

## **Bibliometric Network Analysis (BibTechMon) of the 4th Framework Programme of the EU based on CORDIS data**

(presentation prepared for the workshop of the OECD-NIS Focus Group on  
Innovative Firm and Innovative Firm Networks, Vienna, 18-19 November 1999)

*Alexander Kopcsa (Arcs), Michaela C. Topolnik (Proviso / BMWV), Andreas Schibany (JR)*

### ***Presentation***

- Introduction
- Methodological Approach: BibTechMon
- Description of Data used for the Analysis
- Exemplary Findings
- Issues linked to the Matching and Harmonization of CORDIS Data

### ***Background***

Looking at EU activities in the field of technology and research policy, the Framework Programmes for RT&D are a central instrument to foster and to strengthen RT&D at an European level. The Fourth Framework Programme was the first one in which all 15 EU-members participated as well as a great variety of different scientific and technological areas were covered by thematic and horizontal programmes.

Previous detailed analysis has revealed that the participation in the EU Framework Programmes for RTD does not reproduce existing national RTD profiles. On the contrary, the structure of the Framework Programmes determines national participation whereas the structure and pattern of national R&D systems do not. Hence, each country has a group of firms, institutes etc. which are able to co-operate internationally in RTD, irrespective of the underlying industry and research structure. RTD at European level is thus carried out by transnational research joint ventures. These joint ventures are made up of business firms, research centers and universities and engage in pre-competitive research. The emergence of specific patterns of co-operation can therefore be addressed at several levels of industrial and spatial aggregation.

One central feature of the FPs is collaborative research. In many programmes a central requirement for (successfully) submitting a funding proposal is a well structured consortium consisting of partners from different countries as well as from different (organisational) sectors. Therefore the focus of our analyses of the participation in the EU R&TD Framework Programmes are cooperations and collaborative networks within FP4 involving the industrial, research and education sectors, with special emphasis on identifying specific characteristics of co-operation and patterns of knowledge transfer. This leads to the conclusion that the Framework Programmes as the core of the European research and technology policy have a strong governance effect on the national industry and research structure with new and emerging routes of innovation-related knowledge dissemination in European industry.

## ***Database***

The data used in the analysis were extracted from the CORDIS database through an online query on all the RTD projects financed by the EU within FP4 (1995- 1998). At that time (December 1998) more than 10.000 projects with over 47.000 participations across all the thematic and horizontal programmes of FP4 were listed in the Cordis database. One has to bear in mind that CORDIS only contains projects that are in execution or that are already completed. The projects are filed when contracts are signed and when CORDIS receives the data from the Commission. In selected programmes there is a considerable delay and the data are not always up to date. Therefore most of the projects in the network are from 1996 and 1997, fewer are from the years 1998. A recent query showed that the CORDIS database only had about 10% more projects.

Frequently the name of one and the same partner is spelled differently in the CORDIS database. In order to be able to apply our methodology for the network analysis (BibTechMon) all the data had to be harmonized, i.e. each partner had to be identified by one acronym, organisational type and country. Due to this procedure we reduced the 47.000 participations to 13.000 partners. For the network analysis we selected those organisations that participated in three or more projects (it seems that about three quarters of the partners participated in only one project and probably can not be considered being part of a network).

By applying bibliometrical methods and co-word analysis combined with several statistical methods on data from the CORDIS database, co-operative relationships occurring in the RTD projects executed within the European Union's Framework Programmes were mapped, encompassing the 13.000 partners with a rich set of variables on each: The size of firms, the organisational structure of partners, the geographical distribution of partners and consortia as well as to organisational types grouped in consortia (e.g. industry with university). Finally, the calculated network consisted of 2.700 partners who participated 30.000 times in 8.800 projects; altogether the network showed about 53.000 links. Partners from the 15 EU member states across all sectors (industry, universities, research institutes etc.) were included.

The following section briefly describes the methodology and software we used to draw that network displaying RT&D cooperations in the FP4.

## ***Methodological Approach: BibTechMon***

For the analysis a complex software is applied, based on mathematical and mechanical models including statistics as well as co-occurrence analyses, etc. Derived from the idea of co-word analysis used in bibliometric studies the methodology looks at the co-occurrence of items, i.e. partners, and uses an iteration model for the calculation of co-word (co-item) maps. The iteration model is used for positioning words (i.e. partners) in a two-dimensional plane on the basis of their connections to each other. This procedure enables us to draw the partners in the context of all of their co-partners due to their co-occurrence in the projects. Both the size and the position of a specific partner in the network provides certain information, as will be described below.

## ***Frequency of Partners***

For each partner the frequency of appearance was counted, i.e. the number of projects in which the partner participated in (see. Table 1). This frequency is shown by the size of the point representing the partner in the network.

**Table 1: Frequency of partners, i.e. number of projects per partner (sample of the 20 most frequent)**

<b>Name</b>	<b>Frequency</b>
ku leuven.edu.be	313
csic.ror.es	266
uni cam.edu.gb	226
ntua.edu.gr	207
icstm.edu.gb	184
vtt.ror.fi	181

cnr.ror.it	179
novtno.ncl.nl	168
uni lund.edu.se	158
max_planck.ror.de	157
fu brux.edu.be	150
siemens.ind.de	131
uni_saloniki.edu.gr	129
uni_utrecht.edu.nl	129
tu_delft.edu.nl	128
uni_stuttgart.edu.de	128
cea.ncl.fr	125
luni_wageningen.edu.nl	124
uni london c.edu.gb	123
uni oxford.edu.gb	120

### Co-word analysis

The next step is to group keywords, i.e. partners, in a two-dimensional diagram according to their co-occurrence in articles (projects). Therefore, this co-occurrence matrix contains the co-occurrence of all pairs of partners (Fig. 1).

Fig. 1: Table of co-occurrence

		ku leuven.edu.be	csic.ror.es	uni cam.edu.gb	ntua.edu.gr	icstn.edu.gb	vtt.ror.fi	cnr.ror.it	novtno.ncl.nl	uni lund.edu.se	max_planck.ror.de	fu brux.edu.be	siemens.ind.de	uni_saloniki.edu.gr	uni_utrecht.edu.nl	tu_delft.edu.nl	uni_stuttgart.edu.de	cea.ncl.fr	luni_wageningen.edu.nl	uni london c.edu.gb	
1																					
2	ku leuven.edu.be	313	15	5	9	9	11	10	12	7	8	10	4	3	8	7	6	7	7	7	
3	csic.ror.es	15	288	5	5	4	7	19	8	14	12	8	1	8	9	2	6	2	10	4	
4	uni cam.edu.gb	5	5	228	1	2	2	9	1	2	7	7	1	5	2	4	5	7	1	3	
5	ntua.edu.gr	9	5	1	207	14	13	5	5	4	0	5	5	4	0	5	14	4	2	8	
6	icstn.edu.gb	8	4	2	14	184	10	4	7	3	8	0	6	3	1	11	6	4	4	4	
7	vtt.ror.fi	11	7	2	13	10	191	3	23	8	1	0	2	3	1	5	8	3	8	4	
8	cnr.ror.it	10	19	9	5	4	3	179	1	3	4	3	2	10	4	1	3	5	7	3	
9	novtno.ncl.nl	12	6	1	5	7	23	1	168	2	1	4	3	2	1	9	5	4	5	0	
10	uni lund.edu.se	7	14	2	4	3	6	3	2	158	3	1	3	4	7	9	7	4	13	3	
11	max_planck.ror.de	8	12	7	0	8	1	4	1	3	157	4	1	0	5	1	1	6	3	4	
12	fu brux.edu.be	10	8	7	5	0	0	3	4	1	4	150	1	3	15	3	1	1	1	1	
13	siemens.ind.de	4	1	1	5	6	2	2	3	3	1	1	131	3	1	6	2	2	0	0	
14	uni_saloniki.edu.gr	3	6	5	4	3	3	10	2	4	0	3	3	129	2	6	4	3	4	2	
15	uni_utrecht.edu.nl	8	9	2	0	1	1	4	1	7	5	15	1	2	129	7	2	2	1	3	
16	tu_delft.edu.nl	7	2	4	5	11	5	1	9	9	1	3	6	8	7	128	7	2	0	5	
17	uni_stuttgart.edu.de	6	6	5	14	6	6	3	5	7	1	1	2	4	2	7	128	2	0	6	
18	cea.ncl.fr	7	2	7	4	4	3	5	4	4	6	1	2	3	2	2	2	125	1	1	
19	luni_wageningen.edu	7	10	1	2	4	6	7	5	13	3	1	0	4	1	0	0	1	124	4	
20	uni london c.edu.gb	7	4	3	8	4	4	3	0	3	4	1	0	2	3	5	6	1	4	123	

### ***Jaccard index***

The Jaccard Index, used to normalize the elements of the matrix containing the co-occurrence of each pair of partners, also gives information on the “intensity” of cooperation (co-occurrence):

$C_{ij}$  co-occurrence of partners  $i$  and  $j$

$C_{ii}$  frequency of partner  $i$

$$J_{ij} = \frac{C_{ij}}{C_{ii} + C_{jj} - C_{ij}}$$

### ***Iteration Model***

Finally an iteration model based on a mass point mechanics was applied. Each word is represented by a mass point and the intensity of relation is taken as elasticity between the mass points. A  $n$  dimensional system of differential equations is formulated and solved numerically. (cf. Kopcsa / Schiebel 1998a) The iteration model and a graphical interface allowed to visualise the network structure of the co-occurrence of words and the representation of separate groups of words.

### ***Themes***

Further information can be added to the network by creating “themes”. Information such as type of organisation, country, year of (first) participation, number of participation in each of the specific programmes was attributed to the partner file and included in network. Themes are highlighted by different colours and thus also support the analytical work by allowing to look at different aspects and features of the network in a very flexible way. The number of themes that can be created and added to the set of data is not limited.

Additional information<sup>1</sup> that has not been added yet for international partners includes e.g. industrial or research sector (NACE Codes), turnover, patents, regional affiliation (NUTS) etc. linked to the organisations

The analysis as well as the graphical display of networks of partners participating in European Framework Programmes builds up an information base to answer questions such as:

- Who are the key players in the RTD scene? (those who are most frequent or in the centre of the consortia or networks, “old boys” vs. newcomer).
- What are the developments over time in these relationships? (How and when do firms get involved? Do consortia and/or key players emerge and vanish? Are consortia changing or are they more or less stable over time?)
- How fast do actors from new member-states and associated countries get involved in EU RTD-relationships?
- Are there patterns of cooperations? (interaction between nations and / or with regard to sectors and organisational types)
- Are there specializations of countries occurring at the European level? (national specialisation vs. specialization at the European level)
- What are the regional dimensions of consortia (e.g. are there several countries or partners preferably grouping together)?

---

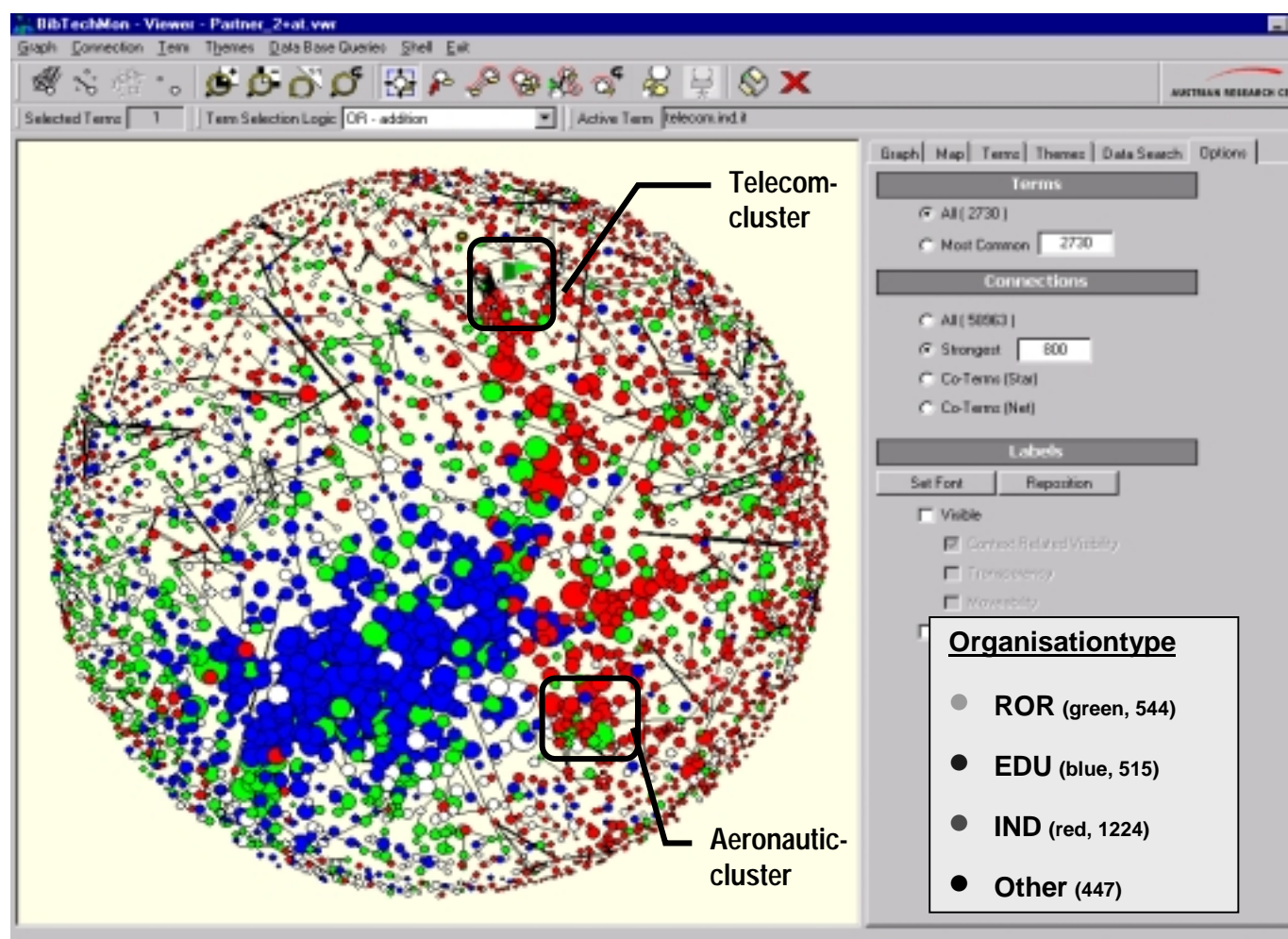
<sup>1</sup> Some - information is already available for many of the Austrian partners participating in FP4.

## Exemplary Findings

To demonstrate the potentialities of the tool the presentation will select some exemplary findings, i.e.

- time-dimension: the development of the network of partners over the years starting with 1995 up to 1998
- sectoral and cluster specialities: the identification of sectoral clusters (telecom, aeronautics); (see Fig. 2)
- features concerning the type of organisation: collaboration between and the distribution of different types of organisations in the network (see Fig. 2)
- sectoral concentration / distribution of partners in FP4 (as well as according to the new programmes of FP5)

Fig. 2: FP4 - Network of partners (2.730) by organisational type



(Source: CORDIS Dec. 1998; BibTechMon)

## Structure of Database

Presently the analysis is based on two different, yet connectable, sets of data (Error! Reference source not found.):

- two tables containing the harmonised CORDIS data with project and partner information
- table of Austrian partners with additional information



CORDIS projects: CORDIS record number, project title, programme acronym, start and end date, funding (only in a few cases) , name and country of prime contractor etc. (for details see **Error! Reference source not found.**)

CORDIS partners: CORDIS record number (same as for the projects), name of partner, type, size (only in a few cases), country, region of organisation (for details see **Error! Reference source not found.**)

Austrian partners: numerator, name of partner, type of organisation, acronym, field of research (universities), industrial sector (NACE ), address, several fields (numerical and text) in order to link to other databases etc. (for details see **Error! Reference source not found.**)

With respect to the Austrian participation and participants a data-base has been built up consisting of organisations and firms which are involved (both those submitting a RT&D proposal and actually participating in a funded project) in the Fourth and Fifth EU RTD Programmes, complementing with additional national data - is in progress.

### ***Issues linked to the Matching of CORDIS Data***

To pursue any kind of network analysis it could be of utmost importance to have a coherent set of data available to ensure reliable results. The permanent problem of incoherent data with respect to the citation of the partners' names was tackled by standardizing the Austrian as well as the international partners. When harmonizing the international partners for the network analysis presented, in many cases the identification of the partners was possible, yet still there might be some indifferences especially concerning the allocation of the partners to the right category of organisational type. To get an even better and more accurate picture of European RT&D networks it would be useful to harmonize the data at a national level to identify the partners correctly and to add the appropriate type of organization as well as a sectoral categorisation (NACE) to the specific partner.

Several other issues still need to be discussed in respect to a match of CORDIS data:

- Source: which source will be used? Updates? (online, CD ?)
- The same organisation / institution is listed under different names, yet each single partner needs to be identified by its name and acronym (same spelling, same organisational typ, same country)
- CORDIS: the same organisation is classified with different organisational types (in accordance with CORDIS classification: industry, education, other, research, non commercial); definitely more accurate if adjusted at national level, as there is no there is no differentiation between SME and large enterprises (Large) in the classification of industry
- To analyse the co-occurrence of partners as well as to allow future updates and comparisons each partner must be conjuncted with its projects (proposal as well as partner key).
- Both the acronym and a long version of the partner's name should be in the database to know who the partner is and to facilitate future updates or comparisons.
- For a systematic standardisation a common syntax of acronyms would be helpful.
- Clarification: which information is already available in other databases, which information needs to be added (missing information)
- To be able to gather additional information the possibility of conjunction to other databases should be possible ( size, region, industrial sectors / NACE-Codes / research sectors)
- How to identify company or institute located at several different places (in one country) → one organisation or differentiate between different locations?

- similar with universities: take 'university' as a whole or differentiate of institutes or faculties (problem: this information is not always available)
- In some specific programmes projects are listed only with the co-ordinator: how to deal with these projects (eliminate?)

### ***Conclusions***

Nevertheless, bearing in mind that the CODIS data are erroneous in some respect, they still provide a good – if not the only - and broad range of information to look at research networks within the Framework Programmes. Combining the analysis at the European level with national data would allow to confront country specific patterns of participation in EU RTD programmes with data on specialisation and firm behaviour at the national level, and to deliver additional input on the effect of the EU innovation and technology policy both on European and national levels as well as on the behaviour of innovative firms and innovative firm networks.

### ***References***

- Kopcsa, A. / Schiebel, E.: Science and Technology Mapping: A new Iteration Model for Representing Multidimensional Relationships. in: Journal of the American Society For Information Science, Jan. 1998, No. 1 / Vol.49, p.7ff
- Kopcsa, A. / Schiebel, E.: Ein bibliometrisches F&TE-Monitoringsystem für Unternehmen, OEFZS-A—4544, Dez. 1998
- Polt et al: Österreichischer Technologiebericht 1999, TIP, Aug. 1999