The Use of Research Funding Databases for Research Assessment Information Systems

Daniel Bovelet[1]
German Research Foundation (DFG)

Summary

The information needs of researchers and policy makers on the performance of research institutions are seldom limited to information from one source, which usually covers an isolated perspective from a particular region or sector of science. This paper introduces the development of indicators suitable for the representation of research activities and achievements based on heterogeneous research funding databases. Thus it becomes possible to assess research institutions from different points of view. Furthermore, the uses and opportunities for the integration into research assessment information systems are exemplified. The explanations refer to the experiences of the DFG. In the scope of the DFG Funding Ranking and its objective of assessing research structures in the German higher education system, information from various research infrastructure and funding databases is consolidated and subsequently used for the derivation of research indicators and the implementation into public information systems.

1 Introduction

Systems of research assessment, such as ranking exercises, are internationally discussed as important strategic management tools for publicly funded research institutions. Furthermore, reports on research assessments are not simply a supplement to the information base, but, more than ever, are becoming determinants for research policy decision-making processes. The DFG has reported on the funding it awards to German universities and non-university research institutions on three occasions (in 1997, 2000 and 2003)[2] as an indicator of research activity and quality.

As the self-governing body of German science and research, the DFG fulfils its central responsibility of funding research in all branches of science and the humanities. With an annual budget of approximately €1.3 billion, the DFG is the central funding agency for basic research in Germany, funding more than 25,000 research projects per year. In this respect, the number and range of DFG-approved projects provide an important performance indicator, because funding is awarded only to those researchers who, with their projects, succeed in the ever-intensifying competition for the best ideas and, in some programmes, also the best infrastructure for research. The strong and

[1] The author would like to thank Juergen Guedler, Lisa Hoppe, Alexis-Michel Mugabushaka and Andreas Schmitz for helpful comments and suggestions.

[2] The previous reports (Deutsche Forschungsgemeinschaft 1997; Deutsche Forschungsgemeinschaft 2000), the current edition (Deutsche Forschungsgemeinschaft 2003) and other related material are available online at http://www.dfg.de/en/ranking.

very positive response to its *Funding Ranking* has encouraged the DFG to extend this form of presentation of its funding activities and establish it as a regular monitoring-style information service.

Nevertheless, the information needs of funding and advisory bodies, science managers, researchers, and other stakeholders, such as society in general, on research activities and achievements of research institutions are very seldom limited to information from one source or information system, which usually represents isolated research indicators or perspectives from a particular region or sector of science. In the context of assessing research structures and institutions, there is a strong need to integrate information from various sources. In the framework of the DFG Funding Ranking, the collaboration with different partner institutions has made it possible to significantly broaden the scope of the reports and information services in comparison to previous editions. On the basis of different research infrastructure and funding databases the current edition includes several types of indicators, revealing the main aspects of research in its analysis (e.g. internationalism and network building).

Furthermore, to meet the strong demand for information on research institutions in Germany, the DFG has developed a web-based information platform based on these databases, containing information about German research institutions. The aim of the information platform is to provide perspectives on the various *research institutions* and the *surrounding regions with their scientific infrastructure* as well as the *networks* between the different institutions. The information on research institutions can be presented in the form of data based reports, including information on their personnel and financial resources as well as their research activities and achievements in different research areas.

This paper reports on the experiences of the DFG in compiling an information service on research structures in the German higher education system, based on heterogeneous sources of information and databases. The paper is organized as follows: Section 2 provides a brief description of the characteristics of research funding databases and the opportunities for the derivation of general research indicators. Section 3 exemplifies the use of research funding databases in research assessment information systems. The explanations refer to the research funding databases, research indicators, and information systems that were developed and used to compile the DFG Funding Ranking. Section 4 summarizes the main points.

2 Development of research indicators based on research funding databases

Funding databases from research funding agencies offer numerous opportunities for the derivation of several types of indicators representing the research activities and achievements of research institutions. In the following, *general characteristics of research funding databases and the derivation of central indicators* for research institutions based on these databases will briefly be presented. The explanations focus on databases covering funding schemes from research funding agencies that allocate their funds on the basis of competition and peer review. Furthermore, *consolidation processes for heterogeneous databases* will be illustrated in order to be able to assess research institutions from various perspectives using the individual properties of different research funding databases. Standardization processes are necessary to refer to uniform classifications, such as in the case of different subject classification systems. In addition to data on the DFG re-

search funding processes, the research funding databases within the scope of the DFG Funding Ranking include information about publicly funded research by the Alexander von Humboldt Foundation (AvH) and the German Academic Exchange Service (DAAD) as well as personnel and financial data on research institutions provided by the German Federal Statistical Office and data on the European Community Framework Programmes collected and cleansed in cooperation with the European Liaison Office of the German Research Organisations (KoWi). The main motive for a consolidation of these databases is to be able to utilize a unique database for statistical analysis and for the development of information systems based on standardized research areas and institutional codes.

2.1 Derivation of indicators

Almost all research funding agencies worldwide base their decisions on proposals they receive on the judgement of (external) experts. Honorary peer reviewers assess the proposals to determine whether they meet the funding criteria. Therefore, production databases of research funding agencies have several characteristics in common. The databases covering research funding processes contain information on (i) *proposals* (personal or institutional data on applicants and the project specifications) and on (ii) *evaluation or peer review processes* (personal or institutional data on reviewers and project appraisals[3]). Furthermore, research funding databases are able to provide information on the respective research areas within the funding processes. These properties of funding databases can be used for the derivation of research indicators suitable for the representation of the research activities and achievements of research institutions differentiated by research areas. In the following, fundamental indicators are exemplified by analysis within the scope of the DFG Funding Ranking.

One of the most important indicators is the *number or volume of approvals* received by research institutions in relation to institutions of the same type or size. In the context of the Funding Ranking, DFG approvals are considered as an indicator for the research performance of an institution. This is justified because DFG funding is awarded only to those scientists and academics who, with their projects, succeed in the ever-intensifying competition for the best ideas and, in some programmes, also the best infrastructure for research and the training of young researchers[4]. Similarly, the *number of reviewers* consulted as experts on each research topic in the stages of decision making can be taken into consideration as an indicator for research activity and quality.

Another characteristic of research funding databases, which can be used for the derivation of research indicators, consists in the coverage of different *research funding programmes and instruments*. The areas of funding operations processed by research funding agencies vary in their

---

[3] This perspective could be extended to distinguish between ex-ante and ex-post evaluations of research projects. However, a consideration of ex-post evaluations of research projects, which could include research results and outputs such as publications or patents, as well as a detailed description of the characteristics of research infrastructure and funding databases and the derivation of indicators would go far beyond the scope of this paper. The focus of this section is on databases covering primarily ex-ante evaluation processes of research funding agencies. See Hornbostel (1997) for an introduction and further information on the development of research indicators as well as their categorizations (e.g. input, throughput and output indicators).

[4] In addition, the number of proposals submitted by a research institution in relation to size measurements of comparable institutions can serve as an indicator for the research activity of an institution.

funding instruments and objectives. Each research funding programme and instrument can be used as basis for the derivation of indicators for different purposes. For example, data on funding programmes for international research cooperation provide a basis for analysis of the *international perception and reputation of research institutions*, while information on research network funding programmes provide a basis for quantitative analysis of the network structure resulting from joint participation by various research institutions in these funding programmes.

In this respect, indicators covering the participation and integration of universities and non-university institutions in research network funding programmes help to identify *research institutions that play a central role in networks of cooperation*. Additionally, network analytical representations can be used to highlight research clusters, which are characterised by especially close cooperation, both on a regional and national level. Section 3.3 discusses possible network indicators, further analysis, and methods of interpretation in more detail and exemplifies a network analytical representation resulting from an analysis of DFG coordinated programmes.

Databases from research funding agencies with funding instruments for international cooperation between scientists and academics and the establishment of international networks can serve as basis for analysis of the internationality of research institutions. Internationality is an integral element of research since it does not stop at national boundaries. In the context of the DFG Funding Ranking, the figures on funding recipients provided by the AvH and the DAAD, as well as the data providing information on the participation of German universities in the Fifth EU Framework Programme, offer a good impression of the international perception and reputation of German research institutions for top-level international scientists and academics.

Alongside the investigation of absolute figures, another important element of ranking or rating studies, which are, after all, designed for benchmarking purposes, is the comparison using figures that put the size of an institution into perspective. For comparative purposes the respective results and values of the various indicators should be classified into ranking groups (e.g. by using standard cluster analysis). Furthermore, the relation to a specific benchmark (e.g. another research institution of the same type and size) should be taken into consideration in the interpretation of the respective values. However, the representativeness of the object of investigation and the comparability of the different units and levels of analysis, especially within the various research areas, are of importance for the derivation of reliable results.

In the scope of the Funding Ranking, scaling of this kind is enabled by the data from the German Federal Statistical Office, which provides data in the form of special reports, compiled and edited in collaboration with the state statistical offices as part of annual surveys carried out at universities and non-university research institutions. This data offers information on the financial capital and the human resources of those research institutions. Details on financial capital are given according to administrative incomes, third-party funding and regular core funds. Details on human resources relate to full-time equivalent scientific staff working at research institutions. Therefore, in connection with qualitative and quantitative information on the management and organization structure as well as the respective institutional environment of the research institution considered, the infrastructural data implemented in these databases can serve as a basis for detailed analysis. After all, for comparative purposes and for specific benchmark analysis a consolidation process for the various databases is necessary, which is illustrated in the following section.

2.2 Consolidation of heterogeneous databases

In order to be able to correlate data from heterogeneous sources, a complicated process has first of all to be used to compile a *concordance within the framework of the different subject classifications and between diverse data collections of research institutions*. The main motive in compiling these databases is to be able to utilize a unique database for statistical analysis and for the development of information systems based on standardized research areas and institutional codes. In addition to the figures provided by the Federal Statistical Office on the financial capital and the human resources of universities and non-university research institutions, information contained in databases of different research funding agencies are implemented within the scope of the Funding Ranking.

The standardization process for the different subject classifications should take place in close consultation with the officers or review committees responsible for the registration of the respective research subjects within the data collection process. As the rules for classification used by different organizations are not universally congruent, especially in the humanities and engineering sciences, some discrepancies can arise between similar statistical analyses by the different organizations. Therefore, this standardization process has to be carried out with great accuracy in order to guarantee adequate compatibility between the different data. In the case of the Funding Ranking the various subject classifications of implemented data sources are standardized and matched to the DFG subject classification. This structure is based on a four-tier subject classification system. It incorporates a total of 201 subjects, 48 review boards, 14 research areas and four scientific disciplines, organized hierarchically. The report makes use of the top two levels of classification, which are shown in figure 1[5].

However, the standardization does not guarantee that the information from the different sources necessarily applies to precisely the same derivation methodology, which has to be taken into consideration when interpreting the various indicators. Thus, a visiting scientist classified by the AvH as belonging to "Food Chemistry" may either have worked at a medical institute, or at an institute designated as belonging to chemistry; or a project funded by the DFG in "Process Engineering and Technical Chemistry" (research area "Thermal and Process Engineering") may have been initiated by a researcher working at an "institute for chemistry", just as a member of a university's administrative staff responsible for compiling staff statistics may allocate a researcher to the subject area "Chemical Didactics" (research area "Chemistry"), although he/she is actually teaching and conducting research at an educational science institute. Nevertheless, it can be assumed that problematic situations, such as those described above, are not the general rule and biases compensate for each other at aggregated levels of analysis. The risk of "apparent accuracy" does, however, increase with the level of detail of classification.

[5] See http://www.dfg.de/en/dfg_profile/structure/statutory_bodies/review_boards/index.html for further information on the DFG subject classification system.

Research Area	Scientific Discipline
Humanities Social and Behavioural Sciences	Humanities and Social Sciences
Biology Medicine Agriculture, Forestry, Horticulture and Veterinary Medicine	Life Sciences
Chemistry Physics Mathematics Geosciences	Natural Sciences
Mechanical and Industrial Engineering Thermal and Process Engineering Material Science and Engineering Electrical Engineering, Computer Science and System Engineering Construction Engineering and Architecture	Engineering Sciences

Figure 1: DFG research area classification system

The standardization process for the diverse data collections of research institutions is of great importance for comparative purposes and for specific benchmark analysis in the scope of the statistical analysis. Therefore, in 2002 the DFG began compiling a database of research institutes, which provides hierarchically structured address data for nearly 20,000 institutes from universities and non-university research institutions in Germany. In general, up to four hierarchy levels are taken into consideration (e.g. institution > department > institute > part of the institute), but in individual cases further sub-division is possible[6]. On this basis, the DFG also began compiling a concordance database to combine several variables, including the institutional codes used by the heterogeneous research funding databases, to refer to a standardized collection of data on the research institutions. This database covers two hierarchy levels for each of the research institutions and serves as a basis for the statistical analysis as well as reporting standard for statistical results.

The final database with standardized data collections can be used not only for statistical purposes but also as a basis for information systems that cover the various research indicators for the research institutions considered. Section 3 exemplifies three fundamental types of features that can be implemented into a research assessment information system, based on research funding and infrastructure databases.

---

[6] This database is integrated into ElektrA, the DFG's proposal process database, and also published in the internet as an information system for international target groups (www.dfg.de/idb). For further details see Guedler (2006).

3 Alternatives for the integration into information systems

In the context of research funding databases several kinds of information systems are imaginable to provide information on the research activities and achievements of research institutions. These information systems can serve not only as general providers of information for the broad public but also as tools for management systems or strategic planning instruments for the research institutions considered. Furthermore, selections for different units and levels of indicators can be incorporated into the information system. The standardization processes for heterogeneous databases offer an opportunity to integrate information from various sources. For instance, using a standardized database, such as one described above, it is possible to refer to the respective research areas and the institutions' names consistently for almost all of the reports, figures, tables, and search functions implemented in information systems. Details on the location and on the respective urban or rural district based on the standardized address data can be drawn on in order to generate maps using a geographic information system.

In the following, the use of three kinds of information systems that can be developed based on research funding databases will be illustrated. *Institutional reports* provide comprehensive information on institutions and their institutional environments, *geographical information systems* allow analysis of regional conditions for research, and *network analytic representations* offer insight into the structures of research networks and clusters. In combination with research indicators, such as those introduced in section 2.2, and other potential (qualitative) information on research activities and achievements of research institutions, these information systems could be implemented into a holistic research assessment information system.

3.1 Institutional reports

Reports on the individual universities and non-university research institutions enable comprehensive information on the parameter values of the research indicators considered to be collected and examined. Thus, several research indicators can be examined in an overall context for each research institution. For the representation and interpretation of the respective values for the different research indicators, the relation to a specific benchmark (e.g. by using ranking groups) should be taken into consideration within such a report. Among quantitative information on research institutions, this type of information system is additionally able to provide qualitative information on the institution and its infrastructure. Figure 2 illustrates an example for elements of a report on the University of Munich in condensed form. This report includes data on DFG approvals differentiated according to scientific discipline and the corresponding funding programme as well as the number of reviewers consulted by the DFG from 1999 to 2001. Furthermore, the report offers information on the financial capital structure, the human resources and indicators suitable for the representation of the internationality of this institution (AvH research fellows and award winners from 1997 to 2001 and DAAD-funded international scientists and academics in 2000 and 2001). This brief report can be extended by introducing numerous additional items (e.g. titles of funded research programmes).

University of Munich

Financial Data

	Regular Expenditure (= total)	Administrative income	Third party funding income	Regular core funds	
Financial resources (in mio. €)	1947.5 (100%)	919.7 (47.2%)	212.7 (10.9%)	815.1 (41.9%)	
	Total	**Humanities and Social Sciences**	**Life Sciences**	**Natural Sciences**	**Engineering Sciences**
DFG approvals (in mio. €)	116.9 (100%)	22.9 (19.6%)	70.0 (59.9%)	21.7 (18.6%)	2.3 (2.0%)
	Total	**Individual Grants Programme**	**Coordinated Programme**	**Direct Promotion of Young Researchers**	**Prizes**
DFG approvals (in mio. €)	116.9 (100%)	40.0 (34.2%)	67.6 (57.8)	5.7 (4.9%)	3.6 (3.1%)

Personnel Data

	Total	Humanities and Social Sciences	Life Sciences	Natural Sciences	Engineering Sciences
Scientists (Total)	5129 (100%)	1270 (24.8%)	3006 (58.6%)	790 (15.4%)	48 (0.9%)
Professors	710 (100%)	306 (43.1%)	274 (38.6%)	122 (17.2%)	7 (1%)
DFG reviewers	309 (100%)	110 (35.6%)	148 (47.9%)	47 (15.2%)	4 (1.3%)

Internationality

	Total	Humanities and Social Sciences	Life Sciences	Natural Sciences	Engineering Sciences
DAAD scientists& academics	103 (100%)	70 (68.0%)	22 (21.4%)	11 (10.7%)	- (0%)
AvH research fellows & award winners	115 (100%)	44 (38.3%)	24 (20.9%)	45 (39.1%)	2 (1.7%)

Figure 2: Example of elements implemented into an institutional report on the University of Munich[7]

[7] The figure shows personnel data based on a special report by the German Federal Statistical Office from 2002: Full-time scientific and artistic staff according to organizational classification, university, fields of teaching and research and staff unit (status: 2000). In the context of the differentiated presentations in the scope of the report in condensed form, one professor and 15 scientists are omitted, because they were not classified by the data of the German Federal Statistical Office. The financial data is based on a further special report of the German Federal Statistical Office from 2002: Regular expenditure, administrative income, third party funding income and regular core funds according to organizational classification, university, fields of teaching and research (1999 to 2000). For further details see Deutsche Forschungsgemeinschaft (2003).

3.2 Geographic information systems

In addition to looking at the parameter values for individual institutions, it is possible to investigate the question of the extent to which these institutions, for example, jointly determine the allocation of approvals for each region. The distribution of approvals and specific values for other research indicators can be documented using cartographical representations (maps) in the scope of a geographic information system.

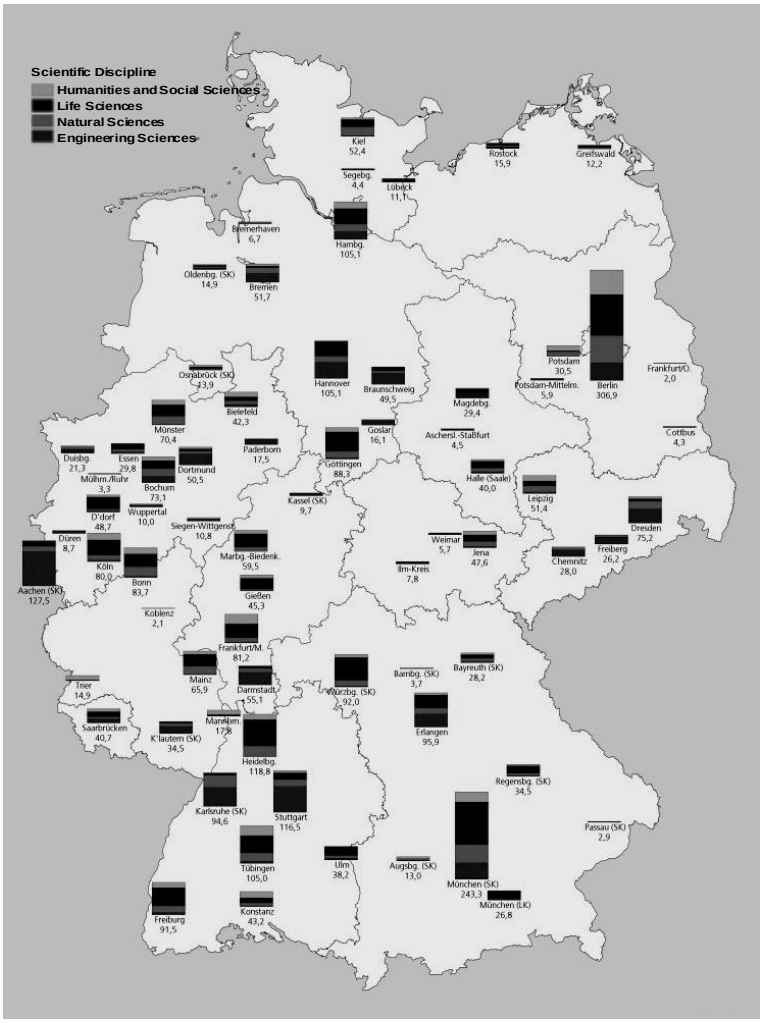

Figure 3: Geographical allocations of DFG approvals received by universities and non-university research institutions from 1999 to 2001 by district[8] and scientific discipline (in millions of euros)

---

8 Only districts that received more than two million euros in DFG approvals in total within the period stated (abbreviations: "SK" for urban districts and "LK" for rural districts).

Figure 3 shows the geographical allocations of DFG approvals by district and scientific discipline using a cartographical representation produced by a geographic information system (GIS). The figures provide insight into the "research regions" and enable regional focal points to be identified. Furthermore, using a GIS in the context of several types of research indicators provides perspectives on the various research institutions and the surrounding regions with their scientific infrastructure.

The methodology used in the DFG Funding Ranking is based on "district codes" provided by the German Federal Statistical Office. The data is processed in such way that conclusions are drawn not just according to postcode areas, but rather according to administrative rural and urban districts. To make this possible, a district code is assigned to each institution implemented in the research institution directory and standardized (address) database for the various research institutions covered by the heterogeneous research funding databases. For institutions located at multiple locations, each site is allocated its own code. Therefore, several levels of analysis can be displayed for the different indicators, such as an overall view for Germany, distributions for the federal states, or allocation by administrative districts. The following section discusses aspects and possible analysis of research networks and clusters. Thus, regional clusters, and especially the structures within the various clusters, can also be determined by using network analytical representations.

3.3 Network analytical representations

Cooperation in networks is a key factor for modern science and research. A metaphor like "research cluster" stands equally for "developing regional focal points" and "internationality", as it does for "disciplinary accentuation" and "interdisciplinarity". The ideal is not generally embodied by an individual researcher working in isolation, but rather by a research team or network with a multitude of national and international relationships, both in an interdisciplinary framework and within their own research discipline. Therefore, the support of inter-institutional and multidisciplinary cooperation is seen as an important instrument for the funding of innovative research. The focus of research network analysis is the issue of to what extent research network funding programmes are taken advantage of in the various research areas and in what way they contribute to networking between the researchers involved at different institutions. In particular, the general significance attributed to these research network programmes and the structures resulting from the joint participation by the various research institutions in these programmes are of interest.

In the scope of the DFG Funding Ranking this is preceded by quantitative analysis of the research network funding programmes spanning multiple institutions within the context of DFG coordinated programmes (such as Collaborative Research Centres, Priority Programmes, Research Units and Research Training Groups), for instance by calculating the number of "partner institutions" with which universities and non-university institutes cooperate. The analysis determined in this way allows conclusions to be drawn on the centrality of institutions in the science system in total or differentiated according to the respective research areas resulting from the relationships within the cooperation networks.

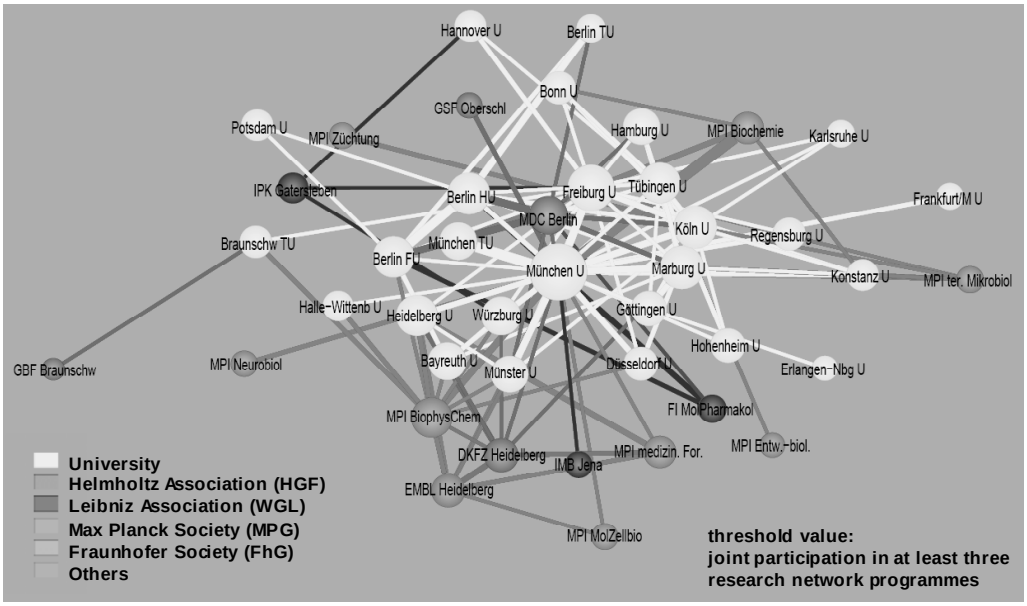

Figure 4: The relationship structure between universities and non-university research institutions in biology (1999-2001)

The network structures can be visualized in the form of network analytical representations. The resulting figures primarily serve to describe the establishment and development of "research clusters". Therefore, structures are considered to be networks or clusters of this kind if they result from joint participation in coordinated programmes in the form of groups of universities and non-university research institutions, which interact particularly intensively. As an example Figure 4 shows the core network in biology[9].

The graphs can be analysed in the following way:

- institutions that are central to the structure are positioned centrally,
- institutions that interact frequently are positioned close to each other,
- the line thickness indicates the frequency of interaction between two institutions and
- the diameter of the circles represents the total number of interactions.

The University of Munich is located in the centre of the network; it is the most important institution in the biology network. The diameter of the circles also indicates the importance of the university. The University of Munich collaborates intensively, for example, with the universities in Freiburg and Muenster. This is shown by the thickness of the lines between these institutions and their relatively close proximity.

---

[9] The network analytical representations were developed by Lothar Krempel, Max Planck Institute for the Study of Societies in Cologne, Germany. For further information on social network analysis see Krempel (2005).

Due to the quantity of institutions this figure represents a complex interplay of various research institutions. The internet version of the DFG Funding Ranking offers an additional application to visualize the research networks. Using scalable vector graphics (SVG), it is possible to view one or more networks[10]. Thus, it enables analysis of a specific network or regional cluster.

4 Conclusion

In the context of assessing research structures and institutions, there is a strong need to integrate information from various sources. Research infrastructure and funding databases in general and the consolidation of heterogeneous databases provide numerous opportunities for the derivation of several types of indicators representing the research activities and achievements of research institutions. In this paper the derivation of central research indicators for individual institutions based on databases covering the funding schemes of research funding agencies is briefly presented. In order to be able to correlate data from heterogeneous sources it is necessary to compile a concordance within the framework of different subject classifications and between diverse data collections of research institutions. Standardization processes for heterogeneous databases allow analysis for a more distinctive assessment of research institutions using the different properties of the research funding databases.

Furthermore, based on these standardized research funding databases, information systems that provide information on the performance of research institutions are also conceivable. For instance, using a standardized database it is possible to refer to the respective research areas and the institutions' names consistently for almost all of the reports, figures, tables, and search functions implemented in information systems. In this paper three information systems are exemplified, which can be developed based on research funding databases. Institutional reports enable comprehensive information on each institution and their institutional environments to be collected and examined. Geographical information systems allow analysis of regional conditions for research. In addition, network analytic representations offer insight into the structure of research networks and clusters. In connection with other potential (qualitative and quantitative) information on research activities and achievements of research institutions, these information systems could be implemented into a holistic research assessment information system.

[10] For further details on static and interactive views of cooperative relationships between different research institutions, see http://www.dfg.de/en/ranking/networks.

5 References

Deutsche Forschungsgemeinschaft (1997): *Bewilligungen nach Hochschulen – Bewilligungsvolumen 1991 bis 1995, Anzahl kooperativer Projekte im Jahr 1996*, Bonn (www.dfg.de/ranking/archiv).

-.- (2000): *DFG-Bewilligungen an Hochschulen und außeruniversitäre Forschungseinrichtungen 1996 bis 1998*, Bonn (www.dfg.de/ranking/archiv).

-.- (2003): *Funding-Ranking 2003 – Institutions - Regions - Networks. DFG Approvals and Other Basic Data on Publicly Funded Research*, Bonn (www.dfg.de/ranking/archiv).

Guedler, Juergen (2006): CRISs as a Management Tool: Experiences and Perspectives at the German Research Foundation (DFG), *Conference on Current Research Information Systems CRIS2006*, Bergen, 10th-13th May 2006.

Hornbostel, Stefan (1997): *Wissenschaftsindikatoren, Bewertungen in der Wissenschaft*, Opladen.

Krempel, Lothar (2005): *Visualisierung komplexer Strukturen. Grundlagen der Darstellung mehrdimensionaler Netzwerke*, Cologne.

6 Contact Information

Daniel Bovelet

Deutsche Forschungsgemeinschaft (DFG)

German Research Foundation

Information Management

Kennedyallee 40

D-53175 Bonn, Germany

e-mail: Daniel.Bovelet@dfg.de

web: www.dfg.de/en/ranking