IST-1999-20350
Deliverable Identification Sheet

Project ref. no.	*IST-1999-20350*
Project acronym	EICSTES
Project full title	European Indicators, Cyberspace and the Science-Technology-Economy System

Security (distribution level)	*Confidential (Restricted to Consortium and Commission)*
Contractual date of delivery	*M12 = December 2001*
Actual date of delivery	*January 10th 2002*
Deliverable number	*D-5.1*
Deliverable name	*Case Studies*
Type	*Report*
Status & version	*Version 2; revised April 2002*
Number of pages	
WP contributing to the deliverable	*WP5; WP8*
WP / Task responsible	*UvA/UvA*
Other contributors	
Author(s)	*Gaston Heimeriks; Peter Van den Besselaar;* *Anne-Marie Oostveen*
EC Project Officer	*Tony Vorlou*
Keywords	*webometrics, case studies, indicators*

Abstract (for dissemination)	This report consists of three case studies about links between websites in a research context. The first exploratory case study aims at deriving Internet based indicators for mapping the development of scientific and technological developments in the information society. The study is designed to obtain information about the usefulness of links as indicators for the dynamics of knowledge production on various levels of analysis. The heterogeneous linking structure of universities might provide immediate information about the context of knowledge production. Our results suggest that the 'research group' constitutes the most meaningful unit of analysis. The second study is a brief exploration of the Triple Helix model of web-based relations between companies, research institutions, and government. It shows that link structures on the web are strongly dominated by universities. At the same time, a factor analysis shows that the positional structure seems regional, more than anything else. The third case study aims at mapping all relevant scholarly communications with respect to a research group operating in an application-oriented techno-scientific environment. Scholarly communication refers to the formal and informal processes by which the research and scholarship of researchers are created, evaluated, edited, formatted, distributed, organized, made accessible, archived, used, and transformed. This study explores the opportunity of using inlinks, outlinks, incoming emails, outgoing emails, project co-operations and co-authored publications of academic, political and commercial institutes as indicators for knowledge production: what is communicated in different media? Can we identify the role of computer-mediated communications in relation to print and other traditional media? Do the communication patterns provide information about the changing nature of knowledge production?

DOCUMENT EVOLUTION (optional)

Version	Date	Status	Notes
0.1	4 June 2001	Draft	
0.2	30 September 2001	Draft	
0.3	25 December 2001	Draft for internal review	
1.0	10 January 2002	Final version delivered	
1.1	15 April 2002	Revised version for internal review	
2.0	25 June 2002	Final version delivered	

Table of Contents

Chapter 1: Introduction

The aim of work-package 5 is to test indicators by in-depth case studies focusing on horizontal and vertical sections of the information society, in order to deepen the understanding of the patterns that emerge from the Internet data, and to clarify the links between changing patterns of interaction in the 'real' networks of the information society, and the 'virtual' networks on the Internet.

According to the Technical Annex, the plan was to start with the so called Triple Helix case study, aiming to study the interaction on the Web between industry, government and universities in selected fields. The design of this case study was to a large extent based on an analysis of WWW-links between these institutions. However, initial work on the triple helix (chapter 3 in this report), the review of bibliometric and webometric methodologies (Deliverable 1.4, state of the art)), and other work within the project (Deliverable 8.1) indicated that 'links' are a difficult type of data. As the mentioned studies showed, linking structures can be found, but are often difficult to interpret. What counts as a link, and what do links indicate?

At the same time, the more detailed planning of the work-package had to be changed, as the various subtasks showed too much overlap and at the same time left several issues not addressed. Therefore, we decided to change the order and organization of the work-package in the following way. It was necessary to start with some case studies that would inform us about the problems and possibilities of measuring link patterns, about the ways to interpret linking structures, and about possibilities of using them as web-indicators. More specific, we were firstly in need of additional information about the value of links as input for the development of indicators for techno-scientific developments, and secondly we needed additional information about the role of electronic communications in relation to other forms of scholarly communication, such as traditional printed communications. Therefore we added the two case studies which form the bulk of the content of this deliverable (chapters 2 and 4). The triple helix case study will now be part of D5.3. This leads to the following structure of WP5.

	Old organization		New organization
D5.1	Triple helix	D5.1	Meaning of links
D5.2	Non-web data	D5.2	Non-web data
D5.3	Web and non-web data: transborder	D5.3a	Web and non-web data: : transborder
		D5.3b	Web and non-web data trans-sector = triple helix
D5.4	Web and non-web data: new economy	D5.4	Web and non web data: new economy

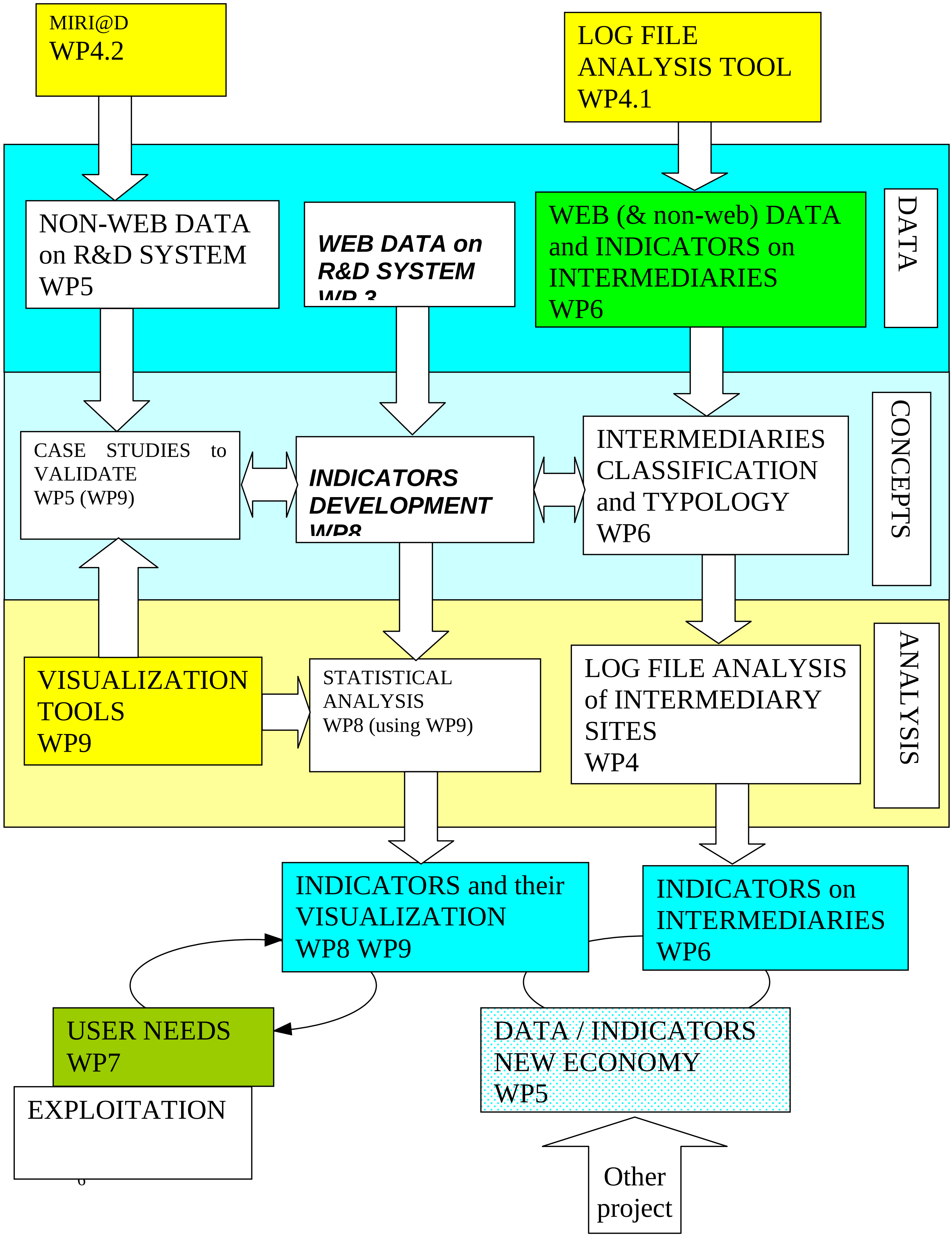
1.1 Integration and Contextualization

In this Deliverable, we will focus on exploring, developing and testing web-based indicators for mapping the development of the science system in interaction with economic, technological and political developments. The central idea is that the enterprise of techno-science in the information society is best captured by the concept of evolving communication systems. Central to this model is the idea that science can be characterized by the distributed communication of 'quanta of knowledge'. Communication should be understood in a broad sense, and refers to various levels of interaction. In other words, we distinguish between number of relevant communication systems.

- the collaborative knowledge production in scientific specialties: coauthoring papers and collaboration in projects;
- the fine-grained interactions (citations) around research topics
- the communication (citations) within the broader discipline and between disciplines
- the user-producer interaction of (sub)disciplines and their commercial and political environment

Institutes and people that are involved in the production and dissemination of knowledge (e.g., universities, research groups, companies, ministries, etc) can be considered the carrying nodes of the communication network. Relevant communications can take place in mutually interacting media that all play a different role in the process of knowledge production and dissemination; scientific papers linked by cited references, project co-operations, hyperlinked web-pages, e-mail communications etc. Consequently, relevant data for developing web-based indicators not only includes web-data, but also non-web-data (research projects, patents and publications).

The conceptual integration of data and methodologies within EICSTES project can be described as follows; Web-data on the European R&D system are collected in WP3 (detailed information on more than 25,000 academic and research institutes). The data will be used in WP8 to develop indicators, in WP9 for visualization and WP5 for comparing web and non-web-data. WP 5 has an important role to play in validating and testing web based indicators, among the testing strategies is comparison of web-data with non-web-data. The log file analysis tool developed in WP4.1 will be combined with web and non-web information collected in WP6 on intermediaries on the web. MIRI@D (WP4.2) will provide us with information about the users and producers of scientific knowledge in the R&D system. The resulting networks are used in WP5 for improving our understanding of the relationship between electronic networks and real-life networks. The case studies in WP5 use data from WP3 (web-data), WP4 (users/producers networks) and indicators from WP8 in order to test the reliability and usefulness of the webometric indicators by comparing the various Triple Helix networks (web and non-web) of governmental, industrial and academic co-operations. The results will feedback into WP8 on developing indicators. Indicators on web-based intermediaries in the R&D system are developed in WP6. Relevant examples of intermediaries are provided by WP5.


MIRI@D
WP4.2
LOG FILE ANALYSIS TOOL WP4.1
NON-WEB DATA on R&D SYSTEM WP5
WEB DATA on R&D SYSTEM
WEB (& non-web) DATA and INDICATORS on INTERMEDIARIES WP6
DATA
CASE STUDIES to VALIDATE WP5 (WP9)
INDICATORS DEVELOPMENT
INTERMEDIARIES CLASSIFICATION and TYPOLOGY WP6
CONCEPTS
VISUALIZATION TOOLS WP9
STATISTICAL ANALYSIS WP8 (using WP9)
LOG FILE ANALYSIS of INTERMEDIARY SITES WP4
ANALYSIS
INDICATORS and their VISUALIZATION WP8 WP9
INDICATORS on INTERMEDIARIES WP6
USER NEEDS WP7
EXPLOITATION
DATA / INDICATORS NEW ECONOMY WP5
Other project

Users will provide feedback in WP7 on the indicators and data developed in WP3 and WP8, as well as the visualization for dissemination developed in WP9. In WP8 indicators are developed using data from WP3 and additional feedback from WP5 and WP7. Visualizations are developed in WP9 using data and indicators from WP3, WP4, WP5. The figure on the previous page shows the relations between the various work packages.

1.3 Levels of Analysis

One of the first question arising in empirical studies of web-data is about the appropriate level of analysis. As a consequence of studying the techno-scientific developments as an evolving communication system, deriving indicators for mapping the development of the science-technology-economy system on the web involves three levels of analysis; the actors, the networks of actors, and the electronic networks in relation to other (non-web) networks.

The first level of analysis deals with identifying the properties of relevant actors. On the Internet, these actors can be web-pages, domains and sites that play a role in production and dissemination of science and technology. Many relevant attributes can be identified on this level of analysis that provide insight in the characteristics and development of the science-technology-economy system. This level of analysis will be reflected in the case study SWI described in this deliverable in Chapter 3.

The second level of analysis is the level of communities of actors. Communications between actors (e.g., between users and producers of knowledge, between collaborators in research projects, citational behavior between scientists in a scientific discipline) are not randomly distributed but take place in clusters or communities. Actors generally communicate more with other actors within a cluster than to actors outside the cluster. These clusters of communications (communities or systems) have properties that cannot be attributed solely to the actors that are part of this network. Evolving communication systems cannot be defined by simply enumerating or tracing the layout of their constituent elements. The definitive attribute of a systemic entity is the set of inter-component relationships which outline its form at any given moment and serve as the core 'identity' which is maintained in spite of dynamic changes over time. It is the relational organization of a system which defines its identity, its properties as a unity, and the frame within which it must be addressed as a unary whole. The methodological consequence of this insight is, that meaningful domains need to be delineated on the bases of their networked relationships. Examples of these communities are scientific disciplines and communities organized around a topic or a project. Chapter 2 deals with this systemic level of analysis.

The third level of analysis is concerned with the position of a communication network in electronic media with respect to other communication networks in the “real world”. Examples include the various communication networks around a research project; networks of email communications, hyperlinks, project co-operations and cited references. This level of analysis aims at providing insight in the relationship of web indicators with traditional indicators for networks of scientific communications. Does the presence of the Internet and the World Wide Web inform us about inter-organizational networks ‘in real life’, or are the real and virtual communication networks relatively loosely coupled? By mapping various communication patterns in Chapter 3, valuable information could be obtained.

1.4 The case studies

In chapter two, we give a report of our case study in which we explored the value of ‘links’ as indicators of the structure and dynamics of knowledge production on various levels of analysis. The heterogeneous linking structure of universities might provide information about the context of knowledge production. However, the question is what the adequate nodes of the linking network are?
The aim of this case study was to gain information about what is communicated by cyber-links and what are the appropriate levels of analysis to study linking patterns? Furthermore, which units of analysis are most appropriate? We analyzed linking patterns on the following four levels of analysis:

1. The academic linking structure between the EU countries. Here we analyze the links between the national systems of higher education within the EU, defined as the set of all universities within a country. On this level, we found geographical and language biases in the academic world.
2. The academic linking structure within the Netherlands. The Dutch universities showed an outlink communication structure different from the inlink structure, and the large general universities are located in the same cluster.
3. The linking structure of a research group. This level consists of the ‘personal network’ of a single research group within a university, that is the network of all institutions linked (above a certain threshold) to this research group, using incoming and outgoing links. The results suggest what this level of the research group constitutes the most meaningful unit of analysis, with analogies with Journal-Journal citation analysis.
4. The level of the individual researcher.

In chapter three, the ‘triple helix of government, industry and university relations’ was studied in the Netherlands. As was described in the state of the art deliverable D1.4, clear methods, concepts and theories related to scientific communications on the web are still lacking. However, an exploratory study was carried out to study the opportunities of mapping the Dutch triple helix structure on the web. For this analysis, a group of Dutch URL’s was selected that contained all Dutch universities, all Dutch ministries, a selection of the most important non-governmental organizations and a list of the largest companies with a stock exchange quotation. It was found that universities contribute to the largest amount of links. Furthermore, the analysis suggested smaller units of analysis (e.g. research groups within the universities) might provide us with more insightful information about ‘Triple Helix interactions’.

The case study in chapter four aims at mapping the various communication patterns of a mode 2 research group; inlinks, outlinks, incoming emails, outgoing emails, project co-operations and co-authored publications. The aim of the study is to compare – in a quantitative and qualitative way – the link structure with other forms of electronic and 'traditional' non-electronic communications.

A remark needs to be made here about the scope of both case studies. The analysis and data collections in WP5a were developed in interaction with our contributions to the development of web-indicators in WP8. This WP aims to derive indicators from complex, large and multidimensional databases that involve the methodological tools of applying a number of quantitative processing. This processing might be bibliometrical, statistical, graph theoretic, following a social network analysis or according to the recent advances of complexity and chaos theories. The indicators generated by these types of processing are expected to capture the subtleties and complexities of the corresponding data and to be firmly embedded in the theory of emergent phenomena and complex systems, which, thus, might be tested as whether they constitute viable conceptual models for the information society. However, as will be clear form D1.4 and D8.1, quite some webindicators have been proposed. Only one of those is used in the two case studies up to now, and that is the *position* of a wet site in the network(s) of weblinks it is embedded in. This will become clear in the next chapters, were we mainly use a factor-analysis based approach of analyzing the positional structure of the link networks. (Burt 1982; Knoke & Kuklinsky 1982) We have chosen to do this because of the methodological nature of the studies up to now. In the following period, we will also use other – relational – webindicators to complete the evaluation of the meaning of links.

1.5 Next steps in WP5

This WP consists of several empirical studies with the explicit aim of uncovering main aspects of the interaction between Internet S&T and Internet economics on the one hand, and S&T and social and economic institutions and geographical space on the other. In this way, the relation between real and web indicators will be clarified. Four subtasks were mentioned: 1) Non-web-data collection; 2) Triple Helix; 3) Trans-border co-operation; 4) Web and non-web indicators of the new economy. The order of these activities needs some change.

The non-web-data collection on the new economy will be based mainly on activities in related projects, focusing on statistics of the new-economy (e.g., the NewKind project), and in relation to WP6, on the economic and informational roles and development of intermediaries. To be able to make this work functional for the main goals of the project, that is the development of Webindicators of the knowledge production system, it will be done a little later in the project than planned.

In the next phase we will focus on three problem fields:

- The non-web data collection needed for the ‘triple helix study’, and for the ‘trans-border co-operation study’
- Beginning with the ‘triple helix study’.
- Beginning with the trans-border co-operation studies.

Chapter 2. The value of 'links' as indicators for knowledge production on various levels of analysis. An exploratory analysis.

2.1 Introduction & theoretical considerations

The Internet is a good example of a socio-technical medium where social networks are developing and interfering with computer networks. Organizations are increasingly communicating through the Internet, and present on the World Wide Web. This may influence the shape and the dynamics of the social networks organizations are part of.

What are meaningful indicators for Internet based inter-organizational relations? What is the relationship of those indicators with traditional indicators for networks between organizations? Does the presence of the Internet and the World Wide Web inform us about inter-organizational networks 'in real life', or are the real and virtual communication networks relatively loosely coupled? Do WWW and Internet based indicators teach us something about the changes of the inter-organizational networks?

In this study we will address some of these questions. We will focus on what is considered to be one of the most important webometric indicators: cyber-links (or links). Scientometric and webometric analysis rest on measuring traces of communicated information. And since links are considered the most permanent and formal communications between web-pages, they provide an obvious starting point for research. Of course, there are many other relevant web-data that can be used for the development of indicators like e-mail communications, characteristics of web sites, web-traffic maps. However, links have often been used in analogy with cited references in print journals, a well established set of methodologies in scientometrics. Elsewhere in this EICSTES project (Polanco, Boudourides 2001) links analyses have been carried out, but the meaning of the found patterns were not easy to determine. This raised the issue of levels of analysis. It was not yet clear what are suitable levels of analysis for conducting link analyses; is it appropriate to analyze a large set of universities? Are other levels of analysis more appropriate?

In this study, we focus on a class of organizations with a relatively developed presence on the Internet: Institutes of higher education in the main European countries and especially in The Netherlands. We will map the Internet relations between institutions of higher education in the mentioned countries, in order to get more insight in the dynamics of the virtual networks within the knowledge producing system. The aim of this study is to map the network patterns of European universities connected by hyperlinks in order to gain insight in the socio-epistemological meaning of linking structures; what is communicated by links? And what patterns are emerging?

The linking patterns of universities and research institutes might provide additional information about the structure context of knowledge production[1], which has to be related to the other communication networks within knowledge production. The most codified level of communication within this context is the system of print media. Many analyses therefore are based on this level of communication, but at the same time this creates a strong bias. Le Pair (1988), for example, found that citations do not provide us with an accurate representation of technological achievements, because knowledge can be built into technological artifacts without necessarily leaving the formal trace of a citation in the scientific literature (Els *et al.*, 1989). Electronic communications provide new opportunities for the knowledge production system, and because this type of communication is heterogeneous and not strongly codified, it may not have an easy equivalent in the printed communication. Communications between universities, and communications between universities and third parties on the Internet may be used to support an interactive mode of knowledge production, an enable us to make this mode accessible to measurement. However, the interpretation of the finding is not straightforward, as the use of new media also changes the object under study.

Here we focus on the mapping of the structural and institutional relations between universities, and a selection of industries and governments. Links are less codified than print based indicators such as cited references. Although the reasons for citing can be very diverse on the level of individual scientists (e.g. the negative citations, or citing friends and colleagues), on the aggregate level of the science system, it generally provides a clear representation of the shared knowledge base of a scientific discipline.[2]

Scientific knowledge that is communicated in electronic media is less codified, more heterogeneous than its print equivalent because the operations that define the cognitive structure and boundaries of the research field do not (or not that fully) apply to electronic communications. This poses an immediate problem for studying the development of science using traces of electronically communicated information. The question arising is, where does the research field leave off, and its (social) environment begins? What is the research field on the web ? Or better, where is it located on the web? Hyperlinks have often been compared to cited references as an indicator of scientific orientation, but they lack the codification of a well-defined environment. In print based communications, scientific papers refer to other scientific papers. Electronic communications, like links, take place in a much less defined domain, consequently, the communications are less codified. We do not know for sure why people on the web link up to other pages. Björneborn and Ingwersen write:

[1] Two problems emerge in using this domain. Firstly universities also have a predominant teaching function, and a large part of the websites of universities and other institutes of higher education may reflect this function. In other words, the electronic link structure is expected to be the composite of two networks: a research network, and an educational network. Secondly, the link structure may also be the composite of various media, as the web is a multimedia environment. For example, the link structure may reflect as well a communication network among collaborating people, as an information space, linking information.

[2] An example that this even works for a social science field is described is Van den Besselaar (2000, 2001).

> Obviously, the breakthrough for everybody to express themselves, practically without control from authorities, to become visible world wide, also by linking to what pages one wants to link, to assume credibility by being ‘there’, and to obtain access to data, information, values and knowledge in many shapes of and degrees of truth, has generated a reality of freedom of information, also in regions and countries otherwise poor of infrastructure.

When we want to study the communication of scientific knowledge on the Internet empirically, the problem of the unit of analysis arises. In print communications, papers are communicated in journals. What would be the equivalent of these information carriers on Internet? And what is communicated? Aguillo suggests a new concept of a site; the presence of research groups, universities and other R&D related institutions on the Internet. ‘A web site comprising hierarchically grouped pages, and represented by the URL address of the highest level, is considered as a unit if it has been defined according to aspects: it is identified as formally different from other Web sites (documentation unit) and is recognizable as representing a research group with the aim of being present at the web (institutional unit). The institutional character of the proposed new unit eases comparative analysis with external data, as usually there are scientometric analysis available following that institutional grouping criterion. Moreover, the global viability of the database increases since a better identification of Web sites is possible if a physical real counterpart is available. (Aguillo, 1998). The design of this study was motivated by these claims; which indicators can we derive on various levels of analysis of communicated links?

2.2 Methodology and Data

Citation analysis has been used to ‘map’ the manner in which information is shared and passed through printed media. In webometrics, *link* analysis could yield similar information about the manner in which information is diffused on the Internet (Clever Project, 1999; Butler, 2000). Rousseau (1997) has called this “sitation” analysis.

However, several problems arise with the application of quantitative analysis to linking structures; the problem of data collection, the boundary problem related to the lack of codification and the problem of the appropriate level of analysis.

A complicating factor in webometric research is that of data collection. Data collection on the web depends on the retrieval features of the various search engines and web robots. Lawrence and Giles (1998) provide a substantial contribution with respect to the commercial search engine coverage of the Web space by introducing the concept of ‘indexable web’. The concept signifies the portion of the web that can be indexed by the engines. The engines do not cover the entire web, the overlap between them is not substantial and their retrieval functions too simplistic for webometric analyses.

The second problem relates to the above mentioned boundary problem. Links are more heterogeneous in nature than cited references, which always refer to communications within the scientific domain. That creates a problem in the interpretation of linking networks. In this study, we will focus on institutes of higher education in Europe. It can be expected that links existing between different universities are generally scientific in nature – still heterogeneously, but probably less than in other domains.

The question of the appropriate level(s) of analysis in studying linking patterns will be answered empirically in this study. In a first set of analysis based on the database of European universities (http://nicomedia.math.upatras.gr/eicstes/webdata.html) the linking network of European universities was analyzed. The aim of this study is to map the network patterns of European universities connected by hyperlinks in order to gain insight in the socio-epistemological meaning of linking structures; what is communicated by links? And what patterns are emerging?

Our raw data refer to about 1000 universities sites from 22 European countries (including all of the EU) plus about 1500 university sites from the USA. These data are collected within the context of the EICSTES project. For all the University sites, we have measured the number of links between any two of them by a software application operating over the AltaVista search engine. A matrix was constructed of all universities within the EU. The focus on EU universities was motivated by previous analyses of the European science-technology system. It can be argued that a European system is emerging on top of the various national systems of innovation. The emergence of a new way of knowledge production on top of the traditional scientific enterprise finds its roots in the intimate interaction between information and communication technologies (ICT), and advanced scientific and technological research and innovation. Furthermore, this development is undoubtedly related to internationalization of modern science. The new mode of knowledge production is reflected and reinforced by transnational research projects (like the EU framework programs) with a strong emphasis on application.

The first level is the academic linking structure between the EU countries. Here we analyze the links between the national systems of higher education within the EU, defined as the set of all universities within a country.

In the next step we focus on the European orientation of Dutch universities and colleges in terms of linking structure. We constructed a matrix of all outlinks from the Dutch higher educational institutes to all other European universities, and another matrix containing all inlinks from European universities to the Dutch sites. Factor analysis of these matrices reveals clusters of Dutch universities with a similar European orientation as indicated by links.

On a lower level of analysis, we focused on all Dutch universities and their national linking patterns. Again we used factor analysis in two dimensions to distinguished between the structure emerging from the inlink pattern and the outlink pattern.

The next analysis focused on the 'research group' as unit of analysis for mapping linking patterns. The SWI group of the university of Amsterdam was selected for several reasons; first of all, this group represents a typical example of mode-2 research; strong emphasis on the context of application, interdisciplinary research focus, strong presence in international research projects and a heterogeneous research output (software applications, papers, deliverables, etc). Secondly, this research group is likely to participate on large scale on Internet because of its focus on computer science and software applications.

Finally, to complete the comparison between the various levels of analysis, a personal homepage of a single researcher of the SWI research group was selected to explore the opportunities of mapping the dynamics of knowledge production on the lowest level of analysis.

As said before, the analysis in this chapter is positional. We use a factor analysis, with orthogonal rotation to determine the position of websites in their 'sitation environment'.

2.3 Results

2.3.1 European University networks

Among other data (collected by automatic intelligent agents scanning various search engines), we are measuring the number of links among almost all of the University sites in the 15 EU countries: Austria (at), Belgium (be), Denmark (dk), Finland (fi), France (fr), Germany (de), Greece (gr), Ireland (ie), Italy (it), Luxembourg (lu), the Netherlands (nl), Portugal (pt), Spain (es), Sweden (su) and UK (uk). Factor analysis of the matrix of links between all European universities on the country level reveals two very pronounced clusters that can be identified as a Germanic and an Anglo-saxon cluster, in addition to some other clusters.[3]

A factor analysis of the country by country 'sitation' matrix was done for the incoming as for the outgoing direction. The factor analysis of the inlinks to European universities suggests a strong geographical and/or language bias. The most pronounced cluster (highest explained variance) is the one containing Germany, Denmark, Austria, Luxembourg add the Netherlands. The second most important cluster contains UK, Ireland, Belgium and Greece. Belgium shows interfactorial complexity; it contributes to all clusters probably because of its bilingual and bicultural nature. Belgium has its highest loading (0.56) on factor B, and its second loading (0.42) on factor D, which also contains France.   The smaller clusters contain Sweden and Finland (cluster C), Spain and France (cluster D) and Portugal and Italy (cluster E).

[3] Detailed results of factor analyses can be found in the appendices.

The following multidimensional scaling plots show the positioning of the EU countries according to the incoming links of their university sites.

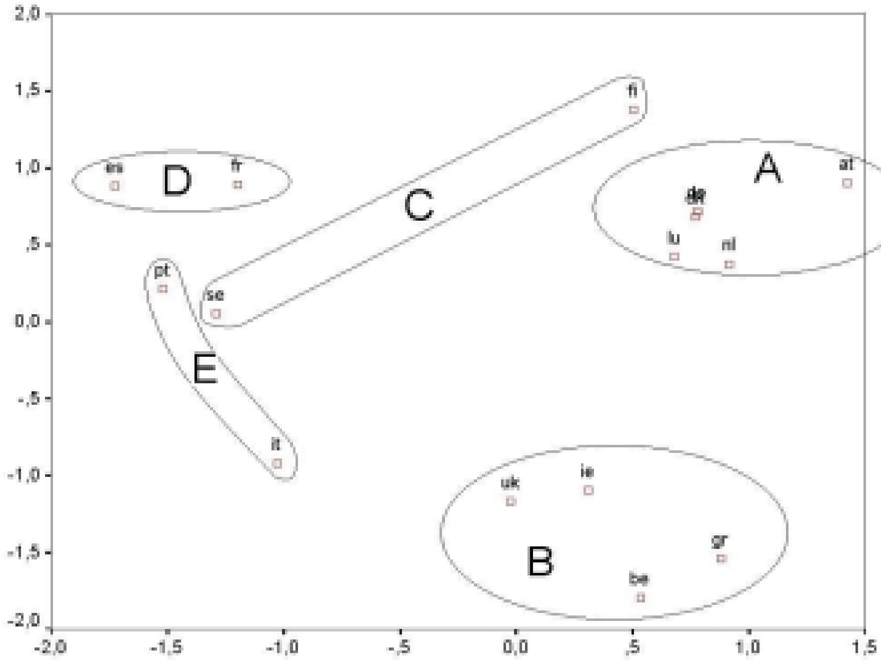


Figure 1 *Clusters of European countries obtained by multi dimensional scaling of factor loadings of the inlink matrix.*

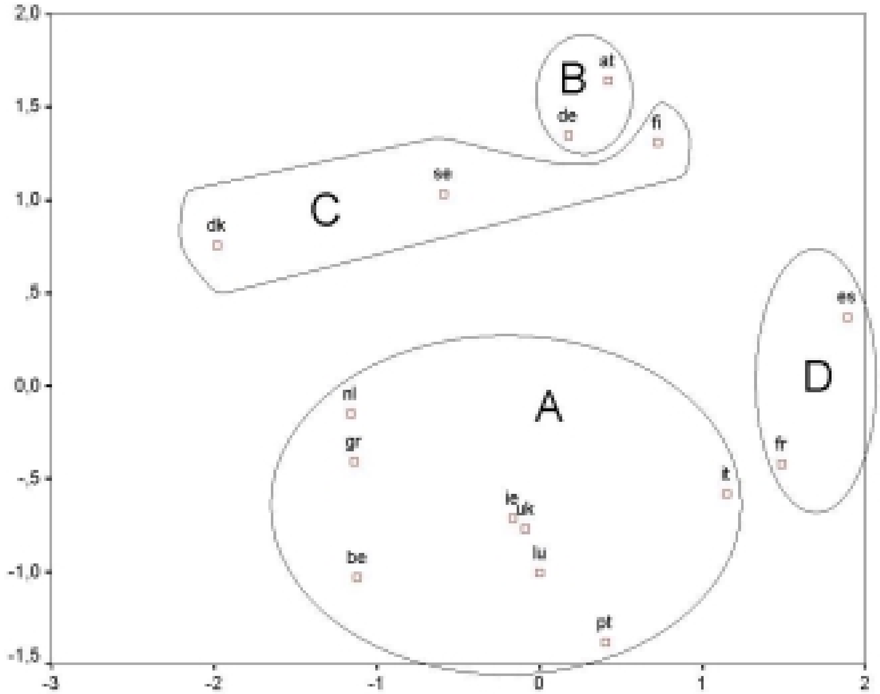


Figure 2 Clusters of European countries obtained by factor analysis of the outlink matrix.

The clusters emerging from the outlink communication patterns shows some resemblance with the inlink structure. However, the most significant cluster contains UK, Ireland, Luxembourg, Greece, Portugal, Belgium, Netherlands and Italy. It seems that the language more than geographical distance is the decisive operation underlying this pattern. This is confirmed by the composition of the other clusters; Austria and Germany (cluster B), Denmark, Sweden and Finland (cluster C) and France and Spain (cluster D).

2.3.2 The position of Dutch universities in Europe

In the next step we focused on the Dutch universities and colleges in relation to the rest of Europe. Exploratory analysis showed us that the number of 'national' links was very large compared to the number of international links of Dutch universities. That is why we decided to study the international and national linking structure separately. We expect that interpretation of the results will be easier is this way. The following tables show the results of the factor analyses. Only the most pronounced factors are included in the tables (as indicated by the scree-plots of the analysis) to make interpretation easier.

Table 1 *Dutch Universities clustered by their inlink pattern.*

INSTITUTE	FACTOR	REGION	FOCUS	TYPE
KUN.NL	1	B	GENERAL	UNIVERSITY
LEIDENUNIV.NL	1	NH	GENERAL	UNIVERSITY
RUU.NL	1	U	GENERAL	UNIVERSITY
RUG.NL	1	GR	GENERAL	UNIVERSITY
UNIMAAS.NL	1	L	GENERAL	UNIVERSITY
UVA.NL	1	NH	GENERAL	UNIVERSITY
EUR.NL	1	ZH	GENERAL	UNIVERSITY
VU.NL	1	NH	GENERAL	UNIVERSITY
HVU.NL	1	U	GENERAL	COLLEGE
KUB.NL	1	B	GENERAL	UNIVERSITY
WAU.NL	1	G	AGRICULTURAL	UNIVERSITY
NIJENRODE.NL	1	U	BUSINESS	COLLEGE
HDH.NL	2	ZH	BUSINESS	COLLEGE
HSHOLLAND.NL	2	NH	GENERAL	COLLEGE
HSDELFT.NL	2	ZH	GENERAL	COLLEGE
HSA.NL	2	NH	GENERAL	COLLEGE
HES-RDAM.NL	2	ZH	BUSINESS	COLLEGE
HVA.NL	2	NH	GENERAL	COLLEGE
UTWENTE.NL	3	O	TECHNICAL	UNIVERSITY
TUE.NL	3	B	TECHNICAL	UNIVERSITY
CHE.NL	3	G	GENERAL	COLLEGE
HRO.NL	3	ZH	GENERAL	COLLEGE
HESASD.NL	3	NH	BUSINESS	COLLEGE

Intellectual focus, type of institute and geographical location.

Again, the analysis consists of two parts; mapping the inlink and the outlink architecture. The clustering of Dutch universities according to their outlinking and inlinking pattern to all European universities (excluding Dutch universities) will inform us how the various Dutch institutes perceive the rest of Europe and how they are perceived themselves by European universities.

The factor analysis of the inlinking matrix indicates that the large universities with a broad coverage of disciplines are perceived as similar by the rest of Europe (factor one). The geographical location doesn't seem to play any role. The clustering has a purely cognitive dimension; the business schools group together in factor two and the institutes with a more technical orientation can be found in factor three.

Table 2 *Dutch Universities clustered by their outlink pattern.*

INSTITUTE	CLUSTER	REGION	FOCUS	TYPE
VU.NL	1	NH	GENERAL	UNIVERSITY
EUR.NL	1	ZH	GENERAL	UNIVERSITY
TUE.NL	1	B	TECHNICAL	UNIVERSITY
RUU.NL	1	U	GENERAL	UNIVERSITY
UVA.NL	1	NH	GENERAL	UNIVERSITY
UTWENTE.NL	1	O	TECHNICAL	UNIVERSITY
UNIMAAS.NL	1	L	GENERAL	UNIVERSITY
WAU.NL	1	G	AGRICULTURAL	UNIVERSITY
HHS.NL	1	ZH	GENERAL	COLLEGE
NIJENRODE.NL	1	U	BUSINESS	COLLEGE
HVA.NL	1	NH	GENERAL	COLLEGE
NHL.NL	1	F	GENERAL	COLLEGE
KUB.NL	2	B	GENERAL	UNIVERSITY
LEIDENUNIV.NL	2	NH	GENERAL	UNIVERSITY
RUG.NL	2	GR	GENERAL	UNIVERSITY
HVU.NL	2	U	GENERAL	COLLEGE
TUDELFT.NL	3	ZH	TECHNICAL	UNIVERSITY
HRO.NL	3	ZH	TECHNICAL	COLLEGE
HANZE.NL	3	GR	GENERAL	COLLEGE
OUH.NL	3	L	GENERAL	COLLEGE
FONTYS.NL	3	B	GENERAL	COLLEGE
HKU.NL	3	U	ART	COLLEGE
HSA.NL	4	NH	GENERAL	COLLEGE
STOAS.NL	4	B	AGRICULTURAL	COLLEGE
HSHAARLEM.NL	5	NH	GENERAL	COLLEGE
VHALL.NL	5	F	AGRICULTURAL	COLLEGE

In addition, factor analysis was carried out on the matrix representing the outlink structure of Dutch universities and colleges to other European institutes of higher education (table 2)

The outlink communication structure shows a different architecture than the inlink structure. The large general universities are located in clusters one and two, indicating a different pattern of international links. Cluster three and four consist mostly of colleges with a less pronounces linking structure.

Table 3 *Clustering of Dutch universities according to their inlink structure (including self-linking)*

INSTITUTE	CLUSTER	REGION	FOCUS
UVA.NL	1	NH	GENERAL
HHELICON.NL	1	U	GENERAL
RULIMBURG.NL	1	L	GENERAL
NIJENRODE.NL	1	U	BUSINESS
TUDELFT.NL	2	ZH	TECHNICAL
THIM.NL	2	U	PHYSICAL THERAPY
KONCON.NL	2	ZH	MUSIC
HRO.NL	3	ZH	GENERAL
KABK.NL	3	ZH	ART
HUYGENS.NL	3	O	ART
DESIGNACADEMY.NL	3	L	ART
WAU.NL	4	G	AGRICULTURE
LOIHOGESCHOOL.NL	4	NH	GENERAL
LARENSTEIN.NL	4		
HVU.NL	5	U	GENERAL
DEHORST.NL	5	U	GENERAL
KPZ.NL	6	O	TEACHER TRAINING
KEMPEL.NL	6	B	TEACHER TRAINING
DOMSTAD.NL	6	U	TEACHER TRAINING
ISELINGE.NL	6	G	TEACHER TRAINING
HANZE.NL	7	GR	GENERAL
HSLEIDEN.NL	7	NH	GENERAL
ICHTHUS-RDAM.NL	7	ZH	GENERAL
TUE.NL	8	B	TECHNICAL
HSBOS.NL	8	B	GENERAL
CHW.NL	8	O	GENERAL
UTWENTE.NL	9	O	TECHNICAL
AKI.NL	9	O	ART
HMTR.NL	10	ZH	MUSIC
AHK.NL	10	NH	ART

2.3.3 The Dutch linking Structure

The analysis of the linking structure between Dutch universities poses a methodological problem. Should the links that refer to other sites (sub-domains) within the same domain be included or not? Factor analysis of the matrix of university-university links, including self-linking, results in small clusters with a strong geographical orientation (table 3). However, factor analysis without self-linking results in one large cluster (table 4). The explanation for this phenomenon could be that all large universities have a similar pattern of linking to each-other. The only way the large universities differ from each-other is by their linking to their local, small neighbor.

The national linking structure clearly shows a strong cognitive dimension; all teacher schools can be found in cluster 6, while the institutes in cluster 1 all have a broad coverage of disciplines. On this level of analysis links are clearly an indicator for disciplinary coverage.

Excluding the self-links results in a cluster of larger universities that have a very similar linking structure, only differing in their links to small, local neighbors. Smaller institutes however, with a more homogenous disciplinary focus group together in thematic clusters.

***Table 4**. Clustering of Dutch universities according to their inlink structure (without self-linking)*

INSTITUTE	CLUSTER	REGION	FOCUS
LEIDENUNIV.NL	1	ZH	GENERA L
RUU.NL	1	U	GENERA L
KUN.NL	1	G	
NIJENRODE.NL	1	U	BUSINES S
KUB.NL	1	B	
ISS.NL	1	ZH	
RUG.NL	1	GR	
EUR.NL	1	ZH	
TUE.NL	1	B	
WAU.NL	1	G	
VU.NL	1	NH	
RULIMBURG.NL	1	L	
HSHOLLAND.NL	1		
HKU.NL	1		
UVA.NL	1		
HVA.NL	1		
HVU.NL	1		
OUH.NL	1		
UTWENTE.NL	1		
HESASD.NL	1		
KONCON.NL	1		
UNIMAAS.NL	1		
THIM.NL	2		
NTH.NL	2		
THRIJSWIJK.NL	2		
HHS.NL	2		
HZEELAND.NL	2		

TIO.NL	2
HSDELFT.NL	2
HRO.NL	2
HDH.NL	2
HEN.NL	3

***Table 4**. (continued)*

AKI.NL	3
EDITH.NL	3
HSIJ.NL	3
CHN.NL	3
VRIJEHOGESCHOOL.NL	3
HSHAARLEM.NL	3
KABK.NL	4
HUYGENS.NL	4
GERRITRIETVELDACADEMIE.NL	4
DESIGNACADEMY.NL	4
AHK.NL	4
SANDBERG.NL	4
HAN.NL	5
ISELINGE.NL	5
HSMAASTRICHT.NL	5
DRIESTAR.NL	5
HES-RDAM.NL	5
LOIHOGESCHOOL.NL	6
LARENSTEIN.NL	6
STOAS.NL	6
CAH.NL	6
HANZE.NL	7
HSDRENTHE.NL	7
NHL.NL	7
HSBOS.NL	8
CHW.NL	8
PSIMONS.BOX.NL	8

At this moment we also explore the opportunities of network visualizations as an important methodology in understanding the network dynamics. This work in progress is related to WP 9. Some preliminary results are included here. However, as the software for including web-data in the visualization tool became available only recently, we cannot report many results at the moment of writing this deliverable. Figure 3 shows clearly that the universities are not homogeneously distributed. It indicates that two factors seems to determine the linking structure of universities; geographical distance and – more importantly – thematic focus. Large institutes are generally too heterogeneous in their intellectual focus to be distinguished in terms of their linking structure.

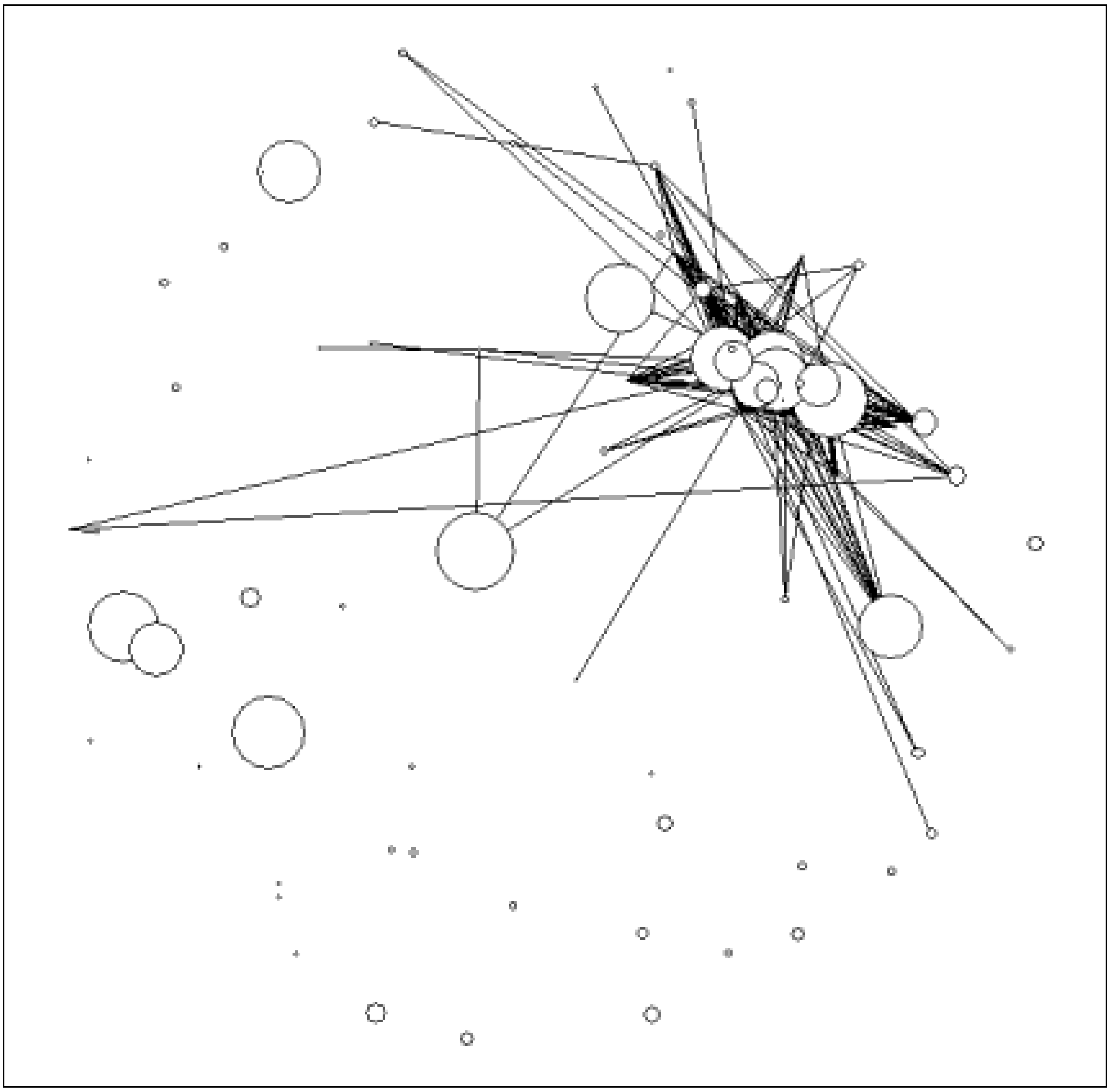


Figure 3. *Network representation of the linking structure of Dutch universities.* (Visualization provided by ARCS)

2.3.5 Linking at the level of a research group

In order to gain more insight in the value of 'links' as indicators a case study was carried out, which focused on the various Internet based and non-Internet communication networks of a single research group SWI.[4] The group was selected because it represents a

[4] We selected our own group Social Science Informatics (SWI) of the University of Amsterdam, as this enables us to interpret the results easily.

typical example of mode 2 research in the information society; emphasis on application, participation in European, transnational projects, and an interdisciplinary research orientation. The case study is reported in the next chapter. Here we use a part of the case study, and only to analyze the link structure at research group level.

Inlinks can be counted in various ways. Here we distinguish between all inlinks to the domain, the inlinks to the home page, and the inlinks to the project pages. The same holds for outlinks. In the next tables 5, 6 and 7, we give the websites that have the strongest linking relation with the SWI website, using a threshold of 3 for the inlinks, and of 59 for the outlinks.

The number of incoming links to the project pages of the SWI department was expected to reflect the heterogeneous environment of application-oriented knowledge production. Inlinks are expected to reflect the typology users of information produced at SWI, while the outlink pattern is expected to show the resources for producing knowledge at the SWI department. The results (table 5) show that universities are dominant in the linking environment of the SWI department. However, a relatively large amount of non-academic organizations has a substantial linking relationship with the project page of the SWI domain. Interesting is the presence of cordis.lu, the organization of European research projects.

Table 5. *Inlinks to the project-page of SWI.*

Number of links	URL	type
353	swi.psy.uva.nl	edu
16	cs.vu.nl	edu
13	aiai.ed.ac.uk	edu
8	cs.toronto.edu	edu
6	cs.utexas.edu	edu
5	www-users.cs.umn.edu	edu
4	id.cbs.dk	org
4	Iiia.csic.es	org
3	courses.cs.vt.edu	edu
3	engronline.ee.memphis.edu	edu
3	odur.let.rug.nl	edu
3	saussure.irmkant.rm.cnr.it	org
3	cmpe.boun.edu.tr	edu
3	cordis.lu	gov
3	cs.uwa.edu.au	edu

The number of the institutes linking to the SWI *home page* (table 6) is larger than the list of institutions linking to the *project page*. Of the seven additional institutions, five are educational. This underlines the fairly academic nature of the inlinks.

In general, sites are positioned in relation to each other. This provides a starting point for analysis. The meaning and significance of sitation/citation is expected to be quite different in the two environments of scholarly publishing and the WWW. If we assume that the WWW is a prototype of the distributed digital library of the future, it would be helpful to know if the tools and techniques developed for the analysis of intellectual structure in paper-based libraries will be able to make the transition to this network-based environment (Wilensky, 1995). Counts of ingoing and outgoing links can be seen as citation and reference analysis respectively. However, due to its dynamic and distributed nature, the Web often demonstrates web pages simultaneously linking to each other- a case not possible in the traditional paper based citation world (Björneborn & Ingwersen 2001). The outlink structure (table 7) is expected to be more resource oriented than the inlink structure. The results support this hypothesis, and next to many links to university sites (resources of knowledge), links to archives, to technical support and to software providers are among the most frequent occurring.

Table 6. *Inlinks to the SWI homepage.*

Number of links	URL	type
443	swi.psy.uva.nl	edu
22	cs.vu.nl	edu
13	aiai.ed.ac.uk	edu
8	cs.toronto.edu	edu
7	cs.utexas.edu	edu
5	cs.umn.edu	edu
5	irit.fr	edu
4	saussure.irmkant.rm.cnr.it	org
4	cs.wpi.edu	edu
4	id.cbs.dk	edu
4	iiia.csic.es	org
4	macs.mines.edu	edu
3	courses.cs.vt.edu	edu
3	engronline.ee.memphis.edu	gov
3	odur.let.rug.nl	edu
3	orgwis.gmd.de	org
3	cmpe.boun.edu.tr	edu
3	cordis.lu	gov
3	cs.uwa.edu.au	edu
3	ilrt.bris.ac.uk	edu
3	uni-hildesheim.de	edu
3	xml.coverpages.org	org

Inlinks as well as outlinks strongly reflect the intellectual environment of SWI. This supports the suggestion that the research group is the most appropriate level of analysis in studying linking patterns; the communications are sufficiently aggregated to exclude the individual preferences, but the communication is sufficiently codified to provide a meaningful indicator for the intellectual environment in which the research group operates. As could be expected, incoming links are more 'application oriented' than outgoing links; indicating the context of application of the knowledge produced at SWI. Outlinks, on the other hand, provide us with an indicator for intellectual and technical resources that are used in the production of knowledge.

Table 7. *SWI home outlinks*

URL	Frequency	Type	'Service'
swi.psy.uva.nl	9503	Edu	university
hypermail.org	2677	Org	archive
gene.wins.uva.nl	1010	Edu	university
aifb.uni-karlsruhe.de	602	Edu	university
ksi.cpsc.ucalgary.ca	321	Edu	university
cs.vu.nl	310	Edu	university
uva.nl	254	Edu	university
carol.wins.uva.nl	251	Edu	university
psy.uva.nl	215	Edu	university
kmi.open.ac.uk	214	Edu	university
eit.com	204	Com	technical support
turing.wins.uva.nl	191	edu	university
ida.liu.se	168	edu	university
iiia.csic.es	123	org	university
cse.unsw.edu.au	108	edu	university
landfield.com	104	com	archive
Home,netscape.com	93	com	software/provider
cs.cmu.edu	84	edu	university
chakotay.ist.org	84	org	university
w3.org	84	org	software
ngi.nl	84	org	platform
delicias.dia.fi.upm.es	83	edu	University
dds.nl	81	org	Internet provider
iit.demokritos.gr	76	edu	University
Cordis.lu	70	Gov	
wins.uva.nl	70	edu	University
arti.vub.ac.be	68	edu	university
hotmail.com	64	com	email provider
geocities.com	62	com	Internet provider
csd.abdn.ac.uk	60	edu	University
sis.port.ac.uk	60	edu	University
java.sun.com	59	com	Software

Table 8*.Clustering of institutes on basis of linking structure.*

URL	Cluster
swi.psy.uva.nl	1
psy.uva.nl	1
kmi.open.ac.uk	1
saussure.irmkant.rm.cnr.it	2
rm.cnr.it	2
ksl.stanford.edu	2
cs.cmu.edu	3
arti.vub.ac.be	3
ics.uci.edu	4
macs.mines.edu	4
informatik.uni-kl.de	5
nathan.gmd.de	5
W3.ORG	6
ACM.ORG	6
smi.stanford.edu	6

In the next step, we constructed the matrix of inlinks and outlinks between all institutes that link with SWI above the usual threshold. Factor analyzing this matrix is expected to reveal clustering of sites with the same position within the linking network. As far as links represent the intellectual focus of a institute, the factors represent clusters of institutes with a similar research and/or service orientation. The whole factor structure is expected to reveal the 'intellectual' environment of the department (table 8). Again, this methodology is very similar to journal-journal citation analysis in scientometrics.

The factor analysis reveals highly pronounced clusters of institutes. The clusters represent scientific specialties relevant to the research conducted at SWI. Factor 1 represents the general SWI environment, and factor 2 is mainly related to ontology based knowledge management. The third factor represents computer sciences groups, active in the field of AI.

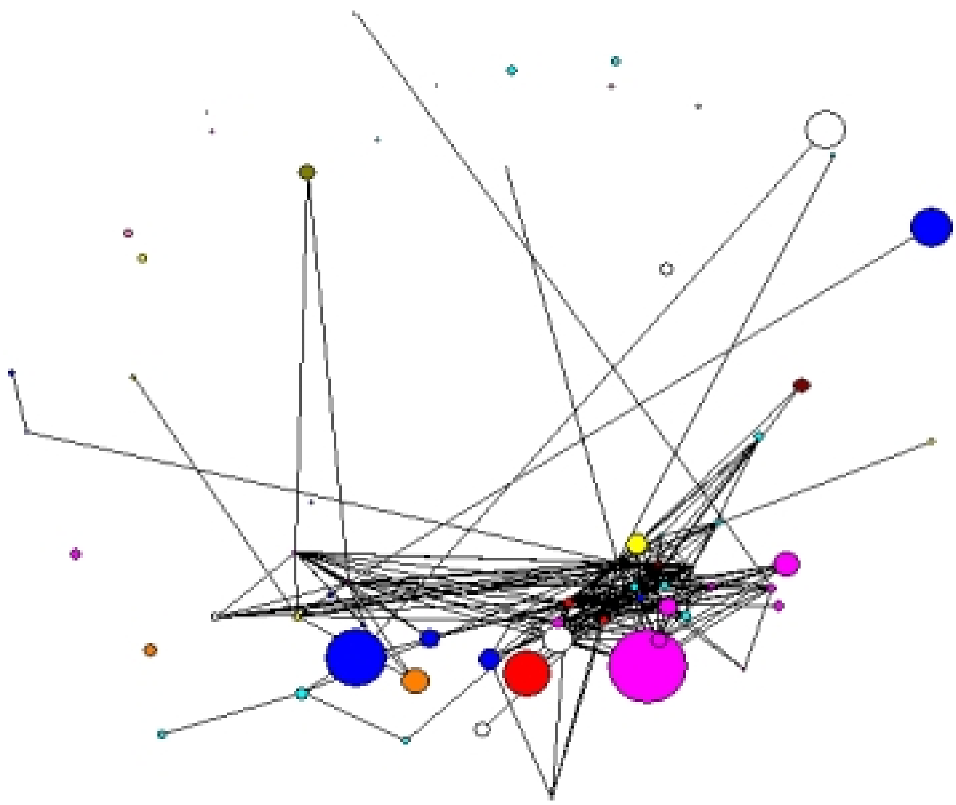


Figure 4*. Network representation of the SWI environment. (ARCS)*

The color code: France (red), UK (blue), Germany (yellow), Spain (orange), Italy (light blue), Greece (pink), SE (light green), NL (mauve) and Austria (dark red).

The figure 4 is a first effort to visualize the data. Again we have not been able to experiment with the visualization. Here the environment of SWI is plotted based on a distinction between the home country of the institutes. Possible alternatives for visualizing are specialization (as represented by factor membership), etc.

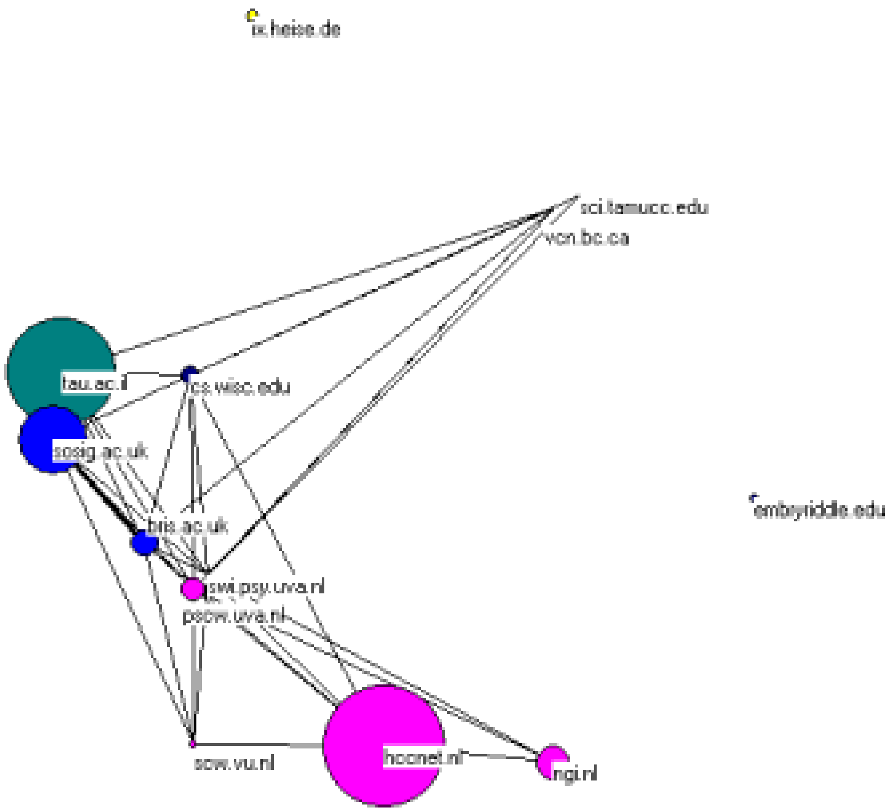


***Figure 5**. Network representation of individual researchers environment. (ARCS)*

The color code: France (red), UK (blue), Germany (yellow), Spain (orange), Italy (light blue), Greece (pink), SE (light green), NL (mauve) and Austria (dark red).

2.3.6 Linking structure at micro level; the individual researcher

A similar methodology was applied to the individual linking structure of an in-residence researcher at SWI. Factor analysis of the matrix of links between all sites that are linked by the researcher's site and all sites that link to this individual researcher's site fails to show any structure. The rotated factor solution does not cluster any websites into a joint factor. In other words, the analysis indicates that at this level of analysis, links represent the heterogeneous profile of individual interests.

Figure 5 above is a first effort to visualize the individual link network of links, again based on country of origin of the linking sites.

2.4 Concluding remarks

Links and link patterns on the web can be used for generating indicators of web presence, web content and web impact. The questions to be answered in order to derive useful indicators are various: Which level of web links are meaningful? What do they mean? What are the meaningful indicators? In this chapter we analyzed the link patterns on different levels. The results of this analysis can be summarized as follows:

- On the highest level of analysis of countries, the emerging structures of hyperlinks between universities provide a measure of European integration in the science system. The integration seems rather geographical proximity and language based, especially in the inlinks dimension. A dynamic analysis may reveal to which extent the Europeanization of knowledge production is occurring. The data we have presented here do not indicate this very much – given the relatively strong dominance of the traditional forces of integration: proximity and language.

- Focusing on the universities in one country with respect to the rest of EU provides us with a measure of internationalization and mutual perception of universities. At this level, the thematic orientation seems to play a rather strong role; the large, general universities have similar positions in the network, as do the technical institutes have. The same holds for the business schools.

- Within the Netherlands, the linking structure between universities provides a measure of intellectual orientation. At this level of analysis, linking has a predominant 'cognitive' motivation. The more specific the focus of the institute, the more pronounced the clustering; art schools together, institutes for teacher training, etc. In other words, the position of similar institutes are also similar in the web link structure.

- At the level of the research group, analyzing the linking architectures provides us with in depth information about the context of knowledge production. On this level of analysis web link analysis bears most resemblance with traditional print media based analysis. Constructing a matrix of linking institutes seems to be comparable to journal-journal citation analysis. Citations are a substantive, but composite unit of analysis: they contain information that may reflect novelty, codification, and evaluation in different respects. However, the aggregated journal-journal citation system here under study is at a level of abstraction such that one does not have to open the "black box" of each citation in order to examine the information which is communicated, but that one analyzes only the number of messages. The same seems to apply to linking patterns of research institutes. The communication are sufficiently codified within a research context to provide a meaningful overview of the context in which knowledge production takes place. It is also a sufficiently aggregated system at a level of abstraction such that one does not have to open the "black box" of each link in order to examine the information which is communicated, but that one analyzes only the number of messages

- The linking structure of an individual researcher provides us with in indicator of personal interests at a micro-sociological level.

In general – and apart from the research group level – it can be concluded that the epistemological value of links is different from cited references in printed scientific papers. Links have a more heterogeneous character than references, which results in more informal and personal link-communications. In some sense, it might be useful to compare links with co-author relations or project co-operations, in stead of citations. Links can have a mutual character (unlike cited references) and often have a more ‘personal’ dimension because of the low codified environment of cyberspace. It can be said, that this poses some methodological problems in delineating an appropriate data set for analysis. As mentioned before, links are not limited to the research domain (unlike cited references) but also included personal, applicational and other dimensions like information resources, and services. Also some doubts have been raised about the reliability of search engine data gathering. These issues will be further discussed in the case study of chapter 5.

Appendix : ***nl2alleu***

Institute	factor																
tudelft.nl	1	0.92	-0.11	0.14	0.21												
unimaas.nl	1	0.89		0.23	0.25					0.12							
RUU.NL	1	0.87	0.29	0.14		0.18		-0.14	0.13	0.13							
KUB.NL	1	0.86	0.14				0.26			-0.14							
RUG.NL	1	0.85	0.16	0.30				0.11								-0.15	
UVA.NL	1	0.83	0.26	-0.29	0.11			0.21									
ISS.NL	1	0.83	0.20		0.12		0.18								0.14		
utwente.nl	1	0.80	0.51						0.15								
WAU.NL	1	0.80	0.32	0.35	0.15			0.14	-0.10								
EUR.NL	1	0.80	0.41		0.22		0.15		0.16	0.11							
KUN.NL	1	0.79	0.34	0.23	0.23			0.20	0.21							0.11	
VU.NL	1	0.77	0.33	0.44													
HHS.NL	1	0.74	0.53						0.13	-0.13					0.11	-0.12	
hsbrabant.nl	1	0.73	0.22	-0.26	-0.27			-0.21		0.13						-0.22	
hshaarlem.nl	1	0.73	0.27	0.52					0.13								
HEN.NL	1	0.72		0.13	0.57		-0.10								-0.12		
HVA.NL	1	0.70	0.28	0.50	0.28				-0.11								
HVU.NL	1	0.69	0.23	0.23	0.22			0.49							0.17		
HKU.NL	1	0.67	0.17	0.32			-0.15	0.40		0.14		0.12				-0.18	
nijenrode.nl	1	0.61	0.13	-0.19			0.29	0.54		-0.15						-0.12	
hzeeland.nl	1	0.61	0.24		-0.15			0.11	0.52	-0.27	-0.10						
STOAS.NL	1	0.52		0.14				0.19	-0.13	0.12			0.49	0.26	0.16	-0.11	
leidenuniv.nl	2	0.40	0.88	0.12													
hsholland.nl	2	0.27	0.84			0.21				-0.12							
thrijswijk.nl	2	0.33	0.75	0.13	0.11				-0.12					0.23			
CHW.NL	2	0.24	0.72	0.61													
EDITH.NL	2	0.30	0.70	0.24	0.45						0.21						0.11
CHE.NL	2	0.46	0.67	-0.21				0.26	-0.13	0.22			0.20		0.14		
HRO.NL	2	0.56	0.63	-0.20	0.21	0.10			0.14	0.18						0.17	
HSA.NL	2	0.48	0.61	0.12				0.24	0.26			0.26		-0.11			0.31
designacademy.nl	3	0.26		0.95													
AHK.NL	3	0.25		0.91													
VHALL.NL	3	0.48		0.51				0.12	-0.15	0.31			0.16	0.31	0.27	0.11	
OUH.NL	4	0.48	0.15		0.71		0.24			-0.19		0.15					
TUE.NL	4	0.59	0.31		0.67		0.11		-0.13								
NHL.NL	4	0.55	0.32		0.62	0.16			0.16				0.11				
iselinge.nl	4	0.12	0.11	0.12	0.61	0.47		0.21	0.12	-0.14	0.33			0.17		-0.11	
dehorst.nl	5					0.94											
HSIJ.NL	5	0.20	0.17		0.10	0.88							0.16				
HAN.NL	5	0.37		-0.15		0.54		-0.34	-0.29	0.17	0.12		-0.19	0.13	-0.17		
HANZE.NL	5				0.30	0.48			0.28	0.30			0.16		0.17	0.43	
rulimburg.nl	6	0.13					0.95										
hsmaastricht.nl	6						0.94					0.18					
fontys.nl	7	0.46	0.10					0.55	0.11	-0.14			-0.11	0.11			
HSBOS.NL	8	0.31		-0.10					0.76	0.19							
hes-rdam.nl	8	0.11			0.21			0.13	-0.14	-0.72							
psimons.box.nl	8	0.42	0.13		0.19	0.18	-0.15	0.23	-0.21	0.42		0.39		-0.14		-0.14	

Appendix (continued)

kempel.nl	9										0.77			-0.13					
domstad.nl	9		0.22	0.14	0.39			0.16	0.21		0.70			0.10					
HSE.NL	10						0.13				-0.10	0.80				0.14			
hlo-ct.nl	10	-0.18				0.31			0.19	-0.11		0.53	0.41						
HDH.NL	11												0.75		-0.17				
hsdelft.nl	11	0.34	0.19		0.26			-0.13		-0.21		0.27	0.36	0.27	0.21	-0.16	-0.15	0.11	
CHN.NL	12	-0.11												0.87					
CAH.NL	13												-0.15		0.83				
driestar.nl	14											0.11			-0.11	0.77			
HMTR.NL	15																0.95		
ichthus-rdam.nl	16																	-0.87	0.13
gerritrietveldacademie.nl	17										-0.11	-0.11						0.14	-0.87
sandberg.nl	18	-0.10								0.13	-0.32	-0.28		-0.21	-0.25	-0.16	-0.20	0.43	0.46

Chapter 3. Triple Helix on the web

The knowledge infrastructure of the New Economy is provided by a "Triple Helix" network of University - Industry - Government relations. In this case study, we explored the opportunities of using hyperlinks to map the Triple Helix of industrial, governmental and academic institutions on a national scale in The Netherlands.

In modern knowledge intensive economies and societies, relations are maintained between research institutions and industrial, governmental, and non-governmental organizations. This is of course the rationale behind concepts as the ‘network economy’ and the ‘new economy’. This idea of network economies is sometimes described in terms of the ‘Triple Helix’ of University - Industry - Government relations. These relations do exist on various levels, and in this chapter we focus on the question whether the ‘Triple Helix’ is also visible on the level of WWW-links. A first exercise of describing the knowledge infrastructure on the Web is carried out by extending the universities linking network with major governmental, commercial en non-governmental sites. What additional information does this provide? We expected that these links show more of the application and policy context of scientific knowledge production. For this analysis a group of Dutch URL’s was selected that contained all Dutch universities, all Dutch ministries, a selection of 25 of the most important non-governmental organizations and a list of about the 200 largest companies with a stock exchange quotation.

Table 1. Frequency of links

	INSTITUTE	NR of LINKS	TYPE
1	tudelft.nl	72712	edu
2	kun.nl	36254	edu
3	uva.nl	34774	edu
4	utwente.nl	31633	edu
5	rug.nl	26919	edu
6	Vu.nl	26014	edu
7	eur.nl	20858	edu
8	leidenuniv.nl	20711	edu
9	tue.nl	20304	edu
10	worldonline.nl	17745	com
11	kub.nl	13718	edu
12	telegraaf.nl	11887	com
13	wau.nl	8558	edu
14	tno.nl	8336	org
15	ruu.nl	6583	edu
16	philips.com	5432	com
17	antenna.nl	5176	com
18	unimaas.nl	3792	edu
19	hvu.nl	3337	edu
20	hva.nl	2618	edu
21	hhs.nl	2426	edu
22	knaw.nl	2390	edu
23	Minvenw.nl	1954	gov
24	Hanze.nl	1803	edu

The first finding is that universities are responsible for the largest amount of links (table 1 above). All Dutch universities are in the top 24 of the institutions with substantial linking behavior, while only 4 companies occur in this list, one ministry and one non-governmental organization. Of the four companies, two are Internet Access Providers, and one is a publisher of newspapers. The fourth is the electronics manufacturer Philips. The non-governmental organizations is actually the large national research facility TNO, and therefore close to the institutions of academic research.

The second set is to analyze the *positional* structure of the weblink network. In table 2, we show the nine strongest factors that result from factor analyzing the matrix of the about 250 institutions selected (see above). Although most of the relevant clusters are dominated by academic institutions, a number of heterogeneous factors are found too: clusters 6, 7 and 9 combine various types of institutions. Partly, the structure seems based on regional proximity (factors 3, 4, 5, 6, 7, 8). Partly, the positional clustering of institutions seems understandable. The various ministries (factor 5) seem to have similar positions in the link network, and the same holds for the media companies (factor 2). On the other hand, it is not clear why the various institutions of higher education exhibit this positional clustering. As linking behavior on the level of a whole university is the aggregate of lower level behaviors, the analysis suggests that smaller units of analysis (e.g. research groups within the universities) might provide us with more insightful information about 'Triple Helix interactions'.

Table 2*. Most significant clusters of institutes by factor analysis of the linking matrix.*

INSTITUTE	TYPE	CLUSTER
OUH.NL	EDU	1
utwente.nl	EDU	1
telegraaf.nl	COM	2
Worldonline.nl	COM	2
wegener.nl	COM	2
UVA.NL	EDU	3
VU.NL	EDU	3
NWO.NL	ORG	3
HRO.NL	EDU	3
tudelft.nl	EDU	4
leidenuniv.nl	EDU	4
minvenw.nl	GOV	5
MINEZ.NL	GOV	5
minfin.nl	GOV	5
nijenrode.nl	EDU	6
RUU.NL	EDU	6
elsevier.nl	COM	6
ICT.NL	COM	7
TUE.NL	EDU	7
philips.com	COM	7
RUG.NL	EDU	8
HANZE.NL	EDU	8
minocw.nl	GOV	9
HVU.NL	EDU	9

The following conclusions can be drawn:

1. Contrary to common belief the universities dominate the linking structure of Dutch cyberspace. Often it has been said that cyberspace is increasingly dominated by commercial institutions. We were able to show that universities are the main actors in establishing linking relationships on the Internet.
2. This analysis confirmed our hypothesis that more information about the complex dynamics of knowledge production and new economy can be obtained by using lower level units of analysis; e.g. research departments, departments at ministries.
3. Within the heterogeneous environment of the Internet, it is important to carefully delineate a sector of activity in order to obtain meaningful insights. The exploratory analysis described above combines a too wide variety of institutions without any common aims and fields of activity. Therefore, the emerging clusters are difficult (or impossible) to interpret.

Chapter 4. The role of electronic communications in science – a case study

This study aims at mapping all relevant scholarly communications with respect to a research group operating in an application-oriented techno-scientific environment. Scholarly communication refers to the formal and informal processes by which the research results are created, evaluated, edited, formatted, distributed, organized, made accessible, archived, used, and transformed. This study explores the opportunity of using inlinks, outlinks, log-files, incoming emails, outgoing emails, project co-operations and co-authored publications of academic, political and commercial institutes as indicators for knowledge production: what is communicated in different media? Can we identify the various roles of computer-mediated communications in relation to print and other traditional media? Do the communication patterns provide information about the changing nature of knowledge production?

4.1 Introduction

The Internet is a good example of a socio-technical medium where social networks are developing and interfering with computer networks. Organizations, institutes and persons are increasingly communicating through the Internet, and present on the World Wide Web. This may influence the shape and the dynamics of the social networks organizations are part of. Over the last few years, the Internet and related computer mediated communication technologies have become increasingly important to scientific work. As a consequence of the information revolution in society in general, and science in particular, new information and communication patterns are emerging. Information has always been central to scientific research, but the introduction of digital information, online databases and ICT have enabled a radical lowering of costs related to information dissemination, both in pure form and black boxed in technologies. This informational turn in science is associated with a new mode of knowledge production with hybrid roles of academic, commercial and governmental institutes, the transformation of economic mechanisms by technology and innovation, and a wider variety of types of research output (e.g., tools, patents, norms and values). Consequently, the processes and operations that maintain and reproduce the science system are changing. In our research we wish to understand and measure the role of electronic communications in science.

What are meaningful indicators for Internet based inter-organizational relations? What is the relationship of those indicators with traditional indicators for networks between organizations? Does the presence of the Internet and the World Wide Web inform us about inter-organizational networks 'in real life', or are the real and virtual communication networks relatively loosely coupled? Do WWW and Internet based indicators teach us something about the changes of the inter-organizational networks?

In this chapter we will address some of these questions. In addition we will study to which extend the various hypotheses about the changing nature of knowledge production can be traced in the communication patterns of the research group.

It might be useful two distinguish between two perspectives on the interaction between ICT and knowledge production: the social and the technological. The technological dimension of knowledge production is associated with the increased use of large (online) databases, increased storage capacity, increased accessibility of data and a wider variety of communication channels. The social developments in relation to ICT and science are related to the systemic properties of the scientific communications; how do the changing local operations affect the global structure of the scientific domain? Are the boundaries of the scientific domain changing? How will electronic information affect the organization of knowledge and the social basis of its production and dissemination? For example, might the combination of electronic information resources with interdisciplinary and multicultural scholarship affect the formal organization of knowledge?

As Langston points out, some assumptions underlying these issues relate to divergent views of the Internet as it relates to scholarly production of knowledge. An optimistic view claims that new technology (as epitomized by graphical, easy-to-use browsers for the world wide web, coupled with sophisticated software that can perform citation analysis and intelligently search through vast amounts of information) will transform scholarly publication for the better by allowing people to quickly seek out information, respond to others, publish electronically at a low cost, and ultimately speed up the typically long cycle of publishing a peer-reviewed article.[5]

In contrast, a less optimistic view points out that scholars do not have time to perform the duties of a publishing house, whether electronic or not (Press 1995, Rowland 1995). Structures and scenarios might price or legislate information beyond the reach of all but a few individuals or institutions, thus decreasing access to information ("Ask Dr. Internet" 1995), or if not so dire, demand subscription fees to access an article database thus firewalling articles (Brent 1995), or not insuring privacy through anonymous access to files (Hickey 1995). Some have noted that the library as it's currently constituted might be bypassed or killed, thus leading to unemployment or drastic restructuring of the jobs for librarians (Cisler 1995, Kling and Lamb 1994). Others note physical harm from the effects of computer technology, a lack of concern for the environment resulting from the manufacturing technology of computer hardware and the increased use of paper (Fuller 1995b). The future reality, of course, will compass most of those viewpoints, at least to a certain extent, (as expressed by more moderate views, such as those of Fuller generally [1995a, b] and Grusin 1994).

As mentioned, the key concept of this chapter is the idea that the enterprise of science is best captured by the concept of an evolving communication system. Evolving

---

[5] See, for example, Harnad (1995a,b); Lanham (1993); Odlyzko (1995); Schwier (1994); Stodolsky (1995); and Taubes (1996a,b).

communication systems are not given or fixed, they are the result of a continuing process during which more complex forms of organization emerge. This undirected evolution is resulting from ongoing interaction between system and environment. This concept not only provides a valuable theoretical framework for scientometric analysis, it allows provides an opportunity to elaborate on the consequence of the informational turn in science. The organization of the communication patterns defines a domain of reflexive interactions in which it can act with relevance to the maintenance of itself (Maturana, 1980). Processes of feedback and of self-reference are vital in maintaining a social system. This insight allows us to articulate some of the operations that construct the boundaries of the science system like peer reviews (Wouters, 1999) and editorial boards of scientific journals (Fujigaki, 1998). As Wouters points out, scientific communication can also be communicated reflexively, by measuring scientometric attributes of the communications, which in turn can be communicated. In terms of the information cycle representation, the indicators appear in the form of a second cycle. This cycle processes information about the primary information cycle, in the form of meta-information (Wouters, 1999). Together, these cyclic communication patterns result in a system that can be described as a self-maintaining evolving communication system with well definable boundaries and well definable boundary constructing processes.

The science system is an example of a self-organizing system. Many investigators have observed social activities and structures, particularly in the science system, that are best described by a power law distribution. A power law is one of the common signatures of a non-linear dynamic process, i.e., a chaotic process, which is at a point of self-organized criticality (Bak, 1991) or residing on the boundary between order and disorder. Such a system is often called a self-organized system because it exhibits structure not merely in response to inputs from the outside but also, indeed primarily, in response to its own internal processes (Krugman, 1996). This perspective has implications for the way the science system can be studied quantitatively by means of indicators.

4.2 Data and Methods

This study aims at mapping the network of networks of scientific communications of a research group in order to gain insight in the relation between the various networks of communications. The goal of this study is twofold, firstly we aim at developing and validating Internet based indicators, and secondly we wish to gain insight in the role of electronic communications in the development of science and technology.

In order to gain more insight in the role of electronic communications in relation to 'traditional' communications this case study was carried out. The single actor approach adopted here means that a variety communications were mapped coming from, or going to- one central actor; the SWI department at the University of Amsterdam (www.swi.psy.uva.nl). The group was selected because it is a typical example of mode 2 research in the information society; emphasis on application, participation in European, transnational projects, and a long tradition in Internet oriented computer science.

(Furthermore, being a member of SWI, the accessibility of data was an argument in favor of choosing this particular research group). The selected traditional communications included co-authored publications in peer reviewed international journals, project co operations as well as a questionnaire to asses the perception of the intellectual environment of researchers in this department. The selected electronic communications include incoming cyberlinks to the SWI homepage, incoming links to the project page (www.swi.psy.uva.nl/projects.html), outlinks from the SWI domain, the incoming emails and outgoing emails.

Information about the recent project co-operations between the SWI department and other research-groups, companies and NGO's could be obtained from the online SWI project pages at www.swi.psy.uva.nl/projects.html. The co-authored publications could be found in the annual reports of the faculty of psychology at the University of Amsterdam. The questionnaire that was handed out to asses the researchers' perception on their intellectual environment (appendix 1) was based on a selection of institutes that were most occurring in the publications, projects, links and emails.

Webometric studies have used large-scale Web search engines such as AltaVista. These allow measurements to be made of the total number of pages in a domain (either a top level domain such as .com or .nz, or a lower level domain such as vuw.ac.nz) or a set of directories, such as the pages in the directories and subdirectories and the links between them. Web search engines provide similar possibilities for the investigation of links between documents to those provided by the citation databases created by the Institute for Scientific Information. We have measured the number of links between SWI and other institutes by using the AltaVista search engine. Incoming links were obtained using the advanced search options of Google. It was unfortunately impossible to use AltaVista since this search engines has a limitation on the number of returned results. On the other hand, Google could not be used to obtain the number of outgoing links of a domain, since it is a page-based search engine that is not working on the higher domain level.

The data on incoming and outgoing email communications were obtained by using data from the log-files of the SWI mail server. For reasons of privacy protection all the information before the @-sign was removed. The content of the messages was not logged in these files. In this way, a frequency distribution of the most occurring incoming and outgoing mail addresses could be obtained.

4.3 Findings

The incoming links to the project pages of the SWI department were expected to reflect the heterogeneous environment of application-oriented knowledge production. Inlinks are expected to reflect the users of information produced at SWI, while the outlink pattern is expected to show the resources for producing knowledge at the SWI department. The results (table 1) show that universities are dominant in the linking environment of the SWI department. However, a relatively large amount of non-academic organizations has a substantial linking relationship with the project page of the SWI domain. Interesting is the presence of cordis.lu, the organization of European research projects.

Table 1 *Inlinks to the project-page of SWI.*

Number of links	URL	type
353	swi.psy.uva.nl	edu
16	cs.vu.nl	edu
13	aiai.ed.ac.uk	edu
8	cs.toronto.edu	edu
6	cs.utexas.edu	edu
5	www-users.cs.umn.edu	edu
4	id.cbs.dk	org
4	Iiia.csic.es	org
3	courses.cs.vt.edu	edu
3	engronline.ee.memphis.edu	edu
3	odur.let.rug.nl	edu
3	Saussure.irmkant.rm.cnr.it	org
3	cmpe.boun.edu.tr	edu
3	cordis.lu	gov
3	cs.uwa.edu.au	edu

The number of the institutes linking to the SWI *home page* (table 2) is larger than the list of institutions linking to the *project page*. Of the seven additional institutions, five are educational. This underlines the fairly academic nature of the inlinks.

Table 2. *Inlinks to the SWI homepage.*

Number of links	URL	type
443	swi.psy.uva.nl	edu
22	cs.vu.nl	edu
13	aiai.ed.ac.uk	edu
8	cs.toronto.edu	edu
7	cs.utexas.edu	edu
5	cs.umn.edu	edu
5	irit.fr	edu
4	saussure.irmkant.rm.cnr.it	org
4	cs.wpi.edu	edu
4	id.cbs.dk	edu
4	iiia.csic.es	org
4	macs.mines.edu	edu
3	courses.cs.vt.edu	edu
3	engronline.ee.memphis.edu	gov
3	odur.let.rug.nl	edu
3	orgwis.gmd.de	org
3	cmpe.boun.edu.tr	edu
3	cordis.lu	gov
3	cs.uwa.edu.au	edu
3	ilrt.bris.ac.uk	edu
3	uni-hildesheim.de	edu
3	xml.coverpages.org	org

In general, sites are positioned in relation to each other. This provides a starting point for analysis. The meaning and significance of sitation/citation is expected to be quite different in the two environments of scholarly publishing and the WWW. If we assume that the WWW is a prototype of the distributed digital library of the future, it would be helpful to know if the tools and techniques developed for the analysis of intellectual structure in paper-based libraries will be able to make the transition to this network-based environment (Wilensky, 1995). Counts of ingoing and outgoing links can be seen as citation and reference analysis respectively. However, due to its dynamic and distributed nature, the Web often demonstrates web pages simultaneously linking to each other- a case not possible in the traditional paper based citation world (Björneborn & Ingwersen 2001). The outlink structure (table 3) is expected to be more resource oriented than the inlink structure. The results support this hypothesis, and next to many links to university sites (resources of knowledge), links to archives, to technical support and to software providers are among the most frequently occurring.

Table 3 *SWI home outlinks*

URL	Frequency	Type	service
swi.psy.uva.nl	9503	Edu	university
hypermail.org	2677	Org	archive
gene.wins.uva.nl	1010	edu	university
aifb.uni-karlsruhe.de	602	edu	university
ksi.cpsc.ucalgary.ca	321	edu	university
cs.vu.nl	310	edu	university
uva.nl	254	edu	university
carol.wins.uva.nl	251	edu	university
psy.uva.nl	215	edu	university
kmi.open.ac.uk	214	edu	university
eit.com	204	com	technical support
turing.wins.uva.nl	191	edu	university
ida.liu.se	168	edu	university
iiia.csic.es	123	org	university
cse.unsw.edu.au	108	edu	university
landfield.com	104	com	archive
home,netscape.com	93	com	software/provider
cs.cmu.edu	84	edu	university
chakotay.ist.org	84	org	university
W3.org	84	org	software
ngi.nl	84	org	platform
delicias.dia.fi.upm.es	83	edu	university
dds.nl	81	org	Internet provider
Iit.demokritos.gr	76	edu	university
cordis.lu	70	gov	university
wins.uva.nl	70	edu	university
arti.vub.ac.be	68	edu	university
hotmail.com	64	com	email provider
geocities.com	62	com	Internet provider
csd.abdn.ac.uk	60	edu	university
sis.port.ac.uk	60	edu	university
java.sun.com	59	com	software

Inlinks as well as outlinks strongly reflect the intellectual environment of SWI. This supports the suggestion that the research group is the most appropriate level of analysis in studying linking patterns; the communications are sufficiently aggregated to exclude the individual preferences, but the communication is sufficiently codified to provide a meaningful indicator for the intellectual environment in which the research group operates. As could be expected, incoming links are more 'application oriented' than outgoing links; indicating the context of application of the knowledge produced at SWI. Outlinks, on the other hand, provide us with an indicator for intellectual and technical resources that are used in the production of knowledge.

However doubts were raised about the extent to which these outlinks resulted from individual actions. The number of outlinks is too large to be explained by individual linking of a research group of around 30 people. It was determined that many of the links were automatically generate by e-mail archives (hypermail.org) and software producers. The outlinking pattern of the homepages section was determined to see if this result differed significantly (table 4).

Table 4. *SWI outlinks from 'homepages'.*

URL	frequency
swi.psy.uva.nl	623
aifb.uni-karlsruhe.de	119
psy.uva.nl	77
ksi.cpsc.ucalgary.ca	62
uva.nl	53
home.netscape.com	46
delicias.dia.fi.upm.es	46
dds.nl	39
iiia.csic.es	39
cs.vu.nl	31
cbl.leeds.ac.uk	30
kmi.open.ac.uk	29
cordis.lu	26
geocities.com	23
info.psy.uva.nl	21
cs.cmu.edu	20
swint1.swi.psy.uva.nl	20
w3.org	19
lri.jur.uva.nl	19
yahoo.com	18
home.pscw.uva.nl	17
inria.fr	15
`isi.edu	15
aaai.org	15

The results of the outlink structure of the homepage section of the SWI domain seems to give a better representation of the aggregated individual linking structure than the previous table. The mail archives are no longer present among the highest ranking institutes. However, the number of outlinks still remains large, further research is necessary.

The e-mail communications provide us with a permanent, archived record of informal and formal communications of science in the making. In this study, we focused on the relational aspects of e-mail communications, i.e. the senders and receivers of messages on aggregated level. The content of the messages was not taken into account. The results (table 5) clearly show the informal character of e-mail communications. There are many non-academic e-mail providers that rank among top 30 of most occurring receivers of SWI mails; hotmail, chello, yahoo, CompuServe, xs4all, etc. This indicates that e-mail communications are not restricted to the academic domain.

Table 5. *Mail from SWI*

Frequency	URL
513	hotmail.com
310	wins.uva.nl
281	psy.uva.nl
262	chello.nl
244	cs.vu.nl
233	Yahoo.com
198	science.uva.nl
189	Compuserve.com
188	Xs4all.nl
183	gmx.net
179	acm.org
178	csd.abdn.ac.uk
177	cs.com
175	ecs.soton.ac.uk
173	teknowledge.com
173	club-Internet.fr
172	zdnetonebox.com
172	ptolemy.arc.nasa.gov
172	netvision.net.il
172	ipisun.jpte.hu
172	In.tum.de
172	imag.fr
172	Ig.com.br
172	dcs.shef.ac.uk
172	daimi.aau.dk
172	cs.ust.hk
172	cs.mu.oz.au
172	bbn.com
172	Arrakis.es

The list of incoming mails to SWI is more academic in nature than the list of outgoing mails (table 6). One of the reasons for this is the presence of a number of mailing list addresses. This makes the list of incoming e-mails a good indicator for the knowledge resources of SWI.

Table 6. *Mail to SWI*

`	URL
1607	returns.onelist.com
1067	suse.com
1010	hotmail.com
887	cs.uu.nl
799	macmail.psy.uva.nl
742	science.uva.nl
703	Yahoo.com
536	SWI
398	nic.surfnet.nl
374	list.uvm.edu
350	sharon.cselt.it
340	hyperion.math.upatras.gr
317	Mail.gmd.de
254	Lists.xcf.berkeley.edu
251	listserv.aol.com
244	psy.uva.nl
239	cs.vu.nl
223	nieuwsbrievenservice.nl
194	Edte.utwente.nl
187	mpi.nl
186	vcn.bc.ca
184	Excite.com
174	Chello.nl
172	eos.dk
159	w3.org
150	Scribe.uwaterloo.ca
138	ncmgroup.com
135	xs4all.nl
130	cs.uah.edu
128	cwi.nl
123	Acm.org

Table 7. *Co-authored publications in international peer-reviewed journals*

URL	Number of articles
few.eur.nl	1
isoco.com	1
jur.uva.nl	1
cs.utwente.nl	1
psychology.nottingham.ac.uk	1
arti.vub.ac.be	1
smi.stanford.edu	1
iiia.csic.es	1
cs.vu.nl	1
cis.ohio-state.edu	1
psy.uva.nl	1
cs.kun.nl	1
ISI.EDU	1
cs.rug.nl	1
cs.umd.edu	1

The role and function of these electronic communications in the process of knowledge production can be further elaborated by comparison to non-electronic communications. A list was obtained of all institutes that participated in research projects together with the SWI department (appendix 2). Furthermore, a list was generated of all institutes that co-authored scientific articles with SWI in international, peer-reviewed journals (table 7).

Additional information about the intellectual focus of the SWI research group was obtained by a questionnaire that was handed out to the in-residence researchers (see appendix 1). With a response rate of 80%, the researchers indicated how close they felt to the institutes mentioned on the list. The list was composed of a selection of institutes that were prominently present in the linking structure, email communications, project co-operations or publication of the SWI researchers. The list is divided in two parts; academic institutions (table 8) and non-academic institutions (table 9).

Table 8. *Results of the Questionaire.*

iiia.csic.es	19
cs.vu.nl	18
kmi.open.ac.uk	16
cs.rug.nl	15
wins.uva.nl	15
cs.cmu.edu	14
aifb.uni-karlsruhe.de	13
psychology.nottingham.ac.uk	13
smi.stanford.edu	13
cs.utwente.nl	11
aiai.ed.ac.uk	10
soc.surrey.ac.uk	10
cs.utexas.edu	9
csd.abdn.ac.uk	9
arti.vub.ac.be	8
cwi.nl	7
cis.ohio-state.edu	6
ksi.cpsc.ucalgary.ca	6
hyperion.math.upatras.gr	5
cs.kun.nl	4
dcs.shef.ac.uk	4
luiss.it	4
ISI.EDU	3
cse.unsw.edu.au	2
few.eur.nl	2
ida.liu.se	2
imag.fr	2
cs.mu.oz.au	1
cs.toronto.edu	1
cs.umd.edu	1
irit.fr	1

Table 9. *Results of the Questionnaire on non-academic institutions.*

isoco.com	14
cordis.lu	12
acm.org	8
W3.org	7
dds.nl	4
teknowledge.com	2
ngi.nl	1
eit.com	0
bbn.com	0

In order to further explore the opportunity of using electronic communications as indicators of scientific activity, we adopted an approach suitable for more quantitative evaluation. First we selected a list of institutes that represents the core set of relevant producers and users of knowledge in the SWI environment (table 10). Exploratory analysis could be performed to determine the correlation between the different communications.

Table 10 . *Overview of most relevant institutions.*

URL	outlinks from user	mail to SWI	publications	questionaire	home inlinks	mail from SWIlinks	projects	projects-inlinks	mail from SWI
acm.org	6	179	0	10	0	12	0	0	123
aiai.ed.ac.uk	9	0	0	14	13	48	0	13	0
aifb.uni-karlsruhe.de	119	43	0	17	0	602	1	0	0
arti.vub.ac.be	8	3	1	12	0	68	0	0	0
bbn.com	0	172	0	0	0	0	0	0	0
cbl.leeds.ac.uk	30	41	0	0	0	55	0	0	7
cica.es	0	171	0	0	0	0	0	0	0
cis.ohio-state.edu	9	0	1	6	0	18	0	0	0
cmpe.boun.edu.tr	0	0	0	0	3	0	0	3	0
cogs.susx.ac.uk	12	0	0	0	0	50	0	0	0
commonkads.uva.nl	7	0	0	0	0	56	0	0	0
cordis.lu	26	0	0	14	3	70	0	3	2
cs.cmu.edu	20	0	0	18	0	84	0	0	0
cs.kun.nl	0	0	1	4	0	0	0	0	0
cs.mu.oz.au	1	172	0	1	0	6	0	0	4
cs.rug.nl	0	5	1	19	0	0	0	0	0
cs.ruu.nl	6	39	0	0	0	42	1	0	0
cs.toronto.edu	0	134	0	3	8	0	0	8	0
cs.umd.edu	7	0	1	1	0	18	0	0	1
cs.ust.hk	0	172	0	0	0	0	0	0	0
cs.utexas.edu	12	39	0	9	7	30	0	6	3
cs.utwente.nl	0	150	1	13	0	6	1	0	20
cs.uwa.edu.au	0	0	0	0	3	0	0	3	0

cs.vu.nl	31	244	1	22	22	310	1	16	239
cs.wpi.edu	1	0	0	0	4	8	0	0	0
csd.abdn.ac.uk	13	178	0	11	0	60	0	0	7
cse.unsw.edu.au	10	39	0	2	0	108	0	0	0
cwi.nl	7	169	0	9	0	22	0	0	128
daimi.aau.dk	2	172	0	0	0	4	0	0	0
dcs.shef.ac.uk	1	172	0	4	0	6	0	0	0
dds.nl	39	62	0	4	0	81	0	0	35
delicias.dia.fi.upm.es	46	39	0	0	0	83	0	0	0
ecs.soton.ac.uk	1	175	0	0	0	2	0	0	3
eit.com	9	0	0	0	0	204	0	0	0
engronline.ee.memphis.edu	0	0	0	0	3	0	0	3	0
few.eur.nl	0	10	1	2	0	0	0	0	51
ftp.aifb.uni-karlsruhe.de	0	0	0	0	0	43	0	0	0
hyperion.math.upatras.gr	0	13	0	5	0	0	0	0	340
ics.uci.edu	4	40	0	0	0	40	0	0	0
id.cbs.dk	0	0	0	0	4	0	0	4	0
ida.liu.se	2	39	0	2	0	168	0	0	0
ifi.unizh.ch	1	44	0	0	0	2	0	0	30
iiia.csic.es	39	17	1	20	4	123	1	4	43
imag.fr	0	172	0	2	0	0	0	0	3
inria.fr	15	39	0	0	0	54	0	0	2
irit.fr	1	39	0	6	5	2	0	0	6
isi.edu	15	39	1	7	0	48	0	0	13
isoco.com	1	148	1	14	0	4	1	0	28
jur.uva.nl	0	112	1	0	0	0	0	0	9
kmi.open.ac.uk	29	0	0	20	0	214	1	0	0
ksi.cpsc.ucalgary.ca	62	0	0	10	0	321	0	0	0
ksl.stanford.edu	1	0	0	0	0	10	0	0	2
luiss.it	0	0	0	4	0	0	1	0	0
nathan.gmd.de	10	0	0	0	0	44	0	0	0
ngi.nl	9	0	0	1	0	84	0	0	0
odur.let.rug.nl	0	0	0	0	3	2	0	3	0
Orgwis.gmd.de	2	0	0	0	3	1	0	0	0
psy.uva.nl	77	281	1	0	0	215	0	0	244
psychology.nottingham.ac.uk	0	3	1	16	0	12	0	0	1
rm.cnr.it	0	0	0	4	0	0	0	0	0
saussure.irmkant.rm.cnr.it	0	0	0	0	4	3	0	3	0
sis.port.ac.uk	4	0	0	0	0	60	0	0	0
smi.stanford.edu	2	8	1	17	0	10	0	0	28
soc.surrey.ac.uk	2	0	0	10	0	8	1	0	0
teknowledge.com	5	173	0	3	0	18	0	0	9
uni-hildesheim.de	0	0	0	0	3	0	0	0	0
w3.org	19	133	0	14	0	84	0	0	159
wins.uva.nl	3	310	0	22	0	70	0	0	33

Visualization of the relative importance of each communication channel provides us with an indication of the relationship between the different channels (figure 1). We intend to use the EICSTES visualization tools for further analysis of the data.

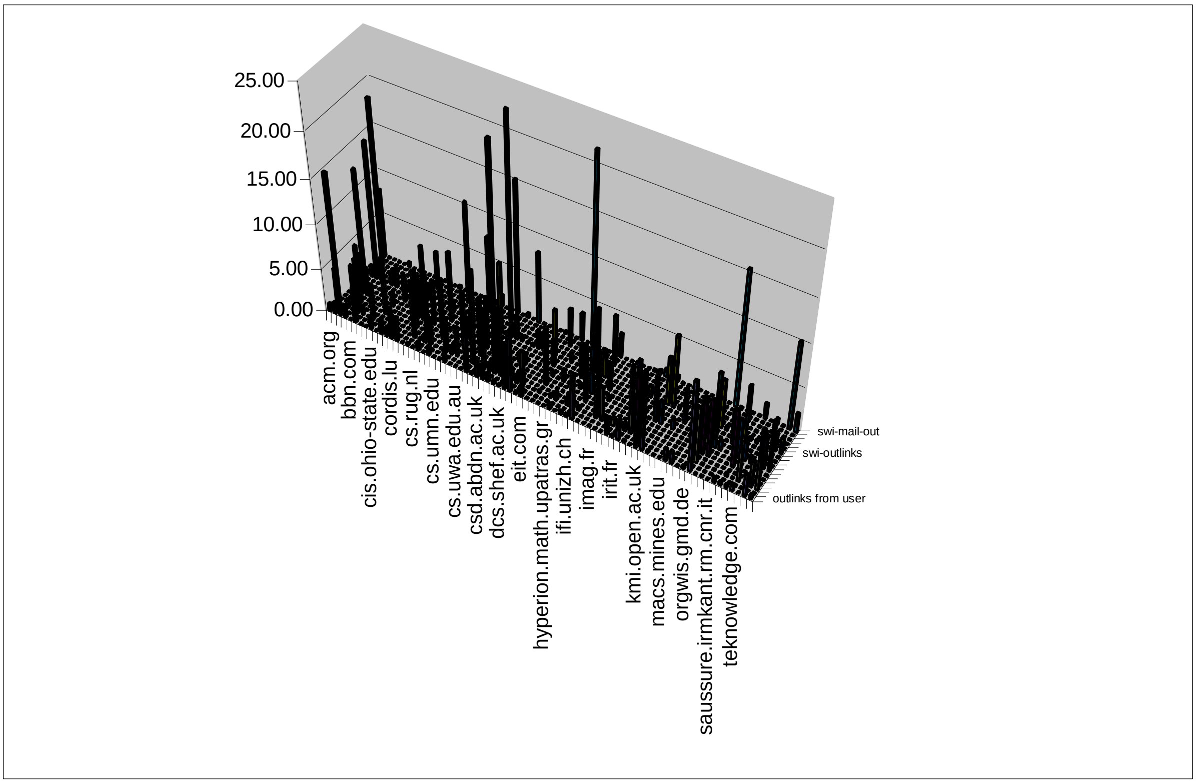


***Figure 1**. Representation of the relative importance of different communications with the most important partners.*

Correlation analysis might reveal the amount of overlap and complementarities between the different communication channels (table 11) in order to gain insight in the function of links and e-mails in the process of knowledge production.

The highest correlation exists between the inlinks to the project page and the inlinks to the home page. The interest of outsiders for the knowledge produced at SWI is strongly reflected in the interest for projects in which SWI participated. In general, the outgoing communications have stronger correlations with other outgoing communications, publications and project communications than incoming communications. Especially the questionnaire is informative in this respect because arguably this is the best indicator for defining the SWI environment.

The Kendall's tau correlation gives an even more pronounced representation; the questionnaire results have high correlations with outlinks, publications and project co operations. This is an encouraging result with respect to the validity of the data, although further research remains necessary.

Table 11 . *Results of the correlation analysis (Pearson) of the most relevant institutions.*

N=68		outlinks from user	mail to swi	publications	questionaire	home inlinks	mail from swilinks	projects	projects-inlinks	mail from swi
outlinks from user	Pearson Correlation	1.00	0.07	0.05	0.30	0.03	0.86	0.28	0.05	0.21
	Sig. (2-tailed)	.	0.57	0.70	0.01	0.81	0.00	0.02	0.71	0.08
MAIL TO SWI	Pearson Correlation	0.07	1.00	0.04	0.14	0.07	0.04	0.05	0.07	0.37
	Sig. (2-tailed)	0.57	.	0.74	0.26	0.58	0.72	0.71	0.60	0.00
Publications	Pearson Correlation	0.05	0.04	1.00	0.37	0.06	0.01	0.21	0.06	0.19
	Sig. (2-tailed)	0.70	0.74	.	0.00	0.63	0.96	0.08	0.63	0.12
Questionnaire	Pearson Correlation	0.30	0.14	0.37	1.00	0.25	0.37	0.45	0.27	0.21
	Sig. (2-tailed)	0.01	0.26	0.00	.	0.04	0.00	0.00	0.03	0.08
home inlinks	Pearson Correlation	0.03	0.07	0.06	0.25	1.00	0.15	0.18	0.95	0.23
	Sig. (2-tailed)	0.81	0.58	0.63	0.04	.	0.21	0.15	0.00	0.06
mail from SWIlinks	Pearson Correlation	0.86	0.04	0.01	0.37	0.15	1.00	0.37	0.15	0.18
	Sig. (2-tailed)	0.00	0.72	0.96	0.00	0.21	.	0.00	0.23	0.15
PROJECTS	Pearson Correlation	0.28	0.05	0.21	0.45	0.18	0.37	1.00	0.17	0.09
	Sig. (2-tailed)	0.02	0.71	0.08	0.00	0.15	0.00	.	0.17	0.49
projects-inlinks	Pearson Correlation	0.05	0.07	0.06	0.27	0.95	0.15	0.17	1.00	0.20
	Sig. (2-tailed)	0.71	0.60	0.63	0.03	0.00	0.23	0.17	.	0.10
MAIL FROM SWI	Pearson Correlation	0.21	0.37	0.19	0.21	0.23	0.18	0.09	0.20	1.00
	Sig. (2-tailed)	0.08	0.00	0.12	0.08	0.06	0.15	0.49	0.10	.

Table 12 . *Results of the correlation analysis (Kendall's tau), most relevant institutions.*

Kendall's tau_b		**outlinks from user**	**mail to swi**	**publications**	**questionaire**	**home inlinks**	**mail from swilinks**	**projects**	**projects-inlinks**	**mail from swi**
outlinks from user	Correlation Coefficient	1.00	0.07	-0.01	0.24	-0.11	0.76	0.13	-0.07	0.17
	Sig. (2-tailed)	.	0.42	0.93	0.01	0.29	0.00	0.22	0.52	0.08
MAIL TO SWI	Correlation Coefficient	0.07	1.00	0.07	0.12	-0.21	0.03	0.05	-0.16	0.46
	Sig. (2-tailed)	0.42	.	0.51	0.21	0.05	0.76	0.65	0.13	0.00
publications	Correlation Coefficient	-0.01	0.07	1.00	0.33	-0.10	-0.01	0.21	-0.04	0.32
	Sig. (2-tailed)	0.93	0.51	.	0.00	0.39	0.95	0.08	0.73	0.00
questionaire	Correlation Coefficient	0.24	0.12	0.33	1.00	-0.02	0.25	0.34	0.04	0.29
	Sig. (2-tailed)	0.01	0.21	0.00	.	0.85	0.01	0.00	0.68	0.00
home inlinks	Correlation Coefficient	-0.11	-0.21	-0.10	-0.02	1.00	-0.15	0.01	0.83	-0.09
	Sig. (2-tailed)	0.29	0.05	0.39	0.85	.	0.14	0.91	0.00	0.40
mail from SWIlinks	Correlation Coefficient	0.76	0.03	-0.01	0.25	-0.15	1.00	0.15	-0.08	0.10
	Sig. (2-tailed)	0.00	0.76	0.95	0.01	0.14	.	0.15	0.40	0.30
PROJECTS	Correlation Coefficient	0.13	0.05	0.21	0.34	0.01	0.15	1.00	0.07	0.06
	Sig. (2-tailed)	0.22	0.65	0.08	0.00	0.91	0.15	.	0.58	0.57
projects-inlinks	Correlation Coefficient	-0.07	-0.16	-0.04	0.04	0.83	-0.08	0.07	1.00	-0.06
	Sig. (2-tailed)	0.52	0.13	0.73	0.68	0.00	0.40	0.58	.	0.59
MAIL FROM SWI	Correlation Coefficient	0.17	0.46	0.32	0.29	-0.09	0.10	0.06	-0.06	1.00
	Sig. (2-tailed)	0.08	0.00	0.00	0.00	0.40	0.30	0.57	0.59	.

4.4 Conclusions

We started this study with the question whether we could identify the role of computer-mediated scholarly communications in relation to print and other traditional media? Scholarly communication refers to the formal and informal processes by which the research and scholarship of researchers are created, evaluated, edited, formatted, distributed, organized, made accessible, archived, used, and transformed. This study explored the opportunity of using inlinks, outlinks, incoming emails, outgoing emails, project co-operations and co-authored publications of academic, political and commercial institutes as indicators for knowledge production: what is communicated in different media? In summary:

- Inlinks provide us with interesting information about the academic environment of the knowledge production as well as the users context (academic and non-academic).

- Outlinks also provide us with information about the academic environment of the knowledge production as well as the knowledge resources and technical resources that are relevant to SWI.

- Incoming emails provide us with a permanent, archived record of informal and formal communications of science in the making. In this study, we focused on the relational aspects of e-mail communications, i.e. the senders and receivers of messages on aggregated level. Clearly, e-mails are more informal than other indicators under study. However, it provides us with valuable information about the knowledge resources (e.g. mailing lists)

- Outgoing emails provide us with an indicator for the personal network in which science takes place.

The next question was whether we could identify the role of computer-mediated communications in relation to print and other traditional media? We answered this question by identifying the research environment of the research group SWI. This environment was defined in terms of communication patterns. This approach seems useful; we were able to identify over 60 research institutes that played an important role in different media. Correlation analysis further confirms our hypothesis; the outgoing communications have relatively strong correlations.

It seems that the communication patterns can provide information about the changing nature of knowledge production. We were able to identify a heterogeneous set of relevant institutes (governmental, industrial and academic) that represent the users and producers context of mode 2 knowledge production. Interesting in this context is the presence of archives, databases, software producers and governmental institutions in the communicational environment of SWI. Striking however is the complete absence of publication in e-journals (that do not have a print version).

Methodological Reflections

However, link analysis remains problematic for two reasons:

- It has often been reported that search engines only cover part of the Internet. Consequently, it is likely that generating link-data using search engines is not 100% reliable (this problem will be discussed in detail elsewhere in this project).
- We found out that not all links are generated by choice. Many (out-)links that were found within the SWI domain were automatically generated. For example, each message in the email archives that are hosted by the research group contains an automatically generated link to the company that provides the infrastructure for archiving (e.g. hypermail.org). Other examples include software tools that were used in producing (online) software application often included a link-references to the company. Probably, this problem is particularly present in the SWI domain with a large number of software applications and e-mail archives. However, it poses yet another methodological problem in using search engines for generating link data.

5 Conclusions and Further Research

From the case studies carried out so far, it could be concluded that that the electronic and print communication patterns can provide information about the changing nature of knowledge production. We were able to identify a heterogeneous set of relevant institutes (governmental, industrial and academic) that represent the users and producers context of mode 2 knowledge production (archives, databases, software producers and governmental institutions). We were able to determine empirically what is the most appropriate level of analysis for mapping techno-scientific developments in the information society. It was concluded that at the level of the research group, in depth information can be obtained by linking architectures about the context of knowledge production. This level of analysis bears most resemblance with traditional print based analysis. Constructing a matrix of linking institutes seems to be comparable to journal-journal citation analysis. The communications are sufficiently codified within a scientific context to provide a meaningful overview of the context in which knowledge production takes place. It is also a sufficiently aggregated system at a level of abstraction such that one does not have to open the "black box" of each link in order to examine the information which is communicated. Web-pages can be compared to articles in journals.

Our case studies showed that the delineation of a domain is of vital importance. Search engines proofed to be unreliable to obtain a linking network that can serve as a starting point for delineation and analysis. For the next stage of the work in WP5, a tool will be developed that can obtain reliable linking information as well as content information by crawling the web. We expect that links should be used to delineate a domain, and content should be analyzed within this domain. Our thesis is that science is best described as an evolving communication system. Systems cannot be defined by simply enumerating or tracing the layout of their constituent elements. The definitive attribute of a systemic entity is the set of inter-component relationships which (a) outline its form at any given moment and (b) serve as the core 'identity' which is maintained in spite of dynamic changes over time. In autopoietic theory, this set of defining relationships is termed a system's organization. It is the organization of a system which defines its identity, its properties as a unity, and the frame within which it must be addressed as a unary whole. The methodological consequence of this insight is, that meaningful domains need to be delineated on the basis of their relationships.

In effect, a system's organization specifies a category, within which there may be many specifically-realized instantiations. Specific systemic entities exhibit more than just the general pattern of their organization - they consist of particular components and relations among them. A systemic unity's organization is specifically realized through the presence and interplay of components in a given space. These comprise the unity's structure. These structural properties of a communication system can be studied by looking at the content of the communications. A unity may change structure without loss of identity, so long as its organization is maintained. The distinction between organization and structure provides a basis for sorting out descriptions of systems into their abstract and concrete aspects.

Scientometric methodologies rest on the assumption that traces of communicated information can be used to map the development of the science system in interaction with its (social, economical, political and technological) environment. These communications can be studied in three different dimensions. The structure of the communication networks can be mapped by quantifying the number and direction of the existing relations in the network. What is communicated (the organization) can be represented (to a certain extent) by mapping the content of these communications (title-words, key-words, etc). The role and behavior of the actors (persons and institutes) that together form a system can be studied as a third dimension.

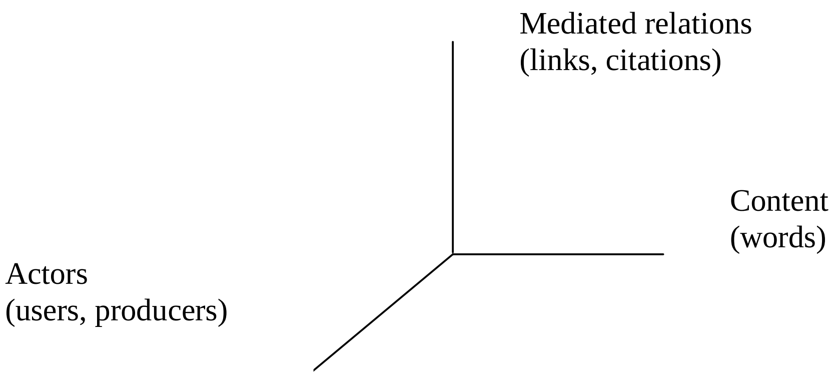


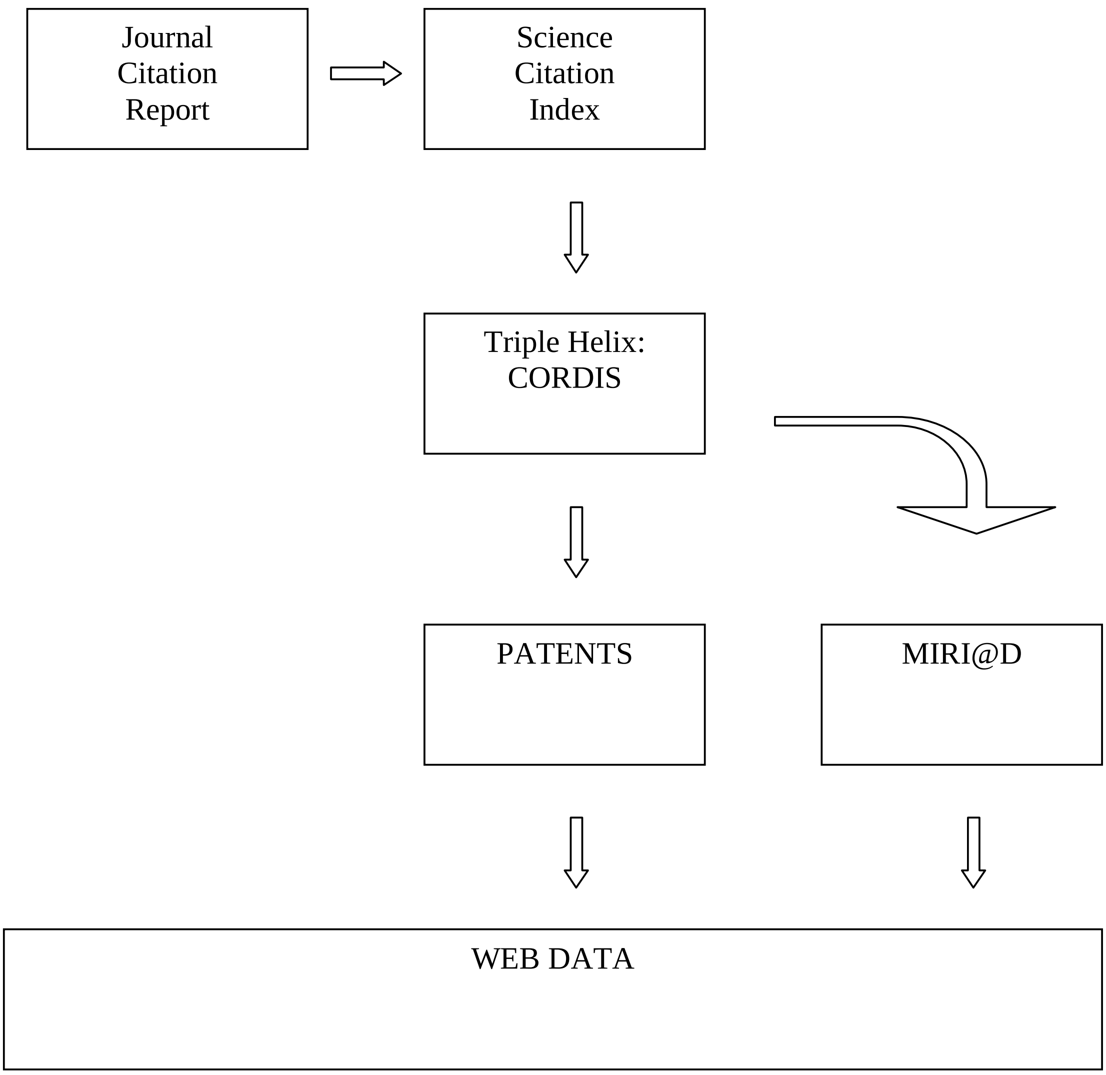
***Figure 1**. The various methodological dimensions of scientometric and webometric analysis.*

The actors-axis represents the dimensions of the relevant actors involved; the type of institution (governmental, industrial, academic), the geographical location (NL, UK, USA, etc), the role of these institutions in the process of knowledge production and new economy (users, producers, suppliers, commissioner).

The relations-axis represents the type (cited reference, hyperlink, project co-operation) and quantity of relations between actors.

The content-axis represents the 'cognitive' dimensions of communications (tile-words, free-text words) and makes it possible to position the various actors in the communication network. If we are interested by knowledge indicators, in the sense of empirically representing knowledge that is contained in the documents, we need to explore these *semantic networks* and *concept graphs*.

Design



1. SCI Delineation Analysis

The first aspect relates to the body of knowledge that can be called 'the mapping of science and technology'. Crucial is here the delineation of the various fields, as this creates the starting point for the analysis of network dynamics. Using the core journal in the field as an entrance point, the citational environment of this journal is defined as all journals that cite or are cited by this journal above a threshold percentage. Factor analysis reveals clearly identifiable clusters of journals representing scientific disciplines. By repeating this procedure using data from previous years, information is obtained about the recent development of the field and its environment.
The non-web data for studying techno-scientific evolving communication networks are provided by the Science Citation Index (SCI). We are interested in the relationship of these print based networks with electronic networks of scientific communications. The data for the fields Biotechnology, Artificial Intelligence and Information Science have been collected for the years 1986, 1992 and 1996. We will add the data for the year 2000 to this set.

In case of biotechnology, factor analysis of the citation matrix for 1996 resulted in 7 separate clusters (scientific specialties). Analysis of the citational environment of *Biotechnology and Bioengineering* in 1992 and 1986 informed us about the development of the discipline biotechnology with respect to content and size. It seems that biotechnology matured as a scientific discipline in the late eighties and early nineties. The structure of the biotechnology factor has been is relatively stable during the eighties and nineties, with a stable core set of journals. Also the environment of biotechnology has remained relatively stable over the years, with microbiology, chemical engineering, biochemistry and water research always present in the citational environment.

The structure of the information science factor has been is relatively stable during the eighties and nineties, and its core is always consisting of *JASIS, Information Processing and Management, Proceedings of the American Society for Information Science* and *Journal of Documentation.* Other journals like the *Annual Review of Information Science and Technology* are in the factor in several years, and also the *Journal of Information Science* is in most years in the core information science factor. S*cientometrics* is sometimes in it, and sometimes in another factor. It seems that two information science factors in the 1996 set actually represent a geographical boundary between the US (the core IS factor journals) and Europe (*Journal of Information Science* & *Scientometrics.*

The transregion co-operation of scientific institutes and the geographical distribution of research topics can then be studied (fig.2). The SCI data will provide with a well defined set of institutes, topics and research co-operations.

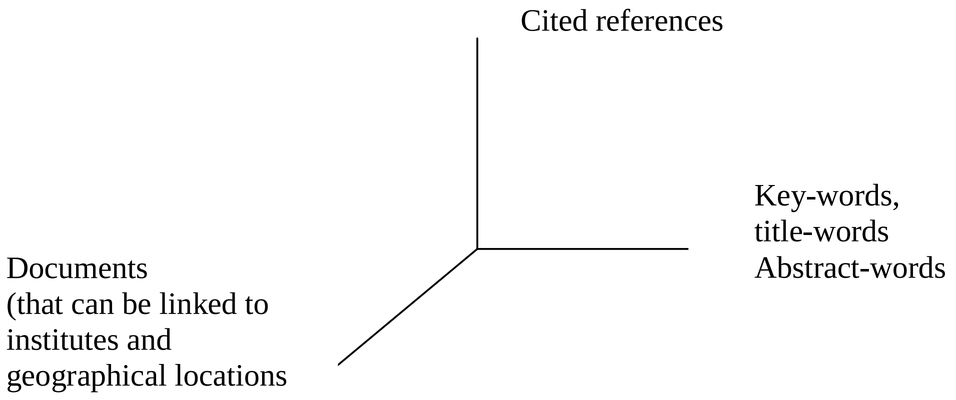


Figure 2. *The dimensions of scientometric analysis of SCI data*.

2. CORDIS

The knowledge infrastructure of the information society is provided by a "Triple Helix" network of University - Industry - Government relations. The CORDIS database of EU projects provides a wealth of data of co-operations between academic, industrial and governmental partners. The list of institutes from the SCI delineation could provide a good starting point. Alternatively, keywords could be used to select a representative set of partners.

There are two ways in which we can link the SCI data to the CORDIS data;

By identifying a core set of (key-)words in the delineated domains of the SCI data. This set of words can be used to map the relevant domain in the CORDIS database.
By identifying a list of institutes and/or authors in the delineated domain of the SCI data. The relevant domain in the CORDIS database is the defined as the list of (commercial, governmental and research) institutes that have participated in projects with those institutes with a publication in the SCI domain.

Our hypothesis suggests that the second method will be more fruitful, but it might be interesting to test this empirically. Generally, the first approach is used in webometrics. Researchers can match pages to an existing topic that they have defined by analyzing the text on the pages. This method is limited because the topic must be known in advance, it is very difficult to retrieve all pages in a timely manner, and because the method depends on the specific words used (e.g., pages in a different language or using different words for the same concept may not be identified). The network of relations provides a more robust way of identifying a community in which a set of concepts and ideas is shared by the actors.

The list of institutes can than be used to study the geographical distribution of co-operations and research topics.

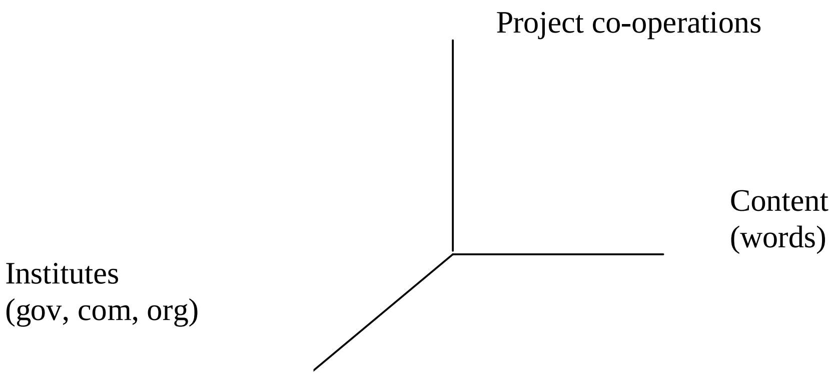


Figure 3*. The various dimensions of scientometric analysis of CORDIS data.*

3. PATENTS

The list companies and research institutes obtained from the CORDIS database could be used to map the technological output of the research programs in the form of patents.

4. MIRI@D Data

The MIRI@D database provides us with a wealth of data on the network of users and producers of scientific knowledge. The set of institutes delineated in the CORDIS and SCI databases can be used as a starting point to determine which institutes requested the documents produced by this core set of institutes. This will provide us with detailed information about the set of relevant users of the knowledge produced in the European triple helix networks in information science, AI and biotechnology.

Information about the producers of relevant knowledge that plays a role in the process of knowledge production within the European triple helix networks can also be obtained from the MIRI@D database.

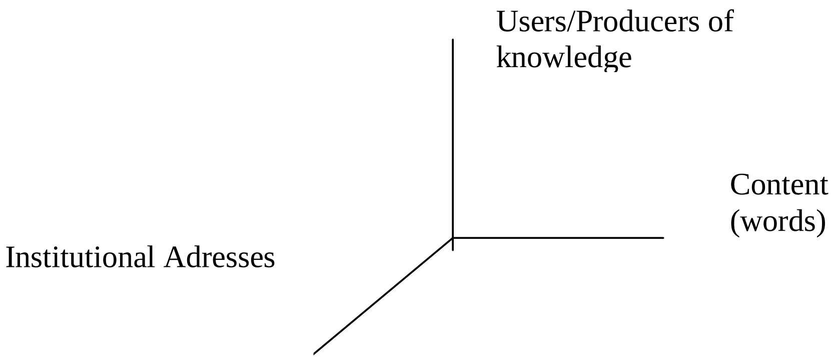


Figure 4. *The dimensions of the MIRIAD analysis*.

5. Web Data

The domains of interlinked universities, companies and governmental organizations provides a starting-point for link and content analysis on the web.

We argued that that a relevant domain is best defined in terms of a set of communicating actors. The analysis of non-web data provides us with a set of actors that have real-life relations. The first step in our analysis is to see if these interactions are represented on the internet. Again, the starting point of the analysis is the list of domains obtained from the CORDIS database. It might be necessary to increase the set of relevant actors by doing a (co-)link analysis starting with the list of CORDIS institutes.

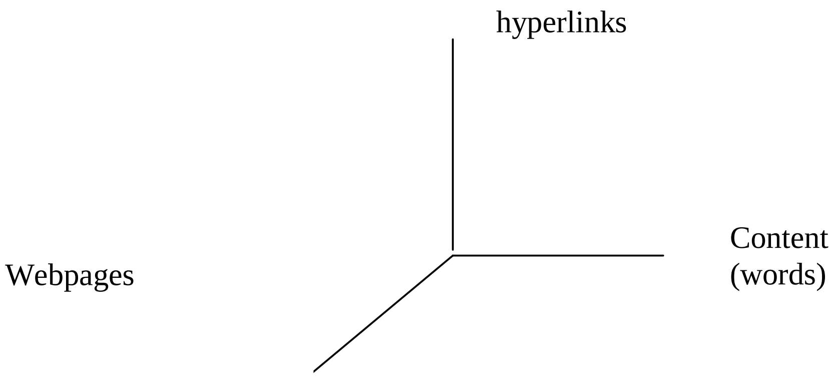


Figure 5. *The dimensions of webometric analysis*.

Two practical issues are important here:

- How do we obtain the linking pattern of the selected set of institutes?
- How do we perform content analysis?

In the next stage, it might be interesting to identify clusters of linked (and linking) pages that do not have a clear equivalent in real world communications: debates on the boundary of scientific, technological and political domains (e.g., genfood). Perhaps a more qualitative approach to content analysis is more fruitful here.

Appendix 1a	Is this institute part of your research environment?	Do you have a BOOKMARK to this domain?	Do you know someone in this institute?	Does someone of this institute belong to your collaboration network?	Do you e-mail with someone in this institute?	
aiai.ed.ac.uk						
aifb.uni-karlsruhe.de						
arti.vub.ac.be						
cica.es						
cis.ohio-state.edu						
cs.cmu.edu						
cs.kun.nl						
cs.mu.oz.au						
cs.rug.nl						
cs.toronto.edu						
cs.umd.edu						
cs.ust.hk						
cs.utexas.edu						
cs.utwente.nl						
cs.vu.nl						
cs.wpi.edu						
csd.abdn.ac.uk						
cse.unsw.edu.au						
cwi.nl						
daimi.aau.dk						
dcs.shef.ac.uk						
ecs.soton.ac.uk						
few.eur.nl						
hyperion.math.upatras.gr						
id.cbs.dk						
ida.liu.se						
ifi.unizh.ch						
iiia.csic.es						
imag.fr						
in.tum.de						
irit.fr						
ISI.EDU						
kmi.open.ac.uk						
ksi.cpsc.ucalgary.ca						
luiss.it						
psychology.nottingham.ac.uk						
rm.cnr.it						
smi.stanford.edu						
soc.surrey.ac.uk						
wins.uva.nl						

Appendix 1b	Is this a service supplier for you?	Do you supply information/services to this institute?	Do you have a BOOKMARK to this domain?	Do you know someone in this institute?	Does someone of this institute belong to your collaboration network?	Do you e-mail with someone in this institute?
acm.org						
bbn.com						
cordis.lu						
dds.nl						
eit.com						
isoco.com						
ngi.nl						
teknowledge.com						
w3.org						

Appendix 2.
Project co-operations

aie.lreg.co.uk	iscn.ie
aifb.uni-karlsruhe.de	isl uk
aifb.uni-karlsruhe.de	isl.co.uk
AMC.uva.nl	Isoco.com
arcs.ac.at	kcl.ac.uk
asadsigns.com.au	kits.edte.utwente.nl
ato.dlo.nl	kls.nl
bfz-essen.de	kmi.open.ac.uk
capsogeti.fr	labein.es
cibit.nl	lhflighttraining.com
cih.hcuge.ch	lr.org
cindoc.csic.es	luiss.it
cinop.nl	Math.upatras.gr
cit.ie	mcu.es
ckb.fgg.eur.nl	met.police.uk
cs.ruu.nl	mitgmbh.de
cs.utwente.nl	Mlnet.org
cs.vu.nl	nec.ac.uk
csi.forth.gr	netinfo.fr
cti.gr	olivetti.com
dcs.napier.ac.uk	physik.hu-berlin.de
dg13.cec.be	regtek.sintef.no
dircon.co.uk	rpk.mach.uni-karlsruhe.de
Eads.net	Sago.it
eclo.org	sics.se
ecn.nl	siemens.com
ecn.nl	smi-web.stanford.edu
Entel.com	soc.surrey.ac.uk
fou.tr.statoil.no	sth.cap.se
ftknowledge.com	SWI uva nl
horizon-sfa.ch	telematica.co.uk
iberdrola.es	tno.it
ibm.fr	tpi.at
idescat.es	ulb.ac.be
iiia.csic.es	uni-bielefeld.deiwt
inist.fr	uv.es
inst-informatica.pt	weiterbildung.unizh.ch

References

Aguillo, I. F. (1998). STM Information on the Web and the development of new Internet R&D databases and indicators. Proc. Online 1998. Learned Information, London

Almind, Tomas C.,and Peter Ingwersen, Informetric analyses on the World Wide Web: A methodological approach to "webometrics", Journal of Documentation, September 1997

Björneborn, Lennart and Peter Ingwersen. Perspectives of webometrics. Scientometrics, 2001, 50(1), 65-82.

Boudourides, Moses A.; Sigrist, Beatrice & Alevizos, Philippos D.(1999). "Webometrics and the self-organization of the European Information Society". Draft Report presented during the Rome Meeting of the SOEIS project. June 17-19, 1999. <http://hyperion.math.upatras.gr/webometrics>

Braam, R. R., Moed, H. F. & van Raan, A. F. J. (1991). Mapping of science by combined co-citation and word analysis. II: Dynamical aspects. Journal of the American Society for Information Science 42(4): 252-266.

Brin, S. and L. Page (1998), the anatomy of large-scale hypertextual Web search engines, WWW7 conference. (http:/www-db.stanford.edu/~backrub/google.html)

Burt, Burt, R.S., *Towards a structural theory of action*. London: Academic Press, 1982.

Callon, Michel (1986a). The Sociology of an Actor Network: The Case of the Electric Vehicle .

Callon, Michel, J. Law, A.Rip, : Mapping the Dynamics of Science and Technology. Sociology of Science in the Real World , 19-34. Macmillan, London.

Cozzens, S. E.. What do citations count? The rhetoricfirst model. Scientometrics, 15:437--447, 1989.

Cozzens, Susan E. and Loet Leydesdorff, Journal Systems as Macro-Indicators of Structural Change in the Sciences, in: A. F. J. Van Raan, R. E. de Bruin, H. F. Moed, A. J. Nederhof and R.

Fujigaki, Yuko & Loet Leydesdorff, "Quality Control and Validation Boundaries in a Triple Helix of University-Industry-Government Relations: ‘Mode 2’ and the Future of University Research," Social Science Information 39(4) (2000) 635-655.

Fujigaki, Yuko (1998) Filling the gap between discussions on science and scientists' everyday activities: applying the autopoiesis system theory to scientific knowledge. Social Science Informatics 37(1), pp 5-22

Fuller, Steve. 1995a. "Cybermaterialism, or Why There is no Free Lunch in Cyberspace." Information Society 11(4): 325-332.

Fuller, Steve. 1995b. "Cyberplatonism: An Inadequate Constitution for the Republic of Science." Information Society. 11(4): 293-303.

Gibbons, M.C., H. Limoges, S. Nowotny, P.S. Schwartzman and M. Trow (1994): The New Production of Knowledge, London, Sage.

Gibson, D., J. Kleinberg, P. Raghavan (1998). Inferring webcommunities from link topology, Proceedings of the 9th ACM Conference on Hypertext and hypermedia. (http:/www.cs.cornell.edu/home/kleinber/ht98.pdf)

Ginsparg, P. "Winners and Losers in the Global Research Village." [http://xxx.lanl.gov/blurb/pg96unesco.html] (1996).

Gläser, J. (2001) Scientific Specialties as the (Currently Missing) Link between Scientometrics and the Sociology of Science. Proceedings ISSI 2001 Australia.

Grusin, Richard. 1994. "What is an Electronic Author? Theory and the Technological Fallacy." Configurations 2(3): 469-483.

Hales, N. Katherine. 1993. "Virtual Bodies and Flickering Signifiers." October 66: 69-91.

Harnad, Stevan. 1995a. "The Post Gutenberg Galaxy: How to Get There from Here." Information Society. 11(4): 285-291.

Harnad, Stevan. 1995b. "Sorting the Esoterica from the Exoterica: There's Plenty of Room in Cyberspace." Information Society. 11(4): 305-324.

Heimeriks, Van den Besselaar, Boudourides, Polanco, Aguillo, ARCS (2001), State of the art in webometrics and bibliometrics. EICSTES deliverable D1.4, part 2.

Helstien, B.A. 1995. "Libraries: Once and Future." Electronic Library 13(3): 203-207.

Hickey, Thomas. 1995. "Present and Future Capabilities of the Online Journal." Library Trends. 43(4): 528-543.

Hitchcock, Steve. "A Survey of STM Online Journals 1990- 1995:The Calm Before the Storm." [http://journals.ecs.soton.ac.uk/survey/survey.html]. 15 January 1996, updated 14 February.

Hulme EW (1923) Stat Biblio Relation

Katz, J. S., The self-similar science system, Research Policy 28 (1999) 501-517

Katz, J.S. and D.Hicks,(1997) Bibliometric Indicators for National Systems of Innovation. Available http://www.sussex.ac.uk/Users/sylvank/best/nsi/index.html.

Kleinberg, J. M., 1998, Authorative sources in a hyperlinked environment, Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms, pp.668-677.

Kling, Rob; Covi, Lisa. 1995. "Electronic Journals and Legitimate Media in the Systems of Scholarly Communication." Information Society. 11(4): 261-271.

Kling, Rob; Lamb, Roberta. 1994. "Envisioning Electronic Publishing and Digital Libraries: How Genres of Analysis Shape the Character of Alternative visions." [http://asearch.mccmedia.com/public/libraries/genres] (1994). To appear in: Academia and Electronic Publishing: Confronting the Year 2000. Ed Robin P. Peek, Gregory

Newby, Lois Lunin. Cambridge, MA: MIT Press. [Note: Broken link removed 4/10/98 by ald]

Knoke, D., & J.H Kuklinsky, *Network analysis*. Beverly Hills: Sage, 1982.

Krugman, P. (1996) "Are currency crises self-fulfilling?", NBER Macroeconomics Annual

Langston, Lizbeth (1996) Scholarly Communication and Electronic Publication: Implications for Research, Advancement, and Promotion http://www.library.ucsb.edu/untangle/langston.html

Lanham, Richard A. 1993 The Electronic Word: Democracy, Technology, and the Arts. Chicago: University of Chicago Press.

Latour, B. (1987): Science in Action, Milton Keynes: Open University Press.

LeFurgy, Bill. "In CinC" Culture in Cyberspace. 1(8), [Note: Broken link to http://www.radix.net/~wlefurgy/cinc08.htm removed 4/10/98 by ald]

Lepair C (1988) The citation gap of applicable science. In:Hdb Quantitative Stu P537

Leydesdorff, Loet & Paul Wouters, Between Texts and Contexts: Advances in Theories of Citation? (A Rejoinder), Scientometrics 44 (1999) 169-182.

Leydesdorff, Loet The Non-linear Dynamics of Sociological Reflections, International Sociology 12 (1997) 25-45.

Leydesdorff, Loet, & Michael Curran, Mapping University-Industry-Government Relations on the Internet: the Construction of Indicators for a Knowledge-based Economy, Cybermetrics 4 (2000), Issue 1, Paper 2 at <http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html>.

Leydesdorff, Loet. Indicators of Innovation in a Knowledge-based Economy. Cybermetrics, 5 (Issue 1), Paper 2, at http://www.cindoc.csic.es/cybermetrics/articles/v5i1p2.html or http://www.cindoc.csic.es/cybermetrics/articles/v5i1p2.pdf"

Leydesdorff, Loet. Words and Co-Words as Indicators of Intellectual Organization, Research Policy 18 (1989) 209-223.

Luukkonen, T. (1997). Why has Latour'ss theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. Scientometrics 38, 27-37.

Maturana, H. R. & Varela, F. J. (1980), Autopoiesis and Cognition: The Realization of the Living, Vol. 42 of Boston Studies in the Philosophy of Science, D. Reidel Publishing Company, Dordrecht, Holland. With a preface to `Autopoiesis' by Stafford Beer. Series editors: Robert S. Cohen and Marx W. Wartofsky.

McCune, David. Guest lecture and discussion, Comparative Literature 283, University of California, Riverside. November 28, 1995.

NIWI Research programme 2000-2004 (http://www.niwi.knaw.nl/us/research/res_prog.pdf)

Odlyzko, Andrew M. 1995. "Tragic Loss or Good Riddance? The Impending Demise of Traditional Scholarly Journals." International Journal of Human-Computer Studies 42: 71-122.

O'Haver, T.C. 1995 "CHEMCONF: An Experiment in International Online Conferencing. Journal of the American Society for Information Science, 46(8): 611-613.

Okubo, Yoshiko Bibliometric Indicators and Analysis of Research Systems: Methods and Examples. STI working papers. OECD 1997.

Polanco, Xavier, Moses Boudourides et. Al., Clustering and Mapping European University Web Sites Sample for Displaying Associations and Visualizing Networks. Paper for the XXX Conference, Crete 2001.

Press, Larry. 1995 "McLuhan Meets the Net." Communications of the ACM 38(6): 15-20.

Price, D. de Solla 1963, Little Science, Big Science, Columbia Univ. Press, New York.

Pritchard, A. 1969, Statistical bibliography or bibliometrics? Journal of Documentation 24, 348-349.

Rand Corporation. Approaches to evaluation of science. (http://ms161u06.u-3mrs.fr/annexe/Cyber0-2804/PAGE/page352.html)

Rifkin, Jeremy ,(2000), The Age of Access: The New Culture of Hypercapitalism, Where All of Life Is a Paid-for Experience. Jeremy P. Tarcher/Putnam.

Rogers, R. (ed) Preferred Placement. Jan van Eyck Editions Maastricht 2000.

Rousseau R. (1997). Sitations. An exploratory study. Cybermetrics, 1 paper 1. ISSN: 1137-5019. (http://www.cindoc.csic.es/cybermetrics/vol1iss1.html)

Rowland, Fytton. 1995. "Electronic Journals: Neither Free nor Easy." Information Society. 11(4): 273-274.

Ruhleder, K. 1995. "Computerization and Changes to Infrastructures for Knowledge Work." Information Society 11(2): 131-144.

Sarup, Madan. 1993. An Introductory Guide to Post-Structuralism and Postmodernism. 2nd ed. Athens: University of Georgia Press.

Schauder, Don. 1994. "Electronic Publishing of Professional Articles: Attitudes of Academics and Implications for the Scholarly Communication Industry." Journal of the American Society for Information Science. 45(2): 73-100.

Schwier, Richard A. 1994. Some Reflections on Scholarly Review and Academic Publication. 13 p. ERIC microfiche collection, ED368317

Selden, Raman; Widdowson, Peter. 1993. A Reader's Guide to Contemporary Literary Theory. 3rd ed. Lexington: University of Kentucky Press.

Stodolsky, David S. 1995. "Consensus Journals: Invitational Journals Based on Peer Review." Information Society. 11(4): 247-260.

Taubes, Gary. (1996b). "Science Journals Go Wired." Science 271(9): 764-766.

Taubes, Gary. 1996a. "Electric Preprints Point the way to 'Author Empowerment'." Science 271(9): 767-768.

Tenopir, Carol. 1995. "Authors and Readers: The Keys to Success or Failure for Electronic Publication." Library Trends 43(4) 571-591.

Van den Besselaar, Peter & Loet Leydesdorff, Mapping Change in Scientific Specialties: A Scientometric Reconstruction of the Development of Artificial Intelligence, Journal of the American Society for Information Science 47 (1996) 415-36.

Van den Besselaar, Peter and Gaston Heimeriks, Codification and Self-Organization in the European STI System. Deliverable D2.5, Task 4, TSER-SOEIS project. Amsterdam: UvA, 2000.

Van den Besselaar, Peter. The cognitive and the social structure of Science & technology studies. Scientometrics 51 (2001) pp 441-460

Van den Besselaar, Peter. The cognitive and the social structure of Science & technology studies. Scientometrics 51 (2001) pp 441-460

van Raan, A.F.J. (Ed.) 1988, Handbook of Quantitative Studies of Science and Technology, North-Holland, Amsterdam.

van Raan, A.F.J. Bibliometrics and Internet: Some observations and expectations. Scientometrics, 50 (1):59-63, January 2001."

W. J. Tijssen (eds.), Science and Technology in a Policy Context (Leiden: DSWO Press, 1993), 219-33.

Wouters, P. 1999, The citation culture, PhD Thesis, Private edition.

Wouters, Paul (1998). The signs of science. Scientometrics 41, 225-241.

Ziman, J. (1994): Prometheus Bound: Science in a Dynamic Steady State, Cambridge, Cambridge University Press.

Ziman, J. M. (1984). An introduction to science studies: The philosophical and social aspects of science and technology. Cambridge: Cambridge University Press