation rate in all scientific domains. However, the difference between the two rates seems to be statistically significant, in addition to Optics and Applied Physics, only in Engineering, Electrical Electronic, and to some extent, also in Computer Science, Hardware & Architecture. In appraising these results, one should bear in mind the observations already made above concerning the fact that scientific articles published in traditional journals are probably less likely to capture the most important diffusion vehicles of new knowledge in fields related to Computer Science.

[19] Please note that since Physics and Materials sciences comprise only one subject category, i.e. Optics and Applied Physics, the results reported in Table II.31 and Table II.32 for these fields are identical.

Table II.31 – Average number of citations received per article, DG INFSO projects vs. benchmarking population, three science fields

Field	Publication year	Articles DG INFSO	Min citations	Max citations	Mean (MOCR)	Std. dev	Articles population	Min citations	Max citations	Mean (MECR)	Std. dev	t test	p value
Computer science	2004	94	0	228	12.1	24.0	20137	0	1298	8.5	18.7	-1.44	0.154
	2005	236	0	158	8.6	14.0	23031	0	701	6.7	12.8	-2.27	0.023
	2006	365	0	42	4.9	6.0	23732	0	452	4.4	8.6	-1.52	0.130
	2007	365	0	30	2.5	3.5	24265	0	120	2.4	4.0	-0.31	0.753
	2008	136	0	11	0.8	1.7	25097	0	31	0.7	1.5	-0.38	0.705
Physics	2004	52	0	81	16.3	15.7	8194	0	280	10.2	15.0	-2.92	0.004
	2005	112	0	65	14.8	12.7	9273	0	346	8.1	11.7	-6.06	0.000
	2006	108	0	34	9.9	8.1	9841	0	127	5.6	7.6	-5.94	0.000
	2007	114	0	40	6.0	6.7	11244	0	135	3.4	4.6	-4.19	0.000
	2008	7	1	13	5.1	4.2	11654	0	29	1.1	1.9	-2.55	0.043
Materials sciences	2004	17	0	82	26.1	22.1	6398	0	323	15.2	19.0	-2.35	0.019
	2005	49	1	53	13.8	11.3	7818	0	937	9.7	16.1	-2.52	0.015
	2006	96	1	34	9.9	7.6	8902	0	92	6.1	7.1	-5.24	0.000
	2007	78	0	25	5.0	5.1	9480	0	62	3.2	3.9	-3.1	0.003
	2008	12	0	7	2.7	2.6	9443	0	63	0.7	1.5	-2.57	0.026

Table II.32 – Average number of citations received per article, DG INFSO projects vs. benchmarking population, six subject categories

Science field		DG INFSO projects					Population						
		(a)	(b)	(c)	(d)	(e)	(a)	(b)	(c)	(d)	(e)	t test	p value
Computer Science, Hardware & Architecture	2004	14	0	228	23.3	59.5	2250	0	328	7.6	15.5	0.985	0.343
	2005	28	0	18	6.1	5.3	2529	0	149	5.6	10.2	0.503	0.619
	2006	46	0	14	2.4	2.9	2747	0	175	3.4	6.6	-2.281	0.027
	2007	46	0	9	1.2	2.1	2791	0	79	1.8	3.2	-1.890	0.065
	2008	18	0	1	0.3	0.5	2889	0	21	0.5	1.2	-2.028	0.059
Computer Science, Information Systems	2004	11	0	22	6.2	6.9	2446	0	1298	8.6	31.5	-1.101	0.297
	2005	37	0	37	7.5	9.8	2778	0	333	7.0	16.3	0.356	0.724
	2006	42	0	26	4.1	5.1	3279	0	452	4.4	13.8	-0.425	0.673
	2007	44	0	21	2.0	3.4	3202	0	47	2.1	3.7	-0.137	0.891
	2008	15	0	1	0.1	0.4	3395	0	18	0.7	1.5	-5.612	0.000
Computer Science, Software Engineering	2004	8	1	16	9.0	6.3	1253	0	282	9.0	15.8	0.021	0.984
	2005	11	0	19	5.9	6.4	1404	0	128	6.7	10.0	-0.415	0.687
	2006	29	0	20	4.4	6.0	1582	0	127	4.4	7.2	-0.013	0.989
	2007	34	0	12	2.3	2.8	1630	0	109	2.1	4.1	0.300	0.766
	2008	7	0	4	0.9	1.5	1927	0	27	0.6	1.3	0.512	0.627
Engineering, Electrical & Electronic	2004	61	0	36	11.0	8.9	14188	0	394	8.6	16.2	2.075	0.042
	2005	160	0	158	9.4	16.1	16320	0	701	6.8	12.7	2.057	0.041
	2006	248	0	42	5.6	6.4	16124	0	246	4.6	7.6	2.371	0.019
	2007	241	0	30	2.8	3.8	16642	0	120	2.6	4.2	0.793	0.429
	2008	96	0	11	1.0	2.0	16886	0	31	0.8	1.6	0.879	0.381
Optics	2004	52	0	81	16.3	15.7	8194	0	280	10.2	15.0	2.925	0.005
	2005	112	0	65	14.8	12.7	9273	0	346	8.1	11.7	6.055	0.000
	2006	108	0	34	9.9	8.1	9841	0	127	5.6	7.6	5.938	0.000
	2007	114	0	40	6.0	6.7	11244	0	135	3.4	4.6	4.194	0.000
	2008	7	1	13	5.1	4.2	11654	0	29	1.1	1.9	2.547	0.044
Physics, Applied	2004	17	0	82	26.1	22.1	6398	0	323	15.2	19.0	2.355	0.032
	2005	49	1	53	13.8	11.3	7818	0	937	9.7	16.1	2.523	0.015
	2006	96	1	34	9.9	7.6	8902	0	92	6.1	7.1	5.242	0.000
	2007	78	0	25	5.0	5.1	9480	0	62	3.2	3.9	3.101	0.003
	2008	12	0	7	2.7	2.6	9443	0	63	0.7	1.5	2.568	0.026

(a) Number of articles; (b) minimum number of citations; (c) maximum number of citations; (d) mean citation per article (MOCR and MECR); (e) standard deviation.

As the impact of a scientific article measured by the number of citations received may depend on several factors, we have carried out a regression analysis to control for some observable characteristics of papers that might affect the number of cites. For example, it might be that the number of authors in DG INFSO articles tends to be larger than the average article in the population, thereby inducing a spurious correlation between the number of citations received and the fact that an article has been produced in the context of an EU project. In particular, the factors that we have considered are:

a) The *size of article*, defined as the number of authors reported in the paper. This variable can be considered as a proxy for the amount of resources devoted to the research that led to the publication. Hence, the expectation is that, all else equal, articles produced by larger teams of scientists should also receive a larger number of citations.
b) The *length of the article* as measured by the number of pages. Although it is unclear why longer articles should attract more citations, the empirical evidence seems to show that this is the case (Abt 2000a). On the one hand, a benevolent interpretation is that longer papers make a more substantive contribution than shorter ones, thereby attracting the attention of other scholars. On the other hand, a more cynical attitude towards this phenomenon might simply point out that longer papers provide more scope for citations, irrespective of their quality.
c) The *number of cited references*. The need to control for this effect is related to the previous point. Examining the patterns of references in six fields of physical sciences, Abt (2000) finds evidence of a linear relation between the number of references and the length of an article (Abt 2000b; Abt 2000a; Vieira e Gomes 2009).
d) The *average impact factor* of the journal in which the article has been published. The impact factor is a measure of the frequency with which the "average article" in a journal has been cited in a given period of time. More specifically, the impact factor for a journal is calculated as the ratio between the number of times articles published in years *t-1* and *t-2* are cited by articles published in year *t* and the total number of published in years *t-1* and *t-2.* For each article, this variable measures the average impact factor over the period 2004-2008 of the journal in

which it has been published. All else equal, one would expect articles published in journal with a higher impact factor to receive a larger number of citations.

e) The *type* of article. Review articles typically receive more citations than other types of articles. There are at least two reasons for this. First, review articles are often authoritative summaries of the existing knowledge in a field, which help other researchers to discriminate among the most fruitful avenues for further research. Second, review articles are normally longer and have a larger number of references than other articles (Abt and Garfield 2002; Moed 2005). In order to account for this effect, we have included in our regression analysis a set of dummy variables capturing the type of articles and distinguishing between articles, reviews, and proceedings papers.
f) As already argued above, the citation norms and procedures may differ across scientific domains. To this purpose, we have introduced in the regression analysis dummy variables for the subject categories (Kaplan 2007).
g) Finally, the regression controls for the publication year of the article, given that older articles are more likely to have accumulated a larger stock of citations than younger ones.

The dependent variable in the regression analysis is the number of citations received by articles up to September 2009. Given the count nature of the dependent variable, the impact of the covariates has been estimated using a negative binomial regression model. For each scientific field, we have estimated two different models. In the first one, we have introduced a dummy variable that takes value 1 if the article has been reported by project coordinators as a scientific output of the sponsored research. In the second one, we have introduced three dummy variables that capture whether the article has been produced by a DG INFSO project funded through the instrument of NoEs, IPs or other funding instruments.

Results are reported in Table II.33, Table II.34 and Table II.35, respectively, for Computer science, Physics and Materials sciences.

Starting from Table II.33, results tend to support the findings based on the simple comparison of the average citation rates. Articles produced by DG INFSO projects in the field of Computer science are no more likely to receive a larger number of citations than other articles. Once we split articles by funding instrument, we observe that papers produced within IPs tend to get a lower number of citations, while the opposite holds for other funding instruments. In both cases, however, the relation is only weakly significant. Moreover, in appraising these results one should always bear in mind that traditional scientific articles in this field represent only one of the vehicles through which researchers transmit and disseminate knowledge. In particular, a more balanced assessment should control whether the same pattern holds for papers published in conference proceedings. All control variables have the expected sign and are statistically significant. In particular, articles published in journals with a higher impact factor, with a larger number of authors, with a larger number of references and with a larger number of pages tend to receive a larger number of citations. Moreover, all dummy variables for publication years are statistically significant suggesting, as expected, that more recently published articles receive a lower number of citations than older articles.

Turning the attention to Physics (Optics), results reported in Table II.34 are broadly similar as far as control variables are concerned. Yet, the dummy for DG INFSO articles is now positive and statistically significant, thereby suggesting that papers produced by EU sponsored research in this field tend to have a greater impact than the average article. In particular, the expected number of citations for DG INFSO articles is 44% higher than the average article, holding all other variables constant. Another way of interpreting the coefficient is by saying that, holding all other variables at their mean, a DG INFSO article gets 1.3 times more citations than the average article in the field. Quite interestingly, all DG INFSO articles receive a larger number of citations, irrespective of the funding instrument, although the magnitude of the coefficients is larger for NoEs and IPs, indicating that NoEs and IPs receive on average more citations than STREPs.

Similar results are found with reference to the field Materials science (Applied Physics). Estimates reported in Table II.35 suggest that a DG INFSO article gets 1.4 times more citations than another article, holding all other variables at their mean value. Yet, differently from Physics, articles produced by NoEs do not seem to command a larger number of citations, whereas the effect seems to be particularly strong for IPs and other funding instruments.

Table II.33 – Negative binomial regression of the number of citations received, Computer science

	Coeff	Std err	Coeff	Std err
Average impact factor (2004-08)	0.595	0.006***	0.595	0.006***
Number of cited references	0.016	0.000***	0.016	0.000***
Number of pages	0.007	0.001***	0.007	0.001***
Number of authors	0.029	0.002***	0.029	0.002***
DG INFSO article	-0.007	0.037		
Funding instrument (only FP6 articles)				
Integrated projects (IP)			-0.178	0.087**
Networks of excellence (NoE)			-0.076	0.060
Other funding instruments			0.107	0.064*
Article published in (baseline=2004)				
2005	-0.238	0.012***	-0.238	0.012***
2006	-0.655	0.012***	-0.654	0.012***
2007	-1.228	0.012***	-1.228	0.012***
2008	-2.445	0.014***	-2.444	0.014***
Constant	0.496	0.022***	0.496	0.022***
Subject category fixed effect	Yes		Yes	
Article type fixed effect	Yes		Yes	
Over-dispersion	1.272	0.007	1.273	0.007
Nr observations	117410		117303	
Log-likelihood	-263056.93		-262717.02	

*, **, *** statistically significant at the 90%, 95% and 99% levels.

Table II.34 – Negative binomial regression of the number of citations received, Physics

	Coeff	Std err	Coeff	Std err
Average impact factor (2004-08)	0.408	0.005***	0.408	0.005***
Number of cited references	0.017	0.0004***	0.017	0.0004***
Number of pages	-0.004	0.001***	-0.004	0.001***
Number of authors	0.058	0.002***	0.058	0.002***
DG INFSO article	0.367	0.053***		
Funding instrument (only FP6 articles)				
Integrated projects (IP)			0.479	0.159***
Networks of excellence (NoE)			0.512	0.088***
Other funding instruments			0.244	0.103**
Article published in (baseline=2004)				
2005	-0.295	0.016***	-0.296	0.016***
2006	-0.621	0.016***	-0.622	0.016***
2007	-1.158	0.016***	-1.158	0.016***
2008	-2.274	0.017***	-2.275	0.017***
Constant	0.718	0.022***	0.718	0.022***
Article type fixed effect	Yes		Yes	
Over-dispersion	0.954	0.008	0.955	0.008
Nr observations	50599		50497	
Log-likelihood	-124721.86		-124361.79	

*, **, *** statistically significant at the 90%, 95% and 99% levels.

Table II.35 – Negative binomial regression of the number of citations received, Materials science

	Coeff	Std err	Coeff	Std err
Average impact factor (2004-08)	0.365	0.010***	0.365	0.010***
Number of cited references	0.015	0.001***	0.015	0.001**
Number of pages	0.021	0.004***	0.021	0.004***
Number of authors	0.045	0.002***	0.045	0.002***
DG INFSO article	0.367	0.064***		
Funding instrument (only FP6 articles)				
Integrated projects (IP)			0.442	0.140***
Networks of excellence (NoE)			0.215	0.145
Other funding instruments			0.421	0.095***
Article published in (baseline=2004)				
2005	-0.465	0.016***	-0.465	0.016***
2006	-0.953	0.016****	-0.953	0.016***
2007	-1.609	0.017***	-1.609	0.017***
2008	-3.079	0.020***	-3.080	0.020***
Constant	0.897	0.047***	0.897	0.047***
Article type fixed effect	Yes		Yes	
Over-dispersion	0.864	0.008	0.864	0.008
Nr observations	42293		42258	
Log-likelihood	-107270.42		-107149.35	

*, **, *** statistically significant at the 90%, 95% and 99% levels.

Overall, the evidence presented here indicates that the scientific output produced in the field of IST by EU funded projects has an impact that is no lower and in many cases is significantly larger than the average article published in this field.

II.4 Patterns of co-authorship

II.4.1 Patterns of international collaboration

In this section, we examine the patterns of international co-authorship of DG INFSO articles. In particular, our main objective is to examine the extent to which articles produced in the context of DG INFSO funded projects tend to exhibit a higher propensity to engage in international collaboration than the average article produced by European authors. Before proceeding, we examine what is the share of articles resulting from DG INFSO projects on the total number of European articles published in the period 2004 to 2008 for the three subject fields used in the benchmarking analysis. Results reported in Table II.36 show that articles produced by European projects account for a share of all articles from the EU15 that ranges from 3.07% in computer science to 1.73% in materials science and to 1.99% in physics. This figures are relatively low, but they are consistent with those reported by Vanececk et al. (2010)[20]. In a study on the impact of FP in the Czech Republic, they show that articles resulting from FP projects account for about 1.46% of all Czech articles published from 2000 to 2007.

Table II.36 – Share of DG INFSO articles in the population of European articles

	Share of DG INFSO articles among all European articles with an affiliation from		
	Computer science	Materials science	Physics
EU 30	2.89	1.78	1.94
EU 27	2.99	1.69	1.90
EU 15	3.07	1.73	1.99

Table II.37 reports the number and share of all articles produced by DG INFSO projects that are internationally co-authored, i.e. in which affiliations from more than one country were reported in the article, for the 15 scientific fields. On average 40% of all ar-

[20] Please also note that we are under-estimating the share of DG INFSO articles on all European articles as the benchmarking analysis includes only the articles received and elaborated in the first batch of data.

ticles by DG INFSO projects were internationally co-authored, though the extent of international co-authorship differs across fields.

Table II.37 – Share of internationally co-authored papers by 15 science fields

Science field	Internationally co-authored articles by DG INFSO projects (2004-2008)	% of all articles
Computer science	*662*	*35.3*
Physics	*359*	*49.3*
Materials sciences	*273*	*46.8*
Chemistry	140	40.8
Biomedical sciences	104	43.3
Neurosciences	59	48.0
Engineering	54	48.6
Clinical medicine	11	31.4
Social sciences	7	28.0
Geosciences	5	45.5
General medicine	3	50.0
Infectious diseases	2	40.0
Agriculture	1	50.0
Ecology	1	12.5
Environmental sciences	1	11.1
All fields	1682	41.0

The table reports the number of fraction of DG INFSO articles with affiliations from at least two different countries. We could not determine the country of affiliation for 27 articles as the affiliation field is blank in the ISI data. Of the 27 articles, 19 are classified in the field of Computer sciences.

Table II.38 reports the number and the share of all articles produced by DG INFSO projects in which at least one author is affiliated with an organisation located in the European Union (27 countries) or in the ERA area (EU27, plus Switzerland, Norway and Turkey). On average 93% and 97% of all articles produced by these projects have an author from the EU27 or from the ERA area. The variations across fields, though existing, are relatively limited. Thus, most of the international research collaboration in the context of EU sponsored projects involves organizations that are located within the EU or the ERA areas.

For assessment purposes, it is of course more interesting to compare the patterns of international collaboration for the sample of DG INFSO articles and for the benchmarking population of articles. In order to benchmark the propensity to collaborate across national borders we have proceeded as follows:

1) For each of three subject fields used for the benchmarking analysis, we have extracted from the population of articles all those papers in which at least one affiliation from an ERA country was reported;

2) For this subset of articles we have then calculated the fraction of articles that report at least another affiliation from an ERA country (or from another geographical area, such as the USA, Japan and so on);

3) The previous calculations have been performed separately for articles generated in the context of DG INFSO funded projects and for other European articles. A z-test has been performed to evaluate the existence of significant differences in the two proportions.

This type of comparison is reported in Table II.39. The first result emerging from this analysis is that the overall propensity to international collaboration differs among scientific fields. As far the field of computer science is concerned, the fraction of articles that involve a European affiliation *and* another affiliation *from any other country* is significantly higher for articles produced in the context of DG INFSO projects than for the average European article, 38.4% vs. 33.0. Weakly significant statistical evidence that DG INFSO articles are more internationally co-authored is available for the field of physics (47.7% vs. 43.2%), whereas no significant difference between DG INFSO articles and other European articles is observed for the field of materials science (50.6% vs. 47.2%).

The second result emerging from the analysis is that DG INFSO articles are much more likely to involve intra-European collaboration than the average European article in the same field. Of all articles with an ERA affiliation in the field of computer science, 28% of all DG INFSO articles involve at least another affiliation from another ERA country; the same fraction for non-DG INFSO articles is equal to 13.6% and the difference between the two proportions is statistically significant at the 99% level. The difference between the two fractions is larger for the field of materials science, 40.2% vs. 19.9%, and for that of physics, 39.0% vs. 19.5%. Overall, one may argue that on the basis of this analysis, FPs have been highly effective in reinforcing the extent of intra-European collaboration in research.

Table II.38 – Share of papers with at least one EU (27 countries) and one ERA affiliation by 15 science fields

Science field	DG INFSO articles with EU27 affiliation (2004-2008)	% of all articles	DG INFSO articles with ERA affiliation (2004-2008)	% of all articles
Agriculture	2	100.0	2	100.0
Biomedical sciences	217	90.4	225	93.8
Chemistry	311	90.7	321	93.6
Clinical medicine	35	100.0	35	100.0
Computer science	*1763*	*94.1*	*1820*	*97.1*
Ecology	7	87.5	7	87.5
Engineering	107	96.4	109	98.2
Environmental sciences	9	100.0	9	100.0
General medicine	6	100.0	6	100.0
Geosciences	11	100.0	11	100.0
Infectious diseases	5	100.0	5	100.0
Materials sciences	*536*	*91.9*	*572*	*98.1*
Neurosciences	101	82.1	113	91.9
Physics	*678*	*93.1*	*712*	*97.8*
Social sciences	20	80.0	22	88.0
All fields	3808	92.9	3969	96.7

Table II.39 – Patterns of international co-authorship, DG INFSO articles vs. benchmarking population

	Fraction of articles with at least one affiliation in the European Union (EU30)								
	Computer science			Materials science			Physics		
	INFSO	Non-INFSO	z-score	INFSO	Non-INFSO	z-score	INFSO	Non-INFSO	z-score
and at least another affiliation in									
Any other country	38.4	33.0	3.81[a)]	50.6	47.2	1.04	47.7	43.2	1.77[c)]
Any other EU30 country	28.0	13.6	13.93[a)]	40.2	19.9	7.77[a)]	39.0	17.5	10.78[a)]
Any other EU27 country	28.0	13.6	13.97[a)]	40.2	19.9	7.78[a)]	39.0	17.5	10.81[a)]
Any other EU15 country	28.0	13.4	14.17[a)]	40.2	19.7	7.91[a)]	39.0	17.0	11.12[a)]
Any other non-EU country	14.2	22.0	-6.32[a)]	14.9	31.7	-5.56[a)]	15.4	30.7	-6.41[a)]
USA	8.2	11.0	-2.99[a)]	7.1	13.9	-3.05[a)]	5.3	11.3	-3.65[a)]
Japan	0.8	1.6	-2.27[b)]	2.1	3.5	-1.19	1.3	2.3	-1.26
Canada	1.4	2.1	-1.74[c)]	0.4	1.4	-1.32	0.8	1.9	-1.53
Israel	0.9	0.5	1.69[c)]	1.7	0.6	2.07[b)]	2.7	1.2	2.63
Russia	0.5	1.1	-1.83[c)]	1.7	3.0	-1.23	2.4	5.1	-2.38[b)]
China	1.0	2.0	-2.32[b)]	0.8	3.5	-2.25[b)]	0.5	2.9	-2.76[a)]
India	0.0	0.5	-2.35[b)]	0.0	1.1	-1.66[c)]	0.0	0.8	-1.77[c)]
Korea	0.2	0.7	-2.15[a)]	0.8	1.4	-0.76	0.3	0.5	-0.71
Rest of world	2.3	5.2	-4.36[a)]	2.5	7.2	-2.82[a)]	3.7	9.2	-3.66[a)]

[a)], [b)], [c)] difference in proportions significant respectively at the 99%, 95% and 90% levels.

The third main result emerging from the analysis is that DG INFSO articles show a relatively lower propensity towards extra-European collaboration than other European articles. For example, the share of DG INFSO articles with a co-authorship from the USA is equal to 8.2% in computer science, 7.1% in materials science and 5.3% in physics; the same fraction for non-DG INFSO European articles is equal, respectively, to 11.0%, 13.9% and 11.3%.

Overall, these results indicate that, while articles produced by DG INFSO projects have a relatively higher propensity to engage in international collaboration than European articles resulting from other sources, the major difference between the two types of articles relates to the *pattern* of collaboration. More specifically, DG INFSO articles seem to show a higher propensity to engage in intra-European collaboration and a lower propensity to engage in extra-European collaboration than the average European article.

To investigate further this issue, we carried out a simple regression analysis. The propensity to engage in collaborative research across countries may be in fact explained by other factors, such as the number of authors that have participated in the published research. In particular, for each of the three fields used for the benchmarking analysis, we estimated three regressions, in which the dependent variables are defined as follows:

(1) For each article we counted the total number of countries present in the affiliations reported in the article. The dependent variable is thus a count variable that takes value 0 if only one country is reported in the affiliations of the paper, it takes value 1 if one additional country is reported in the affiliations of the paper, and so on. The variable thus captures the extent to which an article is the result of an international collaboration. For this regression analysis, we considered the whole population of articles present in our benchmarking data set.

(2) For each article reporting (at least) an affiliation from the ERA (EU30) area we counted the number of other countries from the ERA (EU30) area reported in the affiliations of the article. The dependent variable is a count variable, which takes value 0 if only one ERA country is reported in the affiliations of the article, it takes value 1 if one additional ERA country is reported in the affiliations and so on. The dependent variable thus captures the propensity to engage in intra-European collaboration.

(3) For each article reporting (at least) an affiliation from the ERA (EU30) area we counted the number of extra-European countries (i.e. countries not belonging to the ERA area) reported in the affiliations of the article. The dependent variable is a count variable, which takes value 0 if only one ERA country is reported in the affiliations of the article, it takes value 1 if one additional extra-ERA country is reported in the affiliations and so on. The dependent variable thus captures the propensity to engage in extra-European collaboration.

As far as the explanatory variables are concerned, we included in the regression analysis all control variables already used for the impact analysis (see above section II.3.1). In addition to this, for regression (1) we included a dummy variable which takes value 1 if the article reported (at least) an affiliation from an ERA country, and 0 else. As long as European articles show a higher propensity to result from international collaboration than non-European articles, this dummy variable should present a positive and statistically significant sign. Moreover, we also included in the regression an interaction term between the dummy variable defined above and a dummy variable taking value 1 if the article was produced within a DG INFSO project and 0 else. As long as European articles resulting from DG INFSO project show a higher propensity to come from international collaboration than other European articles, this interaction term should also present a positive and statistically significant sign.

Regarding regressions (2) and (3), we simply included a dummy variable taking value 1 if the article was produced within a DG INFSO project and 0 else. As long as European articles resulting from DG INFSO project show a higher propensity to come from intra-European or extra-European collaboration, this dummy variables should present a positive and statistically significant sign.

Given the limited dependent nature of the dependent variables, we estimated a Poisson regression model. Results are reported in Table II.40, Table II.41 and Table II.42.

Table II.40 – Poisson regression of the propensity to international collaboration, Computer science

	International collaboration (1)	Intra-European collaboration (2)	Extra-European collaboration (3)
Average impact factor (2004-08)	0.175***	0.187***	0.238***
	(0.00862)	(0.0178)	(0.0148)
Number of cited references	0.00411***	0.00279***	0.00442***
	(0.000356)	(0.000741)	(0.000560)
Number of pages	0.00875***	0.00544***	0.00641***
	(0.000851)	(0.00194)	(0.00151)
Number of authors	0.0476***	0.0500***	0.0328***
	(0.000546)	(0.000770)	(0.00127)
European (ERA30) affiliation	1.230***		
	(0.0135)		
European (ERA30) affiliation*DG INFSO article	0.184***		
	(0.0419)		
DG INFSO article		0.775***	-0.509***
		(0.0508)	(0.0793)
Constant	-2.740***	-2.625***	-1.960***
	(0.0294)	(0.0618)	(0.0498)
Publication year fixed effect	Yes	Yes	Yes
Subject category fixed effect	Yes	Yes	Yes
Article type fixed effect	Yes	Yes	Yes
Nr observations	116090	40086	40086
Log-likelihood	-60043.9	-18933.3	-21596.9

*, **, *** statistically significant at the 90%, 95% and 99% levels. Please note that regression (1) includes all articles published in this field, while regressions (2) and (3) include only articles with at least one affiliation from the ERA area.

Table II.41 – Poisson regression of the propensity to international collaboration, Materials science

	International collaboration (1)	Intra-European collaboration (2)	Extra-European collaboration (3)
Average impact factor (2004-08)	-0.0836***	-0.0501	-0.0433
	(0.0156)	(0.0311)	(0.0274)
Number of cited references	0.00226***	0.00192**	0.00127
	(0.000532)	(0.000886)	(0.000945)
Number of pages	-0.00242	0.0132	-0.00240
	(0.00583)	(0.0111)	(0.0102)
Number of authors	0.141***	0.172***	0.0905***
	(0.00284)	(0.00486)	(0.00513)
European (ERA30) affiliation	1.340***		
	(0.0196)		
European (ERA30) affiliation*DG INFSO article	-0.0354		
	(0.0807)		
DG INFSO article		0.566***	-0.843***
		(0.0952)	(0.168)
Constant	-2.499***	-2.376***	-1.584***
	(0.0731)	(0.142)	(0.129)
Publication year fixed effect	Yes	Yes	Yes
Article type fixed effect	Yes	Yes	Yes
Nr observations	42237	13536	13536
Log-likelihood	-24531.7	-7707.5	-9008.6

*, **, *** statistically significant at the 90%, 95% and 99% levels. Please note that regression (1) includes all articles published in this field, while regressions (2) and (3) include only articles with at least one affiliation from the ERA area.

Table II.42– Poisson regression of the propensity to international collaboration, Physics

	International collaboration (1)	Intra-European collaboration (2)	Extra-European collaboration (3)
Average impact factor (2004-08)	0.0777***	0.140***	0.00580
	(0.00868)	(0.0174)	(0.0141)
Number of cited references	0.0132***	0.0129***	0.0101***
	(0.000574)	(0.00104)	(0.000915)
Number of pages	-0.00762***	-0.00187	-0.00781**
	(0.00226)	(0.00406)	(0.00356)
Number of authors	0.0598***	0.0620***	0.0468***
	(0.00103)	(0.00160)	(0.00204)
European (ERA30) affiliation	1.431***		
	(0.0192)		
European (ERA30) affiliation*DG INFSO article	0.132**		
	(0.0643)		
DG INFSO article		0.788***	-0.690***
		(0.0766)	(0.132)
Constant	-2.738***	-2.550***	-1.607***
	(0.0366)	(0.0693)	(0.0564)
Publication year fixed effect	Yes	Yes	Yes
Article type fixed effect	Yes	Yes	Yes
Nr observations	50454	19402	19402
Log-likelihood	-30171.4	-10509.6	-12705.8

*, **, *** statistically significant at the 90%, 95% and 99% levels. Please note that regression (1) includes all articles published in this field, while regressions (2) and (3) include only articles with at least one affiliation from the ERA area.

They tend to confirm the findings of the descriptive analysis. Looking first at regressions (1), we note that European articles, i.e. articles with at least one affiliation from an ERA country, tend to show a higher propensity towards international co-authorship than non-European articles. However, the positive sign of the interaction term between ERA affiliation and DG INFSO article suggests that this tendency is even larger for European articles resulting from projects supported by DG INFSO, with the exception of the field materials science.

Turning the attention to regressions (2), we note that in all the three fields examined articles from DG INFSO projects have a higher propensity to engage in intra-European collaboration than other European articles. On the other hand, from regressions (3) we also note that DG INFSO articles present a relatively lower propensity to engage in extra-European collaboration than other European articles.

II.4.2 Regional patterns of scientific publication

The last step in our assessment concerns the analysis of the regional distribution of the scientific output produced by DG INFSO projects. To this purpose, we have examined the affiliations reported in each article and we have allocated them to the corresponding NUTS2 and NUTS3 regions.

A full (or integer) counting method has been adopted in order to count the number of articles by region. For example, assuming that an article has reported two affiliations, one from region X and one from region Y, one article has been attributed to region X and one article to region Y. However, assuming that an article has reported three different affiliations, two from region X and one from region Y, only one article has been attributed to region X, and one article has been credited to region Y.

From the set of all articles produced by DG INFSO projects, we have extracted the subset of papers in which at least one author was from the EU27 area. For this subset of articles, we have then calculated the share of all articles produced by the top 10 most productive NUTS2 and NUTS3 regions. In addition to this, we have also calculated the Herfindahl index, i.e. the sum of the squared shares, to capture the extent of concentration and asymmetry in the distribution across regions, as well as the number of regions with a non-null scientific publication activity.

Results are reported in Table II.43 for the 15 science fields. Focussing the attention on the three most important science fields, we observe that the share of the top 10 most productive NUTS3 regions ranges from 25% for Computer science to 31% for Physics, while the share of the top 10 most productive NUTS2 regions goes from 31% for Computer science to 34% for Materials science. The fact that the share of the top 10 regions increases only slightly going from the NUTS3 to the NUTS2 levels suggests that the publication activity is diffused across a relatively large number of areas or, put it in a different way, it is not concentrated in few areas within the same NUTS2 regions. This intuition seems to find confirmation in the relatively low values of the Herfindahl index and in the large number of regions with a non-null scientific production.

The share of NUTS3 regions with a non-null scientific production is equal to 24% in Computer Science and to about 13% for Physics and Materials science[21]. However, the share of NUTS2 regions with a non-null scientific production is equal to 69% for Computer Science and to about 48% for Physics and Materials science.

Of course, the interesting issue is to what extent this regional pattern deviates from the pattern we observe in the general population of articles. To this purpose, Table II.44 compares the regional patterns of scientific activity in the sample of DG INFSO articles and in the benchmarking population, for the six subject categories adopted for benchmarking purposes. The most striking result emerging from the data is that the spatial concentration in the sample of DG INFSO articles is significantly higher (around twice as large) than the spatial concentration in the general population of articles. Put it in a different way, there seems to be a large number of regions with a minimal scientific capability that are not represented among DG INFSO projects. Thus, for example, in the field of Computer Science, Hardware & Architecture, there are 88 NUTS3 regions with a non-null scientific production represented within DG INFSO projects, while the number of NUTS3 regions with a non-null scientific production in the general population is equal to 361. In other words, DG INFSO projects are able to include only around 24% of all European regions with a minimal scientific capability.

[21] Taking the EU 27 countries, there are 271 NUTS2 regions and 1303 NUTS3 regions.

Table II.43 – Distribution of scientific articles by regions, 15 science fields (full count)

Field	NUTS 3			NUTS 2		
	C10	H	Nr. regions	C10	H	Nr. regions
Agriculture	100.0	0.50	2	100.0	0.50	2
Biomedical sciences	34.6	0.02	131	38.7	0.02	104
Chemistry	29.3	0.02	155	35.3	0.02	123
Clinical medicine	46.4	0.04	39	50.0	0.04	35
Computer science	24.7	0.01	318	30.8	0.02	187
Ecology	100.0	0.21	6	100.0	0.21	6
Engineering	30.5	0.02	97	36.0	0.02	83
Environmental sciences	83.3	0.08	12	100.0	0.13	10
General medicine	85.7	0.09	12	92.9	0.11	11
Geosciences	73.7	0.07	15	73.7	0.07	15
Infectious diseases	100.0	0.10	10	100.0	0.10	10
Materials sciences	30.1	0.02	175	33.8	0.02	129
Neurosciences	35.6	0.02	76	40.0	0.03	66
Physics	31.0	0.02	167	33.2	0.02	132
Social sciences	48.4	0.05	26	54.8	0.05	24

The table reports the share of the top 10 most productive NUTS2 and NUTS3 regions. It also reports the Herfindahl index, calculated as the sum of the squared shares of all regions, as well as the number of regions with a non-null scientific production.

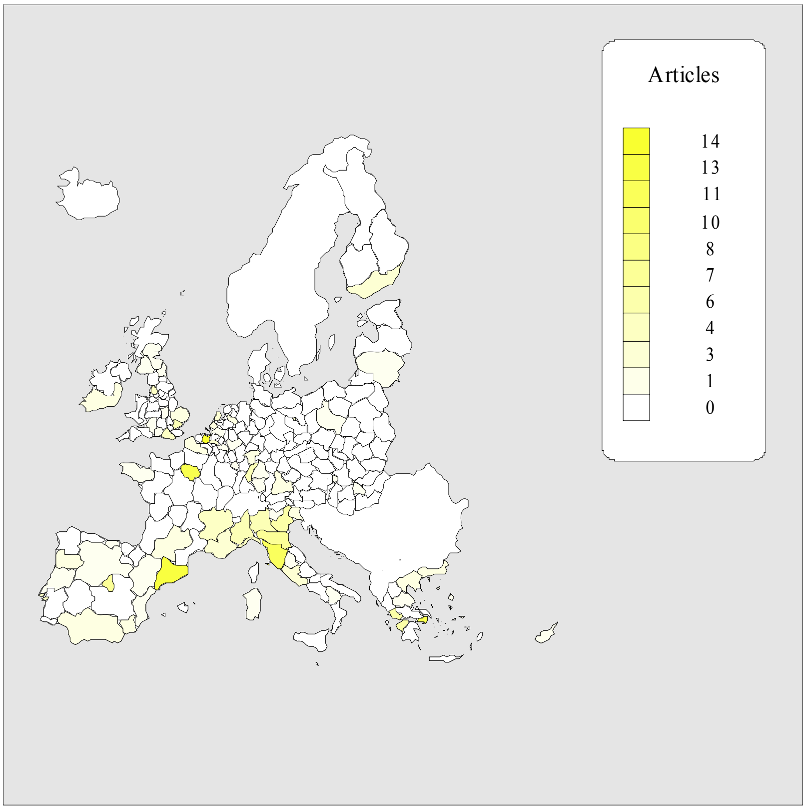
Table II.44 – Distribution of scientific articles by regions, six subject categories science fields (full count)

Field	NUTS 3			NUTS 2		
	C10	H	Nr. regions	C10	H	Nr. regions
DG INFSO articles						
Computer Science, Hardware & Architecture	39.3	0.024	88	42.5	0.027	78
Computer Science, Information Systems	34.2	0.020	96	38.7	0.024	81
Computer Science, Software Engineering	30.7	0.019	77	39.3	0.026	62
Engineering, Electrical & Electronic	30.4	0.016	200	35.9	0.021	137
Optics	32.0	0.018	117	33.7	0.020	99
Physics, Applied	36.0	0.021	114	39.2	0.027	95
Benchmarking population						
Computer Science, Hardware & Architecture	20.7	0.009	361	23.8	0.013	205
Computer Science, Information Systems	17.7	0.007	438	22.8	0.012	229
Computer Science, Software Engineering	18.3	0.007	385	21.5	0.011	220
Engineering, Electrical & Electronic	19.2	0.008	536	24.4	0.013	233
Optics	20.4	0.009	455	25.4	0.014	228
Physics, Applied	22.6	0.010	418	27.1	0.014	215

The table reports the share of the top 10 most productive NUTS2 and NUTS3 regions. It also reports the Herfindahl index, calculated as the sum of the squared shares of all regions, as well as the number of regions with a non-null scientific production.

Figure II.4 – Number of articles by NUTS 2 regions, Computer science, Hardw & Architecture

DG INFSO articles



Benchmarking population

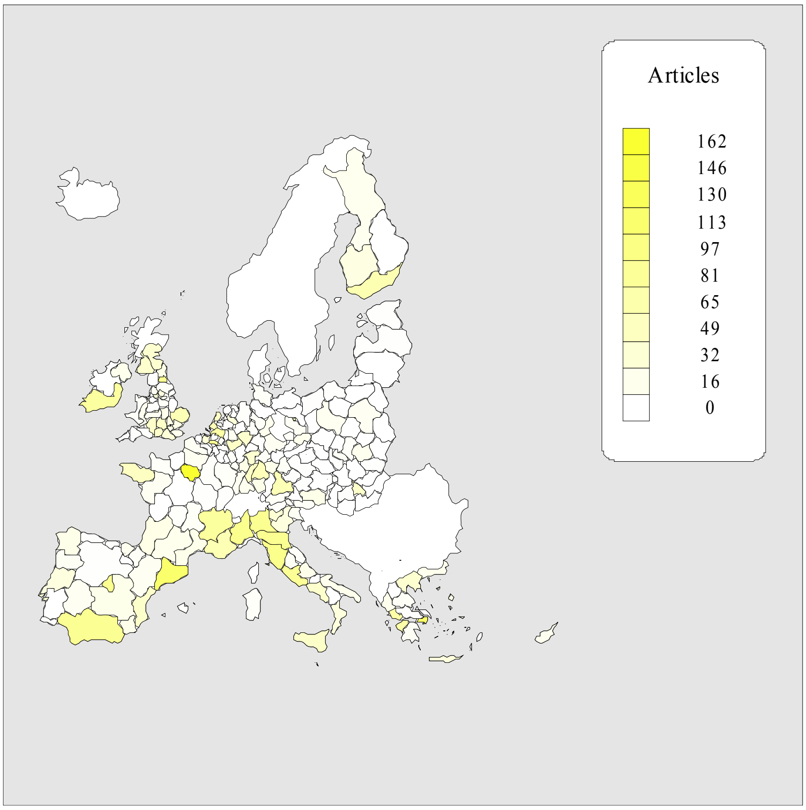
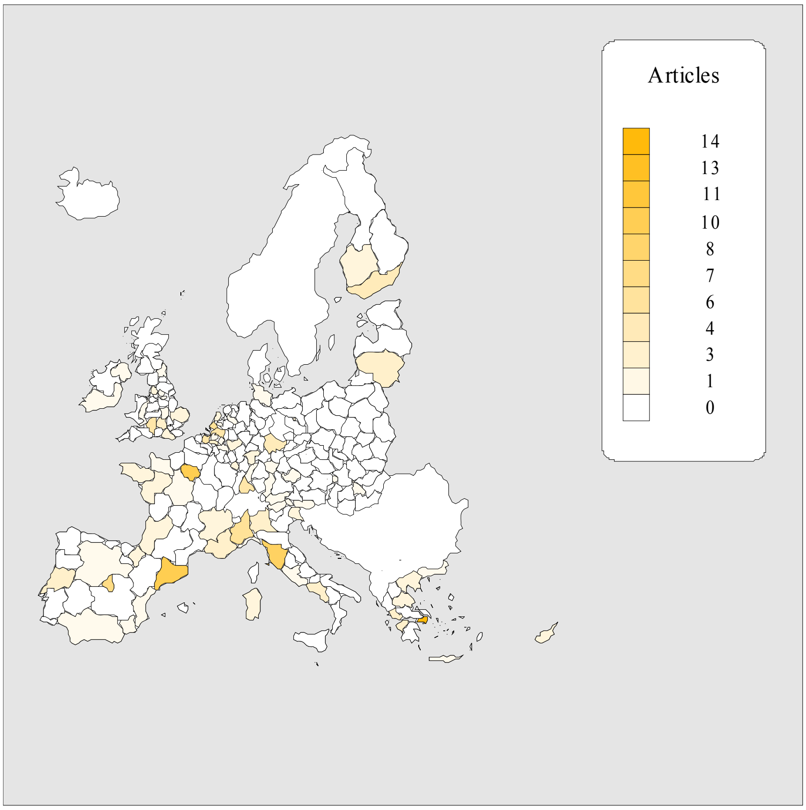


Figure II.5 – Number of articles by NUTS 2 regions, Computer science, Information Systems

DG INFSO articles



Benchmarking population

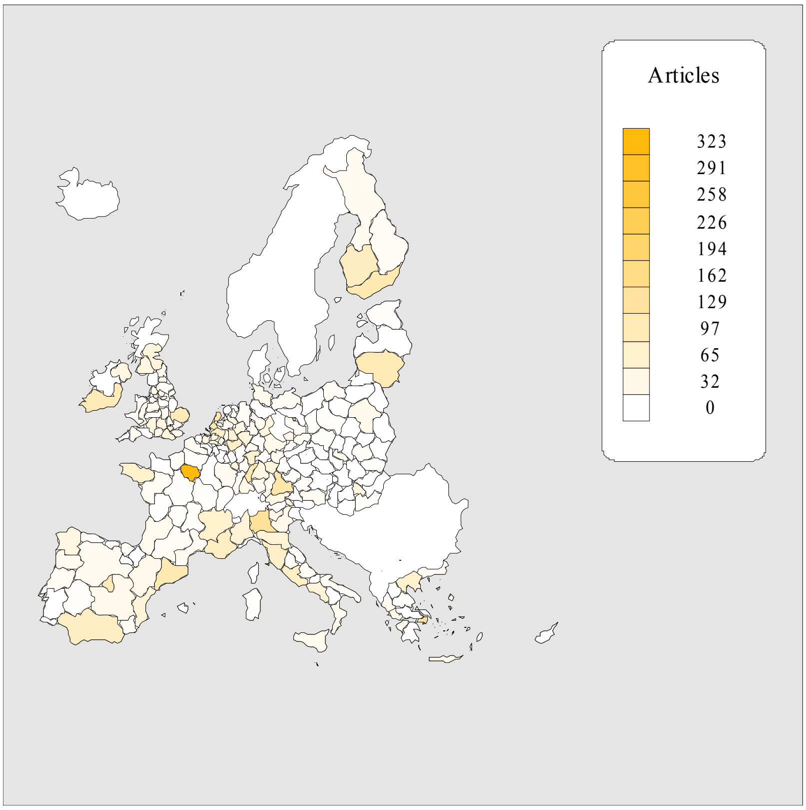
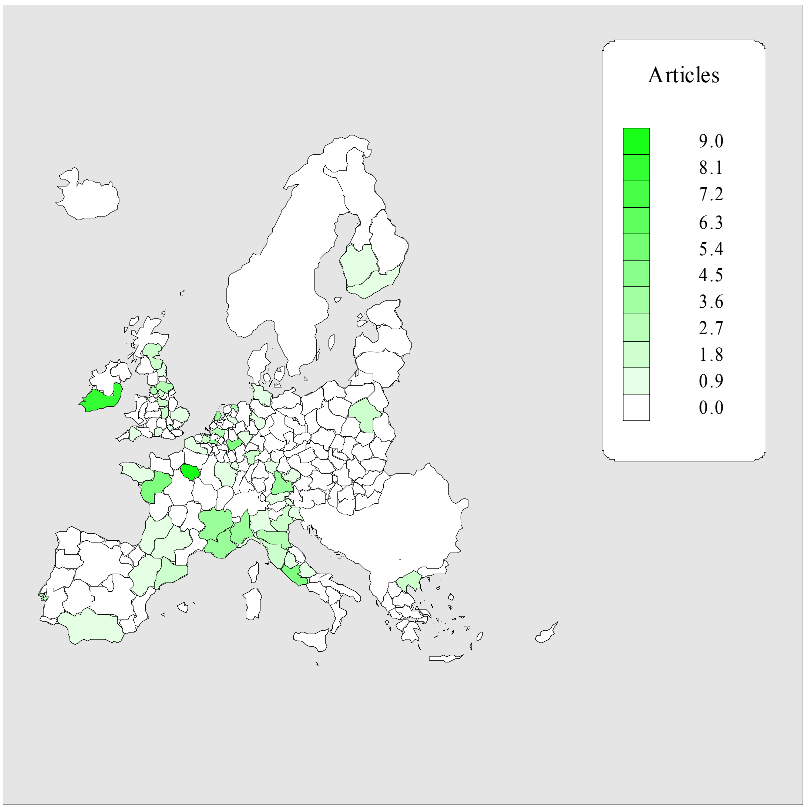


Figure II.6 – Number of articles by NUTS 2 regions, Computer science, Software engineering

DG INFSO articles



Benchmarking population

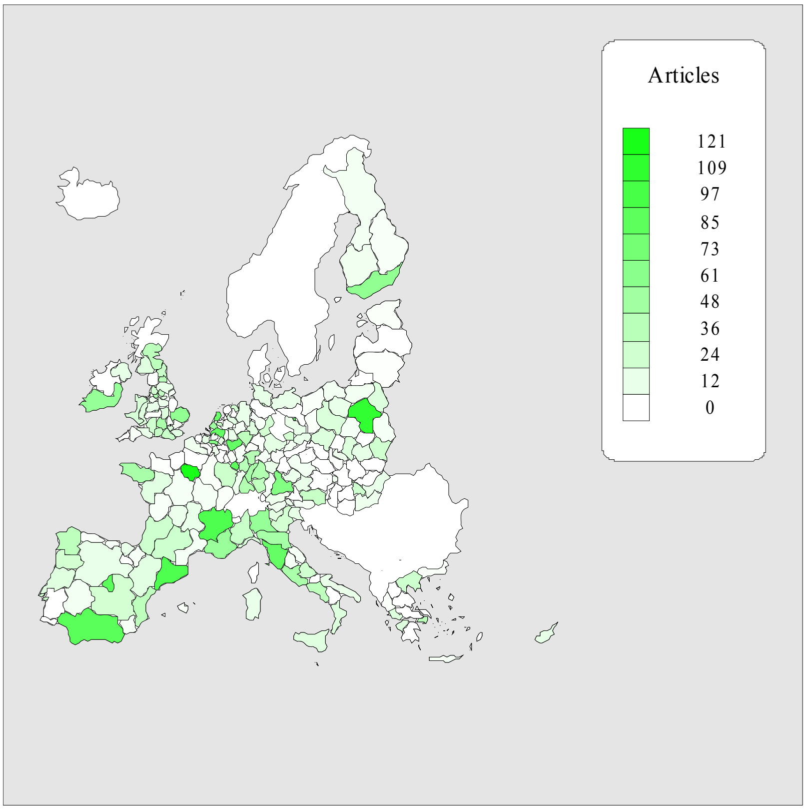
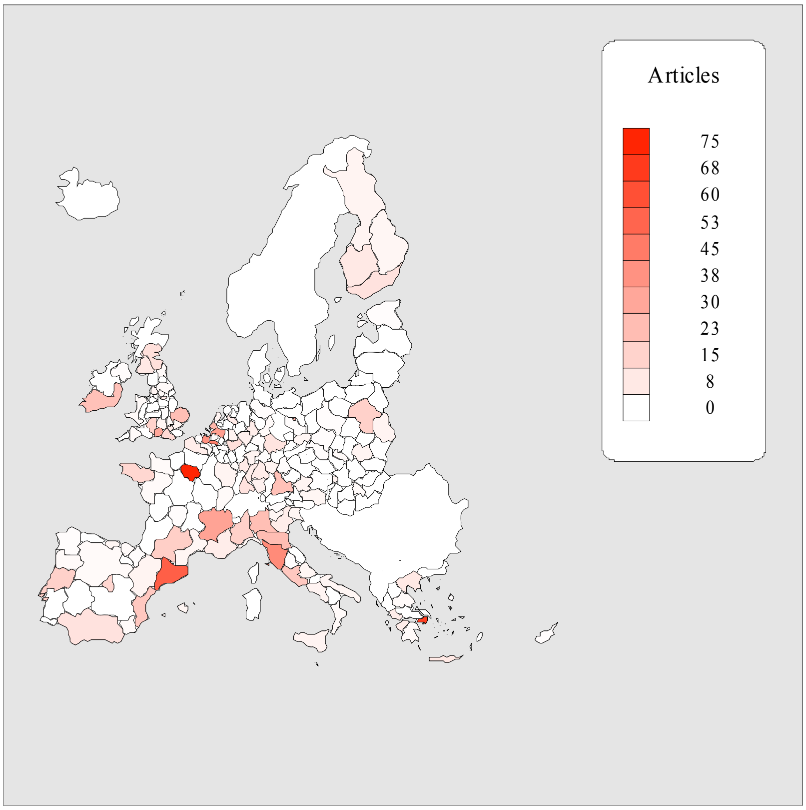


Figure II.7 – Number of articles by NUTS 2 regions, Engineering, electrical & electronic

DG INFSO articles



Benchmarking population

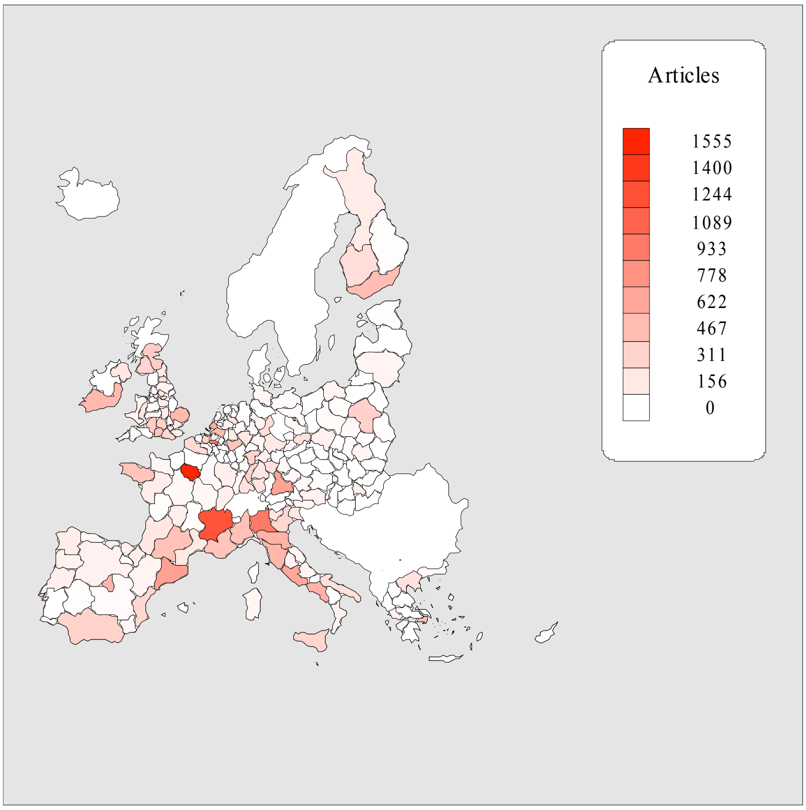
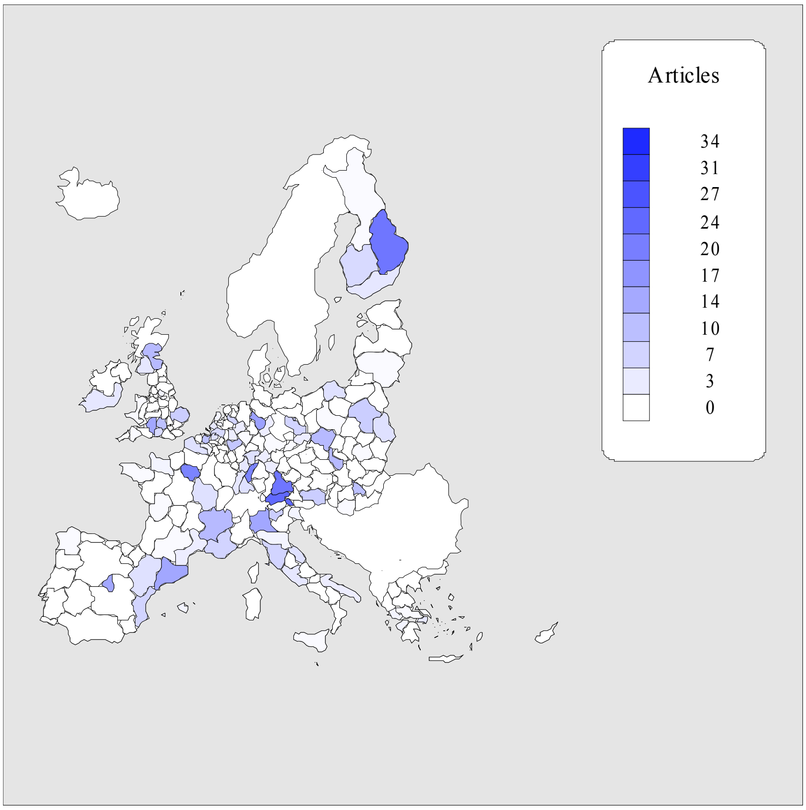


Figure II.8 – Number of articles by NUTS 2 regions, Optics

DG INFSO articles



Benchmarking population

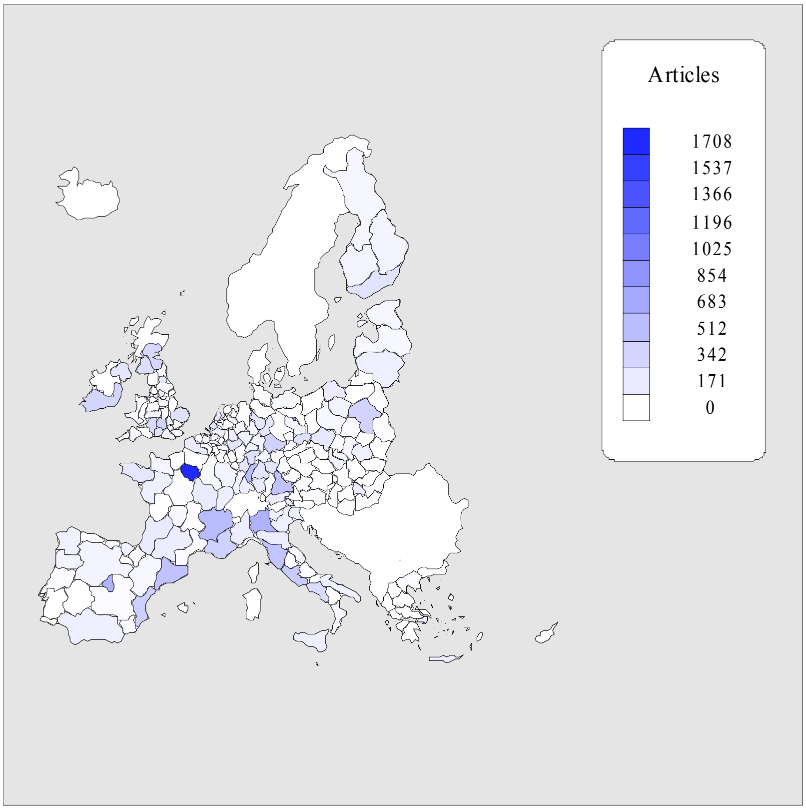
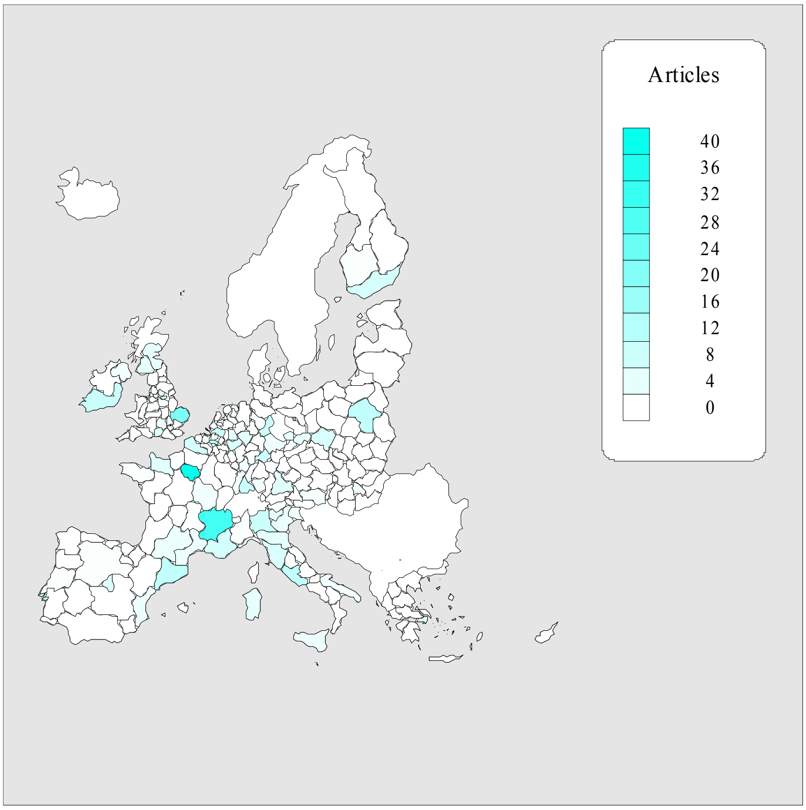


Figure II.9 – Number of articles by NUTS 2 regions, Applied physics

DG INFSO articles



Benchmarking population

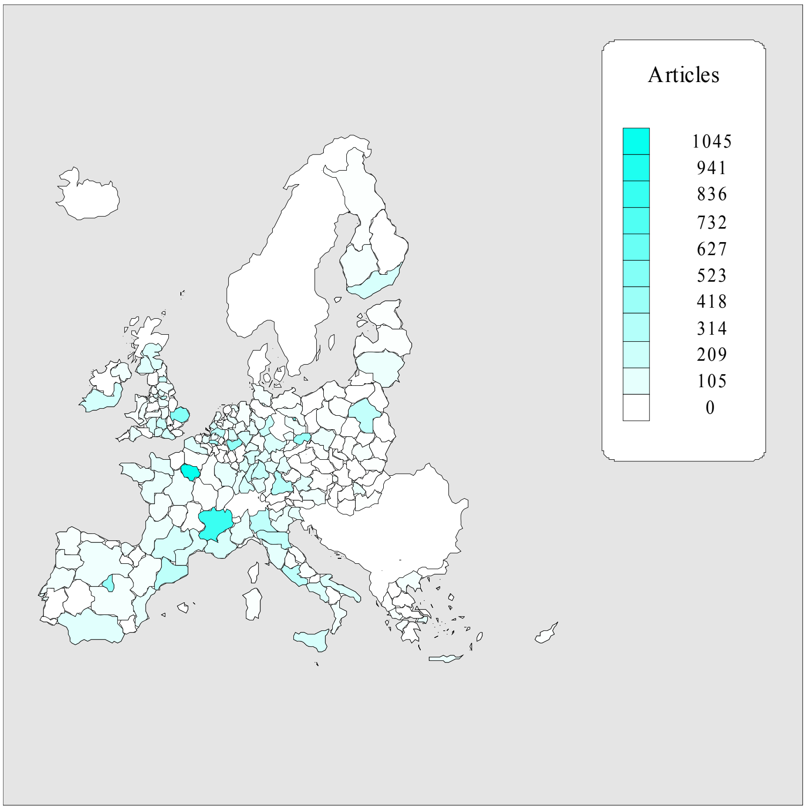


Figure II.4 to II.9 report the spatial distribution of the number of scientific articles, in the case of DG INFSO projects and for the general population, at the NUTS2 levels for the six subject categories used for the benchmarking exercise. The general impression one can derives from observing the maps is that the, despite the higher concentration of scientific output in the sample of DG INFSO articles, the spatial patterns of scientific activity are quite similar. In particular, it is quite interesting to observe both in the context of DG INFSO projects and in the general population, the existence of a sort of regional belt which extends from the south of Spain, to northern and central Italian regions, and to Greece. The similarity in the spatial patterns of scientific activity is confirmed by the calculation of a linear or rank correlation coefficient between the number of articles produced by NUTS2 regions within DG INFSO projects and in the general population. The linear correlation coefficient ranges from 0.62 in Optics to 0.85 in Engineering, Electrical & Electronic, and it is higher than 0.7 in the other fields.

III. Patent analysis

III.1 Database description

The assessment of the technological performance of DG INFSO projects is based upon an analysis of the patent applications reported by the coordinators of these projects. As for the scientific literature, one should be aware of the advantages and limitations of patent data as an indicator of innovative activity. Briefly speaking, the advantages of patent based indicators relate to the fact that they are readily available, they are close to the invention, they cover a broad range of technologies and provide very detailed information on the applicant, inventor, technological fields and so on. On the other hand, the drawbacks of patent based indicators relate to the fact that many inventions are not patented, the propensity to patent and the effectiveness of patents vary across industries, the value of patents is highly skewed and patent regulations differ across countries and patent systems .

It is worth noting that the cross-industry variation in the propensity and effectiveness of patents is less likely to represent a problem in this context given the focus of this study on the ICT industry. As far as the differences in regulations are concerned, for most part of the analysis we will focus on patent applications at the European Patent Office (EPO) in order to minimise the problem.

As explained more in detail in the technical appendix, all patent applications reported by project coordinators as a technological output of the projects have been checked using the Espacenet database[22]. For each patent application, further information on the patent office where the document has been registered, the filing and publication dates, the applicant and inventor names, and the technological class of the invention have been collected. Moreover, equivalent patents, i.e. patents published by different patent offices, but relating to the same invention and sharing the same priority date have been consolidated.

Table III.1 illustrates the basic descriptive statistics of the database on patents reported by DG INFSO projects. Out of 1417 projects, 179 project co-coordinators reported detailed information on at least one patent application[23]. Of the 179 projects that reported at least one patent application, 46 were FP5 projects (i.e. 23.1% of all FP5 projects in the data set), while 133 were FP6 projects (i.e. 14.4% of all FP6 projects in the data set). Overall, we identified 457 unique patents, 89 of which were patent applications by FP5 projects and 368 were applications by FP6 projects. Interestingly enough, no patent application was reported by FP7 projects. Similarly to publications, a few patents were reported as output in different projects. More specifically, four patents that were reported in a FP5 project were also reported in another FP6 project and two patent applications were reported in two different FP6 projects, so that the gross number of patents reported by FP6 projects (i.e. taking into account the double counting) is equal to 374.

It is important to observe that a few project coordinators have reported patent applications whose priority date is prior with respect to the starting date of the project. Therefore, these patents cannot be considered as a result of the funded research. Although one might be tempted to interpret this phenomenon as the outcome of the incentive for pro-

[22] The Espacenet database (http://www.espacenet.com/index.en.htm) provides free access to more than 60 million patent documents all over the world.

[23] Please note that 44 other projects filled the questionnaire on patents with some information. However, the information reported in the survey was not sufficient to identify in a univocal way the patents produced. For example, some project co-coordinators only reported that a certain number of patent applications were filed, but they did not report the patent application or publication number, or the patent office in which the application was filed and so on.

ject coordinators to overstate the achievements of their projects, a casual look at some of these patents reveals that they might be part of the so-called *background knowledge* of the project. According to the FP6 IPR guidelines a project participant must grant access rights to its knowledge (or pre-existing know-how) to another participant if the latter needs such access rights in order to carry out its own work under the project or to use its own knowledge resulting from the project. It is thus likely that most of the patents filed *prior* to the start of the project are those that have been negotiated in the consortium agreement of the project. In particular, of the 374 patent applications reported by coordinators of FP6 projects 273 (73%) have a priority date which is posterior with respect to the starting date of the project, whereas 101 (27%) have a priority date which is prior with respect to the starting date of the project.

Table III.1 – Summary statistics of the patent database

Number of DG INFSO projects	1417 (100%)
of which funded under	
FP5	199 (14.0%)
FP6	927 (65.4%)
FP7	291 (20.6%)
Number of projects reporting application of at least one patent of which	179 (12.5%)
FP5 projects	46 (23.1%)
FP6 projects	133 (14.4%)
Number of unique patent applications	457 (100%)
Number of patent applications by FP5 projects	89 (19.5%)
Number of patent applications by FP6 projects	368 (80.5%)

If we consider as patents generated by DG INFSO projects only those patent applications whose priority date is posterior with respect to the starting date of the project, the number of FP6 projects that reported at least one *new* patent application reduces from 133 to 109. In other words, 24 projects reported only patent applications whose priority date was before the starting date of the project. Likewise, of the 89 FP5 projects that reported at least one patent application in the questionnaire 41 reported patents whose priority was posterior with respect to the starting date of the project, while 48 reported only patents whose priority was before with respect to the starting date of the project. Although the interpretation given above for this phenomenon (i.e. the fact that patents with a priority date before the starting date of the project represent the so-called *background knowledge*) may be reasonable, further enquiry is probably needed to confirm it. More generally, this observation points once again to the need of improving the quality of the data collection process. Project coordinators should be instructed to distinguish between patents that refer to the so-called background knowledge, and that therefore are not an outcome of the project, from patents that can be treated as resulting from the research activities of the project. Moreover, even though the extent of manual data cleaning involved was lower than in the case of scientific publications, due also to the smaller number of items to process, the work of checking the consistency of the data reported and collecting further information on the patent applications has been far from negligible. In the technical appendix, we formulate a concrete proposal on how to improve the data collection process on patents.

In what follows, we will refer to *old* patents to denote those patent applications whose priority date is before the starting date of the project that reported them, whereas we will use the term *new* patents to identify those patent applications whose priority date is after the starting date of the project that reported them.

Table III.2 reports the distribution of the 457 patent applications (both *old* and *new*) by patent office. In this respect, it has to be observed that the same patent can be applied to different patent offices. For example, an application is first filed at the USPO and then extended to the EPO and to the JPO. In this case, one says that the USPO patent has an *equivalent* EPO patent and an *equivalent* JPO patent. In order to calculate the distribution of patents by patent office, one has therefore to take into account this aspect of the patent system and to adopt a rule on how to count patents having equivalents. In the calculations reported in the first column of Table III.2 we have adopted the following rule: patents having *equivalents* have been ordered in the following way: EPO, USPTO, WO and national office. For example, an EPO patent having a USPTO equivalent has been counted only once as an EPO patent, and not as a USPTO patent. Likewise, a USPTO having a German equivalent, but not an EPO equivalent has been counted as a USPTO patent, and not as a German one; and so on. In this way, each patent is counted only once, avoiding double counting when the patent has equivalent patents at different patent offices. Looking at the distribution, we note that the vast majority of patent applications by DG INFSO projects have been filed to the European Patent Office (297 patents or 65% of all patents), followed by the USPTO and the WIPO (PCT) procedure. The following columns of Table III.2 report the number of equivalent patents. Thus, for example, of the 297 patents at the European Patent Office, 162 (55%) were also filed at the USPTO. Please note that the table does not specify how many of the 162 patents were first filed at the USPTO and then extended to the EPO (or vice versa).

Table III.2 – Distribution of patent applications by main patent office

Patent Office	Patents	%	of which having equivalents at					
			US	WO	JP	DE	GB	FR
European Patent Office	297	65.0	162	142	83	37	4	33
United States Patent Office	68	14.8	-	23	17	10	1	2
Wipo (PCT)	59	12.9	-	-	0	7	1	4
German Patent Office	15	3.3	-	-	0	0	0	0
UK Patent Office	3	0.7	-	-	-	0	0	0
French Patent Office	3	0.7	-	-	-	-	0	0
Others	12	2.6	-	-	-	-	-	-
Total	457	100.0	-	-	-	-	-	-

Note: the first column reports the "net" number of patent applications; in order to avoid double counting, patents having equivalents have been ranked in the following order: EP/US/WO/national office. For example, patent EP1552346 having an equivalent WO03102695 has been counted as an EPO patent and not as a WO patent.

Table III.3 reports the average number of *new* patent applications per project, i.e. discarding applications with priority date before the starting date of the project and considering only projects with a non-null patent production. Overall 109 FP6 projects reported 273 *new* patent applications, while 41 FP5 projects reported 73 *new* patent applications. This means that the average number of patents per project goes from slightly less than 2 for FP5 projects to around 2.5 for FP6 projects. These results suggest that a larger fraction of FP5 projects have made *new* patent applications compared to FP6 projects (i.e. 109/927=12% vs. 41/199=21%), but that the patent productivity of FP6 projects is larger than that of FP5 projects. Yet, this comparison has to be interpreted with some caution due to the low number of FP5 projects included in the sample.

Table III.3 – Average number of *new* patent applications per project

	Number of projects with *new* patents	Average number of patents	Standard deviation	Min	Max
FP5 projects	41	1.78	1.29	1	7
FP6 projects	109	2.50	2.84	1	20

Note: new patents are defined as those patents whose priority date is posterior with respect to the starting date of the project that reported them.

Of the 273 *new* patent applications of FP6 projects, 175 are either direct or extended patent applications to the EPO; similarly, of the 73 new patent applications introduced by FP5 projects, 43 are either direct or extended patent applications to the EPO. Overall, FP5 and FP6 projects produced 218 new patent applications to the EPO.

III.2 An assessment of the patent productivity

In this section, we focus the analysis on the patent productivity of FP6 projects that represent 80% of all patents in our sample. Table III.4 reports the distribution of patent applications by funding instrument and by strategic objective. Please note that in doing this exercise we are making the implicit assumption that the patents in our sample are *countable*, i.e. that an EPO patent is comparable to a USPTO or to a national patent. Starting from the top panel of the table, we note the following points:

- Only three instruments have produced new patent applications, i.e. NoE, IPs and STREPs. Projects funded through CA, SSA and I3 have not reported the production of any patent.
- IP projects account for 49% of all new patent applications and around 40% of all projects with new patents. In a similar way, NoEs account for 10% of all new patent applications and 13% of all projects with patents, while STRePs account for 40% of all new patent applications and 47% of all projects with new patent applications.

Table III.4 – Distribution of *new* patents by funding instrument and strategic objective, FP6 projects

	Number of patents	%	Number of projects with patents	Mean number of patents per project
Funding instrument				
Integrated projects (IP)	134	49.1	44	3.05
Networks of excellence (NoE)	29	10.6	14	2.07
Strategic Targeted Research Projects (STREP)	110	40.3	51	2.16
Total	273	100.0	109	2.50
Strategic objective				
Applied IST res. addressing major societal and economic challenges	33	12.1	21	1.57
Communication, computing and software technologies	103	37.7	28	3.68
Components and micro-systems	88	32.2	35	2.51
IST Future and emerging technologies	14	5.2	8	1.75
Knowledge and interface technologies	11	4.0	8	1.38
Nanotechnologies and nanosciences	24	8.8	9	2.67
Total	273	100.0	109	2.50

Note: new patents are defined as those patents whose priority date is posterior with respect to the starting date of the project that reported them.

- The average productivity in terms of new patent applications differs across the three funding instruments. In particular, the average number of patent applications per project (excluding projects with no patents) is larger for IPs (3.05 new patents per project) than for NoEs (2.07) and STREPs (2.16).
- The propensity to patent is apparently higher for IPs and NoEs than for STRePs. Around 23% (i.e. 44/188) of all IP projects and 29% (i.e. 14/49) of all NoE projects reported new patent applications as compared to 10% (i.e. 51/497) of all STReP projects. However, a correct assessment of the different propensity to patent across projects has to take into account different size of the projects funded through the different instruments.

Looking at the strategic objectives, projects funded in the areas *Communication, computing and software technologies* and *Components and micro-systems* present a higher productivity in terms of new patents than projects in other areas. These two areas account for almost 70% of all new patent applications produced by FP6 projects. Not surprisingly, no patent application has been produced by projects in the area of international research co-operation and in the area of policy related research.

As already noted for scientific publications, these productivity differences may actually arise from factors such as the size of the project, the amount of funding and so on. For example, the fact that IP projects report a larger average number of patents than STRePs might be simply due to the fact that those projects have also a larger number of participants and a higher amount of financial resources. In order to control for these confounding factors, we have performed a regression analysis similar to the one carried out for scientific publications (see section II.2.5). In particular, we have estimated a negative binomial regression model in which the dependent variable has been defined has the number of "new" patent applications produced by a project. The covariates included in the analysis are the same used in the regression models for the number of scientific publications (see section II.2.5). In addition to this, we have included in the model the number of "old" patents reported by the project, i.e. the number of patents with priority date before the starting date of the project, which are presumably part of the so-called background knowledge. Before discussing the results, it is worth noting that we had to drop from the analysis all FP6 projects funded through CA, SSA and I3, as well as all projects funded in the strategic areas related to international co-operation and to policy related research[24]. The reason is that *all* projects funded through these instruments/areas have a null patent production, so that the outcome is perfectly determined. Maximum likelihood estimate is in fact not possible when the dependent variable does not vary within one of the categories of an independent variable. After these eliminations, 723 FP6 projects have been considered.

Results reported in Table III.5 seem to confirm also in the case of patents the existence of a curvilinear relationship between project size and productivity, although the coefficients on the number of participants and its squared value are not statistically significant at the conventional levels. The existence of diminishing returns with respect to the scale of projects seems to emerge also from the results reported in column 3. In particular, the value lower than one for the coefficient on the amount of funding suggests that the expected number of patents per project increases less than proportionally for a given increase in the amount of financial resources. Yet, it has to be also observed that the coefficient is not anymore statistically significant once the dummy variables for the funding instruments are introduced in the regression (column 4). Finally, the importance of the project scale in affecting the patent productivity is confirmed by the positive and statistically significant coefficient on the amount of funding per participants (see columns 1 and 2). Not unexpectedly, results also suggest that projects hosting a larger share of universities tend to apply for a lower number of patents. Moreover, it is quite interesting to observe that projects relying on a larger pool of “prior” patents seem to be also more productive in terms of future patents produced by the project. Not unexpectedly, projects hosting a larger share of universities tend to apply for a lower number of patents, though the coefficient is only weakly significant in some regressions and not significantly different from zero in others. As far as the impact of funding instruments is concerned, the coefficient on the dummy variable for IPs is positive but not statistically significant thereby suggesting the absence of differences in patent productivity between IPs and STRePs. On other hand, the positive and statistically significant sign of the coefficient for NoEs indicates that these projects have been on average more productive than STRePs and IPs, once accounting for the size of projects.

[24] In particular, we had to drop 10 STRePs in the area of international co-operation.

Table III.5 – Negative binomial regression of the number of *new* patents, FP6 projects

Variables	(1)	(2)	(3)	(4)
Number of participants	0.0414	0.0471		
	(0.0318)	(0.0374)		
Number of participants [squared]	-0.000956*	-0.00104		
	(0.000575)	(0.000670)		
Funding per participant (log)	1.153***	1.244***		
	(0.411)	(0.433)		
Total project funding (log)			0.723***	0.529
			(0.222)	(0.363)
Asymmetry in funding distribution		0.0221		
		(0.0257)		
Share of universities	-0.00880	-0.0135**	-0.00493	-0.0103*
	(0.00597)	(0.00640)	(0.00527)	(0.00567)
Duration	-0.0118	-0.0214	-0.00296	-0.00987
	(0.0171)	(0.0185)	(0.0175)	(0.0181)
Number of "old" patents	0.542*	0.480	0.632**	0.599**
	(0.282)	(0.312)	(0.303)	(0.294)
Funding instrument (baseline=STREP)				
Integrated projects (IP)	0.137	0.399		0.383
	(0.491)	(0.521)		(0.499)
Networks of excellence (NoE)	1.260**	1.939***		1.005*
	(0.590)	(0.635)		(0.535)
Project started in (baseline=2003)				
2004	0.0347	0.146	0.295	0.295
	(0.469)	(0.486)	(0.409)	(0.424)
2005	0.765	-0.445	-0.552	-0.482
	(0.661)	(0.715)	(0.633)	(0.638)
2006	-0.732	-0.561	-0.496	-0.421
	(0.486)	(0.510)	(0.408)	(0.420)
2007	-2.740***	-2.692***	-2.409***	-2.419***
	(0.855)	(0.952)	(0.826)	(0.825)
Constant	-16.02***	-15.03***	-10.29***	-7.090
	(5.198)	(5.384)	(3.229)	(5.280)
Over-dispersion parameter (log)	1.546***	1.433***	1.517***	1.518***
	(0.199)	(0.197)	(0.193)	(0.190)
Log-likelihood constant only	-482.8	-460.3	-482.8	-482.8
Log-likelihood full model	-459.5	-410.0	-436.9	-435.3
LR test	89.8***	100.5***	91.9***	95.1***
Observations	723	665	723	723

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. 58 projects dropped from regression reported in column 2 due to lack of data on distribution of financial funding among participants. All regressions include dummy variables for the strategic objectives of projects.

This result is to some extent surprising given the focus of NoEs on scientific research activities and the low participation of industry in this type of projects. In this respect, it has to be pointed out that few coordinators reported the production of *new* patents as an output of the project. Overall, only 109 out 723 FP6 projects (i.e. 15% of all FP6 projects) reported the production of one or more *new* patents. In other words, 85% of all FP6 projects reported no *new* patents. In order to corroborate and test the robustness of our findings we proceeded in two ways. First, we re-estimated our regression model by accounting for the presence of a large number of projects with zero patents. In particular, we estimated a hurdle negative binomial regression model. This model distinguishes the choice to patent from the frequency of patents produced. More specifically, it combines a dichotomous model, determining the binary outcome of the count being zero or positive, with a truncated-at-zero negative binomial model for strictly positive outcomes (Winkelmann and Zimmermann, 1995). Secondly, we have examined in a descriptive way the profile of patent applicants by type of projects.

Results of the hurdle regression analysis are reported in Table III.6. They show that a higher amount of funding per participant is associated to a higher probability of patenting, but not to a larger volume of patents. They also show that the probability of patenting is related to the size of the project according to a U-inverted relation and that all else equal NoE projects are more likely to patent than IPs and STRePs, but they are no more likely to patent a larger number of patented inventions. Specifically, the odds of patenting for a NoE project are about 4.5 times greater than IP and STReP projects, holding all other variables constant.

Besides confirming the different propensity to patent across projects (29% of all NoEs reported the production of at least one patent as opposed to 23% of all IPs), the evidence reported in Table III.7 also shows the existence of major differences in the profile of patent applicants. In particular, while companies account for the bulk of patents in the case of IPs and STRePs, universities and public research organisations play a major role in the case of NoEs with 50% of all patents applications. In addition to this, it is also worth remarking that the fraction of jointly owned patents in the case of NoEs is around twice as large as the one observed for IPs and STRePs[25]. These findings are

[25] Even without a formal benchmarking test, it is worth pointing out that the share of patents by universities and PROs and the fraction of co-patents look remarkably larger than in the total population of patents in the corresponding technological classes. Indirectly, this result indicates that the data reported by project coordinators are to some extent reliable and might reflect the true output of funded projects.

likely to reflect both a higher degree of integration and a lower level of conflict among partners around IPR issues.

Table III.6 – Hurdle negative binomial regression: patents

Variables	(1) Logit	(2) Truncated negbin
Number of participants	0.0598*	0.0714
	(0.0316)	(0.0582)
Number of participants [squared]	-0.000913*	-0.00263*
	(0.000507)	(0.00157)
Funding per participant	1.276***	0.156
	(0.407)	(0.487)
Asymmetry in funding distribution	0.0273	-0.0270
	(0.0264)	(0.0401)
Share of universities	-0.0112*	-0.00116
	(0.00656)	(0.00971)
Duration	-0.00699	-0.00254
	(0.0177)	(0.0212)
Number of previous patents	0.0501	0.416**
	(0.216)	(0.188)
Funding instrument (baseline=STReP)		
Integrated Projects (IP)	-0.0214	-0.749
	(0.541)	(0.634)
Networks of excellence (NoE)	1.510***	-0.760
	(0.570)	(0.863)
Project started in (baseline=2003)		
2004	-0.374	0.0379
	(0.535)	(0.589)
2005	0.538	0.378
	(0.662)	(0.821)
2006	-0.876	-0.370
	(0.558)	(0.687)
2007	-2.161*	-0.206
	(1.127)	(1.184)
Constant	-18.31***	-2.395
	(5.064)	(6.327)
Over-dispersion parameter (log)		0.267
		(0.749)
Log-likelihood	-401.3	
Wald $\chi^2(17)$	70.3***	
Observations	665	

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Regressions include dummy variables for the strategic objectives of projects.

Table III.7 – Patent assignees by type of funding instrument

	All projects	IPs	NoEs	STRePs
Companies	65.3	69.9	50.0	64.7
Individuals	0.8	1.1	-	0.6
PROs	14.9	9.3	18.0	20.7
University	16.5	15.9	32.0	12.0
Not assigned	2.5	3.8	-	2.0
Total	100.0	100.0	100.0	100.0
Number of patents	273	134	29	110
Fraction of co-patents	30.3	27.8	55.2	26.4
Number of projects	723	188	49	486
with >0 patents	109	44	14	51
% of all patents	15%	23%	29%	10%
Total EU funding (€m)	3008	1652	256	1100

Note: co-patents are defined as patents assigned to two or more organizations. Patents assigned to individuals or not assigned have been excluded from the calculation of the co-patenting fraction. All projects in the area of International co-operation had no patents. For this reason, they have been omitted from calculations reported in this table.

III.3 Analysis of EPO patents

While the analysis carried out in the previous section has focused on the aggregate number of patents produced by DG INFSO projects, in this section we focus our attention on the subset of *new* patent applications to the EPO. These patents represent about 65% of all new patents produced by DG INFSO projects. However, they are to some extent comparable and can be benchmarked with respect to the population.

III.3.1 Distribution of patenting activity by technological fields

As a first step, we examine the distribution of the EPO patents by technological field. To this purpose, we have adopted the aggregation of the *International Patent Classification* (IPC) codes into 30 fields and 6 areas elaborated by Fraunhofer Gesellschaft ISI, and Observatoire des Science et des Technologies (OST) in co-operation with the French patent office (INPI).[26] The distribution of patents is reported in Table III.8 and Table III.9. Slightly less than 90% of all EPO patents produced by DG INFSO projects

[26] http://www.wipo.int/export/sites/www/ipstats/en/statistics/patents/pdf/wipo_ipc_technology.pdf.

originate in two technological areas, *electrical engineering* and *scientific instruments*. Within the electrical engineering area (fields 1 to 5 in Table III.8) the most important fields are telecommunications and information technology, which together account for 48% of all EPO patents. Within the instruments area (fields 6 to 8 in Table III.8), the three fields of optics, control technology and medical technology account together for 22% of all new EPO patents produced by DG INFSO projects.

On the basis of the observed distribution of patents by technological field, the benchmarking of patenting activity by DG INFSO projects will be carried out with reference to the areas of *electrical engineering* and *scientific instruments*. Moreover, given the low number of patents at the level of *fields*, the benchmarking exercise will be performed at the more aggregated level of technological *areas*.

Finally, Table III.10 reports the distribution of new EPO patent applications produced by DG INFSO projects by priority year. More than 96% of all patents have been applied for the first time after the year 2003.

Table III.8 – Number of *new* EPO patents by technological field, OST30 fields

Technological field	Number of patents	%
1. Electrical machinery	2	0.9
2. Audiovisual technology	22	10.1
3. Telecommunications	75	34.4
4. Information Technology	30	13.8
5. Semiconductors	15	6.9
6. Optics	24	11.0
7. Control Technology	12	5.5
8. Medical Technology	13	6.0
9. Organic Chemistry	3	1.4
10. Polymers	1	0.5
12. Biotechnology	3	1.4
13. Materials	9	4.1
15. Basic Materials Chemistry	1	0.5
18. Materials processing	4	1.8
19. Thermal processes	2	0.9
23. Mechanical Elements	1	0.5
29. Consumer goods	1	0.5
Total	218	100.0

Table III.9 – Number of *new* EPO patents by technological area, OST6 areas

Technological area	Number of patents	%
Electrical engineering	144	66.0
Instruments	49	22.5
Chemistry, pharmaceuticals	17	7.8
Process engineering	6	2.7
Mechanical engineering	1	0.5
Consumer goods	1	0.5
Total	218	100.0

Table III.10 – Number of *new* EPO patents by priority year

Year	Number of patents	%
2001	1	0.5
2002	5	2.3
2003	14	6.4
2004	38	17.4
2005	57	26.2
2006	67	30.7
2007	32	14.7
2008	4	1.8
Total	218	100.0

III.3.2 Distribution of patents by country of applicants and inventors

Our assessment of the EPO patenting activity starts from the national and regional distribution of patent applications. To this purpose, the address of each patent applicant and inventor has been processed and reclassified at the country level, and for the EU 27 countries also at the regional NUTS2 and NUTS3 levels. In computing the distribution of patents by country and region, we have adopted a full counting method. Accordingly, for patents having more than one applicant and/or more than one inventor, each applicant and/or each inventor has been credited a whole patent[27].

[27] Please note that the same analysis has been performed also by using a fractional counting method, according to which each country is credited a fraction of the patent application. For example, for a patent with three applicants, two from country *x* and one from country *y*, the fractional counting method assigns two thirds of the patent to country *x* and one third to country *y*. On the other hand, the full counting method assigns two patents to country *x* and one patent to country *y*. Since results are qualitatively similar, we report here only the results from the full counting method.

Table III.11 reports the distribution of all EPO patents by country of applicants and by country of inventors. Please note that because of the adoption of a full counting method the total number of patents exceeds the actual number of patents produced. Moreover, the total number of patents by country of applicant differs from the number of patents by country of inventor for similar reasons. To facilitate the interpretation of data, we have reported in the same table the unweighted fraction of organisations participating in the DG INFSO projects included in our sample by country. This provides an indication of the participation rate.

Comparing the share of patents and the participation rate, we observe a number of countries whose contribution to patenting is significantly larger than their participation in DG INFSO projects, i.e. Germany, France, Netherlands and Belgium. On the other hand, the opposite pattern is observed instead for the United Kingdom. A few non-EU27 countries, notably Switzerland, United States and Japan, also contribute to a significantly larger fraction of all patents than their participation rate, especially when patents are counted according to the country of applicant (ownership).

If we compare the distribution of patents by country of applicant (ownership criterion) and by country of inventors (invention criterion), we note that the share of patents accounted for by French organizations is significantly larger than the corresponding share measured by country of inventors. This difference is likely to be related to the presence of large public research organizations in France, such as the Commissariat a l'Energie Atomique (CEA), that host researchers from different European countries. On the other hand, the opposite pattern is observed for Belgium, United Kingdom and Italy. The share of patents *invented* in these countries is somewhat larger than the share of patents *owned* by these same countries. In the case of Belgium and Italy, this is likely to be explained by the presence of multinational companies with headquarters located in other countries.

As far as the regional distribution of patenting activity is concerned, we have calculated the share of patents held by the top 10 most patenting NUTS3 regions, the number of NUTS3 regions with a non-null patent production, and the Herfindahl index capturing the spatial concentration of inventive activity. The region of inventors has been used to allocate patents in space and a full counting method has been applied.

Results are reported in Table III.12. In order to facilitate the interpretation of results, we have reported in the same table the corresponding indicators for the scientific articles produced by DG INFSO projects (see above Table II.44).

Table III.11 – Distribution of *new* EPO patents by country of applicants and inventors

Country	Participants	%	Applicants	%	Inventors	%
Germany	725	14.7	53	19.9	122	18.1
France	542	11.0	49	18.4	82	12.2
Italy	493	10.0	20	7.5	67	9.9
United Kingdom	465	9.5	10	3.8	35	5.2
Spain	329	6.7	11	4.1	35	5.2
Netherlands	190	3.9	18	6.7	39	5.8
Greece	166	3.4	1	0.4	5	0.7
Belgium	160	3.3	36	13.5	107	15.9
Sweden	142	2.9	10	3.8	19	2.8
Austria	137	2.8	3	1.1	19	2.8
Poland	108	2.2			6	0.9
Hungary	99	2.0			12	1.8
Finland	91	1.9	6	2.3	16	2.4
Portugal	85	1.7			15	2.2
Denmark	81	1.6	2	0.8	2	0.3
Czech Republic	62	1.3				
Ireland	58	1.2	3	1.1	3	0.5
Romania	50	1.0				
Slovenia	43	0.9				
Bulgaria	41	0.8				
Lithuania	31	0.6				
Slovakia	27	0.5				
Estonia	24	0.5				
Cyprus	24	0.5				
Luxembourg	21	0.4				
Latvia	14	0.3	1	0.4	3	0.5
Malta	8	0.2				
Switzerland	131	2.7	14	5.2	49	7.3
Israel	86	1.7			2	0.3
Norway	81	1.6				
China	64	1.3				
Russian Federation	34	0.7			10	1.5
United States	28	0.6	17	6.4	20	3.0
Brazil	36	0.7			1	0.2
Japan	8	0.2	11	4.1		
Iceland	5	0.1	1	0.4	2	0.3
Korea, Republic Of	4	0.1	1	0.4	3	0.5
Others	226	4.6				
Total	4919	100.0	267	100.0	674	100.0

Results are quite interesting. They show that the degree of spatial concentration of patenting activity is comparable to the one observed for scientific articles. This finding is likely to be related to the complementarity between patenting and publishing already noted above. Regions possessing advanced scientific capabilities are also likely to be the ones able to generate patented inventions.

Table III.12 – Distribution of *new* EPO patents and articles by regions

	C10	H	Nr. regions
DG INFSO patents	30.1	0.016	175
DG INFSO articles			
Computer Science, Hardware & Architecture	39.3	0.024	88
Computer Science, Information Systems	34.2	0.020	96
Computer Science, Software Engineering	30.7	0.019	77
Engineering, Electrical & Electronic	30.4	0.016	200
Optics	32.0	0.018	117
Physics, Applied	36.0	0.021	114

The table reports the share of the top 10 most productive NUTS3 regions. It also reports the Herfindahl index, calculated as the sum of the squared shares of all regions, as well as the number of regions with a non-null scientific/technological production. A full counting method has been adopted to count patents by regions.

III.3.3 Distribution of patenting activity by type of applicant

A relevant issue to explore refers to the type of organizations in the DG INFSO projects that are most likely to generate patent applications. To this purpose, all the 218 *new* patent applications at the EPO have been checked and the names of the applicants have been cleaned and standardised. Moreover, applicants have been classified into four mutually exclusive categories: companies, public research organizations (PRO), universities, and individuals. Regarding this classification, it has to be pointed out that the PRO category is the most heterogeneous one, as it includes (national and international) public research organizations *stricto sensu*, e.g. Commissariat a l'Energie Atomique (CEA), but also semi-public research organizations, e.g. Fraunhofer Gesellschaft, or private non-profit research organizations.

Given that the same patent may be applied jointly by more than one applicant, i.e. co-patents, in computing the distribution of patents by category of applicant, we adopted a full counting method. In other words, if a patent has been co-patented by a company and a university, a whole patent has been credited to the company and a whole patent has been credited to the university.

Table III.13 – Number of *new* EPO patents by type of applicant (whole sample)

Applicant type	Number of patents	%
PRO	44	16.1
Company	185	67.8
University	37	13.5
Individual inventors	7	2.6
Total	273	100.0

Note that the sum of patents exceeds 218 as more than one applicant type may be reported on a single patent document.

The full counting distribution of EPO patents produced by DG INFSO projects is reported in Table III.13. Not surprisingly, the vast majority of patent applications (68%) have been made by corporate organizations. However, it is interesting to observe that 16% and 14% of all patents have been filed, respectively, by PROs and by universities. Of course, we cannot draw any conclusion from this analysis concerning the fact that the share of PRO and university patents is higher or lower than expected. To this end, we need to carry out a benchmarking analysis.

The benchmarking analysis has been performed as follows. From the set of patent applications produced by DG INFSO projects, we have extracted the subset of patents classified in the areas of *Electrical engineering* and *Instruments* and applied after the year 2000. For each of these two areas we have calculated the share of patent applications held by the different types of applicants. The same kind of analysis has been then repeated by taking all patent applications to the EPO in the same technological areas and applied after the year 2000, in which at least one of the patent applicants is from a EU27 country.

Results of this analysis are reported in Table III.14. They confirm that DG INFSO projects are characterised by a significantly larger presence of patenting PRO and universities than in the general population of European organisations patenting in the same technological areas. Not surprisingly, we also observe that the share of patents held by universities tends to larger in the area of Instruments than in the area of Electrical engineering, both in the context of DG INFSO projects and in the general population.

Table III.14 – Number of *new* EPO patents by type of applicant (benchmarking)

	PRO	Company	University
DG INFSO projects			
Electrical engineering	20 (11.2)	144 (80.4)	15 (8.4)
Instruments	12 (22.6)	30 (56.6)	11 (20.8)
Benchmarking population			
Electrical engineering	1442 (2.0)	66815 (96.6)	944 (1.4)
Instruments	2071 (5.5)	34362 (90.9)	1361 (3.6)

The table compares the distribution of EPO patents by type of applicant. A full-counting method has been adopted, i.e. a patent co-applied by a company and a university has been fully assigned to both types of applicants and therefore counted as one company patent *and* as a one academic patent. The comparison is between the distribution of *new* EPO patents by DG INFSO projects and the distribution of EPO patents in the benchmarking population. Note that one DG INFSO patent in the field of Instruments owned only by individual inventors has been excluded from the analysis. No patent application in the field of electrical engineering was owned only by individual inventors.

Table III.15 reports the list of the top 25 patent applicants in DG INFSO projects. The list includes both very large companies, most of which from the telecom industry but also from the software sector, as well universities and large PRO. A rather surprising result is the absence of Nokia in the list of top patent applicants. To investigate further this issue, we examined in detail the "questionnaire" and the "patent" sheets reported by project coordinators. In a few projects where Nokia has been participating we found the following notes: "*Nokia does not report about inventions outside the company as a matter of policy*", "*As a matter of corporate policy, Partner Nokia does not report about patent details outside of the company*", and "*Upon request, Nokia can provide the application number for controlling purposes*".

Table III.15 – Top 25 patenting organisations, *new* EPO patents by DG INFSO projects

Ranking	Name	Patents	%
1	LUCENT TECHNOLOGIES	17	5.2
2	STMICROELECTRONICS	17	5.2
3	COMMISARIAT A L'ENERGIE ATOMIQUE	16	4.8
4	ALCATEL LUCENT	15	4.5
5	SAP	10	3.0
6	THALES	10	3.0
7	THOMSON LICENSING	10	3.0
8	DEUTSCHE THOMSON-BRANDT	9	2.7
9	TELEFONAKTIEBOLAGET LM ERICSSON	9	2.7
10	THOMSON TELECOM	8	2.4
11	KONINKLIJKE PHILIPS ELECTRONICS	7	2.1
12	NTT DOCOMO	6	1.8
13	ALCATEL BELL	5	1.5
14	CENTRO RICERCHE FIAT	5	1.5
15	NTT MOBILE COMMUNICATIONS NETWORK	5	1.5
16	VALTION TEKNILLINEN TUTKIMUSKESKUS	5	1.5
17	ASM LITHOGRAPHY	4	1.2
18	FRAUNHOFER-GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG	4	1.2
19	INTERUNIVERSITAIR MICROELEKTRONICA CENTRUM IMEC	4	1.2
20	MICROSOFT	4	1.2
21	SIEMENS	4	1.2
22	SILICON BIOSYSTEMS	4	1.2
23	UNIVERSITEIT GENT	4	1.2
24	AXSIONICS	3	0.9
25	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS	3	0.9

In addition to this, we also found other notes saying "*Nokia: 3 patent applications during 2007 in relation to MASCOT*", "*Nokia: 4 patent applications during 2008 in relation to MASCOT*", and "*Nokia: 1 patent application during 2009 in relation to MASCOT*". These observations imply that we are probably under-estimating Nokia's patent production in the context of DG INFSO projects. However, since similar notes are also found with respect to other companies, the consequence is that the present study is probably under-estimating the overall patent productivity of DG INFSO projects. As we argue more in detail in the Appendix 1 to this report, the only solution to this type of problems is to improve the whole process of data collection.

A related issue that deserves investigation concerns the patterns of collaboration in the production of patents, as measured by co-patenting activity, i.e. patents jointly owned by two or more organizations. Each patent application by DG INFSO projects has been classified according to the combination of organisational types appearing in the patent as applicants. Results are reported in Table III.16. Not unexpectedly, *single applicant* patent applications account for the vast majority of patents (around 80%). At the same time, the extent of collaboration in patenting is far from negligible. Around 19% of all patent applications registered at the EPO by organisations participating in DG INFSO projects are the outcome of a joint research. Once again, these figures alone cannot tell anything about the extent to which the propensity to collaborate in patent production is larger than expected. To this purpose, a benchmarking analysis has been performed.

Table III.16 – Patterns of *new* EPO co-patenting, DG INFSO projects

Patent owners	Number of patents	%
Only Company	135	61.6
Only PRO	26	11.9
Only University	13	6.4
Only individuals	2	0.9
Company-Company	22	10.1
University-University	6	2.7
Company-University	5	2.3
PRO-PRO	7	3.2
PRO-University	1	0.5
Company-PRO	1	0.5
Total	218	100.0

The methodology used to perform the benchmarking analysis is similar to the one described above. From the set of patent applications produced by DG INFSO projects, we have extracted the subset of patents classified in the areas of *Electrical engineering* and *Instruments* and applied after the year 2000. For each of these two areas we have calculated the share of patent applications according to the combination of patent applicant types appearing in the patent. The same kind of analysis has been then repeated by taking all patent applications to the EPO in the same technological areas and applied after the year 2000, in which at least one of the patent applicants is from a EU27 country. In order to simplify the analysis and reduce the number of possible combinations, applicants corresponding to individual inventors have been excluded from the analysis. Results are reported in Table III.17. As far as the *Electrical engineering* field is concerned, the data suggest a significantly larger degree of collaboration in patenting among the subset of patents generated by DG INFSO projects than in benchmarking population: 21.6% of all DG INFSO patents are the outcome of an inter-organizational collaboration vs. 1.8% in the population of European patents (z-score=17.6, statistically significant at the 99% level). Moreover, even if the collaboration involving two companies clearly represents the most important type of co-patenting both for the subset of DG INFSO patents and for the benchmarking population, it is interesting to note that in the case of DG INFSO patents it does not clearly dominate over other types of inter-organizational collaboration. Co-patents involving at least a university or a PRO account for about 40% of all co-patents produced by DG INFSO projects. The evidence is relatively less strong in the case of *Instruments*: 10.5% of all DG INFSO patents in this area are the jointly application by two organisations vs. 4.5% in the benchmarking population (z-score=2.0, statistically significant at the 95% level).

Overall, we believe that these results are important for two reasons. On the one hand, they suggest that organizations involved in EU funded RTD projects tend to deal with the problem of appropriating the intellectual property rights resulting from their joint research through devising sharing schemes. On the other hand, they also indirectly suggest that the data reported by project coordinators are likely to be correct and reliable. To the extent that the objective of EU funded projects is to encourage collaborative research among different types of organisations, we would indeed expect to observe that a large share of the scientific and technological output resulting from these projects reflects this collaboration. The fact that the data show that this is the case reassures us that we are probably capturing the real output effects of these projects.

Table III.17 – Patterns of *new* EPO co-patenting, DG INFSO projects vs. population

Patent owners	DG INFSO	%	Population	%
Electrical engineering				
Only Company	100	69.4	63747	95.5
Only PRO	13	9	1142	1.7
Only University	-	-	662	1.0
Company-Company	19	13.2	806	1.2
University-University	3	2.1	29	0.0
Company-University	5	3.5	135	0.2
PRO-PRO	3	2.1	35	0.1
PRO-University	-	-	53	0.1
Company-PRO	1	0.7	153	0.2
Total	144	100	66762	100.0
Instruments				
Only Company	24	50.0	31865	89.2
Only PRO	10	20.8	968	2.7
Only University	9	18.8	1326	3.7
Company-Company	3	6.3	770	2.2
University-University	1	2.1	60	0.2
Company-University	-	-	95	0.3
PRO-PRO	1	2.1	35	0.1
PRO-University	-	-	111	0.3
Company-PRO	-	-	491	1.4
Total	48	100.0	35721	100.0

As a final step in investigating the patterns of collaboration, we have examined the extent to which the same individuals are involved in both patenting and publishing. These individuals that we can label as author-inventors play a very important role of social connectors between the domain of open science and the world of private technology. To this purpose, we have compared the names of the authors of scientific articles produced by DG INFSO projects with the names of inventors of patents produced by the same projects. For each instance of a matching, we have checked whether the two individuals are indeed the same person or whether they are simply homonymous individuals. Finally, for each matched individual, i.e. author-inventor, we have collected information on her institutional affiliation[28]. To this end, we have used the SCOPUS database produced by *Elsevier Publishers*. This is bibliographical database similar to ISI-WoS. In addition to information on published articles, however, this database also collects information on authors. For each author, it reports a unique ID code, as well as information on past and current affiliations. For the present analysis, we focused our attention on the current affiliation of authors-inventors as reported in SCOPUS.

Results of our analysis are reported in Table 61. Out of a total of 597 inventors of EPO patents produced by DG INFSO projects, 165 (28%) are also authors of at least one scientific article produced by the same projects. Benchmarking this value is rather difficult, as it would require to generalise this type of analysis to the whole population of EPO patents and scientific articles. However, we can reasonably say that the fraction of authors-inventors looks rather large. As far as the affiliation of these authors-inventors is concerned, not surprisingly the data show that the majority of them (around 85%) are affiliated with either a university or a PRO. However, we think it is remarkable that almost 15% of all authors-inventors have a corporate affiliation. Once again, it is not easy to benchmark this value for the same reasons stated above. Yet, we can similarly conclude that the share of corporate author-inventors in the context of DG INFSO projects is likely to be larger than that found in the general population.

Table III.18 – Number and affiliation of authors-inventors

	Number	%
Authors-inventors	165	27.6
Only inventors	432	72.4
Total	597	100.0
University	75	48.1
PRO	58	37.2
Company	23	14.7
Missing info	9	-
Total	165	100.0

[28] Please note that this analysis has been peformed on the first batch of patent data received in 2009.

IV. Concluding remarks

The report illustrates the results and findings of a pilot study aiming to assess the scientific and technological achievements of collaborative projects funded by the Information Society and Media Directorate General (DG INFSO) of the European Commission in the context of FP6 and FP7. To this purpose, the study has examined a data set on the scientific publications and patents produced by the funded projects. Raw data have been collected by the DG INFSO itself through a MS-Excel survey sent to the project coordinators. KITeS-Bocconi has then undertaken a thorough work of data cleaning and standardisation. Moreover, it has also carried out the work of matching data on scientific publications with the Thomson-ISI Web of Science database and data on patents with the Espacenet and the EPO-Patstat databases. The resulting data set contains for each published article and each patent reported by project coordinators a full set of bibliographical information that allow assessing several aspects of the scientific and technological activity of DG INFSO projects.

Data quality issues

A first set of results emerging from the study concerns the methodological issues related to the process of data collection and data quality. In this respect, it has to be remarked that the work of raw data cleaning and standardisation took a large amount of the time and the resources allocated to the study due to the very poor quality of the data received from project co-ordinators. The experience gained in carrying out this work suggests a few possible ways to improve the process of data collection and the quality of the data. In particular, it suggests the adoption of a web-based platform to collect data from project coordinators and the reduction in the redundancy of information requested from project coordinators through the use of the Digital Object Identifier (DOI) system. The DOI is a unique label that allows identifying in a univocal and persistent way a computer readable object that can be found on the internet. Given that virtually all publishers are nowadays are adopting the DOI system and therefore label the articles and publications they issue with this system, collecting this single information from project coordinators would significantly speed up the process and would also improve the precision and reliability of the data collected. Appendix 1, which accompanies this report, elaborates a concrete and detailed proposal along these lines.

The analytical part of the study has investigated three important aspects of the scientific and technological activities performed by the DG INFSO projects. First, it examined the scientific and technological *productivity* of projects. Second, it evaluated the *quality* of the scientific output. Third, it assessed the extent of *international*, *inter-regional* and *inter-organizational* collaboration. Most of the analytical part of the study has focused on FP6 projects given the relatively low number of FP7 projects, articles and patents in the database. The major results emerging from our analysis are summarised below.

Assessment of productivity

- The distribution of scientific productivity looks rather skewed. Around 50% of all FP6 projects have not reported the production of any scientific articles. This fraction is even higher for FP5 and FP7 projects, though this result is less significant given the relatively low number of FP5 and FP7 projects in the database. The picture changes if one examines the publication of papers in *conference proceedings*. In this case, the fraction of FP6 projects with no publications decreases to around 18%. The two figures are not fully comparable, given the different scientific importance of articles published in peer-reviewed journals and papers published in conference proceedings. Extending the assessment analysis to including conference papers represents a natural and fruitful avenue for further research. At the same time, such an extension would probably require improving the process of data collection from project coordinators.

- The degree of skewness in the distribution of patents among projects is even larger than the one observed for the scientific output. Only 41 out 199 FP5 projects (21%) have applied for at least one new patent and the fraction is even lower for FP6 projects (109/927=12%). Yet, even though the fraction of patenting projects looks lower for FP6 than for FP5 projects, the average number of patent applications per patenting project is larger for the former than for the latter: 2.5 vs. 1.8. Please note that these figures must be interpreted with some caution. Due to the poor quality of data sent back by project coordinators we could not trace information on all patent applications claimed to be produced by projects.

- An important result emerging from the analysis is that the scientific and technological productivity differs across projects funded through different instruments. In particular, our results suggest that Networks of Excellence (NoE) outperform other types of projects in terms of both scientific and technological performance. NoEs have generated on average 29.7 scientific articles per project as compared to 8.7 for Integrated Projects (IP) and 4.6 for Strategic Targeted Research Projects (STReP). Moreover, even though IPs look apparently more productive than STRePs, this higher productivity is almost entirely explained by their larger size in terms of participants and financial resources. In this regard, the preliminary results of the study seem to indicate the existence of diminishing returns with respect to the size of projects. In other words, it is possible that the FP6 has funded "too large" (IP) projects whose larger size has not been matched by a proportionally larger amount of scientific and technological output.

- Concerning the filing of new patent applications, IPs and NoEs seem to exhibit a higher propensity to patent than STRePs: the fraction of all projects reporting at least one new patent application is equal, respectively, to 23%, 29% and 10%. Moreover, considering only projects with new patent applications, the average number of patents per project goes from 3.05 for IPs, to 2.07 for NoEs and 2.16 for STRePs. Once again, these productivity differences arise at least partly from the different amount of resources allocated to projects. In particular, once the number of participants and the amount of financial resources are taken into account, no significant difference remains between IPs and STRePs either in the propensity to patent or in the number of new patent applications. At the same time, results confirm that NoEs tend to exhibit a higher propensity to patent than IPs and STRePs.

- These results are to some extent surprising given the focus of NoEs on scientific research activities and the low participation of industry in this type of projects. An analysis of the profile of patent applicants, however, reveals that while companies account for the bulk of patents in the case of IPs and STRePs, universities and public research organisations play a major role in the case of NoEs with 50% of all patents applications. In addition to this, the fraction of jointly owned patents in the case of NoEs is around twice as large as the one observed for IPs and STRePs. These findings seem to suggest that a higher degree of research integration and a lower level of conflicts among partners around IPR issues have emerged in the case of NoEs. Testing this hypothesis represents, in our view, an interesting avenue for further research.

Assessment of quality

- The assessment of the *quality* of scientific articles produced by DG INFSO projects has been based upon an analysis of the number of citations received in the scientific literature. The quality of the articles resulting from European projects has been benchmarked by comparing it with the quality of the population of all other articles published in the same journals and scientific areas. Specifically, we have examined the three scientific fields more relevant for the DG INFSO projects, i.e. computer science, materials science, and optics.

- Concerning the *scientific impact* of articles produced by DG INFSO projects, results show that the probability they receive forward citations from other articles is generally no lower and often significantly higher than the average article published in the same journal and in the same scientific field. More specifically, the fraction of uncited articles produced by DG INFSO projects is significantly lower than the corresponding fraction in the benchmarking population.

- In addition to this, the mean observed citation rate (MOCR) is systematically larger than the mean expected citation rate (MECR), thereby suggesting that DG INFSO articles are not only more likely to receive at least one citation, but that on average the number of citations received is larger than for the average article in the same journal and scientific field. In particular, DG INFSO articles in the field of optics received on average 1.3 times more citations per article than the average article in this field. Similar results hold for the scientific field materials science, but not in the field of computer science. In this latter case, the omission of conference papers (a major vehicle of knowledge dissemination in this field) is likely to be responsible for the result.

- Overall, the results confirm that DG INFSO projects have been highly effective in attracting top quality researchers and research teams from the research fields relevant for the ICT area.

Assessment of research collaboration

- The propensity of DG INFSO projects to engage in international and inter-regional scientific research collaboration has been tested by examining the patterns of co-authorship and by benchmarking it with the pattern emerging in the population of all European authored articles in the same journals and scientific areas.

- The propensity to engage in international research collaboration is significantly larger for the authors of articles produced in the context of DG INFSO projects than for the average European article in the same scientific fields. In particular, this study finds that DG INFSO articles are more likely to involve *intra*-European collaboration than the average European article. The share of DG INFSO articles involving at least two affiliations from different ERA countries is around twice as large as in the benchmarking population. This result suggests that the FPs have been effective in reinforcing the extent of intra-European collaboration in research. At the same time, this study finds that DG INFSO articles are also characterised by a lower propensity to engage in *extra*-European collaboration compared to other European articles in the same fields.

- The regional (NUTS2 and NUTS3 level) concentration of the scientific output produced by DG INFSO projects looks significantly larger than the spatial level of concentration in the benchmarking population of articles. The number of European regions with a positive number of articles represented within DG INFSO projects is generally much lower than the number of European regions capable of a scientific production in the benchmarking population of articles. As a consequence, the spatial concentration of the scientific output (Herfindahl index) in the case of DG INFSO projects is around twice as large as in the benchmarking population of articles for most of the scientific fields examined here. Despite the larger concentration of scientific output, however, the spatial patterns of scientific activity look rather similar. A linear and a rank correlation analysis of the number of scientific articles produced by NUTS2 European regions within DG INFSO projects and in the benchmarking population confirms the existence of a large (though far from perfect) and positive correlation between the two variables.

- The majority of patent applications produced by DG INFSO projects have been extended to the European Patent Office (EPO). Focusing on this subset of patent applications, the most important technology areas, covering 90% of all EPO patents reported by DG INFSO projects, are electrical engineering (which comprises telecommunications, computer, audiovisual technology and semiconductors) and scientific instruments (which comprise medical technology, control technology and optics). The analysis of research collaboration has been carried out by benchmarking patents produced by DG INFSO projects with the population of other European invented patents in these two technological fields.

- Analysis by type of applicant shows that the share of DG INFSO patents applied by universities and public research organisations (PRO) significantly exceeds the share observed in the benchmarking population of patents in the same technological fields. This suggests that the propensity to patent and/or to engage in research leading to patented inventions by universities and PROs is significantly larger in the context of EU funded projects than in the corresponding population of European institutions.

- Analysis of co-patents, i.e. patents jointly applied by two or more independent organisations, also reveals that the extent of joint patent ownership in the context of DG INFSO projects is significantly higher than in the benchmarking population of patents. This finding probably signals the need for organisations participating in EU funded projects to devise schemes for sharing the intellectual property rights over the results of joint research. Moreover, even though company-company ownership is the most prevalent type of joint ownership, a significant amount of co-patents involves PROs and universities.

- The extent of spatial concentration of patenting activity is comparable to the one observed for the scientific publication activity. The top ten NUTS3 regions in terms of patents account for about 33% of all EPO patent applications produced by DG INFSO projects. This finding can be probably interpreted as evidence of complementarity between patenting and publishing activities.
- Finally, an analysis at the level of individual researchers show that around 27% of all inventors of EPO patents applied by DG INFSO projects are also authors of one or more scientific articles reported by the same projects. These individuals that we can label as author-inventors play a crucial role as social connectors between the domain of open science and the world of private technology. Our analysis also shows that around 14% of such individuals have a corporate affiliation. Even though benchmarking these figures (i.e. assessing the extent to which the observed values are larger or lower than expected) is extremely difficult, we can quite confidently conclude that the degree of interplay between scientific research and technological development achieved in the context of DG INFSO projects looks quite remarkable.

IV.1 Limitations of the study

It is important to point out that the results reported in this study have to be interpreted with some caution due to the limitations of the empirical methodology adopted. First, the study has explored only one thematic area, i.e. ICT, of the FP and the results can be hardly generalised to other fields.

Second, the data on scientific and technological output of EU funded projects come directly from project co-ordinators, they are hardly verifiable in an independent way and therefore are subject to possible reporting bias, even though *a priori* it is difficult to say whether the bias should go in the direction of over-estimating or under-estimating the true output. For example, we noted that several projects reported patents whose priority date is prior with respect to the starting date of the project. We argued that this phenomenon can be probably related to the IPR rules that oblige project participants to grant access rights to their knowledge (or pre-existing know-how) to another participant if the latter needs such access rights in order to carry out its own work under the project or to use its own knowledge resulting from the project. However, this is no more than a conjecture that needs to be verified. More generally, we believe that testing the accuracy of what reported by project coordinators is imperative in order to derive robust conclusions from studies like this one.

Third, the assessment of scientific performance of DG INFSO funded projects has been mostly based on articles published in mainstream scientific journals indexed in the ISI-WoS SCI and SSCI databases. As repeatedly argued in this report, a more correct assessment of the scientific productivity of research projects in the ICT area should take into account that conference proceedings and papers play a particularly important role as a vehicle of knowledge dissemination in this field.

Fourth, even though this report has used state of the art bibliometric and scientometric techniques, one should be cautious before drawing too simple conclusions and policy implications based on the results reported here. For example, one of the conclusions of this study is that the scientific and technological performance differ across projects in terms of funding instruments. In particular, Networks of Excellence outperform IP and STRePs both in terms of production of scientific articles and propensity to introduce new patent applications. Yet, concluding from these results that the superior performance of NoE has to be imputed to the funding instrument would ignore the possibility that the best scientists and research teams are selected into these projects and therefore generate a higher number of articles and are more likely to produced patented inventions. Testing the existence of these potential endogeneity problems is necessary before being able to draw robust conclusions.

Finally, one should point out that, no matter how important, scientific articles and patents represent only one of the several and multifaceted socio-economic benefits that are expected from the FPs.

IV.2 Avenues for further research

The limitations of the study briefly discussed above suggest as many potential avenues for further research. First, we believe that broadening the set of publications considered, in particular taking into account conference proceedings, is essential in order to get a correct assessment of the overall scientific productivity of research projects. More generally, we would recommend extending the bibliometric analysis to other sources of bibliographic information, such as the *Scopus* database, produced by Elsevier, and the *Google Scholar* database.

Along similar lines, for future evaluations it would be useful to complement data on articles and patents with other indicators of performance, such the formation of start-up companies, the mobility and career promoting effect of FP participation and so on. Yet, two problems have to be solved before being able to use effectively such indicators. First, the data collection process should be improved to ensure that comparable, standardised and consistent data are collected. Second, the data should be as far as possible verifiable in an independent way in order to avoid any possible reporting bias.

Despite the importance of collecting other indicators of performance, we believe that further quantitative analyses of data on publications and patents are needed in order to better understand the organization of research activities in the funded projects. Analysis of forward citations received by patents generated by the projects might help assessing the importance of inventions resulting from the funded research, whereas analysis of backward citations could be used to assess the inter-sectoral and international flows of knowledge, e.g. the countries of origin of knowledge "used" by patents. Similarly, further analyses of patents should include an analysis of the so-called "non-patent references" (i.e. scientific publications and other published material) reported on the front pages of the patents. The analysis could help gauge the extent to which European scientific research proves to be relevant in the development of ICT patents resulting from DG INFSO patents.

Although a key result of this study is that articles resulting from projects have a higher than average scientific impact, the interpretation of these results is not entirely clear due to endogeneity problems. Is the production of better articles an outcome of project participation or is it simply the consequence of the fact that the best scientists and research teams participate in FP projects? In a similar way, the study finds that the scientific and technological productivity markedly differ across projects funded through different instruments. In particular, NoEs tend to outperform IPs and STRePs in terms of number of articles produced and propensity to generate new patent applications. Yet, concluding from these results that the superior performance of NoEs has to be imputed to the funding instrument would ignore the possibility that the best scientists and research teams are selected into these projects and therefore generate a higher number of articles and are more likely to produce patented inventions. Solving issues like these requires further efforts of data collection and econometric analysis.

Finally, we would recommend that subsequent evaluation studies include a validation stage of quantitative results. In particular, a key finding of this study is that the scientific and technological productivity markedly differs across projects and funding instruments. Understanding the reasons for such differences calls for the use of qualitative and case-based studies, where selected research projects- both highly successful projects and less successful ones according to their output performance- are sampled for additional data collection. This more qualitative study should assess the impact of project-specific and R&D environmental factors that have impinged on each project's performance.

V. References

Abt, H. (2000a). “Do Important Papers Produce High Citation Counts?” *Scientometrics* 48(1): 65-70.

——— (2000b). “The Reference-Frequency Relation in the Physical Sciences”. *Scientometrics* 49(3): 443-451.

Abt, H.A. and E. Garfield (2002). “Is the relationship between numbers of references and paper lengths the same for all sciences?”. *Journal of the American Society for Information Science and Technology* 53(13): 1106-1112.

Arnold, E. (2004). “Evaluating research and innovation policy: a systems world needs systems evaluations”. *Research Evaluation* 13(1): 3–17.

Arnold Report (2009). “Evaluation of the Sixth Framework Programmes for Research and Development 2002-2006”. European Commission, Bruxelles.

Arnold, E., Clark, J., and A. Muscio (2005). “What the evaluation record tells us about Framework Programme performance”. *Science and Public Policy* 32: 385–397.

Butler, L. (2008). “ICT assessment: Moving beyond journal outputs”. *Scientometrics* 74(1): 39-55.

Butler, L. and M.S. Visser (2006). “Extending citation analysis to non-source items”. *Scientometrics* 66(2): 327-343.

Chubin, D., and E. Garfield (1980). “Is citation analysis a legitimate evaluation tool?”. *Scientometrics* 2(1): 91-94.

Delanghe, H., Muldur, U. and L. Soete (2009). *European Science and Technology Policy. Towards Integration or Fragmentation?*. Edward Elgar: Cheltenham.

Drott, M.C. (1999). “Reexamining the role of conference papers in scholarly communication”. *Journal of the American Society for Information Science* 46(4): 299-305.

EPEC (2009a). “Assessment of the impact of the new instruments introduced in FP6. A Study for the European Commission Research Directorate General”, European Commission, Bruxelles.

EPEC (2009b). “Bibliometric profiling of Framework Programme participants. A Study for the European Commission Budget Directorate General”, European Commission, Bruxelles.

European Commission (2000). “Towards a European Research Area”. European Commission, Bruxelles.

Glänzel, W., Schlemmer, B., Schubert, A. and B. Thijs (2006). “Proceedings literature as additional data source for bibliometric analysis”. *Scientometrics* 68(3): 457-473.

Greene, W.H. (2008). “Functional forms for the negative binomial model for count data”. *Economics Letters* 99(3): 585-590.

Greene, W.H. (2002). *Econometric Analysis*. 5th ed. Prentice Hall: New Jersey.

Kaplan, N. (2007). “The norms of citation behavior: Prolegomena to the footnote”. *American Documentation* 16(3): 179-184.

Leydesdorff, L. and I. Rafols (2009). “A global map of science based on the ISI subject categories”. *Journal of the American Society for Information Science and Technology* 60(2): 348-362.

Lisée, C., Larivière, V. and E. Archambault (2008). “Conference proceedings as a source of scientific information: A bibliometric analysis”. *Journal of the American Society for Information Science and Technology* 59(11): 1776-1784.

Lovaglia, M. (1991). “Predicting citations to journal articles: The ideal number of references”. *The American Sociologist* 22(1): 49-64.

Marimon Report (2004). “Evaluation of the effectiveness of the new instruments of Framework Programme VI” (http://cordis.europa.eu/fp6/instruments_review/). European Commission, Bruxelles.

Meho, L.I., Yang, K. (2007). “A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar”. *Journal of the American Society for Information Science and Technology* 58(13): 2105–2125.

Moed, H. F. 2005. *Citation analysis in research evaluation*. Springer, Berlin.

Vanecek, J., Fatun, M. and V. Albrecht (2010). “Bibliometric evaluation of the FP-5 and FP-6 results in the Czech Republic”. *Scientometrics* 83(1): 103-114.

Vieira, E.S., and J.A.N.F. Gomes (2009). “Citations to scientific articles: Its distribution and dependence on the article features”. *Journal of Informetrics* (forthcoming).

Visser, Martijn S., and H.F. Moed (2005). “Developing bibliometric indicators of research performance in computer science”. In *Proceedings of ISSI 2005, 10th International Conference of the International Society for Scientometrics and Informetrics*, P. Ingwersen and B. Larsen (eds.), Stockholm.

Vonortas, N. (2008). “FP6 participation”. Washington DC, George Washington University.

Winkelmann, R., Zimmermann, K.F. (1995). "Recent Developments in Count Data Modelling: Theory and Application." *Journal of Economic Surveys* 9(1): 1-24.

VI. Appendix 1 – Database construction: methodology

VI.1 Data sources

This appendix provides a detailed description of the work carried out for the construction of the database used in the study “Using performance indicators in monitoring the implementation of ICT research in FP6 and FP7” (SMART 2008/0032). In particular, it focuses upon the methodological issues and problems faced while cleaning and matching the data used for the assessment of EU funded ICT research performance. In order to achieve the objectives of the study, three main sources of information have been used:

a) Output indicators of DG INFSO funded projects;
b) Scientific publications (Thomson-ISI);
c) Patent data (PATSTAT-KITeS).

The three data sets have combined and integrated along different dimensions in order to produce a database amenable to statistical and bibliometric analysis. The key source of information used in the study is represented by the data on patents and publications produced by projects funded by the DG INFSO [point a) above]. This information has been collected by the DG INFSO itself through a mail questionnaire. The survey has been administered to a sample of project coordinators using MS Excel.

Each Excel document contains five spreadsheets:

1. **Questionnaire**: it provides qualitative and quantitative information on the project, e.g. acronym, type of project, number of papers and patents produced.
2. **Patents**: it provides information on the patent applications resulting from the research project.
3. **Publications-Articles**: it provides information on the articles resulting from the project and published in peer-reviewed scientific journals.
4. **Publications-Papers**: it provides information on other types of papers (e.g. presentations to conferences) or other types of documents (e.g. PhD theses), which can be also considered as outcomes of the research project.
5. **Comments**: it contains further qualitative information.

Figures A1.1, A1.2, A1.3 provide screen shots of the Excel documents that were sent to the project coordinators. The templates of these Excel spreadsheets are also available at http://ec.europa.eu/dgs/information_society/evaluation/studies/s2008_01/index_en.htm.

Two important points need to be remarked. First, each spreadsheet has a specific name. In principle, there was no reason for project coordinators to change the name of spreadsheets. However, since names of spreadsheets were not blocked, several of them did so, creating inconsistencies in the naming of spreadsheets that had to be solved. Second, the Excel documents sent to project coordinators provide a precise *template* guiding the process of data collection. Unfortunately, most project coordinators did not respect the template provided. As we will explain in detail below, this created further inconsistency problems.

The raw data on the scientific and technological output of projects have been provided by the DG INFSO in two batches. The first batch has been delivered at the beginning of the study in January 2009 and it refers to data collected through the questionnaires sent out to project co-ordinators in the years 2005, 2006 and 2007. The second batch has been delivered at the beginning of January 2010 and it refers to data collected through the questionnaire sent out in the year 2008. This technical appendix refers to the methodology used to clean and standardise the first batch of data. The second batch of data has been cleaned and integrated with the first batch using the same methodological approach with only a few differences that are highlighted below.

The first batch of raw data collected from DG INFSO project coordinators have been stored and delivered to KITeS-Bocconi on a DVD, containing three folders for the years 2005, 2006 and 2007, i.e. the years for which the collection of data was carried out. Each folder contained a list of MS Excel files, where each file refers to a specific project. Overall, the DVD contained 1301 files[29]. The second batch of data was similarly delivered on a DVD containing 672 distinct files.

29 Please note that data for the same project may be contained in different folders for different years. Some project coordinators have used an incremental procedure for entering data, i.e. for each new year they reported only new documents in that year. Some other coordinators adopted a sort of stock procedure, i.e. each new year they reported new documents produced along with older documents reported in the previous years.

Figure A1. 1 - Template of the spreadsheet Patents

AA	Project partner	Application reference	Title	Number of applications for:								
				Patents concerning protection at:							Trademarks	Registered des
				national level	EU level	US level	EU and Japan level	EU and US level	US and Japan level	EU, US, and Japan		
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												

Source: http://ec.europa.eu/dgs/information_society/evaluation/studies/s2008_01/index_en.htm.

Figure A1. 2 – Template of the spreadsheet Publications - Articles

No	Title of the article	Journal								Authors									
					Pages		Date			Author_1		Author_2		Author_3		Author_4		Author_5	
		Name	Volume	Issues	from	to	Day	Month	Year	Initials	Surname	Initials	Surname	Initials	Surname	Initials	Surname	Initials	Surname
1																			
2																			
3																			
4																			
5																			
6																			
7																			
8																			
9																			
10																			
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			

Source: http://ec.europa.eu/dgs/information_society/evaluation/studies/s2008_01/index_en.htm.

Figure A1. 3 – Template of the spreadsheet Publications - Papers

No	Title of the article	Name of event	Authors									
			Author_1		Author_2		Author_3		Author_4		Author_5	
			Initials	Surname	Initials	Surname	Initials	Surname	Initials	Surname	Initials	Surname
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												

Source: http://ec.europa.eu/dgs/information_society/evaluation/studies/s2008_01/index_en.htm.

Box 1 – How to use Automate macros

The subfolder MACROS of the main folder AUTOMATE contains 5 routines:

1. COMPRESSXLS-ART for importing articles
2. COMPRESSXLS-PAT for importing patents
3. COMPRESSXLS-PAP for importing conferences / papers
4. COMPRESSXLS-QUE for importing questionnaires
5. COMPRESSXLS-COM for importing comments

After installing Automate, the steps to perform in order to import data are the following (We assume that data for 2005 have to be imported. Repeat the same steps for other years, e.g. 2006 and 2007):

1. Make sure that the files to be imported are contained in a specific directory (SOURCE), e.g. C:\\Tender-SMART\Automate\Input_2005
2. Create a directory where you want store imported data (DESTINATION), e.g. C:\\Tender-SMART\Automate\Output
3. Copy in the DESTINATION directory the template of the file to be imported. There are five templates stored in the subfolder Templates:
   (a) A2005.xls for articles
   (b) P2005.xls for patents
   (c) C2005.xls for Comments
   (d) Z2005.xls for Papers/conferences
   (e) Q2005.xls for questionnaire

   For example, to import data for articles, copy the template A2005.xls from the TEMPLATE subfolder to the DESTINATION folder.
4. Double click on one of the five routines, e.g. COMPRESSXLS-ART
5. If default names of the SOURCE and DESTINATION directories are different from the one specified above (points 1. and 2.), change them appropriately. Line 5 of the macro contains the name of the SOURCE directory, Line 9 contains the name of the DESTINATION directory. Double click on these lines and change the name of directories.
6. Press F5 (RUN)
7. The macro will import all the data from the Excel files into a unique file. After the procedure has completed, rename the output file and save it to a different directory.
8. If the macro stops because of errors in the input data, close the xls files currently open without saving them and try to fix the errors. One of the most common errors in the input data is that project coordinators have changed the name of spreadsheets, e.g. from Publications - Articles to Publications_Articles. In such a case, give the correct name to the spreadsheet and save the file overwriting it.
9. Once errors have been fixed, re-run the macro until all files have been processed.
10. It is possible that steps 8-9 have to be iterated depending on the type of errors contained in the input data.

VI.2 Raw data processing

The first step of our data cleaning methodology has been to import the data contained in the separate MS Excel files into a single MS Access database. To this purpose, we have built a specific routine using a software package specifically designed for this type of tasks. In particular, we used *Automate 7*, which is a tool that allows automating manual, repetitive IT tasks[30].

The procedure cannot automatically handle all the errors and inconsistencies mentioned above. Therefore, once imported, data needed a massive work of manual cleaning and parsing of information (see below). The Automate routine developed for this study has been delivered to the Commission in fulfilment of the contractual requirement for the construction of a portable IT tool designed to import data from different MS Office Excel spreadsheets into a single portable database (**Deliverable D1**). A set of instructions on how to use the Automate routines has been also delivered (see Box 1).

It is important to point out that the Automate routine has been built by tailoring it to the specific characteristics of the data provided by DG INFSO. However, the raw data contain such a large variety of errors and inconsistencies that any simple automatic procedure of data import is likely fail in most cases, i.e. data are either not imported or the procedure is unable to parse the correct information into relevant fields (some fields will contain missing information or will not contain correct information). As long as the data collection strategy for the next years will continue to be based on MS Excel files, it is also likely that the variety of errors and inconsistencies introduced by project coordinators while entering the data will change and therefore the importing routine will also have to be adapted.

In particular, the most common problems we could find can be summarised as follows:

- Both "articles" and "papers" spreadsheets contain many errors and inconsistencies. The most common problems we could find using the data are:
  - Changes in the name of spreadsheets (e.g. Publications_Articles instead of Publications - Articles)

[30] http://www.networkautomation.com/automate/7/. Automate 7 is a commercial product. A 30 day trial version is available from the website reported.

  - Wrong match among column names and data reported (e.g. journal publisher in the column for journal title)
  - Format of fields different from template provided to project coordinators (e.g. all authors into a single field)
  - Data reported into a single cell, i.e. not parsed into fields
  - Information for the same article split into several lines
  - Cells containing forced paragraphs within the same cell
  - Reference to appendix files or attached documents for information on articles
- "Articles" spreadsheets should, at least in principle, contain only scientific publications. Actually, they also contain many other types of documents, e.g. newspapers articles, press releases, websites etc. etc, which cannot be considered scientific publications.
- "Papers" spreadsheets (point 3 in the above list) should, at least in principle, contain documents different from scientific publications, e.g. presentations to seminars, conferences and so on. Actually, they sometimes contain also scientific articles strictly defined.
- The number of papers/patents reported in the questionnaire spreadsheet (point 1 in the above list) almost never matches with the number of papers/patents in the corresponding spreadsheets (points 2 and 4 in the above list).
- Patent data are highly incomplete. Patent application number is often missing and only the number of patents produced is indicated, titles of patent applications are either partial or completely missing, patent office where application has been filed is not indicated and so on.

Given the problems listed above, we would strongly suggest to revise the current strategy of data collection based on MS excel spreadsheets and to adopt an alternative approach, such as setting up an on-line database implementing data entry constraints that would force project coordinators to input data in the correct way and would avoid the many problems associated with the data available so far. The following section contains a concrete proposal on how to improve the data collection process.

VI.3 Improving the data collection process: a proposal

Improving the data collection process on the output produced by DG INFSO projects would permit the analyst to shift the allocation of time and resources from the phase of data cleaning and standardisation to the more productive task of data analysis. In order to achieve this objective, we suggest to:

1) *Avoid the use of MS Excel*. The drawback of MS Excel is that it is quite difficult to prevent users that have to enter data from changing some details of the sheet, e.g. name of the spreadsheet, name of the columns etc., thus creating inconsistencies once data are collected back. Moreover, it is also difficult to force users to respect the template provided for entering data. For example, an Excel spreadsheet allows the user to enter the name of the journal in the field for the authors of the article without issuing any error message. On the other hand, implementing such an "error issuing" system seems to be crucial in order to collect consistent information.
2) *Use an online database*. Rather than MS Excel files, we suggest the adoption of a system of data collection based on an online database. The advantage of such a system is that all data entered by project coordinators are stored into a unique database and there is no need to import hundreds or even thousands of separate files into a unique database. Project coordinators can be assigned a user id and a password, which allow them to enter data online at different times. Moreover, they can circulate the user id and password, rather than forwarding files through emails, to other project participants, thus sharing the burden of entering data.
3) *Avoid redundancy, by asking few relevant information from project coordinators*. Entering information on the output of projects can be a very time consuming and burdensome task. Rather than filling all relevant fields, many project coordinators may be tempted to adopt shortcuts, such as cutting and pasting information from other sources, without respecting the templates provided and including in the database irrelevant documents. The data collected so far contain many items that are hardly usable for any bibliometric assessment, e.g. press releases, news, websites and so on. If the objective of the data collection is to focus upon those products of the research that reach a sufficiently high standard of

quality and impact, the attention should be limited to publications in scientific journals and conference proceedings. In order to save time for project co-ordinators, and at the same time ensure that they enter correct information, we suggest that future collection of data requires them to enter the least possible amount of information that allows tracing publications in other databases. In this respect, one concrete possibility to evaluate is to ask project coordinators to enter only a single information, namely the digital object identifier (DOI) of the publication. The DOI is a unique label that allows identifying in a univocal and persistent way a computer readable object that can be found on the internet. More details can be found at the address www.doi.org. Given that virtually all publishers are nowadays are adopting the DOI system and therefore label the articles and publications they issue with this system, collecting this single information from project coordinators would significantly speed up the process and would also improve the precision and reliability of the data collected.

4) *Collect patent application number*. Similar considerations can be made regarding the collection of data on patents produced by DG INFSO projects. The template of the Excel files requires project coordinators to fill in too many and redundant fields. In principle, it would be enough to ask project coordinators to report the application number of each patent as well as the web link pointing to the patent itself into the Espacenet database (http://ep.espacenet.com/). These two information are sufficient to check the consistency of the information reported and to trace any further information on the patent.

VI.4 Cleaning and matching data

As mentioned above, the information imported in the single MS Access database needed a thorough work of cleaning and standardization before the data could be used for performing any bibliometric and scientometric analysis, which is the main objective of the study. The next two sections explain in detail the methodology followed for creating the database of scientific publications and patent applications to be used in such an analysis. Before doing that, Table A1.1 reports the total number of items by year and type of document contained in the single MS Access database. Please note that items may be duplicated, i.e. the same item reported in different years, as some project coordinators adopted a sort of stock procedure of data reporting, i.e. each new year they reported new documents produced along with older documents reported in the previous years. Also note that data related to patents are only rough estimates and must be interpreted with great caution, as information on patent application numbers is missing for a large number of projects (see below).

Table A1. 1 – Number of items in the single MS Access database

Year	Articles	Papers	Patents
2005	1846	5871	303
2006	3991	9476	275
2007	8312	13967	422
2008	5298	12628	528
Total	19447	41942	1528

VI.4.1 Cleaning and matching data on scientific publications

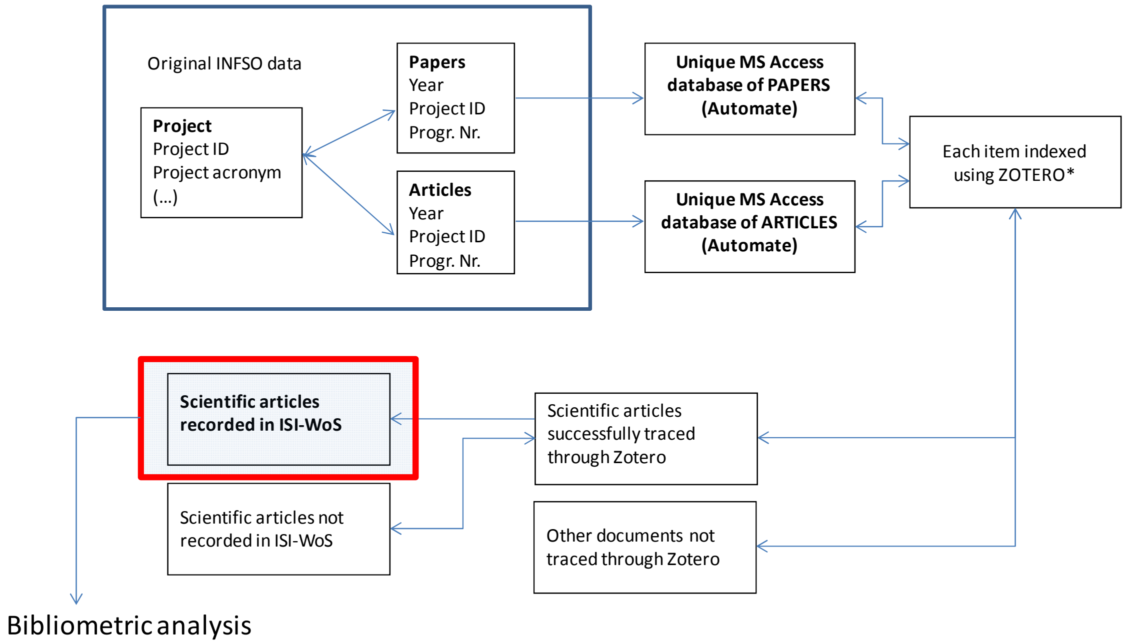
The methodology for constructing the database of scientific publications has been articulated in the following series of steps (see Figure A1.4 for a graphical illustration):

1. Each document contained in the "Articles" Excel spreadsheets and imported into the MS Access database has been manually checked using *Zotero*, which is an open source extension of the Firefox browser that is able to detect books and articles on the web, to store the reference and to manage it through reference management tools, such as Bibtex (http://www.zotero.org). In order to test the reliability of *Zotero* in tracing publications, we have carried out a data quality assessment (see below), which confirmed the validity of the methodology adopted.

2. From the set of "Papers" Excel spreadsheets imported into the MS Access database, we have extracted a subset of documents, through a keyword search strategy. In particular, we have attempted to identify within this set of documents the subset corresponding, at least potentially, to *scientific artic*les. Keywords for Proceedings, IEEE, Journals and variants of them have been used to identify the subset of potential scientific publications. For this subset of potential scientific articles, we have implemented the same Web search strategy through *Zotero*.

Table A1.2 reports the number of "papers" that are potentially scientific publications and that have been indexed through Zotero.

Figure A1. 4 – Scientific publications: data cleaning process



* For "Papers", only items corresponding to "potential" scientific publications have been indexed through Zotero. According to the guidelines provided in the tender specifications, this set of spreadsheets should contain only conference presentations, workshops etc.

3. The outputs of steps 1. and 2. above are two distinct sets of data.
   (a) scientific articles successfully traced through Zotero
   (b) other documents that do not correspond to scientific articles (i.e. not traced through Zotero or simply not checked through Zotero)

4. Scientific articles successfully traced through Zotero have been matched to the Thomson-ISI Web of Science (ISI-WoS) database. In particular, we have used the Science Citation Index (SCI) and the Social Science Citation Index (SSCI) databases included in the ISI-WoS. This allowed us to extract further information, such as the number of times an article has been cited, which is crucial for performing subsequent scientometric analysis. The output of this step thus consists of two further subsets of data:

(a) scientific articles successfully traced through Zotero that have been published in journals recorded in ISI-WoS;

(b) scientific articles successfully traced through Zotero that have been published in journals not recorded in ISI-WoS.

The scientometric analysis carried out in this study has been almost exclusively performed on the subset corresponding to 4.(a), although the subset 4.(b) has been also used for the purpose of statistical reporting, e.g. counting the number of scientific papers produced by projects.

Please also note that for the second batch of data received the same methodology described above has been applied with one difference. Articles and papers have been checked through Zotero in order to parse the title of the journal and to identify those articles published in scientific journals published in ISI-WoS SCI and SSCI databases. However, while the first batch of parsed data has been matched with the source articles to extract further information such as the affiliations, the number of citations received and so on, the same matching process has not been performed for the second batch of data. For these documents, the only information that is available is whether the article has been published or not in a journal indexed in the ISI-WoS SCI and SSCI databases.

Table A1. 2 – Number of "papers" that are potentially scientific publications

Year	All "Papers"	Potential scientific publications
2005	5871	1125
2006	9476	2185
2007	13967	2949
2008	12628	2767
Total	41942	9026

The table provides the number of "papers" in the single MS Access database that are potentially scientific publications Potential scientific articles in the "Papers" spreadsheets are those where the "Event" field (see above Table A1.1) contains keywords such "Journals", "Proceedings", "IEEE" and so on.

5. For all other documents not corresponding to scientific articles- i.e. "papers" that are not potentially scientific publications (point 2. above) and documents that af-

ter checking through Zotero do not correspond to scientific articles (point 3.(b) above), we created a specific database organised into the following fields:

(a) project acronym
(b) project ID
(c) year of data collection
(d) progressive number of document
(e) original fields (i.e. title etc.)
(f) category of document (e.g. website, press release, conference presentation, MSc thesis and so on) (see below).

In summary, the methodology outlined above generated three types of outputs:

1. a database of scientific articles produced by projects and recorded in the ISI-WoS SCI and SSCI databases;
2. a database of scientific publications produced by projects, but not recorded in the ISI-WoS SCI and SSCI databases;
3. a database of other documents produced by projects, organised by category of documents.

Figure A1.5 provides a graphical illustration of the basic structure of the resulting database. For each item contained in the original MS Excel files "Articles" and "Papers" provided by the DG INFSO, a bridge table has been constructed providing information on the type of document. Documents have been classified into the following categories:

a) Books, book series, book chapters (B)
b) Scientific articles published in journals indexed in ISI-WoS (J)
c) Scientific articles published in journals not indexed by ISI-WoS (K)
d) Articles published in proceedings of conferences (indexed through *Zotero*) (C)
e) Articles presented at conferences (not indexed through *Zotero*) (D)
f) Websites of projects, articles in websites (W)
g) Invited talks, seminars, workshops not published (S)
h) Newspapers, press releases, TV news (N)
i) Other documents, e.g. MSc and PhD theses, posters, etc. (O)
j) Missing data, i.e. not enough information to classify item (M).

Figure A1. 5 – Database of scientific publications

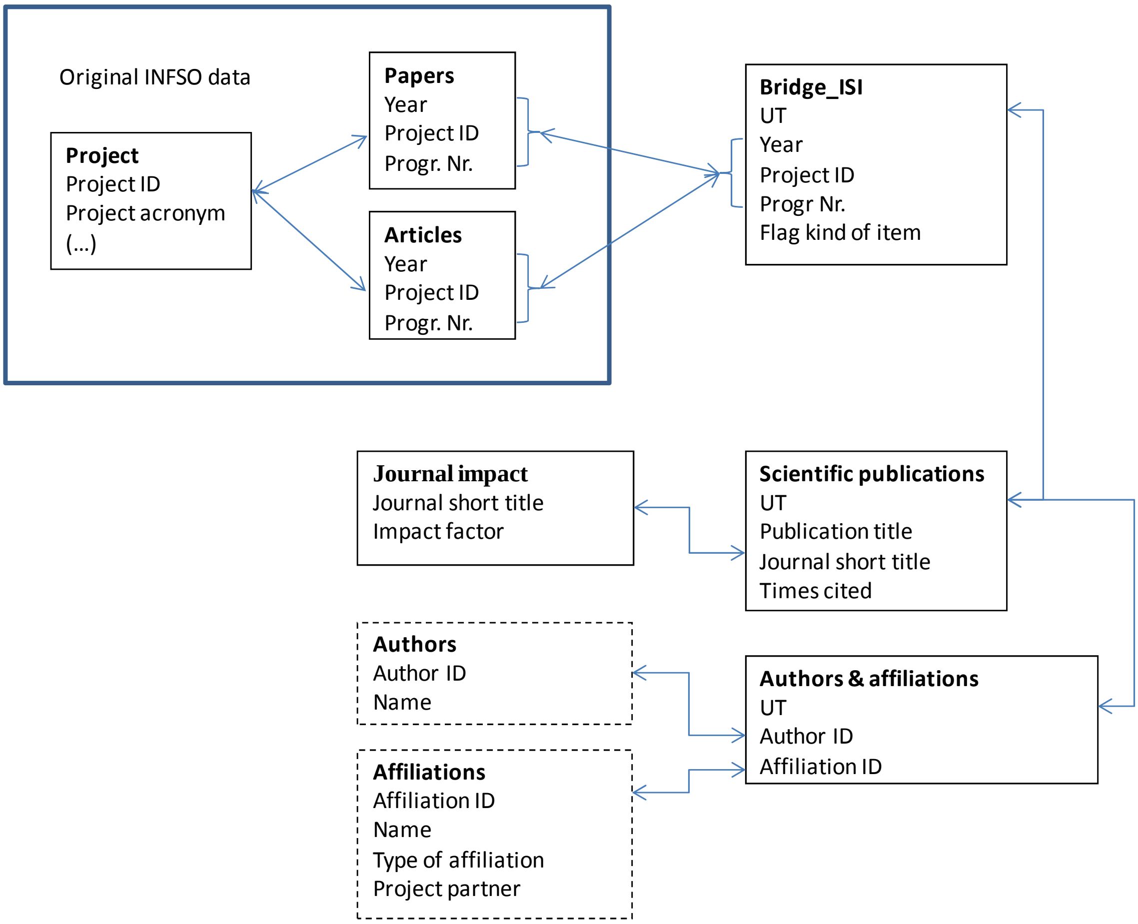


Table A1.3 reports the number of unique documents by category included in the final database. The subset of documents corresponding to category J, i.e. scientific articles indexed in ISI-WoS (SCI and SSCI), has been used for the scientometric analysis of the research performance of EU funded projects. For each article, the following information is available:

- unique ISI identifier (UT)
- standardised name(s) of author(s)
- title of article
- journal name (full and abbreviated)
- journal impact factor
- affiliation(s) of author(s)
- type of affiliation, i.e. company, university, public research organization
- affiliation address of authors, i.e. country and region

– year of publication
– number of citations received (i.e. times cited).

Table A1. 3 – Number of documents by category

Type of document	Number	%
Articles (J)	6783	100.0
Other documents		
Books (B)	1376	3.1
Conference (C)	9772	22.3
Conference (D)	18717	42.6
Non-ISI journals (K)	3248	7.4
Missing data (M)	2598	5.9
Newspapers (N)	715	1.6
Others (O)	2341	5.3
Seminars (S)	4204	9.6
Websites (W)	943	2.2
Total	43914	100.0

Data reported in Table A1.3 suggest that an important fraction of the scientific output produced by DG INFSO projects is not published in traditional scientific journals, but finds its way to publication as conference proceedings papers and articles. They account for around 65% of all other documents produced by DG INFSO projects. In this respect, it is important to point out that for the purposes of the present study, scientific output reported by project coordinators has been matched *only* to the SCI and SSCI databases included in the Thomson-ISI WoS database. The Thomson-ISI WoS database actually consists of several different databases, the most important of which are the Science Citation Index Expanded (SCI-Expanded), the Social Sciences Citation Index (SSCI), and the Conference & Proceedings Citation Index Science (CPCI-S). While the SCI and the SSCI contain information on articles published in peer-reviewed scientific journals that meet specific criteria established by Thomson-ISI (and thereby exclude other journals that do not qualify for inclusion), the CPCI-S provides information on published literature from the most significant conferences, symposia, seminars, colloquia, workshops worldwide.

The choice of matching scientific output of DG INFSO projects with the SCI and SSCI databases, and not with the CPCI-S database, has been dictated by two major reasons. First, the CPCI-S has become available as part of the ISI-WoS only at the end of October 2008 and most libraries do not yet include CPCI-S in their subscription to the ISI-WoS. However, the second and most important reason is related to the very poor quality of the raw data on conference papers provided by project coordinators. For example, in many cases only the acronym of the conference is used or the year of the conference is missing and so on. In these circumstances, even the adoption of sophisticated matching algorithms would still require a significant amount of manual work to ensure the quality of data. Given the very large number of conference papers to clean and match, this work would have far exceeded the amount of resources devoted to this study. Nonetheless, we firmly believe that *including conference papers in the bibliometric analysis could represent a natural and fruitful extension of this study*. At the same time, our experience in carrying out this work also suggests that any future evaluation study of FP projects based on bibliometric indicators would strongly benefit from *improving the methodology to collect scientific output data from project coordinators* (see above). Among other things, this would allow shifting resources away from the trivial, but very time consuming tasks of standardising and cleaning information to the more productive tasks of data analysis and evaluation.

It is also important to point out that *not all* conference papers reported as scientific output by project coordinators of DG INFSO projects are likely to be included in the CPCI-S. On the basis of the methodology adopted here, we could trace 9772 articles corresponding to category C, i.e. Zotero found a website of a publisher or of a library indexing the article and thus reporting information on the authors, title of the article, title of the conference, publication year and so on. It is likely that at least a fraction of such conference articles could be successfully matched with the ISI-WoS CPCI-S. For the other conference papers (18717 articles corresponding to category D), we have not implemented any check using Zotero. However, also in this case, we cannot exclude that a certain fraction of these documents could be found in conference and proceedings indexed by ISI-WoS CPCI-S. Given the weight of conference papers in the overall scientific output reported by coordinators of DG INFSO projects, we believe that extending the bibliometric analysis to such documents could represent a fruitful avenue for further research in order to confirm and generalise the results found in this study.

VI.4.2 Testing the accuracy of Zotero

As explained above, the Zotero tool has been used to identify documents corresponding to scientific articles in the “Articles” and in the “Papers” spreadsheets. For each document reported in those spreadsheets, the title has been searched in Google. If Zotero is able to detect a digital repository in which the document is stored, it reports in the URL bar a specific symbol. By clicking on the symbol, the information on the article is stored in a local database and can then be retrieved and matched to other sources, such as the ISI-WoS. The validity of the procedure adopted thus depends on how accurately Zotero is able to detect an article on the internet. In order to test this crucial issue, we carried out a limited experiment as follows. From the set of around 210,000 ISI-WoS scientific articles used for the benchmarking analysis (see main report), we have extracted a random sample of 100 articles. For each article, the pair formed by the “title of the article” and the ”title of the journal” has been searched on Google and Google Scholar in order to test whether Zotero was able to detect a website where the article was stored. For those articles that Zotero was able to detect, the information was downloaded in a local database. From the information found through Zotero, we then extracted the title of the article and the title of the journal and the pair title of the article/title of the journal was searched again, but this time in the ISI-WoS database. Out of 100 ISI-WoS articles, we could find correct information through Zotero for 92 articles, i.e. we could match the information found through Zotero back to ISI-WoS. For 10 articles present in ISI-WoS, we were not able to trace back them from Zotero to ISI-WoS, for the following reasons: for 2 articles no information was found on the internet; for 5 articles Zotero was not able to detect the title of the journal; for 1 article the title of the journal was found, but it was incorrectly spelled. Even though the experiment has been carried out on a relatively small sample, the test suggests that Zotero represents a fairly reliable and accurate tool for the task of tracing scientific publications on the internet.

VI.4.3 Cleaning and matching data on patents

As far as patent data are concerned, the most effective way of identifying patent documents is *via* the patent publication and/or the application number. Such numbers allow tracing the patent document in official patent databases and to collect relevant information, i.e. patent office, application and priority date, name of applicant, names of inventors and so on. Unfortunately, the vast majority of patent documents reported in the MS Excel spreadsheets lack such information. Several project coordinators only reported titles of the patents and other scattered information. Even worse, many project coordinators only reported that they produced some patents (sometimes reporting only the number of patents produced), without indicating any other information that allows to trace them and collect relevant information. Table A1.4 reports a few examples of the kind of information reported by project coordinators in the field for the title of patents. For none of these cases, other information was available.

Table A1. 4 – Examples of missing information on patents

Title of the patent
Registered software code
Patent application in the field Error resilient Multimedia data transmission
Patent application in the field Scalable video coding
Patent application in the field Transmission of scalable video
Patent application in the field Error protection for scalable video
<details are not yet for publication>
Valve
Nokia: 3 patent applications during 2007 in relation to MASCOT.
See footnote
Not official
Mol Switch Device tba
MPEG21

Overall, we could identify 309 projects that reported *some* information on patents, i.e. for which the Excel spreadsheet on patents was not completely blank. However, for only 179 projects the Excel spreadsheets contained sufficiently detailed information, for some but not all patents reported.

For all patent documents with sufficient information, each patent document has been manually checked using the *Espacenet* database maintained by the EPO in order to verify the consistency of the information reported by project coordinators. After this consistency check, we collected further data on the names of patent applicants, technological codes and so on, following a two step procedure:

1) Each patent was matched to the PATSTAT(KITes) database. PATSTAT (i.e. EPO Worldwide *PATent STATistical* Database) is a single patent statistics raw database, held by the European Patent Office (EPO) and developed in co-operation with the World Intellectual Property Organisation (WIPO), the OECD and Eurostat. PATSTAT provides raw patent data coming from around 73 national patent offices worldwide, including of course the most important and largest ones such as the European Patent Office (EPO), the United States Patent and Trademark Office (USPTO) and the World Intellectual Property Organization (WIPO).
2) Since PATSTAT is updated every six months, the most recent patent applications are not necessarily contained in the PATSTAT release currently maintained at KITeS. For this reason, patent documents not found in PATSTAT were manually checked using Espacenet, in order to retrieve information on applicants, inventors, technological codes and so on.

After relevant information has been collected, raw data on applicants, inventors, addresses etc. have been cleaned and standardised. Overall, we could collect complete information for 457 patent applications, 297 of which are applications extended to the European Patent Office. For fuller details, please refer to the main text of the report. The final database on patents contains for each patent document the following information:

- patent application and publication numbers and dates
- patent offices where application has been filed
- standardised name(s) of patent applicant(s)
- geographical location (NUTS3 level) of applicant(s)
- type of applicant (University, large/small company, PRO)
- standardised name(s) of the inventor(s)
- geographical location (NUTS3 level) of inventors and applicants
- primary and secondary IPC classification codes

Moreover, as described in the main text of the report, a bridge table has been built in order to identify among the set of inventors, the subset of individuals that we have labeled as author-inventors, i.e. individuals who have been both inventors of patents and authors of scientific articles produced by DG INFSO projects.

VII. Appendix 2 – Scientific journals for benchmarking

This appendix reports the list of scientific journals used for the benchmarking analysis. For each subject category used for the benchmarking analysis, we report a table listing the journals selected for the analysis, with an indication for each journal of the number and share of all articles in that subject category produced by DG INFSO projects in the period 2004-2008, and of the total number and share of articles in that subject category published and included in the ISI-WoS database in the same time period.

Table A2.1 - Journals included in the subject category Engineering, electrical & electronic (2004-2008)

Full title	INFSO articles	% (a)	ISI-WoS articles	% (b)	(a)/(b)
IEEE PHOTONICS TECHNOLOGY LETTERS	93	11.3	4001	4.9	2.3
ELECTRONICS LETTERS	69	8.4	4753	5.8	1.4
JOURNAL OF LIGHTWAVE TECHNOLOGY	58	7.0	2214	2.7	2.6
IEEE COMMUNICATIONS MAGAZINE	40	4.8	849	1.0	4.7
MICROELECTRONIC ENGINEERING	37	4.5	2414	2.9	1.5
IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION	33	4.0	2384	2.9	1.4
IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES	32	3.9	2006	2.4	1.6
IEEE TRANSACTIONS ON SIGNAL PROCESSING	29	3.5	2167	2.6	1.3
IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS	26	3.1	2162	2.6	1.2
SOLID-STATE ELECTRONICS	22	2.7	1438	1.7	1.5
SEMICONDUCTOR SCIENCE AND TECHNOLOGY	18	2.2	1665	2.0	1.1
MICROWAVE AND OPTICAL TECHNOLOGY LETTERS	17	2.1	3061	3.7	0.6
IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS	16	1.9	924	1.1	1.7
JOURNAL OF MICROMECHANICS AND MICROENGINEERING	16	1.9	1789	2.2	0.9
SENSORS AND ACTUATORS A-PHYSICAL	16	1.9	2464	3.0	0.6
IEEE TRANSACTIONS ON COMMUNICATIONS	15	1.8	1339	1.6	1.1
JOURNAL OF VACUUM SCIENCE & TECHNOLOGY B	15	1.8	2756	3.4	0.5
IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS	14	1.7	711	0.9	2.0
IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY	14	1.7	1259	1.5	1.1
IEEE TRANSACTIONS ON APPLIED SUPERCONDUCTIVITY	12	1.5	2336	2.8	0.5
MICROSYSTEM TECHNOLOGIES-MICRO-AND NANOSYSTEMS-INFORMATION STORAGE AND PROCESSING SYSTEMS	12	1.5	926	1.1	1.3
SIGNAL PROCESSING	12	1.5	1156	1.4	1.0
IEEE SIGNAL PROCESSING LETTERS	11	1.3	1025	1.2	1.1
IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS	10	1.2	173	0.2	5.8
IEEE JOURNAL OF SOLID-STATE CIRCUITS	10	1.2	1150	1.4	0.9
IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY	10	1.2	579	0.7	1.7
OPTICAL AND QUANTUM ELECTRONICS	10	1.2	415	0.5	2.4
EURASIP JOURNAL ON ADVANCES IN SIGNAL PROCESSING	9	1.1	545	0.7	1.6
IEEE TRANSACTIONS ON ELECTRON DEVICES	9	1.1	2044	2.5	0.4
EURASIP JOURNAL ON APPLIED SIGNAL PROCESSING	8	1.0	661	0.8	1.2
IEEE JOURNAL OF QUANTUM ELECTRONICS	8	1.0	872	1.1	0.9
IEEE MICROWAVE AND WIRELESS COMPONENTS LETTERS	8	1.0	1010	1.2	0.8
IEEE TRANSACTIONS ON BROADCASTING	8	1.0	366	0.4	2.2
OPTO-ELECTRONICS REVIEW	8	1.0	257	0.3	3.1
PROCEEDINGS OF THE IEEE	8	1.0	527	0.6	1.5
IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I-REGULAR PAPERS	7	0.8	989	1.2	0.7
FREQUENZ	6	0.7	211	0.3	2.8
IEEE SENSORS JOURNAL	6	0.7	924	1.1	0.6
IEEE ELECTRON DEVICE LETTERS	5	0.6	1497	1.8	0.3
IEE PROCEEDINGS-COMMUNICATIONS	4	0.5	562	0.7	0.7
IEEE TRANSACTIONS ON MAGNETICS	4	0.5	3691	4.5	0.1
INTERNATIONAL JOURNAL OF COMMUNICATION SYSTEMS	4	0.5	311	0.4	1.3
MICROELECTRONICS RELIABILITY	4	0.5	1144	1.4	0.3
SIGNAL PROCESSING-IMAGE COMMUNICATION	4	0.5	241	0.3	1.7
BT TECHNOLOGY JOURNAL	3	0.4	302	0.4	1.0
IEE PROCEEDINGS-MICROWAVES ANTENNAS AND PROPAGATION	3	0.4	463	0.6	0.6
IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT	3	0.4	1280	1.6	0.2
IEICE TRANSACTIONS ON COMMUNICATIONS	3	0.4	2671	3.2	0.1
IET MICROWAVES ANTENNAS & PROPAGATION	3	0.4	185	0.2	1.6
JOURNAL OF MICROELECTROMECHANICAL SYSTEMS	3	0.4	763	0.9	0.4
IEEE ANTENNAS AND PROPAGATION MAGAZINE	2	0.2	204	0.2	1.0
IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II-EXPRESS BRIEFS	2	0.2	1054	1.3	0.2
IEEE TRANSACTIONS ON CONSUMER ELECTRONICS	2	0.2	853	1.0	0.2
IEEE TRANSACTIONS ON NUCLEAR SCIENCE	2	0.2	2007	2.4	0.1
JOURNAL OF ELECTRONIC IMAGING	2	0.2	433	0.5	0.5
JOURNAL OF ELECTRONIC TESTING-THEORY AND APPLICATIONS	2	0.2	170	0.2	1.2
Other journals	19	2.3	7868	9.6	0.2
Total	826	100.0	82221	100.0	

Note: the table reports the journals in this subject category in which DG INFSO projects have published at least two articles. These are the journals selected for the benchmarking.

Table A2.2 – Journals included in the subject category Computer science, Hardware & Architecture

Full title	INFSO articles (2004-2008)	% (a)	ISI-WoS articles (2004-08)	% (b)	(a)/(b)
ACM TRANSACTIONS ON DESIGN AUTOMATION OF ELECTRONIC SYSTEMS	2	1.3	140	1.2	1.1
ANALOG INTEGRATED CIRCUITS AND SIGNAL PROCESSING	3	2.0	413	3.4	0.6
COMMUNICATIONS OF THE ACM	1	0.7	781	6.5	0.1
COMPUTER	5	3.3	478	4.0	0.8
COMPUTER JOURNAL	5	3.3	258	2.1	1.5
COMPUTER NETWORKS	18	11.8	978	8.1	1.5
COMPUTER STANDARDS & INTERFACES	2	1.3	236	2.0	0.7
COMPUTER SYSTEMS SCIENCE AND ENGINEERING	1	0.7	159	1.3	0.5
DESIGN AUTOMATION FOR EMBEDDED SYSTEMS	1	0.7	27	0.2	2.9
IBM JOURNAL OF RESEARCH AND DEVELOPMENT	1	0.7	268	2.2	0.3
IEEE DESIGN & TEST OF COMPUTERS	2	1.3	213	1.8	0.7
IEEE MICRO	6	3.9	216	1.8	2.2
IEEE MULTIMEDIA	7	4.6	151	1.3	3.7
IEEE NETWORK	6	3.9	187	1.6	2.5
IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS	4	2.6	712	5.9	0.4
IEEE TRANSACTIONS ON COMPUTERS	15	9.9	687	5.7	1.7
IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING	5	3.3	99	0.8	4.0
IEEE TRANSACTIONS ON RELIABILITY	1	0.7	259	2.2	0.3
IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS	3	2.0	542	4.5	0.4
IEEE WIRELESS COMMUNICATIONS	17	11.2	288	2.4	4.7
IEEE-ACM TRANSACTIONS ON NETWORKING	3	2.0	425	3.5	0.6
IEICE TRANSACTIONS ON FUNDAMENTALS OF ELECTRONICS COMMUNICATIONS AND COMPUTER SCIENCES	1	0.7	1741	14.5	0.0
INTEGRATION-THE VLSI JOURNAL	1	0.7	162	1.3	0.5
JOURNAL OF COMPUTER AND SYSTEM SCIENCES	2	1.3	331	2.8	0.5
JOURNAL OF NETWORK AND COMPUTER APPLICATIONS	2	1.3	173	1.4	0.9
JOURNAL OF OPTICAL NETWORKING	13	8.6	180	1.5	5.7
JOURNAL OF SUPERCOMPUTING	2	1.3	321	2.7	0.5
JOURNAL OF SYSTEMS ARCHITECTURE	7	4.6	299	2.5	1.9
JOURNAL OF THE ACM	1	0.7	146	1.2	0.5
MICROPROCESSORS AND MICROSYSTEMS	1	0.7	184	1.5	0.4
MOBILE NETWORKS & APPLICATIONS	6	3.9	252	2.1	1.9
NETWORKS	4	2.6	254	2.1	1.2
PERFORMANCE EVALUATION	3	2.0	322	2.7	0.7
VLDB JOURNAL	1	0.7	154	1.3	0.5
Total	152	100.0	12036	100.0	

Note: all journals in this category have been selected for the benchmarking.

Table A2.3 – Journals included in the subject category Computer science, Information systems

Full title	INFSO articles (2004-2008)	% (a)	ISI-WoS articles (2004-08)	% (b)	(a)/(b)
ACM TRANSACTIONS ON DATABASE SYSTEMS	1	0.7	105	0.7	0.9
ACM TRANSACTIONS ON INFORMATION AND SYSTEM SECURITY	1	0.7	54	0.4	1.8
ACM TRANSACTIONS ON INFORMATION SYSTEMS	1	0.7	70	0.5	1.4
ACTA INFORMATICA	2	1.3	125	0.9	1.5
BELL LABS TECHNICAL JOURNAL	3	2.0	303	2.1	0.9
CLUSTER COMPUTING-THE JOURNAL OF NETWORKS SOFTWARE TOOLS AND APPLICATIONS	1	0.7	59	0.4	1.6
COMPUTER COMMUNICATION REVIEW	2	1.3	379	2.7	0.5
COMPUTER COMMUNICATIONS	18	12.1	1345	9.4	1.3
DISTRIBUTED AND PARALLEL DATABASES	3	2.0	69	0.5	4.2
IEEE PERVASIVE COMPUTING	4	2.7	158	1.1	2.4
IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE	12	8.1	354	2.5	3.2
IEEE TRANSACTIONS ON INFORMATION THEORY	25	16.8	2025	14.2	1.2
IEEE TRANSACTIONS ON MOBILE COMPUTING	4	2.7	277	1.9	1.4
IEEE TRANSACTIONS ON MULTIMEDIA	4	2.7	450	3.2	0.9
IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS	2	1.3	1269	8.9	0.2
INFORMATICA	3	2.0	156	1.1	1.8
INFORMATION AND SOFTWARE TECHNOLOGY	2	1.3	323	2.3	0.6
INFORMATION PROCESSING & MANAGEMENT	2	1.3	351	2.5	0.5
INFORMATION PROCESSING LETTERS	6	4.0	974	6.8	0.6
INFORMATION SYSTEMS	2	1.3	189	1.3	1.0
INFORMATION SYSTEMS FRONTIERS	1	0.7	117	0.8	0.8
INTERNATIONAL JOURNAL OF COOPERATIVE INFORMATION SYSTEMS	1	0.7	94	0.7	1.0
INTERNATIONAL JOURNAL OF DISTRIBUTED SENSOR NETWORKS	1	0.7	36	0.3	2.7
INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE	2	1.3	264	1.9	0.7
INTERNATIONAL JOURNAL OF INFORMATION SECURITY	4	2.7	52	0.4	7.4
INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS	2	1.3	457	3.2	0.4
INTERNET RESEARCH	2	1.3	156	1.1	1.2
JOURNAL OF COMMUNICATIONS AND NETWORKS	3	2.0	200	1.4	1.4
JOURNAL OF INFORMATION TECHNOLOGY	1	0.7	105	0.7	0.9
JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY	2	1.3	549	3.9	0.3
JOURNAL OF VLSI SIGNAL PROCESSING SYSTEMS FOR SIGNAL IMAGE AND VIDEO TECHNOLOGY	5	3.4	274	1.9	1.7
MEDICAL INFORMATICS AND THE INTERNET IN MEDICINE	1	0.7	94	0.7	1.0
METHODS OF INFORMATION IN MEDICINE	5	3.4	422	3.0	1.1
MIS QUARTERLY	1	0.7	159	1.1	0.6
ONLINE INFORMATION REVIEW	1	0.7	167	1.2	0.6
PHOTONIC NETWORK COMMUNICATIONS	5	3.4	262	1.8	1.8
RAIRO-THEORETICAL INFORMATICS AND APPLICATIONS	2	1.3	162	1.1	1.2
SIGMOD RECORD	5	3.4	136	1.0	3.5
WIRELESS COMMUNICATIONS & MOBILE COMPUTING	3	2.0	393	2.8	0.7
WIRELESS NETWORKS	1	0.7	283	2.0	0.3
WIRTSCHAFTSINFORMATIK	1	0.7	211	1.5	0.5
WORLD WIDE WEB-INTERNET AND WEB INFORMATION SYSTEMS	2	1.3	78	0.5	2.5
Total	149	100.0	13706	96.1	

Note: all journals in this category have been selected for the benchmarking.

Table A2.4 – Journals included in the subject category Computer science, Software engineering

Full title	INFSO articles (2004-2008)	% (a)	ISI-WoS articles (2004-08)	% (b)	(a)/(b)
ACM TRANSACTIONS ON GRAPHICS	6	6.7	478	7.2	0.9
ACM TRANSACTIONS ON PROGRAMMING LANGUAGES AND SYSTEMS	3	3.4	161	2.4	1.4
ACM TRANSACTIONS ON SOFTWARE ENGINEERING AND METHODOLOGY	2	2.2	46	0.7	3.2
ALGORITHMICA	2	2.2	337	5.1	0.4
COMPUTER GRAPHICS FORUM	4	4.5	296	4.5	1.0
COMPUTERS & GRAPHICS-UK	3	3.4	400	6.1	0.6
CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE	8	9.0	408	6.2	1.5
EMPIRICAL SOFTWARE ENGINEERING	2	2.2	88	1.3	1.7
FUNDAMENTA INFORMATICAE	8	9.0	444	6.7	1.3
IEEE COMPUTER GRAPHICS AND APPLICATIONS	1	1.1	173	2.6	0.4
IEEE INTERNET COMPUTING	4	4.5	273	4.1	1.1
IEEE SOFTWARE	9	10.1	301	4.6	2.2
IEEE TRANSACTIONS ON SOFTWARE ENGINEERING	6	6.7	297	4.5	1.5
IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS	1	1.1	463	7.0	0.2
JOURNAL OF FUNCTIONAL PROGRAMMING	2	2.2	106	1.6	1.4
JOURNAL OF SYSTEMS AND SOFTWARE	8	9.0	731	11.1	0.8
JOURNAL OF UNIVERSAL COMPUTER SCIENCE	4	4.5	430	6.5	0.7
MATHEMATICAL PROGRAMMING	4	4.5	378	5.7	0.8
SCIENCE OF COMPUTER PROGRAMMING	5	5.6	307	4.6	1.2
THEORY AND PRACTICE OF LOGIC PROGRAMMING	1	1.1	85	1.3	0.9
VISUAL COMPUTER	6	6.7	406	6.1	1.1
Total	89	100.0	6608	100.0	

Note: all journals in this category have been selected for the benchmarking.

Table A2.5 – Journals included in the subject category Optics

Full title	INFSO articles (2004-08)	% (a)	ISI-WoS articles (2004-08)	% (b)	(a)/(b)
APPLIED OPTICS	10	2.5	3772	8.0	0.3
APPLIED PHYSICS B-LASERS AND OPTICS	12	3.0	1476	3.1	1.0
FIBER AND INTEGRATED OPTICS	2	0.5	168	0.4	1.4
JOURNAL OF LUMINESCENCE	2	0.5	1442	3.0	0.2
JOURNAL OF MODERN OPTICS	6	1.5	1182	2.5	0.6
JOURNAL OF NONLINEAR OPTICAL PHYSICS & MATERIALS	2	0.5	198	0.4	1.2
JOURNAL OF OPTICS A-PURE AND APPLIED OPTICS	15	3.8	1131	2.4	1.6
JOURNAL OF OPTICS B-QUANTUM AND SEMICLASSICAL OPTICS	1	0.3	375	0.8	0.3
JOURNAL OF PHYSICS B-ATOMIC MOLECULAR AND OPTICAL PHYSICS	3	0.8	1903	4.0	0.2
JOURNAL OF THE OPTICAL SOCIETY OF AMERICA A-OPTICS IMAGE SCIENCE AND VISION	8	2.0	1358	2.9	0.7
JOURNAL OF THE OPTICAL SOCIETY OF AMERICA B-OPTICAL PHYSICS	14	3.6	1253	2.6	1.3
LASER PHYSICS	2	0.5	1173	2.5	0.2
OPTICA APPLICATA	1	0.3	197	0.4	0.6
OPTICAL ENGINEERING	3	0.8	2288	4.8	0.2
OPTICS AND SPECTROSCOPY	4	1.0	1158	2.4	0.4
OPTICS COMMUNICATIONS	22	5.6	4357	9.2	0.6
OPTICS EXPRESS	116	29.4	7759	16.4	1.8
OPTICS LETTERS	46	11.7	5040	10.6	1.1
OPTIK	1	0.3	540	1.1	0.2
PHYSICAL REVIEW A	124	31.5	10587	22.4	1.4
Total	394	100.0	47357	100.0	

Note: all journals in this category have been selected for the benchmarking.

Table A2.6 - Journals included in the subject category Physics, Applied

Full title	INFSO articles (2004-08)	% (a)	ISI-WoS articles (2004-08)	% (b)	(a)/(b)
APPLIED PHYSICS LETTERS	186	64.6	25565	34.6	1.9
INTERNATIONAL JOURNAL OF MODERN PHYSICS B	1	0.3	2244	3.0	0.1
JAPANESE JOURNAL OF APPLIED PHYSICS PART 1	17	5.9	9033	12.2	0.5
JAPANESE JOURNAL OF APPLIED PHYSICS PART 2	1	0.3	9033	12.2	0.0
JOURNAL OF APPLIED PHYSICS	66	22.9	17654	23.9	1.0
JOURNAL OF LOW TEMPERATURE PHYSICS	5	1.7	1232	1.7	1.0
JOURNAL OF PHYSICS D-APPLIED PHYSICS	4	1.4	3043	4.1	0.3
LOW TEMPERATURE PHYSICS	3	1.0	601	0.8	1.3
PHYSICA C-SUPERCONDUCTIVITY AND ITS APPLICATIONS	3	1.0	3917	5.3	0.2
PLASMA PROCESSES AND POLYMERS	1	0.3	313	0.4	0.8
TECHNICAL PHYSICS LETTERS	1	0.3	1322	1.8	0.2
Total	288	100.0	73957	100.0	

Note: underlined in bold the scientific journals selected for the benchmarking.