# R&D collaboration networks in the European framework programmes: data processing, network construction and selected results

UNITED NATIONS UNIVERSITY

UNU-MERIT

# Working Paper Series

#2007-032

# R&D collaboration networks in the European Framework Programmes: Data processing, network construction and selected results

**Thomas Roediger-Schluga and Michael J. Barber**

United Nations University - Maastricht Economic and social Research and training centre on Innovation and Technology
Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands, Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499, e-mail: info@merit.unu.edu, URL: http://www.merit.unu.edu

# R&D collaboration networks in the European Framework Programmes: Data processing, network construction and selected results

**Thomas Roediger-Schluga[1] * and Michael J. Barber[2]**

* corresponding author

This version: 2007-11-12

## Abstract

We describe the construction of a large and novel data set on R&D collaboration networks in the first five EU Framework Programmes (FPs), examine key features and provide economic interpretations for our findings. The data set is based on publicly available raw data that presents numerous challenges. We critically examine the different problems and detail how we have dealt with them. We describe how we construct networks from the processed data. The resulting networks display properties typical for large complex networks, including scale-free degree distributions and the small-world property. The former indicates the presence of network hubs, which we identify. Theoretical work shows the latter to be beneficial for knowledge creation and diffusion. Structural features are remarkably similar across FPs, indicating similar network formation mechanisms despite changes in governance rules. Several findings point towards the existence of a stable core of interlinked actors since the early FPs with integration increasing over time. This core consists mainly of universities and research organisations. The paper concludes with an agenda for future research.

**JEL Classification:** L14, O38, Z13
**Keywords:** R&D collaboration, EU Framework Programmes, complex networks, small world effect, knowledge creation, knowledge diffusion, European Research Area

**UNU-MERIT Working Papers**

**ISSN 1871-9872**

**Maastricht Economic and social Research and training centre on Innovation and Technology, UNU-MERIT**

***UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at the Centre to stimulate discussion on the issues raised.***

[1] Oesterreichische Kontrollbank AG, Information Services, Strauchgasse 1, A-1010 Vienna, Austria, Phone: +43 53127 2517, email: rot@oekb.at

[2] Department of Technology Policy, Austrian Research Centers GmbH – ARC, Donau-City-Strasse 1, A-1220 Vienna, Austria

## 1. Introduction

In the modern knowledge-based economy, it is well established that knowledge creation and technological innovation rarely happen in isolation. Rather, specialised actors from industry, science and government collaborate in complex ways, not least because of the growing complexity of new technology (see Pavitt 2005). Long viewed as a temporary, inherently unstable organisational arrangement, innovation networks have become the norm rather than the exception in modern innovation processes (see Powell and Grodal 2005). For instance, Hagedoorn and Kranenburg (2003) have documented a steep rise in the number of strategic alliances from the early 1970s to the 1990s.

The central importance of network arrangements is reflected in the various systems of innovation (SI) concepts (see Malerba 2005; Lundvall 1992) that emphasise the paramount importance of strong and efficient linkages between the different component parts of the system. These facilitate interactive learning and rapid knowledge diffusion such that new knowledge is created or existing knowledge is combined in novel ways (see Edquist 2005).

The SI concept has become the key foundations of modern science, technology and innovation (STI) policy. Strengthening linkages between innovating actors, in particular between science and industry, is a core element of regional, national and supranational STI policy (for a discussion of major international examples, see Caloghirou et al. 2002). At the European level, the prime examples are the European Framework Programmes (FPs) on Research and Technological Development (RTD). In these FPs, the European Union has (co-)funded thousands of transnational, collaborative R&D projects; projects aimed at supporting transnational collaboration and coordination in research; and projects supporting transnational mobility for training purposes. Since their inception in 1984, six FPs have been launched and the seventh has commenced in 2007. The main objective of these activities has been to strengthen Europe's science and technology capabilities and to promote the competitiveness of European industry through co-ordinating national policies, integrating national research communities, improving the integration of marginal actors, and bringing together actors with the most advanced resources and capabilities. This has created a pan-European network of actors performing joint R&D.

The EU FPs have attracted many research and evaluation studies, yet comparatively little work has been done on the networks they have induced. Two notable exceptions are Barber *et* *al.* (2006) and Breschi and Cusmano (2004). Barber *et al.* investigate static properties of the R&D networks in FP1-FP4 and find that they exhibit typical properties of large, complex networks. Breschi and Cusmano study the network of collaborations with industry participation in FP1-FP4. They find a rather dense and pervasive emerging network that branches around an "oligarchic core", whose centrality and connectivity has strengthened over programmes.

Further work in this area promises great potential. The EU FPs are the major source of public funding of transnational R&D in Europe. Therefore, the networks that have emerged in the EU FPs provide valuable information on the organisational fabric and social infrastructure of European science and technology. Knowing how the networks look, how the networks have formed and how the networks evolve in response to external stimuli is of great importance for designing, implementing and assessing new policy measures that aim at creating and deepening the European Research Area.

Moreover, the networks that have been induced by the EU FPs are very large for social science standards and quite interesting in that they involve a broader set of actors than other sources on R&D collaborations, in particular data on strategic alliances (for a description of major data sources, see Hagedoorn et al. 2000). Although the networks are moulded by a very particular set of framework conditions and findings thus cannot be generalised naively, the networks provide rich information on a different stage of the innovation process than networks generated from alliance data, patents, scientific publications or surveys.

However, obtaining suitable data is fraught with difficulties. In this paper, we describe the construction of a novel data source on the first five EU FPs, the sysres EUPRO database. To the best of our knowledge, this is the most complete and highest quality dataset currently available[3]. However, because of the heterogeneity and inconsistency of the available raw data, it is far from perfect and we thus discuss its strengths and limitations. We then construct networks, describe key structural features and illustrate their economic relevance. This yields

---

3 We are only aware of one comparable major data source on the EU FPs, the EU RJV database, which is part of the STEP-TO-RJVs database (Caloghirou and Vonortas 2000). It has been constructed in the TSER project 'Science and Technology Policies Towards Research Joint Ventures' and contains information on all projects funded in FP1-FP4 that have at least one participant from the private sector. The 6,300 research joint ventures, however, represent only a subset of the corresponding 20,700 projects with information on more than one participant included in the sysres EUPRO database.

insights on patterns of network formation and the social and institutional infrastructure of European research and technology.

The remainder of the paper is organized as follows. In Section 2, we briefly summarise the rationale, objectives and structure of the European Framework Programmes. In Section 3, we detail the acquisition and processing of publicly available raw data on EU funded projects and participants. In Section 4, we describe the result of this exercise, the sysres EUPRO database, and how we construct R&D collaboration networks from our dataset. In Section 5, we present select results on key properties of the R&D collaboration networks. In Section 6, we summarise our findings and conclude with an agenda for future work.

## 2. Rationale, objectives and structure of the European Framework Programmes

Originally set up to ensure the technological competitiveness of European industry, especially in high-tech industries, the EU European Framework Programmes have evolved considerably since their inception in 1984 (for excellent accounts of the history of EU research policy, see Peterson and Sharp 1998; Guzetti 1995). In particular, the institutional framework and objectives have co-evolved, which has impacted the Programmes' designs.

FP1 to FP3 (running from 1984–1987, 1987–1991 and 1990–1994) were essentially 'technology-push' programmes, deriving their theoretical legitimisation from the well-known 'market failure' argument. According to this line of reasoning, precompetitive R&D is an uncertain, risky and increasingly expensive activity whose results cannot be fully appropriated due to the public good nature of its output. Therefore, subsidies are required to restore private investment incentives and to reap the collective benefits of collaborative R&D in terms of creating critical mass, sharing costs, pooling risk and internalising knowledge spillovers.

In FP4 (1994–1998) and FP5 (1998–2002), diffusion of new technologies, integration of SMEs and user orientation became key objectives. Also, FP4 included for the first time a substantial budget for training of researchers and mobility related measures. Starting with FP5, the EU FPs acquired a stronger mission-orientation as they were expected to deliver responses to the major socio-economic challenges facing Europe.

The theoretical underpinning for this changing focus was the systems of innovation (SI) approach that conceptualises innovation as a complex, interactive learning process that involves a multitude of actors from all societal spheres (see, e.g., Edquist 2005). The systemic model provides complementary and novel directions for STI policy, including additional rationales for supporting collaborative R&D. These include the need to foster interactive learning as a key mechanism for knowledge creation; to optimise linkages between the different (sets of) actors involved in innovation processes that rely on increasingly complex knowledge bases; to diffuse new knowledge and technology rapidly and widely; and to build innovative capacity through equipping workers with the requisite knowledge and skills to thrive in an increasingly dynamic, knowledge-based economy (see, e.g., Lundvall and Borrás 2005).

FP6 (2002–2006) was a major break with the previous FPs. On the one hand, it focused on scientific and technological excellence in a way that resembled the technology-push-oriented FP2 and FP3. On the other hand, it expanded the scope of the FPs and gave them a new role by becoming the financial instrument to make the European Research Area (ERA) (European Commission 2000) a reality.

ERA is intended to overcome the European problems of research fragmentation, underinvestment in R&D, and the lack of co-ordination of national STI policies. This is to be achieved through creating a critical mass and reducing the duplication of efforts by promoting better co-operation and coordination between relevant actors at all levels. FP7, which has commenced in 2007, will continue in this direction, deepening ERA and carrying it further towards the development of the knowledge economy and society in Europe (CORDIS 2007a). It is designed for a period of seven years, until 2013.

The EU FPs have evolved not only in terms of objectives and design, but also in terms of funding and thematic priorities. Figure 1 shows that available research funding has grown from €4bn in FP1 to almost €18bn in FP6. The projected budget for FP7 is a massive €50.5bn. In the early FPs, the main funding areas were energy and ICT. Over time, industrial technologies and life sciences (including food and agriculture), as well as environmental research (including transport) have become increasingly prominent funding areas. Since FP4, R&D activities have been complemented with funding for the training and mobility of researchers, special support for SMEs, networking and exploitation activities.

Figure 1 about here

Despite the evolution of their objectives and their scope (an excellent comprehensive yet concise account is Barker and Cameron 2004), the fundamental rationale of the FPs as mid-term research programmes that support collaborative research in selected technological priority areas has remained unchanged. Moreover, all FPs share a few key structural elements (see Caracostas and Muldur 2001, p. 162). In particular, the EU only co-funds projects of limited duration that mobilise private and public funds at the national level (until FP6, approx. 50% of total project costs),

the focus is on multinational and multi-actor co-operations that add value by operating at the European level (see European Commission 2002),

all projects are proposed by self-organised consortia, and

the selection for funding is based on specific scientific excellence and socio-economic relevance criteria.

Therefore, R&D projects and R&D networks can be compared over time.

## 3. Data acquisition and processing

Given that EU projects are publicly funded, it is surprisingly difficult to obtain detailed information on them. The only publicly available data source is the CORDIS projects database (CORDIS search 2006), which lists information on funded projects and project participants. This database is run by a subcontractor who receives raw data from the different General Directorates (DGs) that co-ordinate the various thematic areas of an FP.

This process results in considerable delays before information on projects and participants becomes available. For instance, a sizeable amount of information on FP6 has only become available in 2006, the last year of its existence. Moreover, there is no information on the strength and duration of a partner's involvement in each project or on partner changes during a project's lifetime – the only way to find out is to retrieve the data regularly from the CORDIS projects database. We have done so for FP5 in 2004 and 2006 and were stunned by the amount of change in the raw data on project participants. In total, FP5 generated 55,430 unique records on project participants. Of these, only about half could be matched with the existing raw data. How many of the remaining records are completely new cannot be ascertained with the current state of data processing.

Most importantly, there is no direct information on project output. The only partially useful data source is the CORDIS Technology Marketplace database (CORDIS 2007b) which was set up during FP5 and seems to have been discontinued in FP7. It contains information on exploitable project output and is intended to be a platform that connects producers and potential users of new technologies and knowledge. Unsurprisingly, it is a highly biased data source, as it only lists results and organisations that were not exploited by project participants. This is mainly the case with participants from science.

### 3.1. Quality of raw data from the CORDIS projects database

The CORDIS projects database contains a great deal of information about EU-funded research projects and project participants. However, many challenges exist in processing the raw data into a usable form. In this section, we critically examine the available data, with specific emphasis on what difficulties the available data present to analysis.

In principle, the CORDIS projects database contains information on project objectives and achievements, project costs and project funding, start and end date, contract type, a standardised subject index, a freely specified index, as well as information on which call of which specific programme each project was funded. In practice, records are rarely complete. Only 88% of project records hold information on subprogramme areas (ideally corresponding to specific calls), 43% on project acronyms, 20% on other indices, ~50% on objectives and general information, 18% on achievements, 92% on contract types, 95% on start and end dates, ~ 45% on project costs and project funding and 8% on project URLs.

On project participants, the CORDIS projects database ideally lists information on the participating organisation, the actual participating department, a contact person, complete contact details, organisation type and URL. Until a fairly recent change of the front end of the database, it also included email addresses, telephone and fax numbers of contact persons.

In practice, the raw data on participating organisations is quite inconsistent. Organisations are spelled inconsistently in up to four languages (in the case of Switzerland) and labelled non-homogeneously. Entries may range from large corporate groupings, such as EADS, Siemens and Philips, or large public research organisations, such as CNR, CNRS and CSIC, to individual departments and labs. Moreover, organisations are subject to change which is reflected in changing organisation names.

Among heterogeneous organisations, only 58% of the records contain information on the unit actually participating. Department labels are completely incoherent, ranging from the organisation name to meaningful subunits like faculties, subsidiaries, institutes, centres, laboratories, to unidentifiable acronyms. Unfortunately, these labels not only represent completely different organisational scales but are apparently self-selected by project participants, resulting in an inconsistent labelling of organisations that partake in multiple projects. Information on older entries and the substructure of firms tends to be less complete.

Unless it lists administrative officials, managing directors, deans, etc., information on contact persons (available for 79% of the records) may be useful for identifying actual participants. The same applies to email addresses, which are available 25% of all records. However, these are no longer available (also for former participants), presumably to protect against spamming. Address details (available for ~90% of all records) can sometimes be used to identify participants, if they are unique and do not refer to some central administrative unit.

Data on organisation types is available for 78% of all records and highly inconsistent, presumably because participants self-select their organisational status. In principle, there are eight organisation types (Educations, Research, Industry, Consulting, Government, Consulting, Non-Commercial and Other). In practice, participants pick the organisation type they deem appropriate and the data is obviously not cleaned. As a result, the raw data on participants frequently lists two to six different organisation types for the same organisation, if this information exists.

Information on organisation sizes has only been available for a brief period of time during FP5, for only 2% of all records. Probably because of the new front end for searching the CORDIS projects database, data on project co-ordinators in FP6 is highly incomplete, frequently only listing the organisation name.

In a nutshell, all these problems mean that the raw data cannot be used for any kind of meaningful network analysis. Therefore, ARC systems research (ARC sys), an institute of the Austrian Research Centers, has embarked on creating a clean and consistent dataset from the raw data that can be retrieved from the CORDIS projects database. The results of this endeavour are organised in the sysres EUPRO database, where EUPRO stands for EU projects.

### 3.2. Data processing

Because of the various shortcomings of the raw data, any fully automated standardisation method is infeasible. Rather, the data need to be cleaned and completed by hand, which is a five step process:

1) Identification of unique organisation name. Organisational boundaries are defined by legal control and entries are assigned to the respective unique organisations, using the most recent available organisation name. Table 1 shows the organisations from science and industry that appear most frequently in the standardisation table of FP1-FP5. The large majority of records can be identified quite easily. Especially among firms, however, organisation names may have changed quite frequently due to mergers, acquisitions and divestitures.
2) Identification of unique organisation type. This is especially important for firms, for which the raw data is highly incomplete and messy. The process is relatively straight-

forward, the only real challenge being the distinction between public and industrial research centres. In the sysres EUPRO database, all for-profit (industrial) research centres are assigned the industry label.

3) Creation of economically meaningful subentities. This is the key step for mitigating the bias that arises from the different scales at which participants appear in the dataset. We would like to use the actual group or organisational unit that participates in each project, but this information is only available for a subset of records, particularly in the case of firms. ARC sys has decided to pragmatically define subentities that operate in fairly coherent activity areas. Wherever possible, subentities are identified at the second lowest hierarchical tier. In other words, each subentity ideally comprises one further hierarchical sub-layer. Thus, universities are broken down into faculties/schools, consisting of departments; research organisations are broken down into institutes, activity areas, etc., consisting of departments, groups or laboratories; and conglomerate firms are broken down into divisions, subsidiaries etc. Subentities can frequently be identified from the contact information even in the absence of information on the actual participating organisational unit. Illustrative evidence is provided in Table 2. Note that subentities may still vary considerably in scale.

4) Identification of the genealogy of participants. The sysres EUPRO database currently covers a period of more than 20 years during which organisations have changed. For static or comparative static analyses, labelling all organisations by their most recent valid name is sufficient. For dynamic analyses, however, we need more precise information, so the genealogy of the main participants in the EU FPs has been traced through internet searches and firm registries.

5) Regionalisation. The dataset has been regionalised according to the European NUTS (Nomenclature of territorial units for statistics) classification system (EUROSTAT 2005), where possible down to the NUTS3 level. Mostly, this has been done via information on postal codes.

Table 1 about here

Table 2 about here

The data cleaning process is labour intensive. Due to resource constraints, only organisation appearing more than 30 times in the standardisation table for FP1-FP5 have been processed so far. This may bias results; however, the networks have a structure (see below) such that the size of the bias is quite low.

### 3.3. Current status

Table 3 shows that the sysres EUPRO database presently comprises information on 43,317 projects over the period 1984 (first project starting dates) to 2012 (last scheduled project end date). At its present state of standardisation, the database includes 42,020 separate organisations that were involved in at least one project. This figure increases to 49,885 when we consider subentities. Data on the first four FPs is complete according to the CORDIS website. In FP5, a handful of R&D projects are missing. Information on these and on projects in FP6 have been retrieved from the CORDIS projects database. The data are currently being cleaned and will be added to the database as soon as possible.

Table 3 about here

Participating organisations are either coded as prime contractor (i.e. co-ordinator) or participant. There is no further information on participants' roles. Information on prime contractors is available for virtually all projects. Although projects by definition comprise at least two partners, information on additional participants is only available for a subset of the projects (see Table 3). However, this subset comprises a sizeable majority of the population for all FPs beginning with the second. This also applies to FP4 and FP5, where the apparent decline in projects with information on multiple partners is due to the addition of training, mobility and supportive measures, which mostly list only the main applicant.

By examining the number of active R&D projects, the above figures can be related to actual research activity over time. Figure 2 shows that the total number of active R&D projects increases until FP4 and reaches about the same level in FP5. Moreover, the figure makes clear that there is considerable overlap between the different FPs that is not only caused by the temporal overlap between them. According to the available data, it takes up to two years after the official beginning of an FP (FP3-FP5) for a sizeable number of R&D projects to kick-off. As the average duration of R&D projects is in the range of 31–35 months in each FP, the number of active projects peaks past the official end date of each FP. Projects funded in the final calls of the FPs may last more than four years past the FPs' official termination.

Figure 2 about here

## 4. Network construction from the sysres EUPRO database

We construct networks from the sysres EUPRO database, under the assumption that the contract data produces networks that reasonably approximate actual patterns of interaction. We start with the affiliation network of collaborative research projects and participating organisations. An affiliation network can be represented by a bipartite graph, which consists of two subsets of nodes with edges existing only between the two sets. In our case, one set are the collaborative R&D projects and the other set are project participants.

To simplify the analysis of bipartite graphs, it is common practice to construct collaboration networks, i.e. a unipartite or one-mode projection that preserves only one type of node and connects all nodes that share a common neighbour in the bipartite graph (see, e.g., Christensen and Albert 2006). We follow this practice and draw an edge between e.g. Alcatel and ABB if and only if these organisations participate in the same R&D project. Edges can be weighted if there are multiple collaborations between the same organisations. Thus, if Shell and the University of Cambridge participate jointly in two projects, the corresponding edge has a weight of two and so on. The *size* of each vertex is its degree in the bipartite graph, e.g. a project comprising ten organisations has size ten, as does an organisation participating in ten projects. The *degree* is defined as the number of direct neighbours in a graph.

In constructing the organisation graphs, we thus assume each project to be a fully connected subgraph, or clique, of organisations. This is an idealised graph type that, although not fully representative, is a reasonable approximation to the actual intra-project structure of all but very large projects. Since the vast majority of projects in our data set have fewer than 15 participants, our construction rule is considerably more accurate than assuming the other idealised type of a star structure, in which each participant is only connected to the project coordinator as central vertex.

To keep our analyses consistent across the FPs, we only consider R&D projects and exclude all training, mobility and accompanying measures in FP4 and FP5. We construct the networks using information on subentities. As described in detail above, this information is not available for all records. As an illustration, Figure 3 plots the total number of records per organisation in FP5 against the number of records for which a subentity is available for the three main organisation types. The figure shows that the frequent participants have been broken down quite successfully into subentities. The data becomes noisier in the middle and the lower range of participations. Plots for FP1 to FP4 are qualitatively similar.

Figure 3 about here

The challenge in constructing a network from such data is to exclude false or spurious links that have a large impact on the network structure while generally retaining as much information as possible and specifically preserving real links. Since participation intensity correlates with organisation size, which in turn correlates with organisational focus, the greatest source of bias are unspecific frequent participants. For instance, project participations by the parent organisation CNRS or EADS do not contain any information as to which subunit actually did participate in the project, but may connect completely unrelated participants. Moreover, because of their frequent participation, these parent organisations tend to create the greatest number of spurious global bridging ties in the network.

With less frequent participants, the problem is less pronounced. First, the information on subentities may actually be correct. For instance, a medium-sized firm may have a few foreign subsidiaries. Second, such organisations tend to be more focused and therefore have more localised links. Third, even if they create spurious ties, these are less frequent and therefore have a smaller impact on the network structure.

For this paper, we have therefore excluded all unspecific entries from organisations participating in more than 50 projects. This translates into removing unspecific entries from the top 1% of organisations in the participation frequency ranking in FP1 to FP5 (see Figure 5).

There is no perfect solution to the potential sources of bias in the data that we have highlighted. However, we have run all the subsequently reported analyses during the continual refinement of the data and have obtained qualitatively and quantitatively similar results, apart from extreme values, e.g. the maximum degree. This makes us confident as to the robustness of our results.

## 5. Empirical characteristics

As formal network formation rules are minimal – a project has to comprise at least two partners from two different countries – we would expect similarly minimal structures. However, the R&D collaboration networks induced by the EU FPs have a great deal of structure that would not be anticipated from the minimal rules.

Prior work by Barber *et al.* (2006) has shown that slightly differently constructed organisation and project networks in FP1 to FP4 are complex networks that share common topological features with many empirical networks in the natural, technological and social domains (see, e.g., Strogatz 2001). A similar result has also been obtain by Powell et al. (2005) for a completely self-organised collaboration network in the life sciences.

We proceed by analysing global structural networks characteristics and by examining the patterns of multiple collaborations between organisations. We then investigate the organisational background of participants and entry and exit patterns over time. The section concludes with the identification of network hubs.

## 5.1. Global network characteristics

In Table 4, we give some basic properties of the organisation networks for FP2–FP5. We exclude FP1 as it was the first program launched and the available data are rather incomplete, making it exceptional in many respects.

Table 4 about here

The increase in the number of vertices *N* shows that a growing number of organisations have participated in subsequent FPs. Most of these are linked to each other. A giant component is present in all FP R&D collaboration networks. In each case, the great majority of vertices and essentially all edges are in the giant component.

The existence of a giant component indicates that two arbitrary vertices are connected, either directly or indirectly through a path of connected vertices, with high probability, ensuring that information or objects can spread or diffuse in a network. This means that even in the absence of other, unobserved communication channels, information can spread in the observed networks.

As expected from prior work (Barber et al. 2006; Breschi and Cusmano 2004), all networks are of small world type (see Watts and Strogatz 1998), including FP5: They exhibit a high clustering coefficient (a measure of local connectedness), a small characteristic path length and a diameter that scales at most logarithmically with the number of vertices.[4] Networks with high clustering coefficients are called *cliquish*.

In terms of what we presently know about knowledge creation and knowledge diffusion in exploration networks (see Cowan 2006), this is a positive result. When path lengths are short, new knowledge can spread rapidly and widely through the population and thus fuel local knowledge creation. Dense local connections facilitate learning. Agents can only learn from each other if they know different things but are sufficiently similar to communicate. As in a barter economy, there must be a double coincidence of wants. This constraint is relaxed if agents can communicate through joint neighbours, of which there are many in cliquish networks.

---

[4] The network-specific *clustering coefficient C* is defined as the average fraction of triangles, i.e. triples of connected vertices, of which vertex *i* is a part and the number of theoretically possible triangles, if all neighbours were connected. The *characteristic path length* $\ell$ is the average shortest path between any two vertices in a connected graph. The longest shortest path in a network is its *diameter*.

Knowledge transmission is limited by absorptive capacity, the ability to make sense of and to leverage new knowledge. Knowledge degrades as it is passed along long chains. If new knowledge is difficult to absorb (e.g. because a considerable part is tacit) or if transmission requires repeated interaction, the redundancy of ties in cliquish networks is again beneficial. It facilitates the validation of new knowledge and the possibility for multiple interactions. This is precisely the case in exploration networks (a term coined by Rothaermel and Deeds 2004), where knowledge is less-codified and there is a great deal of diversity in search activities, such as in the pre-competitive R&D collaboration networks we study.

There is a slight increase in the clustering coefficient from FP2 to FP5. This suggests that integration between collaborating organisations has increased over time, indicating that Europe has already been moving toward a more closely integrated European Research Area in the earlier Framework Programs (on this, see Breschi and Cusmano 2004).

The mean degree in the R&D collaboration networks is roughly constant over time. We interpret this as evidence that organisations have a roughly constant capability to maintain connections to one another. However, the mean degree is not terribly informative as it does not divide the population into two roughly equal halves. Rather, only around a quarter of vertices have a degree higher than the mean, indicating a skewed degree distribution.

Figure 4 displays a log-log plot of the degree distributions of the R&D collaboration networks in FP2-FP5. It shows that the distribution is strongly right-skewed with a heavy tail. This feature has been found in many large real-world networks, in such distinct domains as the technological, biological, psychological and social realms (see Newman 2003). The parts of the distribution that show up as a straight line in a log-log plot are distributed according to a power law. The power-law or Pareto distribution is also called a scale-free distribution and the so-characterised network scale-free (for an illustration, see Watts 2004, p. 251).

Figure 4 about here

This property has important empirical implications. While the majority of vertices have a less-than-average degree, some vertices have a degree that is orders of magnitude larger than the average. These high-degree network 'hubs' are the most visible members and tend to be pivotal for the coherence of the network. Scale-free networks are highly resilient to the random removal of vertices, as most vertices are only linked to a few others (a property referred to as 'attack tolerance' or 'ultraresilience'). However, if the highly connected hubs are removed, the giant component quickly falls apart into smaller, disconnected components, disrupting any global transmission process. Below, we will characterise the hubs of the R&D collaboration networks we study.

Similarly, the vertex size distributions have scale-free characteristics. The distribution of organisation sizes is scale-free for essentially all values (see Figure 5). It is remarkably similar across FPs, indicating that the distribution of organisations able to carry out a particular number of projects has not changed over time. A complementary interpretation of this finding is that the changes in the underlying research activities have not altered the mix of organisations participating in a particular number of projects in each FP.

Figure 5 about here

The distribution of project sizes, as measured by participation, is indicative of a typical range of between roughly 5 and 10 participants. The project size distribution is highly skewed, with over 95% of the projects in FP5 having at most 15 participants while the largest project has over 100 participants. Average project size increases across the FPs, which is consistent with recommendations from evaluation studies and the stated attempts of the EU commission to reduce its administrative burden. The overall shape of the distributions, however, is remarkably similar. This suggests that possible changes in project formation rules – including both formal policies and informal practices – did not affect the aggregate structure of the resulting research networks.

### 5.2. Edge weights and network cores

Thus far, we have focused simply on the presence or absence of links between network nodes. Figure 6 shows the edge weight distribution for FP2-FP5, where edges are weighted by the number of joint project participations of organisations. Surprisingly, there is an almost perfect power-law behaviour for each case, with maximum edge weights of approximately 30.

Figure 6 about here

There are several possible explanations for this result. Organisations may collaborate in greater number of projects because they share a greater commonality or because they have developed trust and joint experience through prior collaborations. Also, organisations that are active in a wider set of complementary activities may have multiple collaborations in different projects. In this case, intra-organisational links and knowledge flows may also be of importance, as search for potential partners may be influenced by the collaboration behaviour of other actors within an organisation. Finally, multiple collaborations in different projects may indicate that only a limited set of organisations is able (or willing) to carry out research in certain areas. Of course, it may also reflect network closure if actor configurations are highly stable over time. However, it is beyond the scope of this work to examine this issue at greater depth.

High edge weights suggest that there is a robust backbone structure of closely interconnected organisations in each FP. As an illustration, we have extracted from the FP5 R&D collaboration network all vertices connected by edges with a weight of at least ten. Figure 7 displays a network map, which has been created with Pajek (Batagelj and Mrvar 2003), a powerful freeware programme for the descriptive analysis and visualisation of large networks. We first applied the Fruchterman-Reingold (1991) graph drawing algorithm and subsequently manually repositioned overlapping vertices.

Figure 7 about here

The figure shows clusters of strongly connected organisations in several thematic areas. The three main ones are geo and environmental sciences in the northeast, aerospace research in the south and automotive and transport research in the west. Within the aerospace cluster, the left half focuses on aerodynamics, while the right half does research on turbines. As well, there are strongly connected organisations in the life sciences, aquaculture, helicopters, informatics, and maritime and naval technology.

In some areas, we find strong national and interorganisational links. For instance, the life science grouping in the north consists exclusively of French organisations, as does the informatics quadruple in the west. Eurocopter France and Germany are strongly linked, as are the four national subsidiaries of Airbus and Rolls Royce with its German subsidiary.

Another interesting feature is the ratio of vertex size to edge weight. Vertex sizes range from 12 to 245. The minimal edge weight is ten, which means that some organisations virtually always collaborate with certain other organisations in R&D projects in FP5.

### 5.3. Organisational background of participants

The networks at hand include a diverse set of actors. We start with analysing actors' identities within FPs. Figure 8 displays the distribution of organisation types for each of the five FPs. The figure on the left is generated from the total set of project participations, while the figure on the right is based on counts of distinct organisations. Both figures show that the vast majority of participants in EU projects are firms, universities or research organisations.

Figure 8 about here

We would like to emphasise that our results do not fully reflect the results of official communications and evaluation studies on EU FP participation. In particular, we find greater industry participation. Unless the raw data published in the CORDIS data base is substantially biased, this may be due to its shortcomings described in Section 3.1. We have devoted great care to cleaning and completing the raw data, which is particularly messy in the case of firms (see Section 3.2). Should the Commission use raw or less processed data in its annual research reports, 5-year assessments, etc., this may explain the difference.

The figure also shows that actors from science have a higher participation intensity than actors from industry. Averaging across all FPs, an industrial actor participates in a mean (standard deviation) number of 1.9 (3.6) projects, a research organisation in 3.8 (8.0) projects and a university in 4.5 (7.1). We interpret this result as evidence for different organisational attitudes towards the kind of research captured by our data. The large variation of our results indicates considerable heterogeneity within the different groups of actors.

Universities and research organisations mainly conduct exploratory research in the 'open science' mode, while firms focus on exploitation governed by the norms of 'proprietary technology' (see Dasgupta and David 1994). The disclosure rules stipulated in the EU FPs work well with exploratory research in the open science mode, but are ill-suited for exploitation. Exploiting existing capabilities that are critical for industrial competitiveness requires secrecy and is therefore typically funded internally. Indeed, research on Finnish firms has shown that these often set up parallel, internal projects in which they exploit results obtained in EU-funded research (Luukonen 2002).

While actors in science mainly conduct exploratory research, it constitutes a smaller part of firms' R&D activities (albeit one that is critical for long-term competitiveness). Accordingly, EU projects are a natural way of funding academic research which is reflected in greater participation intensity by scientific actors. In contrast, firms participate in fewer projects, mainly to acquire knowledge critical for longer-term success or to create future markets, e.g. by setting standards in network technologies.

### 5.4. Exit and entry in the R&D collaboration networks

Next, we explore the identity of actors between FPs. Organisations may not only participate in multiple projects within FPs, but also across them. At the same time, other organisations exit and discontinue their involvement in EU supported research. Figure 9a shows the relative share of organisations returning to the subsequent FP and the share of exiting organisations.

The bar heights sum to 100% for each FP and are broken down into the main organisation types. Interestingly, until FP3, the majority of actors have also participated in the next FP. In contrast, almost two-thirds of the participants in FP4 have not returned to FP5. This shift is due to considerably higher exit rates by firms and other organisations.

Figure 9a about here

The figure shows that across all FPs, a greater number of firms has exited than returned to the subsequent FP. In contrast, the majority of universities and research organisations overlap between FPs, indicating considerably greater stability among these actor groups. This is consistent with our empirical finding on divergent participation intensities as well as our argument on different participation motives.

While Figure 9a displays exit data, Figure 9b shows entry rates. Again, the height of the bars, which sum to 100% for each FP, corresponds to the share of returning and new actors and are broken down into the main organisation types.

Figure 9b about here

Since we are dealing with growing networks, the number of new actors is always considerably greater than the number of actors that overlap. In all FPs, the largest group of entrants are firms. Interestingly, other actors, i.e. governmental organisations, industry associations, non-profit organisations, etc. are the second largest group of new entrants in FP5. This may reflect its greater user orientation compared to the preceding FPs.

Figure 9b shows that the growing size and widening scope of the FPs has attracted a large number of new actors, most of which are firms. At the same time, there is sizeable stability among a subset of participants. 911 participants in FP1 (35%) have also participated in all subsequent FPs. Of these, 47% are universities, 34% are research organisations and only 17% are firms. If we drop FP1, on which data is not as reliable as on the later FPs, the number increases to 2152 (34% of FP2). We obtain similar shares for FP3 and FP4. Thus, about one third of the actors taking part in each of the preceding FPs also participate in FP5. This suggests that there has been a growing organisational and social infrastructure in science and technology over the past two decades that is part of the core of the present day ERA. More detailed dynamic analyses promise interesting insights.

### 5.5. Network hubs

Above, we have noted that scale free networks contain highly connected hubs that are critical to their connectivity. In this section we will identify and characterise the hubs in the FP R&D collaboration networks.

Hubs have many direct neighbours, i.e. the have high degree *d*. Table 5 shows the top ten organisations ranked by degree for FP2-FP5. Comparing maximum degrees across the different FPs, we see a marked rise from FP2 to FP5. This due to the rising average project size and the rising local density of the later FPs.

Table 5 about here

The rankings are dominated by well-known research centres and universities. Different subunits of CNRS consistently rank in the top 10 in all FPs, as does the Faculty of Engineering of the Imperial College London. Other organisations that rank in the Top 10 in more than one FP include the Dutch TNO and the Finnish VTT, the Fiat Centre of Research and perhaps unexpectedly, the Faculties of Engineering of Southampton and Stuttgart University. Apart form Fiat, only few firms rank in the Top 10. In FP3, we find Daimler Benz and in FP2 Bull Europe and Siemens Central Research.

Table 5 also shows the number of projects each organisation has participated in. These numbers show that each of the organisations has been very active in the respective FPs. However, a high participation intensity does not automatically translate into a high degree. This depends on the local structure of the network, i.e. in what kind of projects organisations participate and the density of local links. Thus, a high participation intensity is not a reliable indicator for the network position of an organisation. Indeed, the picture would change even more dramatically if we used other standard centrality measures (see Roediger-Schluga and Barber 2006). The relational information contained in the network linkage structure offers additional insights that cannot be obtained from counting attribute data on the participants.

## 6. Conclusions and directions for future research

In this work, we have described the construction of the sysres EUPRO database, a novel data source on the first five EU FPs that includes all publicly available information on projects and participants. Raw data has been retrieved from the CORDIS project database. However, the raw data has several shortcomings that necessitate extensive, mostly manual work to clean and complete the data.

Projects and organisations form affiliation networks, from which we have constructed organisation projections. These networks are substantial in terms of size, complexity and economic impact. We observe numerous characteristics known from other complex networks, including scale-free degree distributions, small diameters and high clustering. The networks thus exhibit the small-world property, which has been identified in theoretical work as conducive to collective knowledge creation and knowledge transmission in exploration networks. Other features in common throughout the FPs include the typical project sizes and the overall shapes of the various distributions observed. Presumably, network formation mechanisms are similar for all FPs despite changes in governance rules.

Two findings suggest the presence of a stable core of actors in science and technology since the early FPs: there is a significant overlap in participants for consecutive FPs and there is recurring collaboration amongst the same organisations within FPs. These organisations constitute the backbone of the FP R&D collaboration networks and figure prominently in the European Research Area. Moreover, the increasing clustering coefficient suggests that integration between collaborating organisations has increased over time, indicating that Europe has already been moving towards a more closely integrated European Research Area in the earlier Framework Programmes.

Further results stem from investigating vertex properties. We find that the majority of participants are firms, but that universities and research organisations display greater participation intensity and are positioned more prominently in the networks analysed.

Including the relational information contained in the networks shows that vertex size does not imply centrality. This suggests that a policy of creating larger projects may not be fully appropriate to foster networking and the connectivity of R&D collaboration networks in Europe. Rather, projects need to include pivotal actors, which seems to have been the case only partially in the first five Framework Programmes.

The present work points to considerable future work, of both empirical and conceptual nature. At the empirical level, it is clear that we need to refine our understanding of the substructure of the networks, in particular how organisations interact within projects. This includes identifying thematically homogeneous subnetworks and subgroups that are homogeneous in terms of structural properties and organisational mixing patterns. Moreover, the additional information included in edge weights needs to be integrated into structural investigations. Another major route of inquiry is the dynamic analysis of network formation and network configuration. This should also yield information on how the networks have been shaped by external constraints, in particular the governance rules.

Perhaps more fundamentally, there are many open questions at the conceptual level. Networking activities are publicly funded because they are expected to fulfil specific functions, e.g. knowledge creation and knowledge diffusion. Are the network structures that emerge well-suited for these functions? Do different network functions require different network structures? Does this introduce tensions between conflicting objectives, e.g. efficiency and equity? How do structure and function interact? To what extent can complex networks that involve a strong element of self-organised behaviour by decentralised actors be influenced through external stimuli, in particular by governance rules? Isolating relationships between network structure and function and identifying the scope for directing networks towards desirable structures will provide valuable guidance for policy makers in improving existing instruments and designing future ones.

## Acknowledgements

We acknowledge support from the European FP6-NEST-Adventure Programme, contract n° 028875, the Portuguese Fundação para a Ciência e a Tecnologia under Bolsa de Investigação SFRH/BPD/9417/2002 and Plurianual CCM. TRS gratefully acknowledges the support of CCM at Madeira Math Encounters XXX, where portions of this work were done. Earlier versions of this paper were presented at the International Schumpeter Conference 2006 and the Madeira Math Encounters XXXI. We would like to thank participants as well as a referee for helpful comments.

## References

Barber, M., A. Krueger, T. Krueger and T. Roediger-Schluga (2006), 'The network of EU-funded collaborative R&D projects', *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, **73** (3), 036132 (arXiv:physics/0509119).

Barker, K. and H. Cameron (2004), 'European Union science and technology policy, RJV collaboration and competition policy', in Y. Caloghirou, N.S. Vonortas and S. Ioannides (eds.), *European Collaboration in Research and Development*, Cheltenham, UK and Northampton, MA, US: Edwar Elgar, 154-86.

Batagelj, V. and A. Mrvar (2003), 'Pajek – Analysis and visualization of large networks', in M. Jünger and P. Mutzel (eds.), *Graph Drawing Software*, Berlin: Springer, 77-103.

Breschi, S. and L. Cusmano (2004), 'Unveiling the texture of a European Research Area: Emergence of oligarchic networks under EU Framework Programmes', *International Journal of Technology Management*, **27** (8), 747-72.

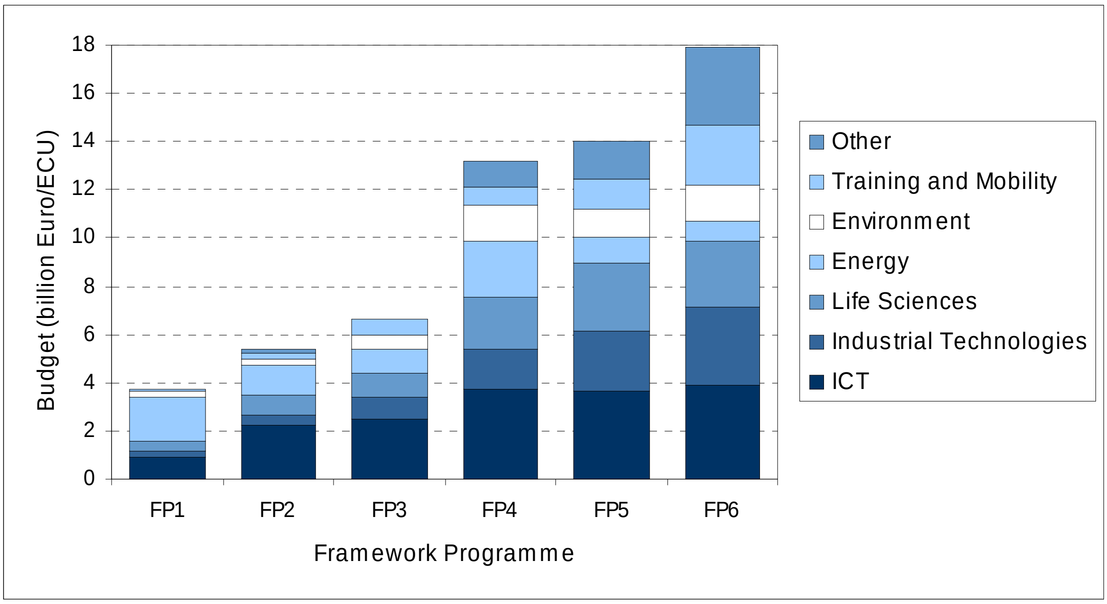
Caloghirou, Y. and N.S. Vonortas (2000), 'Science and Technology Policies Towards Research Joint Ventures', Final Report to the Commission, DG XII, TSER Programme, Contract No: CT97-1075, National Technical University of Athens, Laboratory of Industrial & Energy Economics, available

from http://improving-ser.sti.jrc.it/default/show.gx?Object.object_id=TSER—000000000000097E&_app.page=show-TSR.html.
Caloghirou, Y., N.S. Vonortas and S. Ioannides (2002), 'Science and technology policies towards research joint ventures', *Science and Public Policy*, **29** (2), 82-94.
Caracostas, P. and U. Muldur (2001), 'The emergence of a new European Union research and innovation policy', in P. Larédo and P. Mustar (eds.), *Research and Innovation Policies in the New Global Economy*, Cheltenham, UK and Northampton, MA, US: Edward Elgar, 157-204.
Christensen, C. and R. Albert (2006), 'Using graph concepts to understand the organization of complex systems', *Journal of Bifurcation and Chaos*, (forthcoming).
CORDIS (2007a), 'Understand FP7', available from http://cordis.europa.eu/fp7/understand_en.html (accessed May 25, 2007, last update 2007-05-22).
CORDIS (2007b), 'Technology Market Place', available from http://cordis.europa.eu/marketplace/home.html (accessed May 24, 2007, last update 15/02/2007).
CORDIS (2006a), 'FP6 Budget', available from http://cordis.europa.eu/fp6/budget.htm).
CORDIS (2006b), 'FP5 - EC Programme: Maximum Amounts and Breakdown (1998-2002)', available from http://cordis.europa.eu/fp5/src/budget.htm).
CORDIS search (2006), 'Projects Professional Search', available from http://cordis.europa.eu/search/index.cfm?fuseaction=proj.professionalSearch).
Cowan, R. (2006), 'Network models of innovation and knowledge diffusion', in S. Breschi and F. Malerba (eds.), *Clusters, Networks, and Innovation*, Oxford: Oxford University Press (forthcoming).
Dasgupta, P. and P.A. David (1994), 'Toward a new economics of science', *Research Policy*, **23**, 487-521.
Edquist, C. (2005), 'Systems of innovation – perspectives and challenges', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford: Oxford University Press, 181-208.
European Commission (2006), 'The Fourth Framework Programme', European Commission, available from http://ec.europa.eu/research/fp4.html, last update May 22, 2001).
European Commission (2002), 'What European added value?' European Commission, available from http://ec.europa.eu/research/era/leaflet/en/era04_en.html).
European Commission (2000), 'Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions – Towards a European Research Area', COM (2000) 6, Brussels.
EUROSTAT (2005), 'Nomenclature of territorial units for statistics - NUTS Statistical Regions of Europe', available from http://europa.eu.int/comm/eurostat/ramon/nuts/home_regions_en.html).
Fruchterman, T. and E. Reingold (1991), 'Graph drawing by force-directed replacement', *Software – Practice and Experience*, **21**, 1129-64.
Guzetti, L. (1995), *A Brief History of European Union Research Policy,* Luxembourg: Office for Official Publications of the European Commission.
Hagedoorn, J. and H. van Kranenburg (2003), 'Growth patterns in R&D partnerships: an exploratory statistical study', *International Journal of Industrial Organization*, **21**, 517–31.
Hagedoorn, J., A.N. Link and N.S. Vonortas (2000), 'Research partnerships', *Research Policy*, **29** (4-5), 567-86.
Lundvall, B.-Å. (ed) (1992), *National Systems of Innovation*, London: Pinter.
Lundvall, B.-Å. and S. Borrás (2005), 'Science, technology, and innovation policy', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford: Oxford University Press, 599-631.
Luukonen, T. (2002), 'Technology and market orientation in company participation in the EU framework programme', *Research Policy*, **31** (3), 1459-66.
Malerba, F. (2005), 'Sectoral systems: how and why innovation differs across sectors', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford: Oxford University Press, 380-406.
Newman, M.E.J. (2003), 'The structure and function of complex networks', *SIAM Review*, **45**, 167-256.
Pavitt, K. (2005), 'Innovation processes', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford: Oxford University Press, 86-114.

Peterson, J. and M. Sharp (1998), *Technology Policy in the European Union,* Houndsmill, Basingstoke and London: Macmillan.
Powell, W.W. and S. Grodal (2005), 'Networks of innovators', in J. Fagerberg, D.C. Mowery and R.R. Nelson (eds.), *The Oxford Handbook of Innovation*, Oxford: Oxford University Press, 56-85.
Powell, W.W., D.R. White, K.W. Koput and J. Owen-Smith (2005), 'Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences', *American Journal of Sociology*, **110** (4), 1132-205.
Roediger-Schluga, T. and M. Barber (2006), 'The structure of R&D collaboration networks in the European Framework Programmes', UNU-MERIT working paper, 2006-036, available from http://www.merit.unu.edu/publications/wppdf/2006/wp2006-036.pdf (submitted for publication).
Rothaermel, F.T. and D.L. Deeds (2004), 'Exploration and exploitation alliances in biotechnology: a system of new product development', *Strategic Management Journal*, **25**, 201-21.
Strogatz, S.H. (2001), 'Exploring complex networks', *Nature*, **410**, 268-76.
Watts, D.J. (2004), 'The "new" science of networks', *Annual Review of Sociology*, **30**, 243-70.
Watts, D.J. and S.H. Strogatz (1998), 'Collective dynamics of 'small-world' networks', *Nature*, **393**, 440-42.

## Figures and Tables

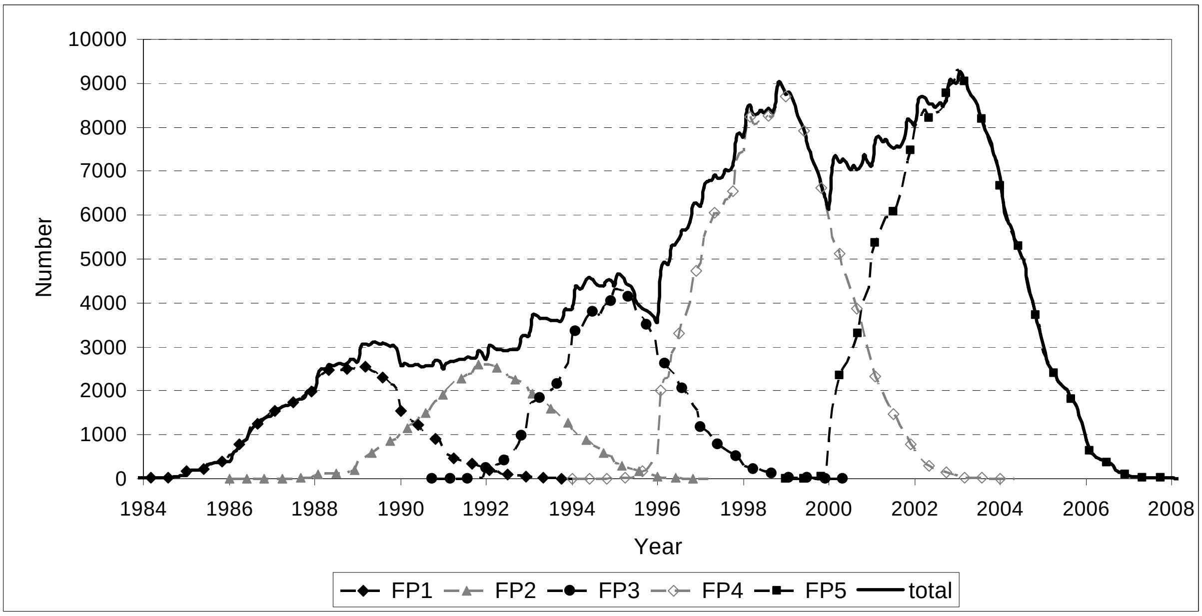
*Figure 1: Budget FP1–FP6: Evolution and share of thematic priorities*




Note: ICT ... information and communication technologies; industrial technologies include materials, aeronautics and space technologies; life sciences include biotechnology, genomics, biomedicine and food; environment includes transport; Other includes support for SMEs, dissemination, demonstration, co-operation with third countries, and ERA related measures.

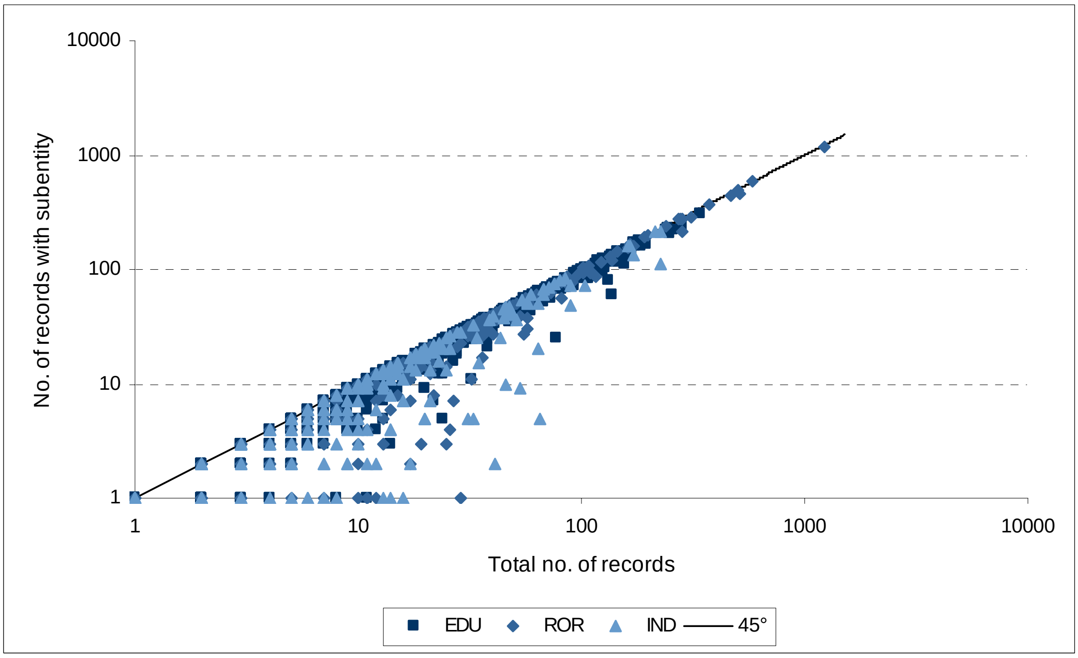
Source: adapted from CORDIS (2006b; 2006a); European Commission (2006); Barker and Cameron (2004, p. 172).

*Figure 2: Number of active R&D projects over time, FP1–FP5*



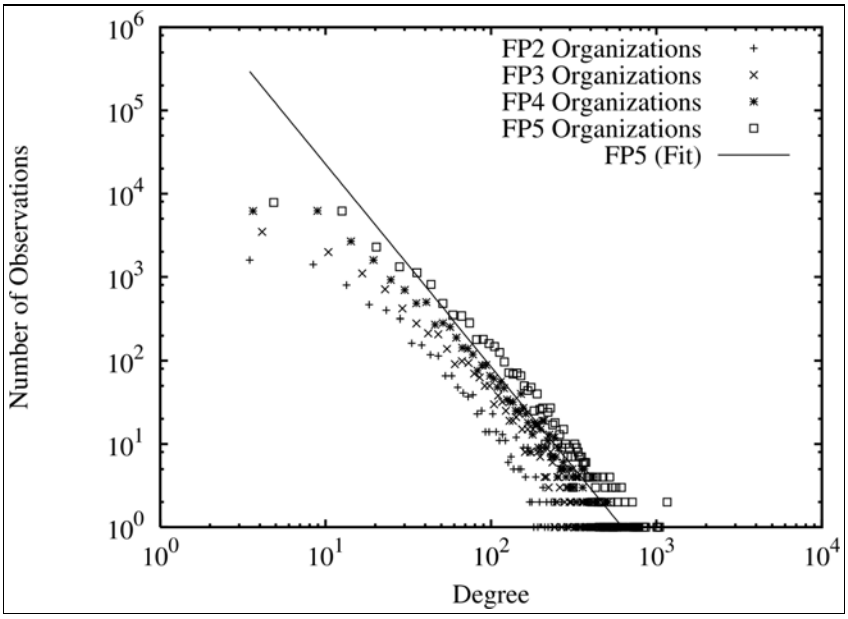

Note: Only indirect actions, i.e. research carried out by third parties and co-funded by the EU, are considered. In the data on FP4 and FP5, non R&D projects are excluded. These are preparatory, accompanying and support measures (ACM), exploratory awards (EAW), Access to Research Infrastructures (LFC), Research network contracts (NET), Research grants (individual fellowships) (RGI), Bursaries, grants, fellowships (BUR), and Research Infrastructure-Transnational access (TA).

*Figure 3: Total number of records versus number of records with subentity (FP5)*



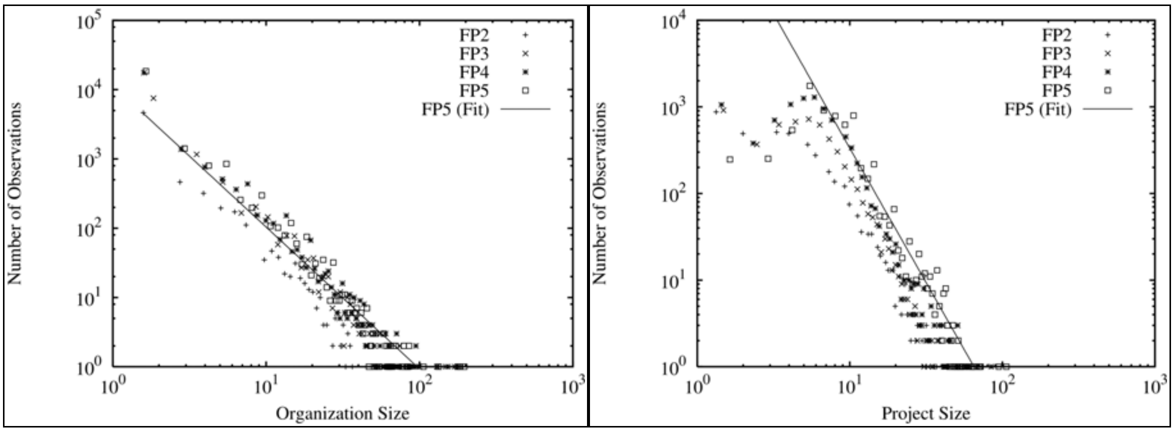

Note: Axes scale logarithmically. Line shows where all records would contain subentity information.

*Figure 4: Degree distribution of FP2-FP5 R&D collaboration network*



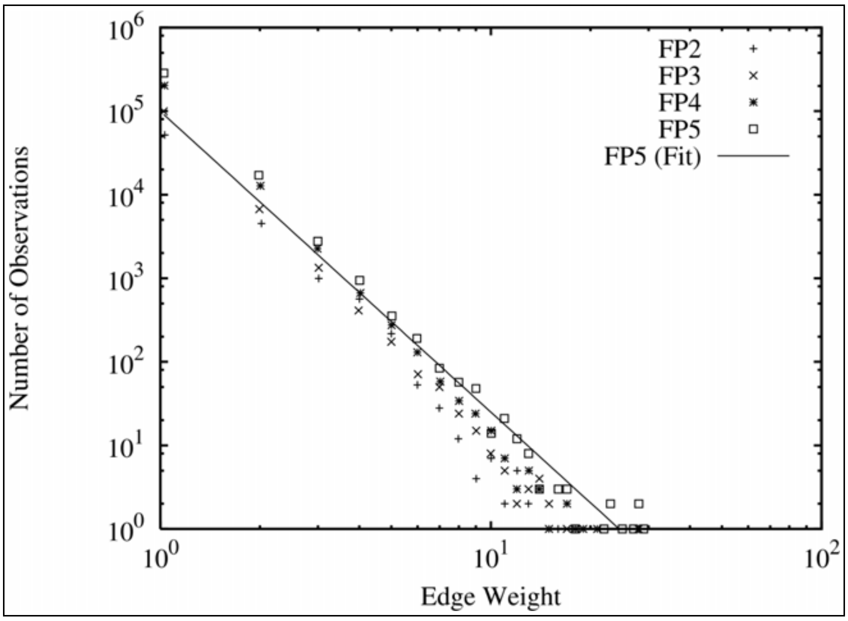

Note: Exponent of the fit line is -2.4. Fit lines for earlier FPs are similar (not shown).

*Figure 5: Distribution of vertex sizes in the FP2-FP5 organisation and project projections*



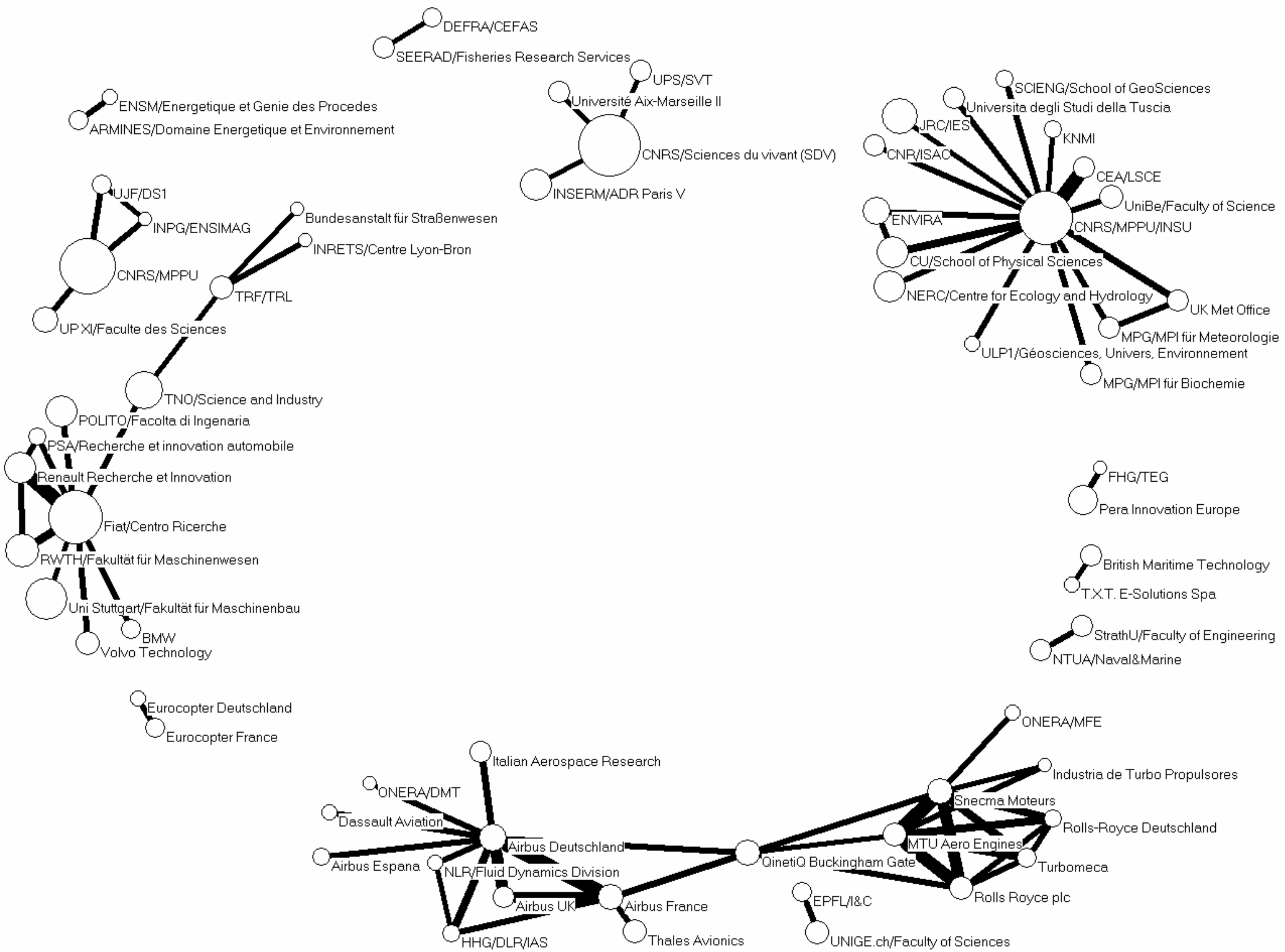

Note: Exponent of the fit line in the organisation projection is 2.05. Exponent of the fit line in the project projection is 3.1. Fit lines for earlier FPs are similar (not shown).

*Figure 6: Edge weight distribution of FP2-FP5 R&D collaboration networks*



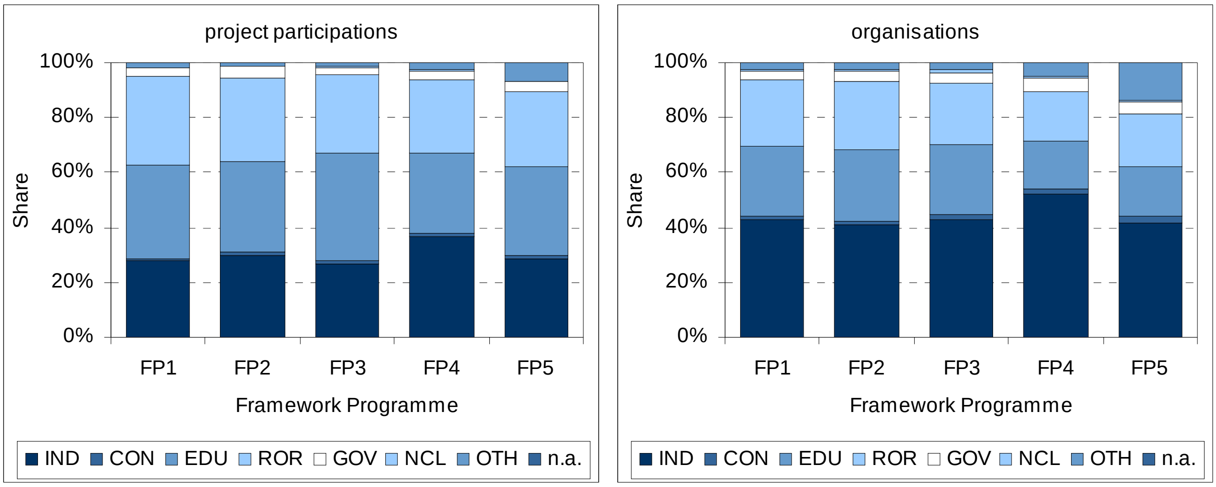

Note: Exponent of the fit line in the organisation projection is = 3.6. Fit lines for earlier FPs are similar (not shown). Observations are binned logarithmically.

*Figure 7: Core actors in the FP5 R&D collaboration networks*



Note: Figure created with PAJEK (Batagelj and Mrvar 2003)
CEA/LSCE … Commissariat à l'Energie Atomique/Laboratoire des Sciences du Climat et l'Environnement, CNR/ISAC … Consiglio Nazionale delle Ricerche/Institute of atmospheric sciences and climate, CNRS … Centre National de la Recherche Scientifique, CNRS/INSU … CNRS/ Sciences de l'Univers, CNRS/MPPU … CNRS/Mathematiques, physique, planete et univers, CNRS/SDV … CNRS/Sciences du vivant, CU … Cambridge University, DEFRA/CEFAS … Department for Environment Food and Rural Affairs/Centre for Environment, Fisheries and Aquaculture Science, ENSM … Ecole nationale supérieure des Mines, ENVIRA … Norwegian Institute for Air Research, EPFL/I&C … Ecole Polytechnique Federale de Lausanne/School of Computer and Communication Sciences, HHG/DLR/IAS … Helmholtz Gemeinschaft/Deutsches Zentrum für Luft- und Raumfahrt/Institut für Aerodynamik und Strömungstechnik, FHG/TEG … Fraunhofer-Gesellschaft/Technologie Entwicklungsgruppe, INPG/ENSIMAG … Institut National Polytechnique de Grenoble/École Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble, INRETS … Institut National de Recherche sur les Transports et leur Securite, INSERM … Institut National de la Sante et de la Recherche Medicale, JRC/IES … Joint Research Centre/Institute for Environment and Sustainability, KNMI … Royal Netherlands Meteorological Institute, MPG … Max-Planck-Gesellschaft, NERC … Natural Environment Research Council, NLR … National Aerospace Laboratory, NTUA/Naval&Marine … National Technical University of Athens/Faculty of Naval Architecture and Marine Engineering, ONERA … Office National d'Etudes et de Recherches Aerospatiales, ONERA/DMT … ONERA/Direction des Grands Moyens Techniques, ONERA/MFE … ONERA/Branche Mecanique des Fluides et Energetique, POLITO … Politecnico di Torino, PSA … Peugeot Société Anonyme, RWTH … Aachen University of Technology, SCIENG … University of Edinburgh/College of Science & Engineering, SEERAD … Scottish Executive Environment and Rural Affairs Department, StrathU … University of Strathclyde, TNO … Netherlands Organisation for Applied Scientific Research, TRF/TRL … Transport Research Foundation/Transport Research Laboratory, UJF/DS1 … Universite Joseph Fourier - Grenoble 1/Direction Scientifique Mathematiques et Informatique, ULP1 … Université Louis Pasteur 1, UniBe … University of Berne, UNIGE.ch … University of Geneva, UP XI … Universite de Paris XI, UPS/SVT … Universite Paul Sabatier de Toulouse III/Sciences de la Vie et de la Terre

*Figure 8: Distribution of organisation types in the organisation networks, FP1–FP5*




Note: IND … industry, CON …consulting, EDU … universities and other higher education, ROR … non-university research, GOV … governmental, NCL … non-commercial (associations, NGOs, etc.), OTH … other, n.a. ... not available.

*Figure 9a: Remaining and exiting actors, breakdown by organisation type, FP1–FP4*

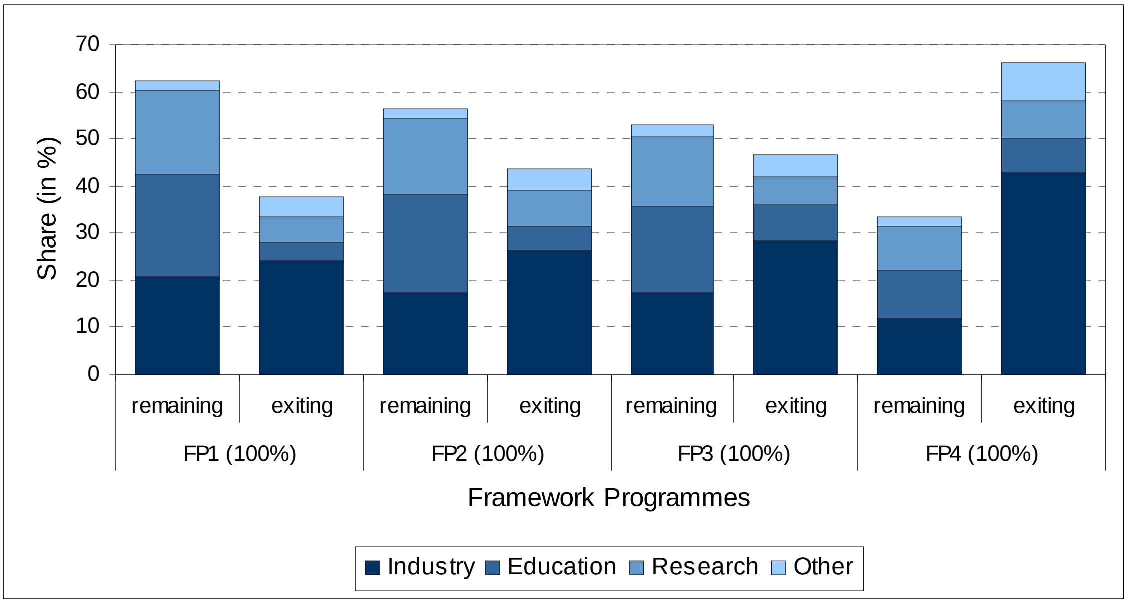


*Table 1: sysres EUPRO database standardisation table participants FP1-FP5 – Illustrative evidence on the quality of raw data on organisation names and organisation types*

| Organisations | No. of unique records | Orgname not evident from raw data (Share in %) | Orgtype | No. of Orgtypes in raw data* | Orgtype incorrect or N/A (%) |
|---|---|---|---|---|---|
| Catholic University Leuven | 656 | 3 | EDU | 3 | 23 |
| National Technical University of Athens | 618 | 5 | EDU | 3 | 28 |
| Imperial College London | 609 | 10 | EDU | 4 | 7 |
| Centre National de la Recherche Scientifique (CNRS) | 2745 | 3 | ROR | 5 | 31 |
| Consejo Superior de Investigaciones Cientificas (CSIC) | 1095 | 4 | ROR | 3 | 23 |
| Fraunhofer-Gesellschaft | 1092 | 18 | ROR | 6 | 4 |
| EADS | 633 | 69 | IND | 3 | 45 |
| Siemens AG | 618 | 1 | IND | 5 | 43 |
| Thales Group | 602 | 8 | IND | 4 | 44 |
| Philips NV | 581 | 15 | IND | 3 | 70 |
| Alcatel C.F. | 452 | 33 | IND | 3 | 71 |

Note: * excluding not available (N/A).

*Table 2: sysres EUPRO database standardisation table participants FP1-FP5 – Illustrative evidence on subentitiess*

| Organisations | Orgtype | No info on department (%) | No subentity (%) | Number of subentities |
|---|---|---|---|---|
| Catholic University of Leuven (KUL) | EDU | 20 | 12 | 11 |
| National Technical University of Athens | EDU | 26 | 22 | 12 |
| Imperial College London | EDU | 15 | 14 | 6 |
| Centre National de la Recherche Scientifique (CNRS) | ROR | 7 | 3 | 8 |
| Consejo Superior de Investigaciones Cientificas (CSIC) | ROR | 14 | 6 | 71 |
| Fraunhofer-Gesellschaft (FhG) | ROR | 25 | 10 | 54 |
| EADS | IND | 47 | 48 | 18 |
| Siemens AG | IND | 45 | 40 | 53 |
| Thales Group | IND | 60 | 14 | 60 |
| Philips NV | IND | 41 | 1 | 68 |
| Alcatel C.F. | IND | 55 | 0 | 62 |

*Table 3: sysres EUPRO database – numbers of projects and organisations described*

| Framework Programme (FP) | Period | Projects | Projects with multiple partners | Organisations | Subentities |
|---|---|---|---|---|---|
| FP1 | 1984–1987 | 3,283 | 1,696 | 1,981 | 2,583 |
| FP2 | 1987–1991 | 3,885 | 3,013 | 4,572 | 6,300 |
| FP3 | 1990–1994 | 5,529 | 4,611 | 7,324 | 10,025 |
| FP4 | 1994–1998 | 15,061 | 11,374 | 19,755 | 24,156 |
| FP5 | 1998–2002 | 15,559 | 10,674 | 22,303 | 27,382 |
| Total | | 43,317 | 31,345 | 42,020 | 49,855 |

Note: EURATOM projects are not listed. Recipients of research grants are not counted as organisations or subentities.

*Table 4: Basic network properties of FP2–FP5 R&D collaboration networks*

| **Graph characteristic** | **FP2** | **FP3** | **FP4** | **FP5** |
|---|---|---|---|---|
| No. of vertices N | 6,364 | 10,125 | 22,117 | 23,198 |
| No. of edges M | 58,463 | 108,637 | 218,651 | 306,695 |
| No. of components | 228 | 465 | 491 | 123 |
| N for largest component | 6,018 | 9,397 | 21,263 | 22,786 |
| Share of total (%) | 94.6 | 92.8 | 96.1 | 98.2 |
| M for largest component | 58,236 | 108,150 | 217,646 | 305,874 |
| Share of total (%) | 99.6 | 99.6 | 99.5 | 99.7 |
| N for 2nd largest component | 11 | 9 | 23 | 12 |
| Clustering coefficient | 0.711 | 0.716 | 0.790 | 0.806 |
| Diameter of largest component | 9 | 8 | 10 | 12 |
| $\ell$ largest component | 3.5 | 3.4 | 3.6 | 3.4 |
| Mean degree | 18.4 | 21.5 | 19.8 | 26.4 |
| Fraction of N above the mean (%) | 28.9 | 25.2 | 24.0 | 25.0 |

Note: $\ell$ … characteristic path length

*Table 5: Central organisations, Top 10, FP1–FP5*

| Rank | FP5 Organisation | Org type | *d* | # | FP4 Organisation | Org type | *d* | # | FP3 Organisation | Org type | *d* | # | FP2 Organisation | Org type | *d* | # |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | CNRS/MPPU | ROR | 1219 | 195 | SotonU/Engineering, Science and Mathematics | EDU | 736 | 94 | CNRS/MPPU | ROR | 717 | 210 | Bull Europe | IND | 372 | 58 |
| 2 | Fiat/Centro Richerche | IND | 1187 | 177 | Fiat/Centro Richerche | IND | 730 | 132 | CNRS/Sciences du vivant | EDU | 603 | 153 | CNRS/Chimie | ROR | 350 | 68 |
| 3 | CNRS/Sciences du vivant | ROR | 1113 | 245 | CNRS/Sciences du vivant | ROR | 707 | 203 | SotonU/Engineering, Science and Mathematics | EUD | 516 | 72 | CNRS/MPPU | ROR | 339 | 102 |
| 4 | AUTH/Faculty of Technology | EDU | 1020 | 86 | KUL/Faculty of Engineering | EDU | 696 | 93 | CNRS/Sciences de l'univers | ROR | 514 | 153 | CNRS/Sciences du vivant | ROR | 319 | 94 |
| 5 | Univ. Stuttgart/Faculty of Engineering | EDU | 859 | 106 | TNO/Science and Industry | ROR | 679 | 70 | UP XI/Faculte des Science | EDU | 503 | 84 | INESC ID Lisboa | ROR | 315 | 45 |
| 6 | ImperialCL/Faculty of Engineering | EDU | 818 | 80 | ImperialCL/Faculty of Engineering | EDU | 653 | 96 | ImperialCL/Faculty of Engineering | EDU | 479 | 84 | Siemens ZFE | IND | 255 | 34 |
| 7 | TNO/Science and Industry | ROR | 790 | 86 | LU/Institute of Technology | EDU | 644 | 67 | CNRS/Chimie | ROR | 448 | 112 | National Institute of Public Health and Environment (RIVM) | ROR | 241 | 37 |
| 8 | VTT/Industrial Systems | ROR | 771 | 61 | Univ. Stuttgart/Faculty of Engineering | EDU | 626 | 95 | OU/Mathematical and Physical Sciences Division | EDU | 425 | 72 | RIKILT – Institute for food safety | ROR | 241 | 31 |
| 9 | CNRS/INSU | ROR | 770 | 180 | CNRS/INSU | ROR | 618 | 191 | Daimler Benz AG | IND | 416 | 59 | ImperialCL/Faculty of Engineering | EDU | 238 | 50 |
| 10 | JRC/IES | ROR | 709 | 85 | VTT/Industrial Systems | ROR | 584 | 72 | CNRS/ST2I | ROR | 407 | 77 | CNRS/INSU | ROR | 230 | 75 |

Note: AUTH … Aristoteles University of Thessaloniki, CNRS … Centre National de la Recherche Scientifique, CNRS/INSU … CNRS/Institut National des Sciences de l'Univers, CNRS/ Sciences de l'Univers, CNRS/MPPU … CNRS/Mathematiques, physique, planete et univers, CNRS/ST2I … CNRS/ Sciences et technologies de l'information et de l'ingénierie, LU … Lund University, INESC … Instituto de Engenharia de Sistemas e Computadores, ImperialCL … Imperial College London, JRC/IES … Joint Research Centre/Institute for Environment and Sustainability, KUL … Katholieke Universiteit Leuven, OU … Oxford University, Siemens ZFE … Siemens Zentrale Forschung und Entwicklung, SotonU … Southampton University, UP XI … Université Paris-Sud XI, VTT … Technical Research Centre of Finland.

**The UNU-MERIT WORKING Paper Series**

2007-01 *Developing science, technology and innovation indicators: what we can learn from the past* by Christopher Freeman & Luc Soete

2007-02 *The impact of innovation activities on productivity and firm growth: evidence from Brazil* by Micheline Goedhuys

2007-03 *Estimations of US debt dynamics: Growth cum debt and the savings glut in Kouri's model* by Thomas Ziesemer

2007-04 *States and Firms on the Periphery: The Challenges of a Globalizing World* by Gabriel R.G. Benito & Rajneesh Narula

2007-05 *How Do Consumers Make Choices? A Summary of Evidence from Marketing and Psychology* by Zakaria Babutsidze

2007-06 *Inter-firm Technology Transfer: Partnership-embedded Licensing or Standard Licensing Agreements?* by John Hagedoorn, Stefanie Lorenz-Orlean & Hans Kranenburg

2007-07 *The Behavior of the Maximum Likelihood Estimator of Dynamic Panel Data Sample Selection Models* by Wladimir Raymond, Pierre Mohnen, Franz Palm & Sybrand Schim van der Loeff

2007-08 *Location and R&D alliances in the European ICT industry* by Rajneesh Narula & Grazia D. Santangelo

2007-09 *How do social capital and government support affect innovation and growth? Evidence from the EU regional support programmes* by Semih Akcomak & Bas ter Weel

2007-10 *The Micro-Dynamics of Catch Up in Indonesian Paper Manufacturing: An International Comparison of Plant-Level Performance* by Michiel van Dijk & Adam Szirmai

2007-11 *Financial Constraint and R&D Investment: Evidence from CIS* by Amaresh K Tiwari, Pierre Mohnen, Franz C. Palm & Sybrand Schim van der Loeff

2007-12 *The Spatial Hierarchy of Technological Change and Economic Development in*

*Europe* by Bart Verspagen

2007-13 *The origins and implications of using innovation systems perspectives in the design and implementation of agricultural research projects: Some personal observations* by Andy Hall

2007-14 *Technology supply chain or innovation capacity?: Contrasting experiences of promoting mall scale irrigation technology in South Asia* by Andy Hall, Norman Clark and Guru Naik

2007-15 *Are firms that received R&D subsidies more innovative?* by Charles Bérubé & Pierre Mohnen

2007-16 *Foreign direct investment and firm level productivity. A panel data analysis* by Geoffrey Gachino

2007-17 *Technological spillovers from multinational presence towards a conceptual framework* by Geoffrey Gachino

2007-18 *Technological capability building through networking strategies within high tech industries* by Wim Vanhaverbeke, Bonnie Beerkens and Geert Duysters

2007-19 *External technology sourcing: the effect of uncertainty on governance mode choice* by Vareska van de Vrande, Wim Vanhaverbeke & Geert Duysters

2007-20 *Exploration and Exploitation in Technology-based Alliance Networks* by Wim Vanhaverbeke, Victor Gilsing, Bonnie Beerkens, Geert Duysters

2007-21 *ICT Externalities: Evidence from cross country data* by Huub Meijers

2007-22 *Knowledge Flows, Patent Citations and the Impact of Science on Technology* by Önder Nomaler & Bart Verspagen

2007-23 *R&D offshoring and technology learning in emerging economies: Firm-level evidence from the ICT industry* by Zhe Qu, Can Huang, Mingqian Zhang & Yanyun Zhao

2007-24 *The Environmental Porter Hypothesis: Theory, Evidence and a Model of Timing of Adoption* by Ben Kriechel & Thomas Ziesemer

2007-25 *Measuring the Effectiveness of R&D tax credits in the Netherlands* by Boris Lokshin & Pierre Mohnen

2007-26 *The productivity effects of internal and external R&D: Evidence from adynamic panel data model* by Boris Lokshin, René Belderbos & Martin Carree

2007-27 *National System of Innovations and the role of demand. A cross country comparison* by M. Abraham Garcia-Torres

2007-28 *The Global Challenges of the Knowledge Economy: China and the EU by* Can Huang and Luc Soete

2007-29 *Redefining Foreign Direct Investment Policy: A Two Dimensional Framework* by Sergey Filippov & Ionara Costa

2007-30 *Redefine the Nexus between Foreign Direct Investment, Industrial and Innovation Policies* by Ionara Costa & Sergey Filippov

2007-31 *Innovation and Competitive Capacity in Bangladesh's Pharmaceutical Sector* by Padmashree Gehl Sampath

2007-32 *R&D collaboration networks in the European Framework Programmes: Data processing, network construction and selected results* by Thomas Roediger-Schluga & Michael J. Barber