SIBIS
IST–2000-26276
Statistical Indicators Benchmarking the Information Society

SIBIS – WP 2:
Topic research and indicator development

Topic Report No.2:
The Internet for R&D

Tasks 2.1 (Update) + 2.2

Report Version:	Final
Report Preparation Date:	Dezember 2001
Classification:	Restricted
Authors:	Franz Barjak and Najib Harabi (University of Applied Sciences Solothurn Northwestern Switzerland) - We thank our external experts, Prof. Ian Butterworth and Dr. Loet Leydesdorff, for providing detailed and insightful comments. This report has benefited from their comments more than the few explicit references would indicate. All remaining insufficiencies and errors remain our full responsibility.
Contract Start Date:	1st January 2001
Duration:	30 Months
Project Co-ordinator:	empirica (Germany)
Partners:	Work Research Centre (Ireland), Danish Technological Institute (Denmark), Technopolis (UK), Databank Consulting (Italy), Stichting RAND Europe (Netherlands), Fachhochschule Solothurn (Switzerland)

Project funded by the European Community under the “Information Society Technology” Programme (1998-2002)

Table of Contents

PART A (D 2.1)

0 Overview over Part A

The present report sets out to give an overview of the most important issues of the SIBIS topic "the Internet for research and development (R&D)". It contains definitions of important terms and concepts and describes how and to what extent the Internet has been put to use within R&D in the few years since its invention and diffusion. The main focus is on the measurement of the use of the Internet by researchers and of its impact on R&D, even though some general issues of Internet applications for research are discussed. The report also points out the major policy goals in regard to the provision (infrastructure) and the utilisation of the Internet for R&D on the European level. Three strands of literature were evaluated to create this report (which is accordingly reflected in its structure, i.e. the sections 2, 3 and 4): scientific literature, policy documents and statistical documents.

There is no doubt that the Internet and R&D are interrelated: The World Wide Web was developed to facilitate scientific collaboration and the Internet has made it easier to access information that would otherwise be difficult to obtain. Asynchronous communication has become less formal and more frequent. R&D collaboration over large distances has been enhanced, as it has become easier to communicate and transmit information even if it is large and bulky. There are essentially three perspectives on the influence of the Net on R&D:

(1) Internet-related ICT infrastructure for R&D

(2) New network technologies and research activities

(3) Computer networks and R&D collaborations

Ad (1) A large body of literature on the Internet-related ICT infrastructure for R&D has been evaluated in this report. Research networks, the first infrastructure element described, exist on different spatial levels. The collection of data and formation of indicators on national networks has recently gained momentum, propelled by TERENA, the Trans-European Research and Education Networking Association. The market penetration of various further on-line information sources and computer-mediated communication tools differs markedly: A few, such as e-mail, have become omnipresent tools for researchers. Some are mainly used in specific research disciplines such as numerical databases. Other applications are still in the market introduction phase, such as on-line meeting tools. The latter also applies to 'grid computing', a new concept of distributed research work which is based on the capacities of ICT infrastructures and tools. The possibilities of evaluating these tools and concepts depend on their market penetration: When usage is not very common, it is impossible or at least difficult to collect quantitative data, establish benchmarks or undertake comparisons at national level. But also when usage becomes widespread, certain data might loose their value: the simple distinctions between 'e-mail users' and 'e-mail non-users' or between 'on-line' and 'offline-only' journals are hardly relevant anymore, as e-mail usage rates or the on-line availability of journals peak. This has significant methodical implications, inasmuch as the scale of an indicator has to be considered closely. Binary scales might be rather short lived.

Some preliminary efforts at quantifying Internet applications in R&D have been undertaken and we found some studies which even attempted to assess the effects of Internet applications on research success. More often than not they found positive effects. These studies and most of the data collection related to Internet infrastructure for research were carried out at the micro-level, i.e. the individual researcher, the individual research paper, the individual research organisation etc. as units of observation. Even with digital libraries, for which some of the most advanced indicator concepts were found, only the micro-level perspective has been taken. Of course, this micro-level approach is very useful for comparing performance, to attribute funds and even to set policy goals. But it has rarely led to indicators which are appropriate for the comparison of national research systems.

Ad (2) The Internet is being used by researchers mainly for collecting data and retrieving scientific information from different virtual storage spaces and for presenting, disseminating and discussing research results. Internet-based data analysis has not yet become commonplace, but new developments and improvements of visualisation and simulation technologies, and especially in regard to grid computing will increase the possibilities in this field.

The literature that deals with the usage of network infrastructure, on-line resources and Internet tools in R&D mainly discusses their advantages and disadvantages for different steps of the R&D process, e.g. the benefits of on-line questionnaires compared to regular methods; or the time and cost savings and distribution effects of e-publishing compared to print publishing. Again, typically a micro-level perspective has been employed and macro-level research questions have been disregarded (e.g. in which countries is advanced on-line research carried out? Where do the authors of e-journal articles come from?). Still, SIBIS can also benefit from the literature that examines how the Internet is being integrated into different steps of the R&D process. The benefit comes mainly from a methodological point of view: Internet-based data collection might constitute a valuable method for gathering data and constructing indicators on different issues of the Internet for R&D.

Ad (3) The European Commission laid down a new research policy in its documents on a European Research Area (ERA). Increasing collaboration among R&D institutes from different EU and CEEC countries in general, as well as creating top-level virtual R&D institutes labelled ‘virtual centres of excellence’, are important goals of ERA. The empirical evidence that worldwide research collaboration has increased within the last decades is strong. Some authors also attribute this to the diffusion of capable ICT infrastructure and Internet applications. The operationalisation of R&D collaboration has been based either on few available output indicators (bibliometric data, patents) or on count data of R&D partnerships. Neither of the two methodologies permits an adequate measurement of the different concepts of R&D collaboration which have been developed in social sciences theoretical literature. The novel concepts of collaboratories and virtual teams, two new and related forms of network-supported long-distance collaborations, have only been described in case studies so far. How to quantify their significance or their geographical spread still remains unsolved.

This topic report also gives some insights from a methodological point of view. Various concepts have been found for collecting and processing data for any of the sub-topics: Surveys by regular mail or via the Internet, the evaluation of WWW pages, the analysis of server log files, the analysis of archives (e.g. of mailing lists) and bibliometrics. These methods might be used as substitutes to obtain one piece of information on a more sound and secure basis. They might also be applied to obtain complementary information for describing more facets of a subject. One of the open tasks of the SIBIS project is to evaluate the available methods for collecting and processing the empirical data to construct the SIBIS indicators.

1 Introduction: main issues and structure of the report

1.1 The Internet for R&D: What the report encompasses and what it excludes

The development of the Internet has affected societies and economies worldwide. Its impact on the research and development (R&D) system must be considered as particularly important: R&D is one of the most important sources of inventions and technical progress and is therefore a major motor of economic growth. The SIBIS reports on the Internet for R&D set out to develop indicators which are appropriate for measuring the extent to which the Internet has been integrated into R&D and the effects of this.

The present report is a literature review on the Internet in R&D. It sets out to

- provide a structure for the topic,
- define the most important terms,
- give an overview of the most important concepts related to science, politics and statistics
- and describe the concepts and approaches that exist regarding measurement.

Even though some general issues of Internet applications for research are discussed, the main focus is on the measurement of the use of the Internet by researchers and of its impact on R&D. This report assesses what concepts and indicators have been developed and implemented to date. It will also point out the major policy goals in regard to the provision (infrastructure) and the utilisation of the Internet for R&D. The present report will neither identify the gaps within measurement concepts nor develop any new indicators itself. These tasks will be fulfilled within the second report of workpackage 2, the Indicator Report.

The OECD defines R&D as creative work which is undertaken on a systematic basis in order to increase the stock of knowledge and the use of this stock of knowledge to devise new applications.[1] This definition of R&D makes use of two elements: The first is on an input level, stating that R&D requires creative and systematic work. The second is on an output level, as R&D has to create new knowledge or, in other words, find solutions for problems that cannot be answered with the available knowledge and techniques.

The second part of the definition which focuses on the goal of R&D is of special importance for understanding its uniqueness. Creative and systematic work can be carried out for similar but nevertheless different goals. For example, an artist usually is very creative and he might also be very systematic in his work, generating a new opus in the field of music, literature or the fine arts. However, an artist's work is (usually) not targeted at answering natural, social or technical questions nor solving problems in one or more of these areas. The creative work is justified by itself, whereas the creative work of a scientist always has to pursue the goal of increasing knowledge. This doesn't mean a researcher's work and its results have to be immediately applicable and useful to society. They can also deal with basic problems which for the time being 'only' increase the understanding of nature, society or a technical field.[2] But they have to be based on a properly defined problem and an outline for dealing with it scientifically.

A scientist also pursues different goals from the ones of a business manager who introduces new products into the market to raise his profit, market-share or gain a dominant position in the market. Many innovations have their roots in R&D results, but innovations are not the

---

1 OECD (1994): The Measurement of Scientific and Technological Activities: Proposed Standard Practice for Surveys of Research and Experimental Development - Frascati Manual 1993. 5th Edition. Paris, p. 29.

2 This kind of research is usually called 'basic research'. The OECD defines it as follows: "Basic research is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundations of phenomena and observable facts, without any particular application or use in view." 'Applied research', in contrast, is directed primarily towards a specific practical aim or objective, see OECD (1994), op. cit., p. 68-69.

same as R&D results: the latter can be scientific publications or presentations, inventions which may never be used, used only once or used everywhere in the world, whether available free of charge or only for high licence fees due to patent or trademark protection. Innovations, on the other hand, are new products, processes or forms of organisation which have been introduced into the market.[3] For this market introduction, additional activities are typically indispensable, such as market research on available products, competition, possible returns, optimal sales strategies and other activities that are not associated with R&D.

R&D is also not the same as education, though scientists often work in both functions. While R&D aims at extending the boundaries of knowledge, education primarily has the objective of teaching the important things within these boundaries. Thus, education is the foundation of self-reproduction of science and doubtless the borders between research and education are anything but clear; for example the insights gained in the process of teaching often constitute inputs into research. Nevertheless, in this part of the study it is supposed that research and education can be divided analytically. A separate research topic of SIBIS will deal with education issues.

The study will not focus on informatics research or even more specifically on Internet research, but it will take a more interdisciplinary approach: it investigates the utilisation of the Internet and its effects on research in social sciences and humanities as well as in natural sciences and engineering. Of course, it might not be possible to cover the entire world of science within one project and, consequently, it might be necessary to use single areas of research as examples. But it should be possible at least to some extent to generalise the findings and use them as pilot results for future studies.

It is not important for the current analysis whether an R&D organisation is public or private: Universities, public research institutes, R&D departments of large enterprises or the single researcher-innovator who develops a new product and founds a new company, all are included.

1.2 A brief preface on knowledge production

The academic systems of the industrialised countries have exhibited a marked growth over the last century. The scholarly tasks of creating and evaluating knowledge, preserving information and disseminating this information to others have gained importance as their ‘products’ (e.g. educated workers and inventions) have met rising demand. Higher education and research and development expanded within the academic system through the creation of new universities, faculties, research and teaching facilities etc. The success of the academic system also contributed to its growth beyond its own limits: increasingly non-academic training and research organisations on both a profit and a non-profit basis have sprung up to challenge the academic knowledge monopoly.[4] Additionally, the information technology revolution has facilitated the production and distribution of codified knowledge to such an extent that the university has lost the advantage of being the only location integrating research, teaching and access to preserved information (libraries).[5]

If the global volume of knowledge grows but the learning and knowledge capacity of individuals remains more or less constant, or only increases at a slower rate, then specialisation (i.e. everybody knows only a fraction of the total societal knowledge) is the logical consequence. Specialisation applies to research disciplines as well as to individuals, research departments, institutes and entire universities. The growth of the knowledge producing system has been a motor for specialisation.

---

3 OECD (1997): Proposed Guidelines for Collecting and Interpreting Technological Innovation Data: Oslo Manual. 2nd Edition. Paris, p. 31.

4 See Gibbons, M. et al. (1994): The New Production of Knowledge. Sage: London; Thousand Oaks; New Delhi.

5 See Noam, E. M. (1995): Electronics and the Dim Future of the University, in: Science, vol. 270, pp. 247-249.

Moreover, the number of locations where research and development (and teaching) are carried out has also risen. This is mainly due to the inherent advantages of decentralised knowledge production: the academic system benefits from a rich hinterland that provides students, problems and funds for applied research and development, interaction and discussion with the society etc. In turn, the economy benefits from adequately educated job entrants, the possibility of contracting specific problems to the academic system and the interaction between scholars and business people which may generate an innovative climate.[6] The impact of the knowledge producing system on the economy would not necessarily have led to decentralised knowledge production if all knowledge could be codified and transmitted over long distances. But some parts of knowledge remain tacit, because they complement codified parts of knowledge,[7] the costs associated with codification are disproportionately high[8] or codification would need a codebook which does not yet exist.[9] Hence, the increase in the number of knowledge producing sites was also inherent to the growth of the knowledge producing system.

Both specialisation and the increased number of locations where R&D is carried out have contributed to increased demand for communication within the academic system and across its boundaries. Communication, the transmission of information, may be considered as the necessary step between the creation of a ‘one-man secret’ (or subjectively new knowledge) and ‘socially new knowledge’:

> *“the production of new knowledge … is not really complete until it has been transmitted to some others, so that it is no longer one man’s knowledge only.”*[10]

Communication is therefore essential to knowledge production. Researchers have to keep abreast of advancements in their fields of research in order to avoid duplication of efforts and to be able to effectively use the findings of others. The codification and communication of research findings is also a necessary condition for criticism and quality control. Communication has been labelled the ‘essence of science’:

> *“The main effort of individual scientists is manufacturing new information either by describing new data or by formulating new concepts or conceptual integrations of data (theory). In order for these formulations to be successful contributions to science, they must be communicated in such a form so as to be comprehended and verified by other scientists and then used in providing new ground for further exploration. Thus, communicability becomes a salient feature of a scientific product since its recognition by peers as a unique contribution is essential to establishing a scientist’s success in science.”*[11]

The specialisation within knowledge production has also created an incentive for collaborative R&D: more often than not, research problems are complex and looking at them from different points of view with distinct disciplinary backgrounds enriches the amount and quality of solutions which can be found. The division of labour can be considered as the counterpart to specialisation. Collaborative R&D contributes to the rising demand for communication needed to co-ordinate the research and exchange information on all the different parts and stages of an R&D project.

This necessity to communicate has been accounted for in various ways. The existence of scientific conferences and publications may be attributed to the quest for presenting,

---

6 As some have noted, this has also changed the internal logic of the academic system which has opened up towards social demands, see Gibbons, M. et al. (1994), op. cit.

7 See Nonaka, I. (1994): A dynamic theory of organizational knowledge creation, in: Organization science, vol. 5, pp. 14-37.

8 See Nelson, R. R.; Winter, S. G. (1982): An evolutionary theory of economic change. Cambridge and London, p. 82.

9 See Cowan, R.; David, P. A.; Foray, D. (2000): The explicit economics of knowledge codification and tacitness, in: Industrial and corporate change, vol. 9, no. 2, p. 225.

10 Machlup, F. (1962): The production and diffusion of knowledge in the United States. Princeton, p. 14.

11 Garvey, W. D. (1977): Communication: the essence of science. Facilitating information exchange among librarians, scientists, engineers and students. Pergamon Press: Oxford, pp. 1-2.

disseminating, discussing and evaluating research results; research fields that deal with similar topics or use similar theoretical concepts are grouped into disciplines so that communication can take place between like-minded people who know the codebook(s); research departments and institutes constitute locations where scientists work, meet and hence can communicate face-to-face.

The new information and communication technologies have provided additional communication channels in R&D. They have accelerated the long-term tendency towards more knowledge codification.[12] As ICT have made the transmission of information simple and inexpensive, they have created an incentive to codify knowledge. Also, new techniques of knowledge production related to ICT provide more of the knowledge in a codified form. The large computing power of computers has made an empirical revolution possible, with more and more knowledge expressed in quantifiable models and supplemented by indicators and data for empirical measurement. Moreover, ICT have created additional incentives for collaboration, as they have lowered co-ordination costs and transaction risks, i.e. "the possibility of opportunistic behaviour by another party to the relationship, leading to uncertainty surrounding the level and division of the benefits from the increased integration of decisions and operations."[13] Hence, the hard- and software to digitise and process information, the resources to store and retrieve it, the communication networks to transmit the digitised information objects – all are the result of the increased demand for information and communication and at the same time are the driving forces of its further growth.

This interrelationship must be borne in mind when analysing the interactions between the Internet and research and development. While Internet tools and applications might seem to be the cause of certain processes and developments in R&D, they may themselves originate in developments in the R&D sector and not be strictly exogenous.

1.3 Structure of Part A

Material from three main sources was identified and evaluated in the creation of PART A of this report:

- Scientific literature published in journals and books, published and unpublished conference papers and Internet documents,
- Policy documents from multinational governmental institutions: primarily the European Commission, international organisations such as research associations and national institutions (governments, associations in the field of R&D),
- Documents from statistical offices, other organisations generating data such as the OECD and private enterprises.

The middle part of Part A is structured along the lines of these different approaches (chapters 2, 3 and 4). A synopsis of the different approaches and a summary is presented in the final part (chapters 5 and 6).

Chapter 2, a review of the scientific literature on the Internet and R&D, *first* looks at the different Internet-related infrastructure elements which are necessary to carry out modern research and development (2.1). Research is supposed to become increasingly network dependent: high-performance research networks are a prerequisite for transmitting large amounts of information; a broad range of Internet applications has fundamentally changed the ways by which scientists collect data, retrieve information and communicate. Still in its infancy is a set of new protocols and organisational routines, known under the heading of 'grid computing', which aims at increasing the available computational power and making resource

[12] See David, P.; Foray, D. (1995): Accessing and expanding the science and technology knowledge base, in: STI-review, no. 16, pp. 13-68.

[13] Clemons, E. K.; Row, M. C. (1992): Information Technology and Industrial Cooperation: The Changing Economics of Coordination and Ownership, in: Journal of Management Information Systems, vol. 9, no. 2, p. 15.

sharing and synchronous work possible. The literature review then takes a *second*, process-oriented perspective and analyses how the net has changed and more often than not enriched R&D processes (2.2). Three different steps or tasks of an R&D project (data collection and information retrieval, data analysis, publication and dissemination of research results) are distinguished and discussed regarding their relation to the new network technologies. Subsequently we employ a *third*, institution-oriented perspective (2.3). By reducing communication costs significantly, the Internet has created a strong incentive to replace other inputs with communication. Therefore, new and more intertwined forms of collaboration have appeared (known as *collaboratories* or *virtual teams*) which are still in their infancy but should not be ignored. A final chapter of the literature survey presents a brief summary and conclusions (2.4).

Since SIBIS is an applied research project it aims at developing indicators which are relevant for policy formulation and evaluation, particularly in respect to the *e*Europe initiative. Therefore, it was essential to examine the goals and measures of European R&D policies in respect to the Internet. Chapter 3 provides an overview of the most relevant documents from the Commission and from other multinational and national sources. As very few approaches and indicators to assess policy impacts have been proposed in this strand of literature, the present report concentrates on documentation of the goals and measures.

The third source of information evaluated were reports and publications from statistical offices and other organisations involved in the creation and publication of R&D data. Chapter 4 gives a summary of this literature. Again, the analysis has focussed on approaches and indicators which reflect the usage and impact of the Internet on R&D. But as specific results have been scarce, we decided to describe what R&D statistics measure in general, thus aiming to outline possible paths towards enriching the existing data with new, more Internet-oriented indicators.

Chapter 5 pools the results from the three previous chapters 2-4 and presents an overview of indicators on the Internet for R&D which have been developed in the different sources as well as a summary of the main findings of this topic report. The chapter focuses on the existing indicators, their policy relevance and factors influencing their consistency over time. The methods which are needed to obtain the primary data as well as some additional measurement issues and possible problems are briefly mentioned. A closer examination will take place in deliverable 2.2 (Part B of this report).

2 Internet for R&D: survey of the literature

2.1 Internet-related ICT infrastructure for research activities

The existence of an appropriate ICT infrastructure has become a pre-condition for an efficient R&D system. It is impossible to think of modern science without ubiquitous PCs, supercomputers for tasks which require large computing power, campus networks, National Research Networks and supranational connections, a multitude of computer programmes that carry out sophisticated calculations, model schemes and designs, or 'simply' store text and other data and facilitate its retrieval in a user-friendly manner. These are only some examples of currently available ICT infrastructure that has changed research during the past decades.
The present section deals in particular with Internet-related infrastructure including both hardware and software. It also extends to 'humanware', the computer skills available to R&D projects which may be embodied in the researchers themselves, in the technical staff they employ or in services they buy. To include these differing topics we have to use a very broad understanding of infrastructure that is not expedient for many other research questions. But this approach is well suited to our purposes, as we undertake this analysis in particular to develop indicators which are suitable to monitor the impact of computer networks on R&D and to devise policies which can optimise these impacts.[14]

We will begin with research networks and look at data and information resources which are available on-line in section 2.1.2. Section 2.1.3 continues with a discussion of the currently available tools for computer-mediated communication such as e-mail and on-line meeting tools. Most of these infrastructure elements have been around for some years and it is possible to assess their usage levels and some effects on R&D output.

Things change somewhat in section 2.1.4. Among scientists, science managers and science policy makers, a broad consensus prevails that R&D systems and ICT infrastructures for R&D will undergo fundamental changes in the near future. These changes are demand-driven as well as technology-driven: On the one hand the needs of researchers to collaborate and jointly use specialised and expensive instruments or databases irrespective of their geographic location are growing; on the other hand the increase of computing power and transmission capacity in networks create new options for computational analysis, such as distributed computing, which further fuel the demand for higher ICT performance levels.

"There is a need to redefine the ICT "architecture" of science, such that hypotheses formation, problem solving, information processing and computation become a network task, while the main challenge for the individual workstation is to provide appropriate human-computer interaction facilities to support adequately specific research problems."[15]

Therefore section 2.1.4 will give a brief overview of the Grid concept. The last section 2.1.5 takes into consideration the fact that hardware and software alone are not enough for performing Internet-based R&D successfully. Human abilities are an indispensable prerequisite, too. We will not go into detail here, as concepts and data on the technical staff for Internet-based R&D are scarce and as the SIBIS project includes an entire workpackage that analyses "work, skill and employment" in the information society.

---

[14] See as an even broader definition of ICT infrastructure in relation to R&D: "From that perspective, one must consider infrastructure as broadly defined, notably including hardware, software, multi-use and multi-user databases, and the social organisations that create and apply them." Bainbridge, W. S. (1999): Information infrastructure issues in the social sciences, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, p. 124.

[15] European Technology Assessment Network (1999): Transforming European science through information and communication technologies: challenges and opportunities of the digital age. Final version, p. 34.

2.1.1 Research networks

The objective of this section is to assess the nature of network infrastructure that is necessary for cutting edge R&D Internet applications, and to compare the endowment of different European countries. Switzerland is used as an example to describe some features of research networks in more detail. A research network (RN) is conventionally defined as

"...a production network... which supports various types of domain specific application research. This application research is most often used to support the sciences and education but can also be used in support of other areas of academic and economic endeavour."[16]

Different types of RN exist at different spatial or organisational levels:

- An Institutional Research Network (IRN) is a network that supports universities, institutes, libraries, data warehouses, and other 'campus' like networks.
- National Research Networks (NRNs), such as the Swiss SWITCH or Germany's DFN networks, support IRNs or affinity-based networks.
- Pan National Research Networks (PNRNs) interconnect and support NRNs (e.g. Dante's Ten-155; GÉANT and the NORDUNET).

With the widespread use of computers in offices, businesses and educational institutions in the early 1980s, the volume of data transferred not only exploded but it also became necessary to transfer various kinds of data: multimedia, video, voice, and data. The introduction of optical fibres in the place of copper wires allowed higher rates of data transfer, and to some extent solved problems associated with varied bandwidths necessary to transfer multimode data. It then became imperative to develop a technology which would transmit, in as short a time as possible, different kinds of information that required varying bandwidths. The need to meet these requirements gave birth to Asynchronous Transfer Mode (ATM) technology. A switching technology that is circuit-based, ATM uses optical fibre network technology,[17] and can allocate the available bandwidth in a flexible manner.[18] In this method, small blocks - called cells - are used to transmit data from one point to the other. Each of the cells is 53 bytes long. The cell contains only a virtual circuit identifier (VCI) and a virtual path identifier (VPI) but not the addresses of source and destination as in Internet Protocol (IP) packets. The technology is also quite flexible in the sense that it can support a very broad array of technologies like DSL, IP Ethernet, Frame relay, SONET/SDH and wireless platforms.[19]

Since its inception, ATM technology has been adopted in over 80 percent of the networks worldwide and commands a global market worth billions of US dollars (see footnote 19). As seen in table 1, only two out of all the European countries listed do not use ATM. Even though ATM has proved to be quite a popular technology in network architecture, it does have some drawbacks to overcome which lead network specialists to transfer Internet Protocols (IP) directly over Synchronous Digital Hierarchy (SDH) technology or over ATM (see footnote 19). Currently the Internet Protocol used in most of the networks is the Internet Protocol Version 4 (IPv4). Many networks are switching (or planning to switch) to IPv6 because of its many advantages like 128 bit address size, etc.[20]

16 Aiken, R. A. (2000): New Frontiers for Research Networks in the 21st Century, in: Wouters, P.; Schröder, P. (eds): Access to publicly financed research. The Global Research Village III, Amsterdam 2000. Background papers. Amsterdam, p. 92.

17 See http://sunsite.nstu.nsk.su/sunworldonline/swol-04-1997/swol-04-connectivity.html

18 See ftp://ftp.cordis.lu/pub/etan/docs/ict-report.pdf

19 See http://www.atmforum.com/glossaryfs1.html

20 See http://playground.sun.com/pub/ipng/html/INET-IPng-Paper.html#CH4

Table 1: Comparison of the current status of the ICT infrastructure in Switzerland and in selected countries of the European Union and Central and Eastern Europe

Country	Name of the network	Technology used	Transmission capacity
Switzerland	SWITCH http://www.switch.ch	POS, ATM	2.5 Gbps Up to 155 Mbps
Countries of the European Union:			
Austria	ACOnet http://www.aco.net/acoges2.htm	ATM	10 to 155 Mbps
Belgium	BELNET http://www.belnet.be/main_uk.html	ATM	34 to 1000 Mbps
France	Renater http://www.renater.fr	IP/ATM/SDH IP/ATM/WDM	155 Mbps and 2.5 Gbps
Germany	DFN http://www.dfn.de	DWDM/SDH	Up to 2.5 Gbps
Greece	GRNET http://www.grnet.gr/index_en.html	ATM	Up to 155 Mbps
Ireland	HEANET http://www.heanet.ie	ATM	Up to 155 Mbps
Italy	GARR http://www.garr.it/garr-b-home-engl.shtml	ATM ISDN	34 to 155 Mbps
Netherlands	SURFnet http://www.surfnet.nl/en	ATM Protocol: IP over DWDM, IPv4 IPv6 POS	Up to 10 Gbps
Nordic Countries[a]	NORDUnet http://www.nordu.net	ATM, DWDM Technology	Up to 2.5 Gbps
Portugal	FCCN http://www.fccn.pt	ATM VPN-IP Frame relay Dedicated circuits	Up to 180 Mbps
Spain	ReDIRIS http://www.rediris.es/index.en.html	ATM ISDN backup Protocol: IP Multicast: IRIS-MBONE	34 – 155 Mbps
United Kingdom	JANET: http://www.ja.net UKERNA: http://www.ukerna.ac.uk/aboutukerna.html	ATM, IP over ATM, SDH, DWDM	2.5 Gbps

Table continues on the next page

Table 1

Country	Name of the network	Technology used	Transmission capacity
Central and Eastern European accession countries to the European Union:			
Czech Republic	CESNET2 http://www.cesnet.cz/english	Pocket over SONET (POS)	2.5 Gbps
Hungary	HUNGARNET http://www.hungarnet.hu/english/netserv/summary.html	ATM FDDI, Ethernet, microwave	Up to 34 Mbps
Poland	NASK http://www.nask.pl/english	ATM	155 Mbps
Slovenia	ARNES http://www.arnes.si/english	ISDN	64 Kbps – 2 Mbps

a Denmark, Finland, Iceland, Norway, Sweden

Source: FHSO compilation.

The other technology that is being used in some of the ICT networks (for example, the DFN network in Germany and the NORDUnet in the Nordic countries) is DWDM technology. Dense Wavelength Division Multiplexing (DWDM) technology makes it possible to simultaneously transmit multiple data streams over a single fibre at data rates as high as the fibre capacity. This technology makes it possible to increase the existing capacity, for example, an existing 2.4 Gbps system can be increased sixteen fold (see footnote 18).

Almost all the research networks in Europe are members of the TEN-155 network. Trans European Network (TEN) offers its members a transfer rate of 155 Mbps. Built and managed by DANTE (Delivery of Advanced Network Technology to Europe), a Cambridge-based company, TEN-155 has nodes in Switzerland as well as in Austria, France, Germany, Italy, the Netherlands, Sweden, and the United Kingdom.[21] The network combines IP and ATM technologies. Member countries can select between ATM or SDH over IP access. While TEN–155 supplied 155 Mbps data transfer rate, its successor GEANT will make data transfer rates in the range of a few Gbps possible. The aim was to initially provide - early 2001 - operational speeds of 2.5 Gbps, with the intention of doubling this speed every year over a four-year period.[22] Still, Europe significantly lacks behind the United States, where transmission capacities in the Gbps range were already employed in 1999.[23] The development of a European research network with large transmission capacities was on the one hand hampered by the telecommunication monopolies which prevailed in many countries until recently. On the other hand, the decentralised political structure within Europe has led it to put more emphasis on the breadth of access to a network and less on network capacities.[24] Of course, this can also be viewed as an advantage: most of the traffic on research networks is domestic or even local, and European NRNs manage to integrate a large part of their potential customer base.[25] Also users and (national) funding bodies overlap geographically and detrimental externalities are avoided.

---

[21] See http://www.dante.net/ten-155/brochure2.html

[22] See http://www.dante.net/geant/geant-brochure-nov00.html

[23] See Axmann, H.-P.; Payr, S. (1999): A Global Research Infrastructure for the Global Research Village: The European Perspective, in: in: STI Review No. 24: Special Issue on “The Global Research Village”. Paris, p. 40.

[24] Ibid., pp. 29-47.

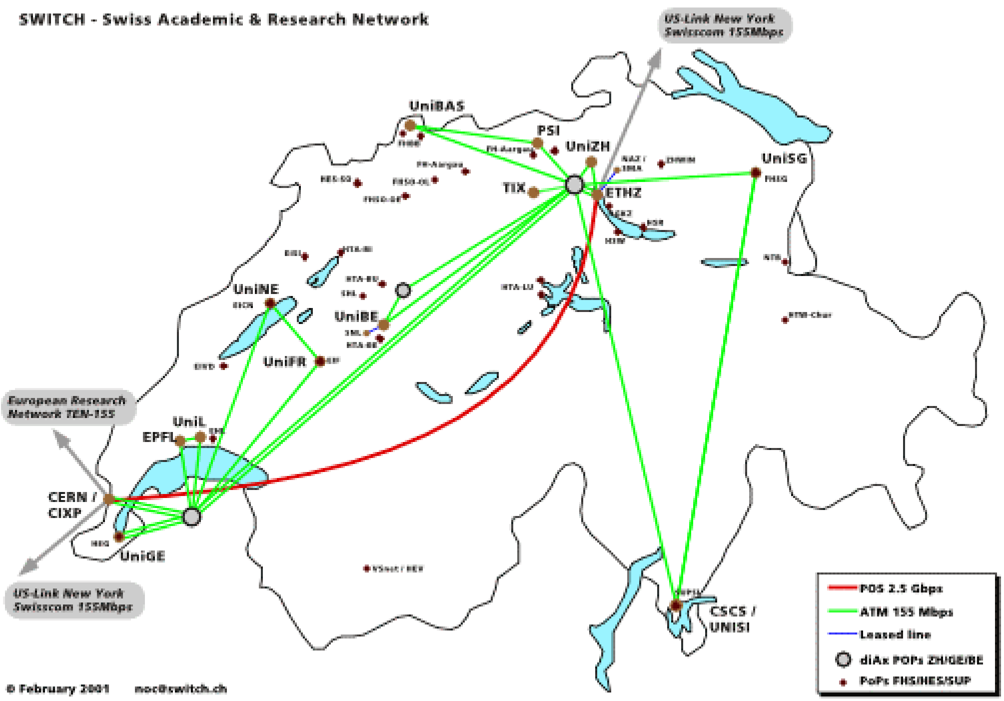
[25] See Vietsch, K. (2000): Research Networking in Europe, in: Scimitar Electronic Journal, vol. 17. (http://www.iihe.ac.be/scimitar/J1000/rns.html).

The ICT infrastructure scenario in Switzerland

The Swiss Academic and Research Network ([SWITCH](http://www.switch.ch)) is the only network that provides ICT architecture for academic research in Switzerland. Founded in 1987 by the Swiss Confederation and eight university cantons of Switzerland, SWITCH set out to establish and operate academic and research networks by providing modern data transmission facilities. In 1998, diAx/Ascom was selected to establish the national network. As of April 1999, connection to USA using a bandwidth of 24 Mbps established with the support of WorldCom was possible.[26] Since its establishment, SWITCH has used Asynchronous Transfer Mode (ATM) methodology as the core of its network, below which is diAx's Synchronous Digital Hierarchy (SDH) infrastructure.

The Universities of Applied Sciences were connected to the SWITCH network using the services of Cablecom Media AG. A bandwidth of 2 to 16 Mbps were made available to the Universities and Universities of Applied Sciences - a total of 29 institutions.[27] A mesh of static virtual circuits (VCs) are used to provide the necessary Internet Protocol (IP). A bandwidth between 34 and 155 Mbps is made available to the universities connected to SWITCHlan. The following figure shows the topology of SWITCHlan's backbone.

Figure: Topology of the SWITCHlan backbone



Source: http://www.switch.ch/lan/national.html

In June 2001 it was announced that Sorrento Networks' GigaMuxTM optical transmission system had been selected for the deployment of SWITCH's national optical backbone network which is to provide 10Gbps service to the member institutions.

---

[26] See http://www.switch.ch/ng/map/networknow.html

[27] SWITCHjournal (2000), p. 10

No statistical indicators have so far been officially established to evaluate the performance of the ICT architecture of the various networks available. An initiative carried out by the Trans-European Research and Education Networking Association (TERENA) produced a broad range of valuable indicators on NRNs.[28]

A useful and accessible indicator for evaluating network performance is the maximum data transfer rate of an NRN. Some NRNs like those in the Netherlands, France, the United Kingdom , the Czech Republic and the Nordic countries offer transfer rates in the Gigabit per second (Gbps) range. Others still have to be content with Megabit per second (Mbps) capacities (see table 1). However, it is questionable whether the maximum transmission capacity is really a representative measure of the average service level of an NRN as usually not all of its clients are connected with the same bandwidth. The maximum transmission capacity might be considered as a measure of the upper level of data transfer within a national research system, though it does not give any information about the average transmission capacity and the level of service that is available to the average research site. The average transmission capacity, regardless of how it is calculated, should be more representative for the service level of an NRN and therefore more appropriate for comparing countries.

Additional information on the performance could also be obtained by looking at the transmission capacities to large sites within the NRN,[29] though this indicator is only useful if a meaningful international definition of a 'large site' can be established. This definition is not provided by TERENA. The available bandwidth to other European countries and the US, another indicator, takes into account the interfaces of NRNs and Pan-National Research Networks (PNRNs).[30] Though the largest volume of traffic usually takes place within the national networks,[31] it is also useful to take the international connectivity of an NRN into consideration as international collaborative research has increased (see section 2.3.2, p. 46). Taking the maximum transmission capacity and multiplying it by the length of the network is used an indication of the absolute network size.[32]

Each of these indicators covers the supply side of NRNs but ignores the demand at least to some extent. This does not necessarily constitute a problem, when supply and demand are aligned to each other. However, TERENA also constructed an indicator that can be considered to take the demand into consideration: the data transmitted ('traffic') in an NRN per student and per year. Though this figure benefits from its straightforward approach, it also has some major shortcomings for our purpose: the number of students might correlate to the number of university teachers but it is certainly not a very good indicator of the research efforts within a national research system. On the contrary, a large number of students may lower the research capacities of university scholars and lead to a diversion of basic research to other entities. The numerator 'traffic' is also far from optimal as we cannot tell if the measured traffic is created through teaching, research or private activities on the WWW.

In addition to transmission capacities and transmitted data, TERENA set out to find out more details about the client structure of NRNs.[33] They wanted to determine the number of institutions connected and the group to which they belong: "Industry, commerce and industrial research", "University education and research", "Higher and further education" or some additional types of institutions, as a percentage. As TERENA states on its web page, the survey questions were phrased somewhat ambiguously and some respondents instead provided their market penetration, e.g. what percentage of universities and research institutes were connected to the NRN, rather than the percentage makeup of the groups.[34] This question is certainly more difficult to answer, but for the purpose of comparing the utility of an

28 See http://www.terena.nl/compendium/ToC.html.

29 See http://www.terena.nl/compendium/xscap.html.

30 See http://www.terena.nl/compendium/cap.html.

31 See Vietsch, K. (2000): Research Networking in Europe, in: Scimitar Electronic Journal, vol. 17. (http://www.iihe.ac.be/scimitar/J1000/rns.html).

32 See http://www.terena.nl/compendium/size.html.

33 See http://www.terena.nl/compendium/instit.html.

34 Ibid.

NRN to its national research system it is probably also more appropriate than the client structure.

Another chapter of the TERENA questionnaire dealt with the financial and budgetary issues of NRNs. It asked for absolute budget sizes which were subsequently used to calculate budget per student ratios. This information might depend, as TERENA itself states, on the wealth of a country. It should also depend on the volume of clients that receive NRN services (e.g. in some countries NRNs deliver services to secondary and primary education, in other countries only to the higher education level) and on the range of services that is delivered. Therefore an indicator would have to be produced that is based on the expenses for a fixed and comparable set of services, related to the potential customer base and measured in purchasing power parities.

TERENA also included a section on staffing in its questionnaire to national research network providers: Technical staff and total staff were counted in full-time equivalents.[35] The staffing depends on NRN size and services and the structure of service delivery. To take as a minimum differences in the structure of service delivery into account, the entire input of working hours (in person-months per year) was also requested. There might still be a considerable bias in the data due to different work tasks: while the core task of the NRNs is to provide network access and data transmission, they might also engage in considerable additional services which increase their personnel.[36] Therefore, as with budgetary data, some control mechanisms regarding service categories and the size of the customer base have to be included for personnel data in order to make the data comparable.

Research networks are an important part of the ICT infrastructure of a research system, but they are definitely not the only important part. The next section will take a look at the different information sources which have been made available on-line.

2.1.2 On-line information sources

Information sources are needed within different steps and for different purposes of R&D activities. For example, information, in numerical, textual or other forms is a basic input for solving most research questions; information is also necessary for communicating research results, e.g. information on the journals that exist in a research field, on their focus or on publishing standards. The present chapter discusses the major types of network-based information sources that exist for scientific users.

Traditionally three types of databases have been distinguished: bibliographical, full-text and numerical databases. The first two can by pooled which leaves us with two generic types, one containing textual information, the other containing numerical information. Before the advent of the World Wide Web, the retrieval of information from scientific databases had been measured by invoices from the host to the subscribing researcher or research institution. Such records usually were kept in terms of minutes spent connected to the database, the number of textual documents or the amount of numerical data retrieved or delivered in hardcopy. Since the advent of the Web, however, the amount of information available via the computer has multiplied. Also, free access became a competitive alternative to databases with subscription and access fees, so that billing records then became sparse and could no longer serve as a reliable indicator of access frequency.

The next two parts of this section will discuss textual and numerical databases which have existed – and continue to do so – in electronic and non-electronic forms. We will show that the Web created new applications for databases and we discuss how these new applications in R&D have been measured so far. The third part of this section deals with the Internet in general, looking at it as a huge but less structured collection of scientific information.

35 See http://www.terena.nl/compendium/staff.html.

36 See http://www.terena.nl/compendium/intro.html.

Textual databases

The published results of scientific research have typically been stored in textual databases (we will use 'libraries' as a synonym). Thus, other researchers, politicians, entrepreneurs or other interested individuals can retrieve, evaluate and exploit them for their own purposes. The usage of libraries was facilitated by different search devices such as card indexes or more recently microfiches. The invention of computers and electronic storage resources has created electronic databases. Later on the invention of the Internet and the WWW has changed the search facilities and also created the possibility of remote access to electronically stored bibliographical and full-text databases. These new services are called 'electronic library services' or 'digital libraries'. A useful definition of digital libraries has been developed by the D-Lib working group on digital library metrics:[37]

"The Digital Library is the collection of services and the collection of information objects that support users in dealing with information objects and the organization and presentation of those objects available directly or indirectly via electronic/digital means."[38]

The authors stress that a digital library not only contains information objects in different formats (ranging from 'traditional' textual documents to live objects such as sensor readings) but it also delivers management, storage, retrieval and other services.[39] Furthermore, in addition to purely digital information objects, analogue media may be stored as a digitised representation in an electronic form.

Interesting examples of digital libraries go beyond converting analogue information objects into digital information objects (text files, digital catalogues etc.) but add additional value by making previously unattainable information available:

- the "arXiv" (formerly Los Alamos archive), a library which contains scientific papers in pre-print form which means they are made available in advance of their publishing in scientific journals,[40]
- the Virginia Polytechnic Institute and State University's (Virginia Tech) Electronic Theses and Dissertations library which contains graduate students' work which varies in quality and therefore publication value,[41]
- connected libraries at different locations such as the California Digital Library which increased the number of databases and electronic journals available to its users at the different campuses while at the same time saving valuable funds due to systemwide licenses,[42]
- reprint archives such as the documents and citations database ResearchIndex[43] which contains published articles for simplified electronic retrieval; or JSTOR which contains digitised versions of older volumes of research journals,[44]
- the British Association for University Research and Industry Links (AURIL) invention database with a clearing-house offering technology available for licensing across the university sector.[45]

---

37 D-Lib is a project within the US Digital Libraries Initiative, see http://www.dlib.org.

38 Leiner, B. (1998): The Scope of the Digital Library. (http://www.dlib.org/metrics/public/papers/dig-lib-scope.html).

39 This somewhat blurs the distinction between textual and numerical databases which we made above. Nevertheless, we will keep it for the sake of clarity.

40 See Ginsparg, P. (1994): First steps toward electronic research communication, in: Computers in Physics, vol. 8, no. 4, pp. 390-396. (also available from: http://arxiv.org/blurb).

41 See http://www.ndltd.org/ and Fox, E.A. (1999): Networked Digital Library of Theses and Dissertations, in: Nature web matters, 12 Aug 1999 (http://www.nature.com/nature/webmatters/library/library.html).

42 See http://www.cdlib.org/news/progress.html.

43 http://citeseer.nj.nec.com/cs

44 See http://www.jstor.org and Guthrie, K. (2000): Revitalizing Older Published Literature: Preliminary Lessons from the Use of JSTOR. Paper for the conference "The Economics and Usage of Digital Library Collections", 23-24 March 2000, Ann Arbor, Michigan. (http://www.si.umich.edu/PEAK-2000/guthrie.pdf).

45 See http://webdb2.patent.gov.uk/auril.

- Information gateways and science portals (see p. 24)

This list only aims at giving an idea of the variety and capability of digital libraries. The last decade has witnessed an incredible amount of programmes and projects supporting the digitisation of scientific libraries and the creation of digital collections.[46] Associations and working groups have been formed which deal with the many open issues accompanying this development and we cannot even attempt to give an overview of the important topics within this brief report. But, in line with the aim of the SIBIS project we will focus on the discussion on measuring the performance of digital libraries. Many research projects and initiatives have started in the second half of the nineties to develop a set of useable indicators to make the performance of digital libraries transparent.[47] They have developed indicators which can be grouped into four different fields:

1. Indicators of patron-oriented resources and services,
2. Indicators of user demand and usage,
3. Indicators of user characteristics,
4. Cost indicators.

Table 2 gives an overview of the indicators proposed for measuring the performance of digital libraries. They can be a starting point for statistics. Some indicators can be detailed to sub-categories, e.g. counting all titles in a digital collection, or only the e-journals, e-books and other electronic resources such as CD-ROMs. Some normalisation is necessary to make the data comparable across different organisations or countries. The Equinox project uses two methods of normalisation: relating an indicator to its target group (“target population”, “population to be served”) or calculating it as a fraction of regular non-electronic resources and services (e.g. electronic resources and services as a fraction of total resources, costs of electronic resources and services as a fraction of total costs).[48]

Table 2: Indicators for measuring the performance of digital libraries and electronic library services (ELS)

Indicator	Sub-category	Standardisation	Source
1. Indicators of patron-oriented resources and services			
Titles in a digital collection	all titles, e-journals, e-books, other electronic resources	in relation to the target group, as percentage of all titles	#
Library web pages in service			#
Public workstations available		in relation to the target group	+, #

[46] Some of the most important ones shall nevertheless be mentioned:

In the U.S.: the Digital Library Initiative, see http://www.dli2.nsf.gov and the summaries in Interagency Working Group (IWG) on Information Technology Research and Development (IT R&D) of the National Science and Technology Council (2001): Information Technology: The 21st Century Revolution, pp. 63-72. (http://www.itrd.gov/pubs/blue01/).

In Europe: the Telematics for Libraries programme under the EU's Third and Fourth Framework Programmes for Research and Technological Development funding over 100 actions from 1990-1998, see http://www.cordis.lu/libraries/en/ intro.html and PricewaterhouseCoopers (1999): Impact of the Telematics for Libraries Programme under the Fourth Framework Programme. (ftp://ftp.cordis.lu/pub/ist/docs/digicult/impact.pdf). See also the multitude of European national programmes and initiatives at http://www.cordis.lu/libraries/en/natpol.html.

[47] The following descriptions draw especially on three initiatives: International Coalition of Library Consortia (1998): Guidelines for Statistical Measures of Usage of Web-based Indexed, Abstracted, and Full-Text Resources. (http://www.library.yale.edu/consortia/webstats.html). – Shim, W. et al. (2000): ARL E-metrics project: developing statistics and performance measures to describe electronic information services and resources for ARL libraries. (http://www.arl.org/stats/newmeas/emetrics/phaseone.pdf). – Brophy, P. et al. (2000): EQUINOX - Library Performance Measurement and Quality Management System: Performance Indicators for Electronic Library Services. (http://equinox.dcu.ie/reports/pilist.html). Other initiatives that to our knowledge have not published any results as of November 2001 are undertaken within the D-Lib Forum Working Group on Digital Library Metrics (http://www.dlib.org/metrics/public/index.html) and supported by the Digital Library Federation (http://www.diglib.org/use.htm).

[48] See Brophy, P. et al. (2000); op. cit.

Indicator	Sub-category	Standardisation	Source
Library staff developing, managing and providing ELS and user training		as percentage of all staff	+, #
Classes and training lessons on ELS			#
Population reached by ELS		in relation to the target group	+
2. Indicators of user demand and usage			
Number of sessions (logins)		in relation to the target group, as percentage of all information requests	*, +, #
Number of searches and menu selections	Searches, selections from alphabetical or subject menus	in relation to the target group, as percentage of all information requests	*, #

Table 2

Indicator	Sub-category	Standardisation	Source
Number of documents/items examined	Documents viewed, marked or selected, downloaded, e-mailed or printed Citations, full text	in relation to the target group, as percentage of all information requests	*, +, #
Number of rejected sessions because of requests exceeding the simultaneous user limit		as percentage of all attempted sessions	*, +
Library computer workstation use rate			+
Attendance at ELS training lessons			+
User satisfaction with ELS			+, #
3. Indicators of user characteristics			
Location of user	On-campus, remote		+, #
Status of user	Faculty, undergraduate, graduate, staff		#
4. Cost indicators			
Costs for ELS	For different types of resources (e-journals, e-books, video files etc.) one-time, subscription rates	as percentage of total expenditure on acquisitions	+, #
Costs in relation to some kind of usage measure	Per document or entry examined, per session, per search		+, #

*Sources are *) International Coalition of Library Consortia (1998), op. cit.; +) Brophy, P. et al. (2000), op. cit.; #) Shim, W. et al. (2000), op. cit.*

Source: FHSO compilation.

Data acquisition is possible from three sources, the publisher, the digital library and the user, by means of two different methods: survey and log file analysis (see table 3). We will show below that it is still difficult to generate usable log file data, especially if the objects of comparison are countries and not servers (see the section on non-interactive data collection by means of computer networks, p. 37). Nevertheless, many ongoing initiatives on digital library metrics (see footnote 47) and the broad discussion involving both publishers and librarians give some confidence that solutions and comparable statistics will be available in the future.

Table 3: Data acquisition on digital libraries

	Log file analysis	Survey
Publisher	e.g. number of sessions, searches, documents examined, rejected logins	–
Digital library	e.g. number of sessions, searches, documents examined	e.g. data on patron-oriented resources and services, cost figures, capacity utilisation
User	–	e.g. user characteristics (location, status), user satisfaction

Source: FHSO compilation.

While some alternative quantitative indicators for textual databases have been proposed, the qualitative dimension within digital library initiatives is only reflected in a user satisfaction indicator (see table 2). Other analyses use different approaches, such as a comparison of the advantages and disadvantages of electronic vs. printed documents. According to a survey among the German Max Planck Society, researchers see the advantages of electronic documents as their access-on-demand and timeliness, while the disadvantages include the dependency on electronic infrastructure, incompleteness and the unsolved problems of archiving and long-term access.[49] Some of the specific benefits of e-publications, such as full text search and integrated multimedia content, were obviously not used widely by the respondents to the survey. But the majority of those using these tools pointed out that they were advantageous rather than disadvantageous compared to printed journals. An investigation among American scholars produced similar results.[50]

Numerical databases

Throughout the Nineties the use of personal computers as part of the new information and communication technologies revolutionised the use of databases for R&D activities. Early in the decade one scientist, David R. Lide, a physicist at the [U.S.] National Bureau of Standards, observed that "One class of scientific information where the new technology promises to have major impact is the hard factual data, usually numeric in nature, which form the lifeblood of science and are an essential ingredient in the transfer of scientific knowledge into useful technology."[51]

Lide pointed to three main areas of research which were using numerical databases: the physical sciences, geosciences, and biosciences:

- Spectroscopy and crystallography have been the prime areas in physics, with work being performed by the National Institute of Standards and Technology in the United States, and with products emerging such as the Wiley Register of Mass Spectral Data. The largest database of carbon-13 in NMR spectrometers is maintained by the FIZ Karlsruhe and STN International.[52] STN also offers numerical databases such as the Handbook of

---

49 See Rusch-Feja, D.; Siebeky, U. (1999): Evaluation of Usage and Acceptance of Electronic Journals. Results of an Electronic Survey of Max Planck Society Researchers including Usage Statistics from Elsevier, Springer and Academic Press (Full Report), in: D-Lib Magazine, vol. 5, no. 10. (http://www.dlib.org/dlib/october99/rusch-feja/10rusch-feja-full-report.html).

50 See Lenares, D. (1999): Faculty use of electronic journals at research institutions, in: Proceedings of the ACRL Ninth National Conference. (http://www.ala.org/acrl/lenares.pdf).

51 Lide, D. R. (1994): The impact of information technology on the access to science, section 3, in: Expanding Access to Science and Technology: The Role of Information Technologies, edited by Ines Wesley-Tanaskovic, Jacques Tocatlian, and Kenneth H. Roberts. Proceedings of the Second International Symposium on the Frontiers of Science and Technology, Kyoto, Japan, 12-14 May 1992. (http://www.unu.edu/unupress/unupbooks/uu07ee/uu07ee05.htm).

52 See http://www.fiz-karlsruhe.de/

Data on Organic Chemistry (HODOC) database, Beilstein, Gmelin, DECHEMA, one from the Thermodynamic Research Center (TRC) at Texas A&M University and the NIST Chemical Thermodynamics Database. In keeping with the trend in the Nineties, STN now features Web access to its products. Geophysics benefited from a plethora of new data from satellites, and new projects were initiated employing ICT such as the International Geosphere–Biosphere Programme IGBP.[53]

- In the area of geosciences the World Meteorological Organization WMO has made significant contributions to climate change research. Now peer technology or "distributed computing" (see section 2.1.4), makes use of idle computers in the Climateprediction.com experiment.[54] Another example of geoscientific applications is the metadatabase GlobalChange Master Directory.[55]
- Chemical structures likewise became more easily useable in digitised form. New drugs were developed with the aid of numerical databases in the Human Genome Project.[56] Other examples of bioscientific applications include the Protein Data Bank and the GenBank sequence database.[57]

Of course, databases are also used in other research disciplines to an increasing extent: nearly every large international organisation or statistical office at national and sub-national level maintains a database and more and more provide access via the Internet on a subscriber basis or for free. The Luxembourg Income Study is one of the frequently cited examples in the social sciences. It is a cooperative research project which is mainly funded by the national science and social science research foundations of its member countries and contains income data on 26 countries.[58]

We did not find any attempts at measuring input-related data (database producers, contributors, volume of data input) or usage-related data (subscribers to databases, amount of usage) on a country level. A possible source for such an undertaking could be the Gale Directory of Databases (GDDB), a proprietary database of databases which is available from the Gale Group. It contains more than 13,000 databases of all types and in all subject areas and includes information on the producers and vendors.[59]

Internet sites in general

Information is not only available in structured and managed data sets. In addition, the Internet and especially the World Wide Web offer huge amounts of less structured but also highly informative data. A study of the contents of the World Wide Web estimated that in February 1999 about 6% of the information on the publicly indexable web was stored on scientific or educational web servers, whereas 83% of the information was found on servers with commercial content.[60] At first sight the scientific content seems to be small. But with an estimated 800 million web pages (and today probably double that) the scientific content added up to roughly 50 million pages. Additionally, not only the content of academic websites is relevant for R&D: Governments and foundations provide valuable information on the funding of R&D projects, private enterprises offer information on their R&D and related issues, software firms make software available which is used for different research-related tasks. If the scientific searcher has no prior knowledge of where to look for a specific information item, two types of tools are available to locate it:

a) On-line search engines

---

53 See http://www.igbp.kva.se/cgi-bin/php/frameset.php

54 See http://www.climateprediction.rl.ac.uk/

55 See http://gcmd.gsfc.nasa.gov

56 See Waldorp, M.M. (1995): On-Line Archives Let Biologists Interrogate the Genome, in: Science 269, 8 September.

57 See OECD (1998), op. cit., p. 29.

58 See http://www.lis.ceps.lu/.

59 See http://www.galegroup.com/pdf/facts/gdod.pdf.

60 See Lawrence, S.; Giles, C. L. (1999): Accessibility of information on the web, in: Nature, vol. 400, 8 July 1999, pp. 107-109.

b) Scientific portals

Ad a) Usually search engines use ‘crawler programs’, programs which start at one page, index it, follow the links on that page, index the pages found and so on. When an indexed term is entered in a search routine, the search engine software goes through the index and retrieves all matches according to their (supposed) relevance for the searcher. Based upon this automated routine, search engines have crawled and indexed hundreds of millions of web pages.[61] Nevertheless, they fail by far to cover the entire WWW.

One reason for that might be the structure of the WWW which was recently compared to a bow tie: With a central core and large amounts of links both leading to that core and away from it, covering in total about 70% of the web pages in simulations, with another 30% considered as dead ends or disconnected sites.[62] These pages will often not be found by standard web crawlers and are therefore not indexed. Another reason that only part of the WWW content will be included in search engine indexes is the duplication of information on different websites or under different addresses.[63] It would only increase the amount of stored data and not render any additional benefit to the user if the contents of mirror sites were to be included in the indexes. A third reason might be diminishing marginal returns associated with increasing the coverage of a search engine index, as most queries can be satisfied with a relatively small database.[64] The formation of a large database also creates additional costs, e.g. for storage resources and maintenance. Lawrence and Giles mention another limitation of search engines: the time lag between the publication of a new page or the update of an existing page and its crawling and indexing by the search engine might be several months.[65] Various research projects have also shown that the completeness of document retrieval not only depends on the search engine used, but even varies within one search engine depending on the day or the exact query formulation.[66]

All mentioned problems limit the use of search engines for scientific purposes. But the biggest hindrance to using search engines efficiently is a combination of limited human capability and the lack of intelligent techniques that make up for this. Any search engine will have problems finding the most relevant websites if imprecise and insufficient search terms are submitted. Formerly the relevance was largely based on counts of the search term on the retrieved page and, was vulnerable to manipulation by the author and owner of a page. The inventors of Google, Sergey Brin and Lawrence Page introduced a new method of assessing the relevance of a web page by counting the number of links that point towards it. The more links that lead to a page, the higher its rank or ‘authority’. Thus, the ranking of web pages according to their importance could be improved. However, the search engines’ ability to find the most relevant information for an insufficiently completed or difficult search is still rather limited and often the searcher has to browse a large amount of retrieved results which is both tiring and time consuming. Various research projects have been started to solve this problem and increase the efficiency of web searches or the storage of web contents for scientists.[67] So

---

61 As the amount changes day by day it is of little use to include any figures. Figures can be found e.g. on the Searchenginewatch website, where they are updated regularly, see http://www.searchenginewatch.com/reports/sizes.html.

62 See the joint AltaVista, IBM, Compaq study from Broder, A. et al.: Graph structure in the web. (http://www9.org/w9cdrom/160/160.html)

63 See Searchenginewatch (2000): Numbers, Numbers – But what do they mean, in: The Search Engine Report, March 3, 2000. (http://www.searchenginewatch.com/sereport/00/03-numbers.html)

64 See Lawrence, S.; Giles, C. L. (1999), op. cit.

65 Ibid., p. 109.

66 See Bar-Ilan, J. (1999): Search engine results over time – A case study on search engine stability, in: Cybermetrics, vol. 2/3, no. 1. (http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html). - Rousseau, R. (1999): Daily time series of common single word searches in AltaVista and NorthernLight, in: Cybermetrics, vol. 2/3, no. 1. (http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html). - Mettrop, W.; Nieuwenhuysen, P. (2000): The reliability of Internet search engines: Fluctuations in Document Accessibility, in: Proceedings of National Online Meeting 2000, May 16-18, 2000, New York, edited by Martha E. Williams, Information Today Inc, Medford, N. J., pp. 271- 281. - Thelwall, M. (2001): The responsiveness of search engine indexes, in: Cybermetrics, vol. 5, no. 1. (http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html). – Beheshti, J. (2000): Looking for the needle in a haystack: retrieving information from the web. Presentation at the SSGRR International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, L'Aquila, July 31 - August 06, 2000. (http://www.ssgrr.it/en/ssgrr2000/papers/206.pdf).

67 See Butler, D.: Souped-up search engines, in: Nature, vol. 405 (2000), pp. 112-115.

there is a real chance that searches on the Internet will be revolutionised within the next few years or that at least improved tools for scientific searches will be available.
However, search engines have already become an important resource for researchers and are used fairly often to retrieve information; for example a UK study carried out in spring 2000 returned the result that more than 50% of academic staff frequently used search engines. This means that they are the most frequently used tool of a total 18 different electronic information services included in the survey.[68]

Ad b) Science portals or information gateways set out to facilitate the network-based access to resources in a specific discipline or a limited field of research. Resources are organized and classified, their quality is checked and the resources are then supplemented with additional descriptions and information by librarians and information professionals. Users know that they are using quality controlled information sources. Examples of such gateways include SOSIG and EEVL in the U.K., DutchEss in the Netherlands or the Finnish Virtual Library Project.[69]

Important European research projects on information gateways are:

- The DESIRE project, which has developed a handbook to facilitate the setting up of information gateways in addition to an on-line tutorial for learning how to evaluate the quality of an WWW information resource. The project has also performed work on caching and information sharing.[70]
- The Renardus project, which sets out to create a European meta-gateway providing a single point of access to selected quality Web resources for academics. Users should benefit from enhanced services and existing gateways should benefit from shared solutions.[71]

Though in some countries such information gateways are strongly supported,[72] they do not seem to be used as much as search engines. The above cited JUSTEIS survey among academics only ranked scientific portals 11th out of 18 electronic information sources, with less than 10% of respondents being frequent users.[73] However, a literature survey of user satisfaction analyses carried out as part of the Renardus project concludes: “users appreciate quality resources: the evaluation and categorization of resources is of great importance to them. The actuality of electronic data is also important.”[74] This indicates that gateways provide different services than search engines and might be relevant for different user groups or types of information.

Of course, search engine sites and portal sites only constitute possible entrances to the ‘warehouses of scientific information’. The warehouses themselves are the Internet pages of digital collections, databases, electronic publishers, universities, other research organizations, individual researchers etc. While some of these have been discussed above, we have not yet mentioned academic websites. The European Survey of the Information Society (ESIS) II assessed the percentage of high schools and universities with Internet websites in Central and Eastern European countries.[75] The amount and type of information which is displayed on the Internet can be assessed by means of web content analyses (see section 2.2.1, p. 35). An

[68] See Armstrong, C. J. et al. (2000): JUSTEIS - JISC Usage Surveys: Trends in Electronic Information Services, table 6.15. (http://www.dil.aber.ac.uk/dils/research/justeis/cyc1rep6.HTM#6_8).

[69] See http://www.sosig.ac.uk, http://www.eevl.ac.uk, http://www.kb.nl/dutchess and http://www.uku.fi/kirjasto/virtuaalikirjasto.

[70] See http://www.desire.org and Place, E. (2000): Recent Results from the DESIRE project, in: Scimitar, vol. 16 (http://www.iihe.ac.be/scimitar/J0100/yuri_d2-bristol-ejpaper.html).

[71] See http://www.renardus.org/index.html and Heery, R.; Carpenter, L.; Day, M. (2001): Renardus project developments and the wider digital library context, in: D-Lib Magazine, vol. 7, no. 4 (http://www.dlib.org/dlib/april01/ heery/04heery.html).

[72] We owe this opinion to Ian Butterworth, one of our external experts who reviewed the draft of this report.

[73] See Armstrong, C. J. et al. (2000), op. cit.

[74] See Becker, H. J.; Klaproth, F.; Lepschy, P. (2000): Survey of end user surveys. Renardus Deliverable 6.6. (http://www.renardus.org/deliverables/d6_6/D6_6_final.pdf).

[75] See http://www.eu-esis.org/esis2www/synthCEEC7.htm.

evaluation of the educational web content of university sites was carried out by e.g. Chen et al.[76]

If a researcher has located the piece of information he/she needs, then there is usually little or no problem in obtaining it. Standards for storage formats and transfer routines have been developed: WWW browsers are equipped to read, print and download files, other file formats such as Portable Data Format (pdf) and Postscript (ps) can be accessed and used by means of widely available software, transfer of data is facilitated by other applications such as File Transfer Protocol (ftp) clients. The utilisation of these software tools has been assessed in various empirical studies (see p. 38).

2.1.3 Tools for computer-mediated communication (CMC)

Whereas the previous sections have been dealing mainly with information and different types of digital collections, the present section discusses the Internet-based tools for social communication between scientists. The term 'social communication' is used here for the intended transmission of information between human beings in which the receiver has been personally addressed by the sender. Other forms of communication, such as mass communication, where the general public or a selected part of it functions as the receiver, will *not* be considered in this section (however, section 2.2.3 on the publication and dissemination of research results is based on this form of communication).[77] For the new ways of social communication which have emerged since the beginning of the network revolution the term 'computer-mediated communication' (CMC) was coined. Sudweeks and Allbritton define CMC as:

> *"… human communication between two or more individuals through the use of central computers that store and process message content, and are connected to users in a communication network."*[78]

For our purposes we would have to add to this definition that CMC is '*intended* human communication between two or more individuals *in which the receiver has been personally addressed by the sender* through the use of central computers…'.

Empirical research has found some evidence that CMC has lowered many barriers to communication that are associated with geographic distance such as time, costs and language.[79] Thus it has led to an increase of communication among scientists. The most important and widely used new way of communication is electronic mail (e-mail). In addition, other forms of CMC such as mailing lists and on-line meetings are also used in R&D. They will be discussed subsequently.

E-mail

Some empirical investigations on the usage and effects of e-mail among scientists have been carried out over the last years based on questionnaires and interviews. The samples were usually rather small and the results should be considered as tentative and preliminary. In summary the following issues seem to be important:

a) E-mail is widely used for communication

[76] See Chen, C. et al. (1998): How did university departments interweave the Web: A study of connectivity and underlying factors, in: Interacting with computers, vol. 10, pp. 353-373.

[77] On the problems of finding a generally acceptable definition of communication see e.g. the 160 definitions listed in Merten, K.: Kommunikation. Eine Begriffs- und Prozessanalyse. Opladen 1977, pp. 168-182.

[78] Sudweeks, F.; Allbritton, M. (1996): Working together apart: Communication and Collaboration in a Networked Group. (http://www.it.murdoch.edu.au/~sudweeks/papers/acis96.html), also published in: C. D. Keen, C. Urquhart and J. Lamp (eds), Proceedings of the 7th Australasian Conference of Information Systems (ACIS96), Vol. 2, Department of Computer Science, University of Tasmania, pp. 701-712.

[79] See Walsh, J. P.; Bayma, T. (1996): The Virtual College: Computer-Mediated Communication and Scientific Work, in: The Information Society, vol. 12.

b) E-mail is considered as an important means of communication that adds value
c) E-mail utilisation has been found to be correlated to researchers' productivity

Ad a) The most robust result seems to be that e-mail has become a universal and important communication tool among scholars. Surveys which were carried out in the first half of the 90s already revealed high e-mail usage figures. In six small universities/colleges in the south-eastern U.S. 64% percent of the faculty claimed to be e-mail users as early as in 1993/94.[80] A survey carried out within a faculty in Israel in June 1995 returned an e-mail usage rate of about 78%.[81] But some contrary indications also exist: in a local survey among the faculty of a U.S. university more than 60% of the respondents stated that they never used e-mail for communication with colleagues.[82] However, this survey specifically investigated the extent of e-mail use for teaching. It had a rather small sample and it was carried out in the beginning of the 90s.

Taking the network externalities of e-mail into account (the benefits increase with the number of users), the percentage of scientists using e-mail should have increased notably in the meantime and we should expect usage rates of nearly 100%. A more recent survey from the U.K. also points in this direction.[83]

But empirical analyses also stress the finding that e-mail and face-to-face communication complement each other and are by no means substitutes.[84]

Ad b) Nearly all e-mail users consider it as important for their work.[85] The increased contact with scholars at other institutions, information about conferences, awareness of researchers at other institutions, research productivity, awareness of calls for papers and the amount of collaboration were stated as important benefits of e-mail use.[86] Advantages of e-mail compared to other means of communication are: answers to e-mail messages can be written whenever it is convenient, problems of conversing across different time zones or with people who travel a lot or spend much of their time in meetings are lower. Non-native speakers feel more comfortable when using a foreign language in written than verbal communication. Distinctions of status or age are less noticeable in written communication.[87]

Both issues listed under a) and b) are important regarding the hypothesis advanced a few years ago and that has become known under the term of '*balkanization of science*'.[88] It

---

80 See Liebscher, P.; Abels, E. G.; Denman, D. W. (1997): Factors that influence the use of electronic networks by science and engineering faculty at small institutions. Part II: Preliminary use indicators, in: Journal of the American Society for Information Science, vol. 48, no. 6, p. 499.

81 See Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997): Internet use by faculty members in various disciplines: a comparative case study, in: Journal of the American Society for Information Science, vol. 48, no. 6, p. 512.

82 See Mitra, A.; Hazen, M.D.; LaFrance, B.; Rogan, R.G. (1999): Faculty Use and Non-Use of Electronic Mail: Attitudes, Expectations and Profiles, in: Journal of Computer-Mediated Communication, vol. 4, no. 3 (http://www.ascusc.org/jcmc/vol4/issue3/mitra.html).

83 See Day, J.; Bartle, C. (1998): The Internet as an Electronic Information Service: Its Impact on Academic Staff in Higher Education. Proceedings IRISS '98 International Conference: 25-27 March 1998, Bristol, UK. (http://sosig.ac.uk/iriss/papers/paper06.htm).

84 See Koku, E.; Nazer, N.; Wellman, B. (2000): Netting Scholars: Online and Offline, in: American Behavioral Scientist, vol. 43. (also available at: http://www.chass.utoronto.ca/~wellman/publications/nettingscholars/scholnet-abs9a.pdf) – Rallet, A.; Torre, A. (1999): Is Geographical Proximity Necessary in the Innovation Networks in the Era of the Global Economy, in: GeoJournal, vol. 49, no. 4, pp. 373-380.

85 See Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997), op. cit., p. 512. - Lubanski, A.; Matthew, L. (1998). Socio-economic Impact of the Internet in the Academic Research Environment. Proceedings IRISS '98 International Conference: 25-27 March 1998, Bristol, UK. (http://sosig.ac.uk/iriss/papers/paper18.htm.

86 See Walsh, J.P.; Maloney, N. G. (2001): Computer network use, collaboration structures and productivity, in: P. Hinds and S. Kiesler (ed.): Distributed work. Cambridge, Mass. – Cohen, J. (1996): Computer mediated communication and publication productivity among faculty, in: Internet Research: Electronic Networking Applications and Policy, vol. 6, no. 2/3, pp. 55.

87 See OECD (1998), op. cit., chapter 'Communication among scientists'. – Sanderson, D. (1996): Cooperative and collaborative mediated research, in T. M. Harrison and T. D. Stephen (eds): Computer Networking and Scholarly Communication in the 21st Century University, New York, pp. 106-107.

88 See Van Alstyne, M.; Brynjolfsson, E. (1996): Could the Internet Balkanize Science?, in: Science, vol. 274, no. 5292, pp. 1479-1480.
We thank our external reviewers, Loet Leydesdorff and Ian Butterworth, for commenting on this issue.

proposes that local and discipline-transgressing communication between scholars is increasingly replaced by long-distance communication within the discipline or is even restricted only to the specific specialisation.[89] However, it is far beyond the scope of this review to investigate whether a connection exists between e-mail as a form of long-distance social communication and the interrelations among academics at one location, and what the causes of changes might be.[90]

Ad c) The relation between a scholar's e-mail use and his/her individual productivity and productivity in collaborations (measured as the number of co-authored publications) has been found to be positive.[91] But there exists multiple feedback to the social structure of a collaboration and reverse causality might constitute a problem in the cited analyses (i.e. more productive researchers might have larger professional networks which require them to communicate more).[92]

Mailing lists

Mailing lists are lists of electronic addresses of individuals who are interested in receiving and/or providing information on a specific subject or a set of subjects attributed to the topic of the list via electronic mail.[93] These electronic blackboards for information can simply serve to announce events and distribute information such as new publications, job openings or calls for papers, or they can function as platforms for exchanging support or advice and for discussing scientific problems. In principle synchronous communication and immediate response to a message is possible. But more often communication will be asynchronous and each receiver will read and respond to a message at his/her own convenience. Mailing lists are located on the intersection between social communication and mass communication and it is often difficult to define whether they belong to one or the other.[94] This depends on the intention of the list owner, the kind of e-mails that are sent to the list and also the underlying relations between the subscribers. Gresham lists a variety of names "electronic conferences, e-conferences, computer conferences, mailing lists, lists, listservs, electronic forums, online discussion groups, scholarly discussion groups, special interest groups, news groups, and netgroups" and uses 'computer conferencing' as a generic term.[95] We will use 'mailing lists' as this term is the best description of what is generally taking place, i.e. the sending of e-mails to a list of subscribers.

Empirical analyses of mailing lists are still rather scarce. Matzat summarises some insights from case studies which show that users of mailing lists evaluate them as useful for the creation of new research ideas, for the transfer of information and for getting into contact with peers.[96] The low quality of the discussions within mailing lists has been criticised. They apparently are not very well suited to the discussion of controversies and established researchers hesitate to use them.[97] Rojo's empirical results indicate that the number of participants, the percentage of regular contributors and the commitment of the list owner are

---

89 Ibid. – Noam, E. M. (1995): Electronics and the Dim Future of the University, in: Science, vol. 270, pp. 247-249.

90 Interesting contributions on this issue are the above cited articles from Van Alstyne and Brynjolfsson, Noam, and the book from Gibbons, M. et al. (1994), op. cit.

91 See Walsh, J.P.; Maloney, N. G. (2001), op. cit. - Cohen, J. (1996), op. cit. - Hesse, B. W. et al. (1993): Returns to science: Computer networks in oceanography, in: Communications of the ACM, vol. 36, no. 8, p. 97.

92 See Walsh, J.P.; Roselle, A. (1999): Computer Networks and the Virtual College, in: STI Review No. 24: Special Issue on The Global Research Village. Paris, p. 66.

93 See the definition in Matzat, U. (1998): Informal Academic Communication and Scientific Usage of Internet Discussion Groups. Proceedings IRISS '98 International Conference: 25-27 March 1998, Bristol, UK. (http://sosig.ac.uk/iriss/papers/paper19.htm)

94 Rafaeli (1993) has developed a concept of collaborative mass media for these forms of communication, cited in: Rojo, A. (1995): Participation in scholarly electronic forums. (http://www.digitaltempo.com/e-forums/Oframes.html)

95 Gresham, J. (1994): From invisible college to cyberspace college: computer conferencing and the transformation of informal scholarly communication networks. (http://www.helsinki.fi/science/optek/1994/n4/gresham.txt).

96 See Matzat, U. (2001): Academic Communication and Internet Discussion Groups. Their Spread, Use and 'Survival' within Academic Communities. (http://www.ppsw.rug.nl/matzat/short-project-description.html).

97 Ibid.

important determinants of a list's success.[98] Better participation at information flows and integration into the scientific community were perceived as benefits from the participants whereas the lists obviously seldom raise the interaction and collaboration among scholars.[99] Gresham notes that scholars use mailing lists to informally exchange different types of information and advice.[100] Zelman and Leydesdorff compare project-related, (research) field-wide and intermediate (national) mailing lists and make the distinction between formal and informal communication and whether this is transferred to CMC.[101] They also look for self-organizational properties of mailing lists.

The empirical evidence on mailing lists has usually been based on one or both of the two following sources:

- questionnaire-based surveys and
- email messages within the mailing lists which can be obtained from the list servers.

On-line meetings, seminars and conferences

On-line meetings, seminars and conferences can be considered as a new form of synchronous social communication. On the one hand their objective is to create a 'virtual meeting room' with participants from geographically distributed locations being present at the same virtual location. Or as awareness tools they can tell the members of distributed work groups if their colleagues are present and available.[102] On the other hand, as a substitute for personal meetings on-line meeting tools must also offer at least similar possibilities of using various media and aids: shared textual, numeric or visualised information, including a real time display of modifications, addition of further information sources (without having to transfer data, software or other constituents first) etc.

Until now only case studies on the strengths and weaknesses of on-line meetings exist. Some of these use NetMeeting, a shareware that can be downloaded free from Microsoft. It enables the participants of an on-line meeting to simultaneously view and modify documents or other applications. Furthermore, it provides different features such as a whiteboard, chat, file-transfer, and audio and video transmission capabilities, though the latter of course requires additional apparatus such as microphones and webcams. The video and audio features were not utilised in the two cases that will be described subsequently. Instead, telephone conferencing was employed as the channel for synchronous communication. Finholt et al. state as a result of their case study on the implementation of NetMeeting in a software development group that participants to the field trial were "moderately satisfied" (ranking of 6.5 on a 10 point response scale) with NetMeeting.[103] A number of technical weaknesses such as lack of speed and specific characteristics of the program and organisational problems, e.g. delays in having the software installed, lack of a default server or incompatible program versions created inconveniences for the users. Mark, Grudin and Poltrock also received rather positive feedback from their participants at Boeing, giving NetMeeting a 4.8 rating averaged across satisfaction questions (on a scale of 1 to 6, strongly disagree to strongly agree).[104] The decrease of travelling activities can be regarded as additional proof of

---

98 Rojo, A. (1995), op. cit.

99 Ibid.

100 See Gresham, J. (1994), op. cit.

101 See Zelman, A.; Leydesdorff, L. (2000): Threaded Email Messages in Self-Organization and Science & Technology Studies Oriented Mailing Lists, in: Scientometrics, vol. 48, no. 3, pp. 361-380.

102 See Wellman, B. (2001): Computer networks as social networks, in: Science, vol. 293, 14. Sep. 2001, pp. 2031-2034.

103 See Finholt, T.A.; Rocco, E.; Bree, D.; Jain, N.; Herbsleb, J.D. (1998): NotMeeting: A field trial of NetMeeting in a geographically distributed organization, in: SIGGROUP BULLETIN, vol. 20, no. 1, pp. 66-69. (http://www.bell-labs.com/org/11359/colab_prod/).

104 See Mark, G; Grudin, J.; Poltrock, S. E. (1999): Meeting at the Desktop: An Empirical Study of Virtually Collocated Teams, in: Proceedings of ECSCW'99, The 6th European Conference on Computer Supported Cooperative Work, 12-16 September 1999, Copenhagen, Denmark, pp. 159-178. (http://research.microsoft.com/users/jgrudin/)

the usefulness of the on-line meeting tool.[105] Nevertheless, besides technology-related shortcomings this study also highlighted problems which come from not being collocated, such as co-ordinating the interaction and maintaining the users' attention.

Studies on the effects of video in different work settings and group constellations are ambiguous in their results regarding whether, how and under what circumstances video broadcasting adds to the quality of work and improves the output.[106]

The case studies and descriptions of different tools for on-line meetings and remote collaboration create the impression that there are still several technical shortcomings relating to hardware and software and that the proficiency of technical staff and users with the new technology is rather low. It might be too early to generalise the findings and to try substantiate them with quantitative data.

2.1.4 Grid technologies

The different elements discussed in the previous three parts of this section combine to form a powerful new development in the field of scientific computing which has become known under the term of 'grid computing' (synonyms are metacomputing, seamless scalable computing, global computing).[107]

The history of grid computing started in the 80s with the clustering of individual computers as a technique to solve the problem of insufficient computer power.[108] Though large capacities could be created in this way, it remained a relatively expensive and technically demanding way which was not feasible for many smaller and less wealthy research organisations. The Entropia network offered a new solution by linking idle computers via the Internet to solve computational problems.[109] It was established only in 1997 but it benefited from earlier pioneering projects such as Condor.[110] More distributed computing applications followed, such as SETI@home (search for extraterrestrial intelligence), FightAids@home, Compute-Against-Cancer, GIMPS (search for large prime numbers) and many more.[111] Of course, the success of these distributed computing projects has been based on their appeal to the general public and its willingness to dedicate their computers' spare time for scientific calculation.

Drawing from the experience with distributed computing, grid computing has begun to emerge as a new important field of scientific computation. It can be distinguished from conventional distributed computing by "its focus on large-scale resource sharing, innovative applications, and, in some cases, high-performance orientation."[112] The Grid has been called prosaically a

105 This finding of the investigation cited above from Mark, Grudin and Poltrock should, however, not be interpreted as an indication of a general decrease in travelling due to CMC. On the contrary, particularly in R&D it is expected that long-distance collaboration with a lot of CMC actually increases the incentives to travel, see e.g. Rallet, A.; Torre, A. (1999), op. cit., pp. 375-376.

106 See Olson, G.M.; Olson, J.S. (2002): Distance matters, in: John M. Carroll (Ed.): Human-Computer Interaction in the New Millennium (forthcoming), pp. 139-179.

107 See Baker, M.; Buyya, R.; Laforenza, D. (2000): The Grid: international efforts in global computing. Presentation at the SSGRR International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, L'Aquila, Jul 31 - Aug 06 2000. (http://www.ssgrr.it/en/ssgrr2000/papers/268.pdf).

108 See on the following issues: Foster, I.: Internet Computing and the Emerging Grid, in: nature web matters, 7 December 2000. (http://www.nature.com/nature/webmatters/grid/grid.html)

109 See http://www.entropia.com and.

110 See http://www.cs.wisc.edu/condor.

111 See http://setiathome.ssl.berkeley.edu; http://www.fightaidsathome.org; http://www.parabon.com/cac.jsp; http://www.mersenne.org/prime.htm

112 Foster, I; Kesselman, C.; Tuecke, S. (2001): The Anatomy of the Grid. Enabling Scalable Virtual Organisations, to appear in: International journal of Supercomputer Applications, p. 1. (also available at: http://www.globus.org/research/ papers/anatomy.pdf).

"seamless, integrated computational and collaborative environment".[113] It is based on four different types of infrastructure elements:[114]

(1) Grid Fabric: geographically distributed computational, storage and network resources, special scientific instruments, databases and operating systems,
(2) Grid Middleware: software that manages the usage of the Grid Fabric layer, e.g. remote process management, storage access, information (registry), security, authentication and Quality of Service,[115]
(3) Grid Development Environments and Tools: offer programmers the opportunity to develop applications for the Grid and permit brokers to manage and schedule computations and other procedures on the Grid,
(4) Grid Applications and Portals: the user interface where jobs for the Grid can be submitted and the results be collected.

Grids are relevant at all stages of a research project: they can be used for remote data collection, analysis and computation, presentation and discussion of results. Their basic aim is to create the same high-level research environment for distributed research groups that is available to collocated groups (or even better considering that all resources are linked up to each researcher's desk). Some authors distinguish between two different types of Grids:[116] computational Grids which offer access to almost unlimited computing and distributed data resources, and access Grids which provide a group collaboration environment.

The innovative character of the Grid results from its novel communication, authentication, information and management protocols and the new organisational forms of scientific research they support (see section 2.3.1). Early research projects that have produced important Grid middleware were carried out in the U.S.: The Globus Project developed the 'Globus Toolkit', software tools that make it easier to build computational Grids and Grid-based applications. This includes tools and libraries for solving problems in various areas, e.g. information infrastructure, communication, resource and data management etc.[117] The Legion project has a similar goal but it employs a different, object-oriented approach.[118] A multitude of further Grid R&D and implementation projects is currently underway.[119] The EU included the support of Grid technologies in its *e*Europe initiative[120] and supports Grid research and development projects within the IST programme. Important European Grid projects are:

- the DataGrid project which aims to develop an environment to support globally distributed scientific exploration involving very large data sets. It focuses on Grid applications for three research fields: high energy physics, earth observation sciences and bio-informatics;[121]
- the EUROGRID project which aims to establish and operate a Grid infrastructure, develop software components and demonstrate distributed simulation codes from different

---

113 Smarr, L. (2000): Infrastructure for Science Portals, in: IEEE Internet Computing, January/February 2000, pp. 71-73. (http://www.computer.org/internet/v4n1/smarr.htm).

114 See Baker, M.; Buyya, R.; Laforenza, D. (2000), op. cit.

115 This layer is sometimes further differentiated into a 'communication layer' containing the core communication and authentication protocols and a 'resource layer' containing the management of individual resources, see Foster, I; Kesselman, C.; Tuecke, S. (2001), op. cit.

116 See Smarr, L. (2000), op. cit.

117 See http://www.globus.org/toolkit/default.asp

118 See http://legion.virginia.edu/index.html

119 See for an overview Baker, M.; Buyya, R.; Laforenza, D. (2000), op. cit. and the link lists at http://www.gridforum.org, http://www.gridcomputing.com and http://www.globus.org/about/related.html; for European projects: European Research Consortium for Informatics and Mathematics (ERCIM): Grids: e-science to e-business., www.ercim.org) News, No. 45, April 2001.

120 See European Commission (2000b): *e*Europe 2002: An Information Society for all. Action plan prepared by the Council and the European Commission for the Feira European Council, 19 and 20 June 2000, p. 9.

121 See http://www.eu-datagrid.org and for a description: Segal, B. (2000): Grid Computing: The European Data Grid Project, presentation to the IEEE Nuclear Science Symposium and Medical Imaging Conference, Lyon, 15-20 October 2000. (http://web.datagrid.cnr.it/pls/portal30/docs/1442.DOC).

application areas (biomolecular simulations, weather prediction, coupled CAE simulations, structural analysis, real-time data processing);[122]

- the DAMIEN (Distributed Application and Middleware for Industrial Use of European Networks) project which deals with the problem of how to develop applications for computational Grids.[123]

The scientific discussion on Grid computing focuses on the further development and improvement of the Grid layers. We have not found any approaches which tried to evaluate the participation of different countries within Grid computing and development.

2.1.5 Computer skills for R&D

Anyone who has had problems convincing his/her computer to carry out the submitted tasks knows that it isn't sufficient to have capable hardware and software: to use computers successfully you also need 'humanware', i.e. expertise and experience of working with them.

The lack of ICT skills in research has been remarked upon repeatedly. At its second conference on the Global Research Village in Sintra in 1998, the OECD agreed upon the promotion of training of scientists in the use of ICT as one of the tasks which governments should fulfil to maximise the benefits of ICT for science.[124] A high-level consulting group on ICT and European science stated that "there is a need to train scientists in these new tools and techniques in order to fully exploit the potential benefits of increasingly sophisticated ICT."[125] The European Commission included researcher awareness-raising and training regarding the possibilities created by ICT among its European Research Area (ERA) policy goals, which aim to make better use of the potential offered by electronic networks.[126] The second communication on ERA introduced an even broader approach which highlighted the necessity "to support the development of the expertise and skills needed in a knowledge-based society."[127]

The computer skills available to R&D may be assessed by means of four different approaches:

a) Assessment of the computer skills levels of R&D personnel
b) Evaluation of the training activities on computers for R&D personnel
c) Assessment of the specialised computer staff available for R&D
d) Assessment of the computer skills bought from external providers

We do not know of any operationalisations or investigations regarding points a) and b). The possible methods for performing this will be discussed in the next report, deliverable 2.2.

Ad c) Though computer staff might be indispensable for performing a research project, the rather narrow OECD classification in most cases leads to an exclusion of computer staff from R&D personnel in the private business sector. The OECD proposes a rule of thumb for deciding whether to include supporting scientific and technological (secondary) activities within R&D: When a secondary activity is primarily undertaken in the interests of R&D, it should be included; if it is designed essentially to meet needs other than R&D, it should be

---

122 See http://www.eurogrid.org.

123 See http://www.hlrs.de/organization/pds/projects/damien.

124 See Aubert, J.E.; Bayar, V. (1999): Maximising the Benefits of Information Technology for Science: Overview and Major Issues, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, pp. 24-25.

125 European Technology Assessment Network (1999), op. cit. p. 57.

126 European Commission (2000a): Towards a European research area. Communication from the Commission COM 2000 (6). Brussels, 18 January 2000, p. 11. (http://europa.eu.int/comm/research/ area/com2000-6-en.pdf).

127 European Commission (2000b): Making a reality of the European research area: guidelines for EU research activities (2002-2006). Communication from the European Commission COM(2000) 612 final. Brussels, 4 October 2000, p. 17. (Http://europa.eu.int/comm/research/area/com2000-612-en.pdf).

excluded.[128] Consequently, in-house computing services to R&D provided by specific R&D computing units should be included among R&D personnel, but specific services to R&D provided by central computer departments should be excluded.[129] The exclusion of computer staff at central computer departments might produce a bias in the R&D personnel data depending on the organisation structure of its computing inputs.[130]

Implementations of the OECD classification are not sufficiently detailed to contain differentiated data for the proportion of R&D personnel in the private business sector who provide specific computing services.[131] To quantify the staff responsible for computer services in universities is particularly difficult. There are many units that provide these services and the provision even within an individual university might be very decentralised.

Ad d) Of course, to employ specialised computer staff is only one of various solutions to acquire the computer services which are needed for R&D. Other solutions might be to obtain these services from external providers and other ancillary institutions: For example, the transmission of data on the national or international level is usually managed by research network organisations (see section 2.1.1) and R&D institutions have to pay if they want to be serviced. Another example are centralised electronic library services (see section 2.1.2 on textual databases), whose staff makes on-line information resources available.

No attempt to quantify these external and computer-related services to R&D in total has yet been published.

While in this section we discussed the different existing elements of Internet infrastructure for R&D as well as new developments, effects on R&D output and available approaches for measuring endowment and performance, we will take a different perspective in the next section.

2.2 Integration of the new network technologies into research activities

The present section develops a process-oriented view of the usage of the Internet for R&D. It looks at separate steps of the research process and specifies where researchers use the Internet to carry out tasks related to their research work. For this purpose three different steps will be distinguished:

- Data collection and information retrieval,
- data analysis and technology development,
- publication and dissemination of research results.

We would like to stress that we do not assume an underlying linear model of R&D. Multiple relations exist between the three different stages and some loops may be necessary before a final research result can be produced. For example, a natural sciences research project may construct a model for analysing a research problem (probably based on existing models which have been evaluated before), build an apparatus for data collection (maybe itself based on the gathering of information on existing instruments and their capabilities), collect data and feed it into the model and take one or more loops to improve the model, apparatus or collected data. They may then publish a draft of intermediate results, receive criticism and

---

[128] OECD (1994), op. cit., p. 37.

[129] The approach is broader, when we look at R&D expenditures: the OECD proposes the inclusion of the costs of services to R&D provided by central computer departments as overhead costs, see OECD (1994), op. cit., p. 79.

[130] The OECD itself concedes the existence of such a bias in another context mentioning an underestimation of researchers in the United States due to the exclusion of military personnel in the government sector, see OECD (2001), op. cit., p. 50.

[131] E.g. the German Stifterverband differentiates in its national R&D surveys only among scientists and engineers, technical staff and other R&D personnel, see: Grenzmann, C.; Marquardt, R.; Wudtke, J. (1999): Forschung und Entwicklung in der Wirtschaft 1997 - 1999. Bericht über die FuE-Erhebungen 1997 und 1998. Essen 1999, Appendix, p. 35.

improve the model (or the data or the data collecting apparatus) and then publish the next version of results etc. This is just an example intended to highlight the fact that the R&D process is much more complex than the structure of the present section implies. However, it is necessary to reduce this complexity for analytical purposes and we will therefore continue with this rather under-determined scheme of the research process.

Our analysis will concentrate on two issues: On the one hand it looks at the indicators that are used to measure and describe the usage of Internet infrastructure and applications. On the other hand it investigates the effects of the Internet usage on R&D output, e.g. whether on-line data and information retrieval generates positive impacts on R&D. Such positive impacts could be an increase in researcher productivity, a decrease in R&D costs etc. Recent studies have shown that the use of the Internet among scientists is increasing.[132] In general it is assumed to improve the quality of research[133] though some problems are also mentioned such as the overwhelming quantity of information, its partially poor quality, the danger of plagiarism, the low speed of access and the large degree of commercialisation.[134] A positive attitude towards Internet applications, promotion of their further development and acceptance by R&D policy makers can only be justified, if positive effects can be confirmed.

2.2.1 Data collection and information retrieval

Research projects typically start with a review of the scientific literature on their research topic. Thus they monitor the current knowledge and create an information base which serves

- to formulate the research problem(s) and question(s),
- to formulate hypotheses in order to answer the research question(s),
- to select available methods appropriate for collecting and evaluating the data needed and
- to assess their peers' research results.

The task of information collection is necessary to formulate relevant and precise research orientations and to avoid the excessive duplication of research. The above listed third step is typically only needed in empirical research projects, i.e. research that also aims at confronting its theoretical assumptions with empirical data. To understand the structure of the present chapter it is necessary to point out that we make a distinction between 'information' and 'data'. Cowan, David and Foray define information as follows:[135]

"We find it useful to operationally define an item of information as a message containing structured data, the receipt of which causes some action by the recipient agent – without implying that the nature of that action is determined solely and uniquely by the message itself."[136]

So, information is more than written, verbal or visual records or bits and bytes in an electronic data storage device. It is *structured* data that also needs some additional knowledge, usually in the human receiver's mind, to be understood and dealt with properly. We will use this differentiation between data and information to structure this chapter. We start with a discussion of data collection by means of computer-networks. In the second section we will

---

[132] See Lubanski, A.; Matthew, L. (1998), op. cit.

[133] Ibid.

[134] See Day, J.; Bartle, C. (1998), op. cit.

[135] The definition resembles a lot Porat's definition in his seminal work on the information economy: *"Information is data that have been organized and communicated. The information activity includes all the resources consumed in producing, processing and distributing information goods and services."* Porat, M. U. (1977): The information economy: definition and measurement. OT special publication 77-12 (1).
On alternative definitions see Machlup, F. (1962), op. cit., p. 8. - Shapiro; C.; Varian, H. (1999): Information rules. A strategic guide to the network economy. Boston, p. 3.

[136] Cowan, R.; David, P. A.; Foray, D. (2000), op. cit., p. 216.

take existing information as the starting point and see how the Internet has affected the storage and the retrieval of information that is used for R&D.

Interactive and non-interactive data collection

Network-based data collection might employ (a) different forms of *interactive* methods, i.e. the respondent knows that he is disclosing data as he answers some sort of question. It can also use (b) *non-interactive* research methods which are carried out using computer networks. In these cases the supplier of information may or may not know about the data collection, e.g. depending whether she/he has been informed in advance.

Ad (a) It is possible to *collect data interactively* via the Internet in various ways: by means of e-mailed or on-line questionnaires, on-line experiments or chat interviews.[137] The on-line methods have some advantages compared to traditional off-line methods: They can reach a larger audience and be less expensive as no costs for mailing or expensive interviewers accrue – but these savings might be upset by large programming expenses, so all in all the costs vary markedly. The research results might be less biased by the setting, as a laboratory situation is avoided and it can be carried out double blind. Also, the results are immediately accessible and further data processing is cheaper. A lot of debate and research has been going on regarding how representative and valid on-line methods are compared to their off-line counterparts. But as a high utility has been found,[138] on-line methods have generally been accepted.

Therefore, it is not astonishing that the popularity of these methods has increased rapidly with the spread of the information society. Not only those having a professional interest in the new media such as computer scientists and communication researchers embraced interactive on-line research methods, but also psychologists, social scientists, economists and last but not least the guild of market and opinion researchers. Computer programs and tools have been developed to support the realisation of on-line data collection taking into account the different research disciplines' interests.

The variety of on-line research methods and their usage in different research disciplines has not been a good precondition for the development of indicators which permit a comparison of national research systems in this respect. An example of how the usage of on-line research methods could be assessed was carried out by the owners of the 'German Internet Research' (gir-l) mailing list. They asked their subscribers in an e-mail-based survey to estimate their past usage of different on-line research methods.[139] The results indicate that surveys (web-based and e-mailed) were the most important methods. Log file analysis, that will subsequently be discussed as a non-interactive method, has also been performed many times.

Ad (b) *Non-interactive data collection* from computer networks makes use of the fact that data is stored on different kinds of servers. Webopedia defines a server as:

> *"A computer or device on a network that manages network resources. For example, a file server is a computer and storage device dedicated to storing files. Any user on the network can store files on the server. A print server is a computer that manages one or more printers, and a network server is a computer that manages network traffic. A database server is a computer system that processes database queries."*[140]

[137] Regarding the form of on-line surveys, see: MacElroy, B. (1999): Comparing seven forms of on-line surveying, in: Quirk's Marketing Research Review, article no. 510. (http://www.quirks.com/articles/article_print.asp?arg_articleid =510).

[138] See Krantz, J. H.; Dalal, R. (2000): Validity of Web-based psychological research, in: M. H. Birnbaum (Ed.): Psychological Experiments on the Internet. San Diego, CA.

[139] See Bosnjak, M. et al. (1998): On-line Forschung im deutschsprachigen Raum. Erste Ergebnisse einer Umfrage unter Mitgliedern der 'German Internet Research' Mailingliste. (http://www.or.zuma-mannheim.de/inhalt/projekte/or_expert/girl98_1.pdf).

[140] http://www.webopedia.com/TERM/s/server.html

Both (ba) the content on and (bb) the usage of Internet servers (the 'traffic') are appropriate data sources for on-line research. Different methods are needed to collect data from these resources.

Ad (ba) Internet *content* is presented in different networks of which the World Wide Web is the largest. The possibility of collecting data from the WWW via browsers and search engines has led informetrics, the scientific discipline which investigates and measures issues around information, to focus attention on the new medium. 'Webometrics' appeared "… which covers research of all network-based communication using informetric or other quantitative measures."[141] Informetric methods until then had been used for a variety of research questions and objectives such as statistical aspects of word and phrase frequencies in communications and sociological research, publication and citation indexes to assess the productivity of scientists or the impacts of scientific publications.[142] In general informetric methods applied to the World Wide Web make it possible to find out how many times a certain piece of information is contained in any available data set from the WWW. Such data sets might be individual hosts, web pages or any database that mirrors (parts of) the WWW such as the index database of a search engine (for more on search engines see below).
Interesting examples of webometric analyses have chosen the latter approach. They implement different algorithms that query the host, title, text, links or other recorded features of web pages. They investigate

(1) the presence of countries, important centres of education, scientific domains, document types and content on the WWW,
(2) the calculation of 'web impact factors' by counting the amount of links that point to country domains and hosts on the web or to scientific journals,
(3) the importance and development of university – industry – government relationships in national innovation systems (as reflected by search engines).

Ad (1) Almind and Ingwersen took two search engines' (Lycos, OpenText) web indexes to evaluate how different items of information are represented on the WWW.[143] For our purposes, the methods employed to analyse the presence of countries, centres of education and scientific domains are particularly interesting.

Ad (2) The 'web impact factors' (Web-IF) calculated by Ingwersen are based on the assumption that the impact of a web page is reflected by the number of pages that link to it.[144] This parallels a common approach of scientometric citation analysis. To calculate an appropriate indicator for a country's impact on the WWW, search routines were developed that retrieve all web pages that link to that country. The number of link pages was then divided by the number of pages found in its domain. Ingwersen thus calculated comparable figures of the web-presence of a country for a certain period of time using the AltaVista search engine (see table 4). By differentiating between pages that link to other pages within a country and pages that link to them from abroad, self-link and external-link Web-IFs are computed. The global Web-IFs in the bottom line of table 1 serve as a benchmark. A similar analysis was carried out by Wormell for assessing the impact of scientific journals.[145]

[141] Almind, T.; Ingwersen, P. (1997): Informetric analyses on the World Wide Web: methodological approaches to 'webometrics', in: Journal of Documentation, vol. 53, no. 4, p. 404.

[142] See Tague-Sutcliffe, J. (1992): An introduction to informetrics, in: Information Processing & Management, vol. 28, no. 1, pp. 1-3.

[143] See Almind, T.; Ingwersen, P. (1997), op. cit.

[144] See Ingwersen, P. (1998): The calculation of web impact factors, in: Journal of Documentation, vol. 54, no. 2, pp. 236-243.

[145] See Wormell, Irene (1998): Informetric Analysis of the international impact of scientific journals: How 'international' are the international journals?, in: Journal of Documentation, vol. 54, no. 5, pp. 584-605.

Table 4: Selected national impact factors for the WWW: Web-IF (20 Aug. – 21 Sep. 1997)

Domains in rank order	Web impact factor	Web-IF self link	Web-IF external link	Number of web pages
.no (Norway)	1.113	0.49	0.62	218,141
.uk (United Kingdom)	0.994	0.46	0.53	1,046,961
.fr (France)	0.886	0.42	0.46	454,822
.dk (Denmark)	0.886	0.52	0.37	144,433
.se (Sweden)	0.866	0.51	0.36	153,267
.fi (Finland)	0.823	0.43	0.39	317,829
.jp (Japan)	0.404	0.31	0.09	1,826,051
.gov (government)	1.472	0.42	1.05	646,585
.org (organisations)	1.186	0.40	0.78	1,677,934
.com (business)	0.942	0.59	0.35	12,084,719
.edu (academic)	0.807	0.47	0.33	5,390,097
Total domains of sample	**0.899**	**0.51**	**0.39**	**22,497,477**

Source: Ingwersen (1998), op. cit., p. 240.

Ad (3) The recent analysis from Leydesdorff and Curran can be taken as a good example for the appliance of webometric methods to analyse issues of research and development.[146] Based on the hypothesis that two different modes of production of scientific knowledge exist, a traditional and disciplinary one ('Mode 1') and a novel, trans-disciplinary, application- and problem-oriented one ('Mode 2'),[147] the spread of the latter was analysed by means of webometric methods. To count the university-industry-government relations (Triple Helix) complex queries were launched using various features of the AltaVista search engine.[148] They concluded that it was possible to measure Triple Helix relations via the Internet and that the WWW can be considered as a less codified counterpart of traditional scientometric sources (e.g. scientific journals).

Another type of web content analysis was carried out by Chen et al. who evaluated the quality of the computer-based course material of selected US American universities using a pre-established assessment scheme.[149]

Ad (bb) *Traffic:* The usage of the Internet and other networks is recorded in so-called log files. They record all requests to a server to disclose information or to collect it via a network. These log files are located at the requesting server (client-side log), at the responding server (Web server log) or at an intermediary server that is used to improve performance or to filter access to the network (proxy server log). Log files have been used for scientific analysis in different contexts, e.g.:

- to evaluate the design of Web sites,[150]
- to analyse the frequency of usage of different Internet applications,[151]

---

[146] See Leydesdorff, L.; Curran, M. (2000): Mapping university-industry-government relations on the Internet: the construction of indicators for a knowledge-based economy, in: Cybermetrics, vol. 4, no. 1, paper 2 (http://www.cindoc.csic.es/ cybermetrics/articles/v4i1p2.html).

[147] See Gibbons et al. (1994), op. cit.

[148] Such useful features are domain names (e.g. 'domain:ch' for Switzerland), text within links ('link:fhso.ch' to retrieve all pages that link to the University of Applied Sciences Solothurn), language and date entries or simple text entries ('text:university').

[149] See Chen, C. et al. (1998), op. cit.

[150] See Burton, M. C.; Walther, J. B. (2001): The Value of Web Log Data in Use-Based Design and Testing, in: Journal of Computer-Mediated Communication, vol. 6, no. 3. (http://www.ascusc.org/jcmc/vol6/issue3/burton.html).

- to assess the usage of digital libraries and electronic publications.[152]

However, log file analysis for purposes other than gauging the demand on a server is still in its infancy. Before comparable statistics from log files can be calculated, many problems will have to be solved: some are of a technical nature, such as the caching of pages on browsers or proxy servers which improves the performance but leads to lower web server usage figures; or the activity of search engine crawlers and robots which inflates the usage figures of those sites that don't screen them or prohibit access. Other problems derive from a lack of standards in web server logging tools. These are programs that automatically analyse the server log files and process the data for human interpretation. The different commercial and shareware programs which are in use perform different tasks, employ differing definitions and return results which are not comparable. Finally, a last category of 'problems' results from the participation of human beings (!): Individuals might access digital collections from different addresses, downloaded files can be distributed via e-mail, search frequencies might reflect the amount of data absorbed or the user's inability to handle the search routines etc. All these problems might lead to log files that contain only fragments of the real usage of Internet sources by scientific users.

Other problems such as the protection of privacy and the methods of data collection by statistical bodies haven't been mentioned yet. So it is not astonishing that log files have not been used for comparing the usage of the Internet in different national research systems. Many challenges have to be faced before this will be possible.

The scientific discussion on interactive and non-interactive data collection methods concentrates on their applicability, strengths and weaknesses and details regarding proper implementation. The meta-perspective, i.e. to what extent they are employed in different research communities, has not yet been adopted. Therefore, the description above did not yield large returns for constructing a SIBIS indicator on R&D data collection. However, it was worthwhile from a methodological point of view: the SIBIS project itself might use the methods described for gathering empirical data to build its indicators. This option will be discussed in more detail in the next report, deliverable 2.2.

Information retrieval

There are two different starting points for information retrieval from the net: Either the scientific searcher knows where to look for the required information when he/she enters the Internet due to his/her education, professional experience, and real-world information environment (books, journals, colleagues etc.) or he/she does not and consequently starts trying to locate a relevant source. If the location of a piece of scientific information is known, then different applications such as WWW-browsers, ftp-clients, pdf- and ps-readers are available to retrieve it. If the location is not known, search engines are an indispensable tool for finding information items on the Internet (see section 2.1.2, p. 23).

Two approaches have been carried out to analyse the extent and the effects of information retrieval from the net on R&D:

a) Assessments of the significance of information retrieval from the Internet in general
b) Investigations of the citation frequencies for e-journals compared to off-line journals

Ad a) Different surveys confirm that the Internet has facilitated the access to information in general and to publications in particular and, that it has increased the awareness of relevant research and shortened the time required to obtain it. In a London School of Economics'

---

[151] See Kaminer, N.; Braunstein, Y.M. (1998): Bibliometric analysis of the impact of Internet use on scholarly productivity, in: Journal of the American Society for Information Science, vol. 49, no. 8, pp. 720-730.

[152] See Bauer, K. (2000): Who Goes There? Measuring Library Web Site Usage, in: Online, January 2000. (http://www.onlineinc.com/onlinemag/OL2000/bauer1.html). – Luther, J. (2000): White Paper on Electronic Journal Usage Statistics, in: Journal of Electronic Publishing, vol. 6, no. 3. (http://www.press.umich.edu/jep/06-03/luther.html). – Rous, B. (2001): Usage statistics for online literature, in: Professional/Scholarly Publishing Bulletin, vol. 2, no. 1. (http://www.pspcentral.org/bulletins/spring2001_bulletin_pdf.pdf). – Odlyzko, A. (2001): The rapid evolution of scholarly communication. (http://www.dtc.umn.edu/~odlyzko/doc/rapid.evolution.pdf).

(LSE) survey on the socio-economic impact of the Internet in the LSE researchers' research environment 72% of the respondents considered the WWW important for locating research.[153] A similar percentage indicated that the Internet in general has actually increased their ability to locate relevant research. A survey among the research communities of construction information technology and construction management and economics returned the result that the Web is one of the most important sources for searching for and retrieving information.[154]

An analysis from Kaminer and Braunstein related the frequency of information retrieval from the Internet with scientists' productivity measured in terms of publications. They applied two methods to create the Internet use data: a questionnaire-based survey and an analysis of client-side log files, i.e. they counted different Internet processes and their duration within their sample of researchers by means of recording their Internet accesses. Kaminer and Braunstein found positive effects on productivity for different ways of information retrieval such as ftp, Kermit (a group of file transfer, management and communication software programs) and Telnet-connections to local library catalogues.[155] However, they used an ad-hoc model and do not provide any explanation for the results.

Another assessment of the significance of the Internet on information collection was carried out in the EC Community Innovation Survey II: there the enterprises were asked for the importance of computer networks as source of innovation-oriented information (see also section 4.3).

Ad b) Studies from the mid-nineties still returned lower citation frequencies for e-journals compared to off-line journals (using citation data from the Institute of Scientific Information).[156] A more recent analysis compares the citation frequencies of conference articles making a distinction between articles available on-line and those that are not. On-line articles have been cited a lot more often: the median is 7.03 citations per article compared to 2.74 for off-line articles.[157] Of course, this kind of analysis cannot pin down the causality between on-line availability and citation rates. It might very well be that better articles are made available on-line to secure a high impact. On the other hand it is also possible that articles available on-line are cited more often due to a higher visibility and a greater spread. Estimating the reading of on-line articles from electronic archives (arXiv, JSTOR) and e-journals and comparing it with that of printed journals Odlyzko concludes:

"The general impression from the statistics quoted above is that articles in electronic archives and electronic journals may not yet be read as frequently as printed journal articles, but are getting close."[158]

The author also reports large growth rates for electronic sources, some well over a 100% annually. Another analysis that compares citation levels of on-line only and printed articles from a biomedical journal also showed on average 2.6 times higher citation rates for the printed articles (articles from 1997 through to 1999 were studied along with their citations "within the biomedical literature" as of January 2000).[159]

---

153 Roughly 60% also considered it important for locating people and data, see Lubanski, A.; Matthew, L. (1998), op. cit., appendix d (http://cep.lse.ac.uk/iriss/appendixd.html).

154 See Björk, B.-C.; Turk, Z. (2000): How scientists retrieve publications: an empirical study of how the Internet is overtaking paper media, in: The Journal of Electronic Publishing, vol. 6, no. 2. (http://www.press.umich.edu/jep/06-02/bjork.html).

155 See Kaminer, N.; Braunstein, Y.M. (1998), op. cit.

156 See Harter, S. P. (1996): The Impact of Electronic Journals on Scholarly Communication: A Citation Analysis, in: The Public-Access Computer Systems Review, vol. 7, no. 5. (http://info.lib.uh.edu/pr/v7/n5/hart7n5.html) – Harter, S. P.; Kim, H. J. (1997): Electronic Journals and Scholarly Communication: A Citation and Reference Study, in: The Journal of Electronic Publishing, vol. 3, no. 2. (http://www.press.umich.edu/jep/archive/harter.html).

157 See Lawrence, S. (2001): Free online availability substantially increases a paper's impact, in: Nature, vol. 411, no. 6837, p. 521. (also available as 'Online or invisible?' at: http://www.neci.nec.com/~lawrence/papers.html).

158 Odlyzko, A. (2001), op. cit.

159 See Anderson, K.; Sack, J.; Krauss, L.; O'Keefe, L. (2001): Publishing Online-Only Peer-Reviewed Biomedical Literature: Three Years of Citation, Author Perception, and Usage Experience, in: Journal of Electronic Publishing, vol. 6, no. 3 (http://www.press.umich.edu/jep/06-03/anderson.html).

Most of the studies referenced employ similar methods to assess the impact of on-line publications (e-journals, electronic literature archives, conference papers available on-line): the citation levels of on-line publications are compared to those of printed publications. However, Guthrie criticises the fact that citations do not provide a complete picture of the potential usefulness of a journal article: some of the most viewed articles in a digital library were rarely cited.[160] For another reason it has become difficult to compare ‘on-line only’ and ‘off-line only’ journals: the diffusion of electronic versions of printed journals. Also it is impossible to deduce from a citation if the author has used the electronic or the printed version of a publication.

Odlyzko uses the number of downloads of on-line articles as a measure of their reading frequency and doubts that direct comparisons of traditional journals or libraries with electronic collections are really relevant, as the electronic infrastructure is improving rapidly and the growth rates of the usage of electronic publications are large.[161] However, this view of an unconditioned growth of electronic publishing is not unrivalled: communication conventions differ among research disciplines and consequently communication media will continue to be different.[162]

These are only tentative and not brand new results on the effects of information retrieval from the Internet on scholarly research and innovation. They are based on surveys among researchers, citation analyses and in one case (Kaminer and Braunstein 1998) on the analysis of client-side log files.

2.2.2 Data analysis and technology development

The scientific analysis of empirical data has usually taken place on some kind of local computer (PC, server, mainframe, supercomputer) or other specifically configured instrumentation. Increased transmission capacities and more capable software support network-based forms of remote or distributed data analysis. New developments in this area are currently the target of many research projects. Grid computing is the most comprehensive approach towards establishing a collaborative, computer-based multi-media research environment (for a brief discussion see section 2.1.4).

In addition, some further impacts of existing technologies on data analysis have been noted; e.g. e-mail has helped to align analytical work that needs some amount of synchronisation and congruency, such as chemists’ and biologists’ experiments.[163] The ETAN report lists a couple of cases in which the Internet or other communication technologies were used to control research equipment and carry out research tasks from distant locations.[164] A long list of US projects and current developments in different disciplines is included in the IWG on IT R&D report.[165] Many initiatives are underway which are targeted at the visualisation of data and the transmission of visualised objects.[166]

---

160 See Guthrie, K. (2000), op. cit.

161 See Odlyzko, A. (2001), op. cit.

162 See Kling, R.; McKim, G. (2000): Not just a matter of time: Field differences and the Shaping of Electronic Media in Supporting Scientific Communication, in: Journal of the American Society for Information Science, vol. 51, no. 14, pp. 1306-1320. (also available at: http://xxx.lanl.gov/ftp/cs/papers/9909/9909008.pdf). – Kling, R.; McKim, G. (1998): The Shaping of Electronic Media in Supporting Scientific Communication: The Contribution of Social Informatics. Paper presented at the "European Science and Technology Forum: Electronic Communication and Research in Europe" Darmstadt/Seeheim, 15 to 17 April 1998. (http://www.slis.indiana.edu/kling/pubs/seeheim.htm).

163 See Walsh, J.; Roselle, A. (1999), op. cit., p. 58.

164 See European Technology Assessment Network (ETAN) (1999), op. cit., p. 46.

165 See Interagency Working Group (IWG) on Information Technology Research and Development (IT R&D) of the National Science and Technology Council (2001), op. cit.

166 See the list in Sandstrom, T. A.; Watson, V. R. (2001): Collaborative Expeditions through Remote Scientific Data via the Internet, SSGRR 2001 International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, L'Aquila, Aug 06 - Aug 12 2001. (http://www.ssgrr.it/en/ssgrr2001/papers/Velvin%R.%Watson.pdf)

The sources examined contain lots of projects and developments in different fields of science. To describe only a fraction of them would not only be far beyond the possibilities of this report, it would also not render much added value. Our objective is not to describe in detail the cutting edge uses of the Internet in science, but to establish a stable basis for measuring innovative Internet technologies and computer network-based routines which are diffused to some extent. However, in the field of data analysis and technology development we have not found any meta-studies which employed a similar perspective, not to mention rudimentary statistics or measurement concepts. At large this also applies to the publication and dissemination of research results, though the situation is slightly better. We will discuss this in the next section.

2.2.3 Publication and dissemination of research results

Whereas we employed the user perspective on digitised information in section 2.2.1, we now employ the authors' perspective: research results have to be communicated to colleagues, principals, specific stakeholders or the general public in order that they can be reviewed by peers, used for further research, turned into new products and processes, implemented for new policy measures etc. (see section 1.2). While the publication of a finished research paper formerly was the publisher's work, the Web has created a new medium for publication and it has changed the process of publishing in many aspects.

E-publications

The Internet has created new media for communicating research results. Besides the dissemination of research papers as e-mail attachments or by posting them on a WWW homepage, more formalised forms have appeared. They can be summarised with the generic term 'electronic publications' (e-publications) and the activity of using the Internet as a publication media can be labelled 'e-publishing'. New computer-based e-publications for scholars and other researchers are:

- *Electronic pre-print publications and archives:* these make research results available *before* printed journals have completed publishing and sometimes before quality controls (referee processes) have taken place. The basic model is the well documented arXiv in Physics (formerly Los Alamos archive).[167]
- *Electronic journals and electronic books:* some only exist as such, others are published before, at the same time as, or after, the printed versions but usually by the same publisher.[168]
- *Electronic reprint publications and archives:* these contain selected articles, usually those with high impact and citation rates or those submitted by the authors, which are published again in electronic form or integrated into electronic databases for simplifying retrieval.[169]

Numerous publications and conferences reflect the ongoing discussions on the advantages and disadvantages of e-publications compared to printed publications. Extensive debates, often making use of the new media themselves, have appeared on the following questions:[170]

167 See Ginsparg, P. (1994), op. cit.

168 See as examples of e-journals that also deal with topics of the current study: The Journal of Electronic Publishing (http://www.press.umich.edu/jep/index.html). – Cybermetrics. International Journal of Scientometrics, Informetrics and Bibliometrics (http://www.cindoc.csic.es/cybermetrics/cybermetrics.html). – Journal of Computer-Mediated Communication (http://www.ascusc.org/jcmc).

169 See the chapter on digital libraries and as examples the reprint section of The Journal of Electronic Publishing (http://www.press.umich.edu/jep/index.html) or Steve Lawrence's documents and citations database ResearchIndex (http://citeseer.nj.nec.com/cs).

- Are on-line publications quicker in making research results available? (see also next section)
- Are e-journals less expensive in the production process and easier to distribute than printed journals?
- Do the established means of quality control such as peer review make sense for e-journals or should they be replaced by other forms of quality control? (see also next section)
- How can intellectual property rights be protected when research results are published electronically?
- What kind of storage and archives of electronic media is possible and adequate in the light of frequently changing software standards?
- Are e-publications going to replace printed publications or are they going to be an additional information source?

This report cannot aim to give a comprehensive picture of these debates and the arguments put forth, as it would either require many pages or be incomplete and incorrect. We discussed above to what extent scientists use e-publications for retrieving information (see p. 38) and we presented some evidence on the advantages of on-line versus printed publications (see p. 21). Now we will look at the evidence on how they use it for disseminating their work.

The reading frequency of scientific papers available in both print and on-line form has been shown to be higher than that of purely printed papers (see p. 38). Some hypothesise that this will lead to the increasing attractiveness of electronic publications for scientific authors, as a larger audience can be reached and as there is a danger that “anything not on the Web will be neglected”.[171] However, investigations of the factors influencing the choice of a journal for publication indicate that the reputation is probably more important than the size of the readership.[172] A recent analysis of a biomedical journal revealed that authors still have a clear preference for printed journals (given the choice between on-line only and off-line journals), and that the amount of requests and citations which they received according to their perceptions justified this preference.[173] However, the analysis also showed that these perceptions of a lower impact and citation of on-line articles might be misplaced.

New publishing processes and methods of quality control

The Internet has not only created new publication media, it has also changed and accelerated the publishing process. The entire process of communication between author, editor, publisher and peer can take place by means of e-mail. This reduces the time needed for transmitting an article compared to regular mail. Even more important time savings might accrue from publication in an electronic journal that publishes its contents ‘on demand’, that is whenever a publishable article is available. Comparisons of the different publication scenarios calculated a time reduction of 10-15% for a journal that transmits its articles by e-mail and 50-70% for e-journals with individual articles.[174] Even more time can be saved if no previous reviewing of the paper takes place, which is common in some pre-print archives in the natural sciences. In this case, the time reduction for publishing might be more than 99% (3 days instead of 128-308).

---

170 See on these debates e.g. The Journal of Electronic Publishing (http://www.press.umich.edu/jep/index.html). – Nature: Web debate: Future e-access to the primary literature (http://www.nature.com/nature/debates/e-access/). – Tenopir, C.; King, D.W. (2000): Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers. Washington, D.C.: Special Libraries Association. – The Global Research Village III (2000): Access to publicly financed research. Background papers. Amsterdam. (http://www.minocw.nl/english/conferentie/deel4.pdf) – Butterworth, I. (1998): The impact of electronic publishing on the academic community an international workshop org. by the Academia Europaea and the Wenner-Gren Foundation. London [etc.]: Portland Press.

171 Odlyzko, A. (2001), op. cit.

172 See Björk, B.-C.; Turk, Z. (2000), op. cit.

173 See Anderson, K.; Sack, J.; Krauss, L.; O’Keefe, L. (2001), op. cit. – See on the latter also Björk, B.-C.; Turk, Z. (2000), op. cit.

174 See OECD (1998), op. cit., p. 53.

However, the latter scenario in particular might raise some doubts with regard to quality control. If an article is not reviewed, who is going to guarantee that its contents meet scientific standards? Besides the doubts that might be raised regarding peer reviews as a valuable method of quality control,[175] it might be less of a problem if the research group is sufficiently large and its expertise constitutes the current state of knowledge on the topic.[176] Furthermore, new ways of reviewing e-publications have appeared which can be considered as alternatives to traditional peer review.[177] They usually combine two approaches, public reviewing and reviewing after publication.

- For public reviewing the draft of a research paper is posted on an Internet site and interested readers have the opportunity to attach comments to this draft. These comments are part of the review and the author has to consider them when revising his/her article for final publication. Bingham describes a pilot project at the Medical Journal of Australia (MJA) which devised this approach. Though few readers accepted the invitation to comment, some of the remarks contributed to the modification of the articles.[178]
- Reviewing after publication takes place, if the original article is supplemented with comments and critique from its readers. This system makes use of the technological possibilities of the Internet. It connects the article and the review paper or even single chapters and arguments by means of hyperlinks. This might lead to scientific discussions that take a similar form to those of scientific mailing lists.[179]

The literature on new publishing processes and peer reviews largely draws on the experiences of journal editors and publishers, on case studies of different journals and pilot projects. Comparable indicators on the integration of the Internet into the publishing and reviewing processes have not been developed on an international scale or for different research fields. Whether this should be performed will be addressed in the second SIBIS report on indicator development, deliverable 2.2.

2.3 Computer networks and R&D collaborations

A third approach for assessing the effects of the Internet on R&D is the analysis of its relation to the organisation of R&D activities. By harnessing the possibilities of the Internet the overall costs of communication will decrease and the demand for communication will increase in institutions which perform R&D work (see also section 1.2). Communication resources will replace other resources and the structures within organisations will become more communication-intensive.[180] Of course, this increase in communication takes place within each single institution, university department, R&D corporation, R&D department of a private business enterprise, government research institute etc. This effect has been be covered in section 2.2 which described how Internet applications have affected different phases of R&D activities. The current section focuses on the effects of the Internet on the organisation of R&D activities and more specifically on the development of collaborative R&D.

As various authors point out, and the results of empirical research confirm, the increase of R&D collaboration in recent years cannot exclusively be attributed to cheaper and faster

175 See Smith, R. (1998): The Future of Peer Review, in: OECD: The Global Research Village: How Information and Communication Technologies Affect the Science System, Paris, pp. 255-265. (http://www.oecd.org/dsti/sti/s_t/scs/prod/global.pdf).

176 See OECD (1998), op. cit., p. 54.

177 See Bingham, C. (1998): Peer review on the Internet: a better class of conversation, in: The Lancet, vol. 351 (suppl I), pp. 10-14. (http://thelancet.com/journal/vol351/isss1/full/llan.351.s1.reviews.13114.1).

178 Ibid.

179 Examples of e-journals that employ this kind of review are the Journal of Interactive Media in Education (http://www-jime.open.ac.uk) and Psycoloquy (http://www.cogsci.soton.ac.uk/psycoloquy).

180 Malone and Rockart state these as impacts of information technology in general on co-ordination costs, see Malone, T.W.; Rockart, J. (1992): Information technology and the new organisation, in: Proceedings of the 25th Hawaii International Conference on Systems Sciences, vol. 4, pp. 636-643.

modes of electronic communication, but they certainly belong to the factors that account for it.[181] Recent research from Smith and Katz using bibliometric data for the U.K. has shown that research collaboration decreases with distance. The median collaboration radius – the average radius around an institution within which 50% of its collaborators are located – measures 60-80 km.[182] On the other hand they have also shown that the distance between collaborators has tended to increase over the past 20 years. The latter tendency might be due to improved transportation, information transfer and communication.

2.3.1 Forms of R&D collaboration

The scientific literature on R&D collaboration contains a variety of different categorisations:

(1) Hagedoorn and others on the one hand characterise research partnerships in terms of the members and make a distinction between private and public participants.[183] A similar but more elaborated approach is applied in the literature on 'triple helix' relations, a term that was coined for the relations between universities, business enterprises and government institutions.[184] Sometimes even more R&D collaboration partners are studied, e.g. Godin and Gingras consider universities, hospitals, industry, federal and provincial governments and others.[185]

(2) Another dimension of categorisation is used by Hagedoorn et al. taking into account the organisational structure of research partnerships. From this point of view they differentiate between informal and formal partnerships. They further describe formal agreements as either equity-based, i.e. research corporations that are founded by two or more independent partners, or non-equity-based, such as R&D pacts or consortia or project-based R&D contracts.[186]

(3) Between these two approaches lies yet another categorisation which is for example made by Dodgson and which could be labelled as 'functional'. He distinguishes between vertical collaboration along the chain of production from customer to supplier and horizontal collaboration between partners at the same level in the production process.[187]

(4) Another scheme was employed by Smith and Katz to categorise collaborative partnerships in higher education in the U.K. They identify three ideal-typical forms of collaboration:[188] Corporate partnerships which are set up between research institutions such as universities or research departments aimed at improving access to external resources; team collaboration of a semi-formalised nature below the corporate level and driven principally by the need for multi-disciplinary skills and experience; inter-personal collaboration which is intellectually driven, mostly discipline-based and does not depend on institutional affiliation but rather migrates with the researcher.

(5) Sanderson took the intensity of collaboration as the defining criteria. He distinguished between collaborative research, in which participants jointly identify common research questions and coordinate efforts to create new knowledge, and cooperative research, in which participants simply exchange information or data.[189]

---

181 See Walsh, J. P.; Bayma, T. (1996), op. cit., pp. 343-350. - Smith, D.; Katz, S. (2000): Collaborative Approaches to Research. Fundamental Review of Research Policy and Funding. Final Report. The Higher Education Funding Council for England (HEFCE), pp. 27-28. - Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997), op. cit., p. 513.

182 See Smith, D.; Katz, S. (2000), op. cit., p. 59.

183 See Hagedoorn, J.; Link, A. N.; Vonortas, N. S. (2000): Research partnerships, in: Research Policy, vol. 29, p. 568.

184 Etzkowitz, H.; Leydesdorff, L. (eds.) (1997): Universities in the global knowledge economy: A triple helix of university-industry-government relations. London.

185 See Godin, B.; Gingras, Y. (2000): The Place of Universities in the System of Knowledge Production, in: Research Policy, vol. 29, no. 2, pp. 273-278.

186 See Hagedoorn, J.; Link, A. N.; Vonortas, N. S. (2000), op. cit., p. 569.

187 See Dodgson, M. (1996): Technological collaboration and innovation, in: Dodgson, M.; Rothwell, R. (eds.): The handbook of industrial innovation, Cheltenham and Brookfields, p. 285.

188 See Smith, D.; Katz, S. (2000), op. cit., pp. 71-72.

189 See Sanderson, D. (1996), op. cit., pp. 97-98.

The Internet has made a new type of R&D collaborations possible which depend on the electronic transmission of information. Instead of working together in one place or meeting frequently these collaborations access geographically distant resources, exchange data and information, carry out their analytical work and document the results by means of the Internet. They not only use the Internet but actually could not exist without it. Therefore it seems justifiable to distinguish these new Internet-based collaborations from other forms of R&D collaboration.

Two new organisational forms have been documented in the literature: collaboratories and virtual teams. Neither could exist without state-of-the-art research infrastructure such as high-performance computers, networks with high transmission capacities and a variety of computer and network tools (see section 2.1). They will further benefit from new developments in the field of Grid computing which are specifically designed to facilitate the collaboration of distributed research teams (see section 2.1.4). Collaboratories and virtual teams are very similar and sometimes the terms are used synonymously. In general the former is predominantly used in an academic environment and the latter in a private business setting.
According to Finholt the term '*collaboratory*' was coined in the late eighties and is a hybrid of 'collaborate' and 'laboratory'. He also provides an early definition:

> *"…a center without walls, in which researchers can perform their research without regard to physical location – interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries."*[190]

The novelty of a collaboratory is not that it institutionalises collaboration over large distances, but that this kind of collaboration is only possible because of the capabilities of the new information and communication technologies and that it is organised to harness them as effectively as possible. Examples have been documented in many different research fields, such as biology, physics, medicine and communication science, and are usually related to basic research.[191] They jointly use large databases, scarce and expensive high-performance computing power or other instrumentation and they pool the experiences and expertise of scientists from different locations. In this way resources are increased for the participants.

The term *virtual team* has been used for arrangements where R&D work is not collocated but distributed over geographically distant sites. As in collaboratories, coordination and collaboration are largely computer- and network-based. E-mail, video conferencing and other tools for communicating, planning and scheduling, as well as document and application sharing, are used to overcome the problems that result from not working at the same place and sometimes not even at the same time. The creation of virtual teams is a response to globalisation, with business enterprises becoming increasingly globalised through mergers and acquisitions of foreign firms or an extension of their sales areas to a worldwide level. Also, the scarcity of highly qualified technical staff in many industrialised countries and the promise of round-the-clock development on different continents in different time zones can be considered as additional supporting factors to the establishment of virtual teams.[192] They tend to be found predominantly in applied research and development rather than in basic research.

---

190 Wulf 1989, p. 19 cited in Finholt, T.A. (2001): Collaboratories, in: B. Cronin (Ed.): Annual Review of Information Science and Technology, vol. 36. (also available at: http://intel.si.umich.edu/crew/technical_reports_alphabetical.htm).

191 See the extensive list in Finholt, T.A. (2001), op. cit. and the examples in European Technology Assessment Network (ETAN) (1999), op. cit., pp. 40-42. - OECD (1998): The Global research village: How information and communication technologies affect the science system. Paris. (http://www.oecd.org/dsti/sti/s_t/scs/prod/global.pdf)

192 See Herbsleb, J. D. et al. (2000): Distance, Dependencies, and Delay in a Global Collaboration, in: Proceedings of CSCW 2000, Philadelphia, PA, Dec. 2-7, 2000 (http://www.bell-labs.com/org/11359/colab_prod/).

Many examples are documented from software development[193] but they also are common in other industries.[194]

The empirical literature on collaboratories and virtual teams is normally based on case studies, i.e. interviews or survey-based investigations among the members. It should still be considered as mainly exploratory and the results are preliminary. They can be generalised only to a limited extent and they do not permit a cross-country comparison of participation in these ICT-based R&D collaborations. The main findings of these investigations into the consequences of collaboratories and virtual teams for R&D activities are:

- They have increased the number of participants in scientific projects and the diversity in terms of experience and expertise.[195] Also research results seem to derive from a broader empirical base.
- There is some evidence that collaboratories have changed scientific work, e.g. in respect to the use of data and a more equal treatment of younger and less experienced researchers.[196] Other impacts are that theoretical and empirical research seem to be integrated more easily, e.g. due to visualisation techniques.
- They seem to increase the speed and efficiency of *some* scientific tasks.[197] However, some investigations of applied research and development projects find negative impacts of non-collocated R&D on work speed and project duration compared to single-site R&D.[198] Delays are obviously caused by a multitude of factors. These range from problems of communication due to few overlapping working hours between America, Europe and Asia, through to difficulties in coordinating meetings and obtaining informal input from distant colleagues.[199]

If indicators can be found that measure the significance of new, Internet-based forms of R&D collaboration for the R&D systems in Europe, this will create additional value for monitoring European research policy. The European Commission introduced the goal of increasing pan-European networks within the R&D system as well as across its boundaries as part of its new strategy of creating a European research area, laid down at the Lisbon summit.[200]

2.3.2 Indicators of R&D collaboration

A major problem of each of the ‘traditional’ categorisations of R&D collaborations is that the attainable information on R&D partnerships by far lacks the sophistication needed to accurately measure the incidence of the different categories. Two different methodologies have been used:

- Bibliometric methods and patent analysis
- counts of R&D partnerships based on publicised information or questionnaires.

---

193 See Herbsleb, J. D. et al. (2000), op. cit. - OECD (1998), op. cit.

194 See for example on the automotive industries: Wierba, E.E; Finholt, T.A.; Steves, M. (2001): "What Have You Done for Me Lately?" A Case Study of Barriers to Collaborative Tool Adoption in a Manufacturing Engineering Setting. (http://intel.si.umich.edu/crew/Technical%20reports/Wierba_Finholt__What_have_you_done_for_me_lately_05_31_01.pdf)

195 See Finholt, T.A. (2001), op. cit.

196 Ibid.

197 Ibid.

198 See Herbsleb, J. D. et al. (2000), op. cit., pp. 5-6. – Wierba, E. E; Finholt, T. A.; Steves, M. (2001), op. cit., pp. 8-9. – Finholt, T. A. (2001), op. cit.

199 See on the problems of distributed work and the benefits of collocated work: Olson, G.M.; Olson, J.S. (2002), op. cit.

200 See European Commission (2000c): Towards a European research area. Communication from the Commission COM 2000 (6). Brussels, 18 January 2000.

Bibliometric methods and patent analysis

When using bibliometric methods, research is assumed to have been carried out collaboratively when a scientific publication has more than one author. Depending on the provenance of the authors the collaboration can be in-house, between different organisations within a country or international. Though this quantification has some shortcomings, for example the fact that a co-author belongs to a different institution doesn't give any information about the quality of a collaboration or the factual input of the authors listed, it is a generally accepted measure.[201]

Based on this technique, empirical evidence is strong that international R&D collaboration has increased:[202] the percentage of internationally co-authored papers has increased worldwide and for all scientific disciplines: from 7.8% in the period 1986-88 to 14.8% in the period 1995-97 (see table 5). Smaller countries in particular exhibit high rates of international co-authorship. Many times over 30% of all published papers had co-authors from other countries; in Portugal it was even 50.8% and in Switzerland 48.1% during the period 1995-97.

Table 5: Percentage of co-authored and internationally co-authored scientific and technical articles by country 1986-88 and 1995-97

Country	**Percent co-authored**		**Percent internationally co-authored**	
	1986-88	**1995-97**	**1986-88**	**1995-97**
EU member states[a]				
Austria	48.6	66.1	27.1	43.6
Belgium	53.1	66.8	31.2	46.6
Denmark	57.3	68.1	25.9	44.3
Finland	54.5	70.2	20.9	36.1
France	51.8	64.8	22.2	35.6
Germany	39.3	54.9	20.7	33.7
Greece	42.2	61.1	27.6	38.3
Ireland	48.3	60.1	28.9	41.9
Italy	61.4	71.6	24.0	35.3
Netherlands	48.0	64.4	21.3	36.0
Portugal	53.5	70.0	37.6	50.8
Spain	43.1	58.7	18.8	32.2
Sweden	56.8	66.8	24.0	39.4
United Kingdom	39.4	53.9	16.7	29.3
Others				
Japan	39.5	54.0	8.1	15.2
Switzerland	48.6	62.7	34.5	48.1
United States	46.4	56.8	9.8	18.0
World	***38.6***	***50.1***	***7.8***	***14.8***

a No data available for Luxembourg.

Source: National Science Board (2000), op. cit., p. A438-A449.

The underlying assumption applied to patent analysis is very similar to the bibliometric approach. R&D collaboration is assumed to have taken place when there is more than one owner of a patent or author of a patent application. The OECD integrated the percentage of

[201] See Leydesdorff, L. (2001): The challenge of scientometrics. The development, measurement, and self-organization of scientific communications. 2nd edition, p. 232.

[202] National Science Board (2000): Science & engineering indicators 2000. Arlington, Va., pp. A438-A449.

patents with foreign co-inventors in its science and technology indicators database.[203] The country comparison for patents with foreign co-inventors reveals again that larger countries collaborate less with foreign partners (see table 6). Exceptions are the U.K., with a rather high percentage of foreign co-inventors, and Finland and Sweden with rather low percentages. The OECD also mentions a 'home advantage' factor, i.e. a preference to patent on the home market leading to a bias within the patent data.[204] Hence, we should expect somewhat higher figures for the U.S. and Japan and lower figures for the European countries, if we used patent data from USPTO or JPO. However, a thorough explanation of the differences shown in table 2 would require a detailed analysis which is outside of the scope of this report.

Table 6: Index of patents with foreign co-inventors in 1995-1997[a]

Country	Index	Country	Index
Austria	10.6	Netherlands	7.8
Belgium	15.2	Portugal	22.4
Denmark	9.9	Spain	7.4
Finland	5.9	Sweden	6.0
France	5.6	United Kingdom	8.9
Germany	4.7		
Greece	12.4	**European Union[b]**	**3.2**
Ireland	16.8	Japan	1.5
Italy	4.3	Switzerland	14.4
Luxembourg	35.4	United States	4.8

a The index is calculated as the average percentage of foreign co-inventors for all co-invented patents, e.g. 14.4 for Switzerland means that 14.4% of the co-inventors were from abroad (and consequently 85.6% domestic). Due to a different method of calculation, a comparison with the figure for the previous time-period 1993-95 is not possible.

b Intra-European co-inventions are netted out, i.e. only patents with co-inventors outside of Europe are counted.

Source: OECD (2001), op. cit., p. 196, based on data from the European Patent Office.

R&D partnerships

A different approach to measuring R&D collaboration has been implemented by Hagedoorn et al.:[205] Their CATI database contains new cooperative research- and technology-oriented agreements between independent organisations on an annual basis. This database also has some weaknesses and limitations, especially because it is a 'literature-based alliance counting' that only integrates alliances publicised in newspapers, journals etc. It does not cover any of the various forms of collaborative research that are outside the scope of these formal alliances, such as informal knowledge sharing, and hence it can hardly provide representative data on collaborative research in general.[206] It does, however, create the possibility of comparing the participation of different countries and regions in formalised research partnerships. Counts of R&D partnerships based on the CATI database also indicate that their incidence has increased over the last twenty years.[207]

[203] See OECD (2001): Science, technology and industry scoreboard 2001. Towards a Knowledge-based Economy. Paris, pp. 112-113, 196. – OECD (1999): Science, technology and industry scoreboard 1999. Benchmarking Knowledge-based Economies. Paris, pp. 80-81, 164.

[204] OECD (2001), op. cit., p. 60.

[205] See Hagedoorn, J.; Link, A. N.; Vonortas, N. S. (2000), op. cit. - Hagedoorn, J. (1996): Trends and patterns in strategic technology partnering since the early seventies, in: Review of Industrial Organisation, vol. 11, pp. 601-616.

[206] See Foray, D.; Steinmueller, E. (1999): Collective invention and European policies "COLLINE" 1997-1999. (http://www.dauphine.fr/imri/COLLINE/exesummary.html).

[207] See for an example: National Science Board (2000), op. cit., p. A121.

Surveys targeted at the assessment of innovation and R&D also frequently pose questions on the incidence and significance of inter-firm collaboration. The innovation surveys in the European Union member states (CIS II) use a core questionnaire that contains a question on innovation co-operation, which was defined as "active participation in joint R&D and other innovation projects with other organisations."[208] Foyn reports the percentage of innovators involved in co-operation with the different countries covered by CIS II.[209] The figures show that, especially in the Scandinavian countries, business enterprises undertake a lot of innovation co-operation, whereas in some southern and central European countries such as Italy, Spain, Portugal, Germany and Austria, co-operation is less common (see table 7).

Table 7: Percentage of innovators involved in co-operation in 1994-1996

Country	Manufacturing	Service
Austria	23	18
Belgium	32	45
Denmark	57	66
Finland	71	60
France	35	35
Germany	24	17
Ireland	36	23
Italy	11	-
Luxembourg	29	46
Netherlands	29	28
Portugal	20	23
Spain	21	-
Sweden	59	48
United Kingdom	32	28
European Union	**26**	**24**

Source: Foyn, F. (2000), op. cit., p. 2.

Another approach that also might be useful for assessing R&D collaborations has been employed by Leydesdorff and Curran: they counted references on websites to evaluate the relations among universities, firms and governmental institutions (see section 2.2.1, p. 36).

2.4 Summary and conclusions

This survey assessed the scientific literature in various research disciplines (communication research, computer science, social sciences) as well as a large body of interdisciplinary literature to investigate the following questions:

a) What are the important Internet infrastructure prerequisites for modern R&D and how has their spread and impact been measured?

b) How has the Internet affected the conception and the results of different steps of the R&D process?

c) How has the decrease in communication costs, due to the spread of the Internet, affected R&D collaborations and how has this been assessed empirically?

208 Core questionnaire of the Community Innovation Survey (CIS) II according to Muzart, G. (1999): Description of national innovation surveys carried out, or foreseen, in 1997-99 in OECD non-CIS-2 participants and NESTI observer countries. OECD STI working papers 19999/1, Paris, p. 59.

209 See Foyn, F. (2000): Community Innovation Survey 1997/98 - final results. Eurostat (Ed.): Statistics in Focus: Theme 9 Research and Development, no. 2/2000.

Ad a) Research networks are the basic prerequisite for performing research that involves the transmission of large volumes of data. They are usually organised on the local (campus), national and supra-national level with ports to the next level/other networks. The capacities on each level determine, how well the data transmission works between two servers. Highly useful information on transmission capacities, other supply features and performance levels of national research networks (NRN) have been collected by the Trans-European Research and Education Networking Association (TERENA). It addressed questionnaires to NRNs for this purpose.

On-line information sources are another element of the Internet infrastructure which are useful for R&D. A large body of literature has been published on electronic full-text and bibliographical databases. Innovative versions of digital text collections and e-publications have been implemented and the empirical evidence that their usage has been rising can hardly be questioned. The concepts of measuring the provision and use of digital libraries are well advanced. Different indicators have been proposed which can be described as either resource-oriented, usage-oriented, user characterising or cost-oriented. As potential data sources, server log files and questionnaire-based surveys have been singled out. Surveys constitute the only reliable source of information to date. Some studies indicate that the WWW in general has become a vital resource for many researchers and that it has even increased research productivity. The empirical evidence for this is again based on surveys.

The basic data source regarding different tools for computer-mediated communication is also the questionnaire-based survey. E-mail has become omnipresent in social communication for scholarly research and a positive feedback on research productivity has been tentatively assessed. Due to privacy concerns an automatic recording of e-mail frequency or receiver addresses is unthinkable. Concerning mailing lists, an appropriate and much frequented location for discussing topics which are not too demanding, the analysis of messages within the list is usually also possible. But this method does not provide much insight into the value of this communication form for the individual researcher and it is not particularly useful for benchmarking countries regarding their participation in mailing lists with scientific content. Finally, on-line meetings are a rather new and, as some case studies reveal, technologically somewhat immature form of communication. Therefore, it seems too early to quantitatively assess their spread and impact.

Grid computing denominates new developments to increase computational power and create virtual spaces for multifaceted collaboration. A multitude of projects exist, targeted at both the implementation of Grids and further R&D on their various layers. We did not find any approaches which assess the participation in grid computing and development on a national level.

The existence of technical staff with Internet and other network skills is also a prerequisite for completing R&D projects successfully. The OECD defines a rather narrow concept of R&D personnel and excludes most of the ancillary computing staff from its definition of R&D personnel. The available data sources (and most of the surveys that could form the basis of international comparisons) are not sufficiently detailed to single out the network administrators and webmasters at R&D organisations. Some general data on staffing of NRN providers has been published by TERENA.

Ad b) The Internet is being used by researchers mainly for collecting data and retrieving scientific information from different virtual storage spaces and for presenting, disseminating and discussing research results. Internet-based data analysis has not become commonplace yet, but new developments in the field of grid computing will increase the possibilities in this field.

Network-based data collection employs different interactive methods such as on-line surveys, on-line experiments or chat interviews. As far as we know no studies have yet been compiled which compare the usage of these new on-line research methods among different research communities, universities or countries. Other novel, non-interactive forms of Internet-based data collection are the informetric analysis of web contents and log file analysis. The former in particular has been employed for comparing the presence of countries on the WWW, for calculating Web impact factors and for tracking relationships in national R&D systems. Log file analysis is still implemented mainly for gauging the demand on a server. Multiple problems related to technological features, a lack of standardisation and the human usage of the Internet make it impossible to use the recorded log data for other purposes (such as assessing the usage of Internet applications in different countries). Interactive as well as non-interactive methods of data collection might become relevant for the current analysis, especially because of their methodical approach. However, this must be investigated in more detail in the further course of the project.

Various Internet tools exist for the retrieval of scientific information from specific Internet sources and the WWW in general. Surveys and empirical analyses back the hypothesis that the Internet facilitates information retrieval and that Internet usage is positively correlated with R&D productivity. However, the models are set up on a rather ad-hoc basis and fail to explain whether it is the utilisation of the new technology, the improved access to more information or the added value of collaborating that leads to higher productivity figures.

Besides the simple distribution of research papers as e-mail attachments and postings on WWW pages, e-publishing has become the most important innovation in presenting and disseminating research results. When e-journals accounted for only a fraction of the annually published journal literature and when there was a clear distinction between non-electronic and electronic publications, the comparison of citations from bibliographic databases was an appropriate and valid method for investigating the impact of e-journals. But, as the distinctions have become blurred and as many traditional print journals are also made available on the Net, the indicator has lost most of its value. The acceleration of communication and the new ways of giving comments and critique have also changed the review and quality control part of scientific publishing. This applies especially to on-line media and, to a somewhat lesser extent, also to off-line publications. The discussion has been focused very much on the advantages and disadvantages of different review procedures, and their implementation on a country level has as yet not been evaluated.

Ad c) Empirical evidence is strong that research collaboration has been increasing. Some analyses also point to increasing geographical separation between the collaborators, leading to the (much criticised) conclusion of a ‘death of distance’[210] due to the ICT. The social sciences literature also contains many different conceptualisations of R&D collaborations. But the operationalisation of these concepts has been deficient and neither the available output indicators nor the various count data of R&D partnerships permit an adequate measurement. Collaboratories and virtual teams as new forms of network-supported collaboration have only been described in case studies so far. How to quantify their significance or the geographical spread of their participants still remains an unresolved issue.

[210] Cairncross, F. (2001): The Death of Distance, 2nd edition, Boston, Mass.: Harvard Univ. School Publications.

3 The Internet in European R&D policy documents

3.1 Overview of policy documents on R&D

As the aim of SIBIS is to develop indicators which can be used to formulate and evaluate policy measures, documents in the field of R&D policy constituted an important source of information. A large number of documents from the Commission, the EU member states and Switzerland was retrieved and analysed with regard to their objectives and the measures directed at enhancing the usage of information and communication technologies in R&D (see table 8 for an overview). It should be stressed that we did not set out to conduct a thorough analysis of European and national research policies. This would have considerably exceeded the scope of the present investigation. The main sources were the web pages of government entities in R&D and annex 1 therefore lists the web addresses of the documents as they were found in mid-2001. Some of the material is only available in the national language.

Table 8: Overview of policy documents on the Internet for R&D

No	Title of document (Date)	Region	Pub. date	Type of doc.
1.	Towards a European research area	EU	2000	Communication
2.	Innovation in a knowledge-driven economy	EU	2000	Communication
3.	Making a reality of the European research area	EU	2000	Communication
4.	*e*Europe: An Information Society for all	EU	2000	Communication
5.	*e*Europe 2002: An Information Society for all. Action plan prepared by the Council and the European Commission for the Feira European Council	EU	2000	Action Plan
6.	Proposal for a decision of the European Parliament and the European Council concerning the multiannual framework programme 2002-2006	EU	2001	Communication
7.	A European Research Area for Infrastructures	EU	2001	Working doc.
8.	First Report on Progress towards the European Research and Innovation Area	EU	2001	Report
9.	Resolution Future of European Research	EU	2000	Resolution
10.	Faster Internet - Member States' Actions	EU	2001	Report
11.	Access to Global Research as a Political Issue	OECD	2000	Report
12.	The Need for High Bandwidth Computer-based Networking in Europe	Europe	2000	Resolution
13.	Green Paper on Austrian Research Policy	Austria	1999	Green Paper
14.	Research report 2000 (available only in German)	Austria	2000	Report
15.	Austrian Memorandum on the Commission Communication "Towards a European Research Area"	Austria	2000	Communication
16.	Declaration of the Federal Government on Current Issues in Research and Technology Policy	Austria	2000	Action Plan
17.	Promotion of IT Research and Education	Denmark	1999	Report
18.	Danish Council for Research Policy Annual Report 1999	Denmark	2000	Report

No	Title of document (Date)	Region	Pub. date	Type of doc.
19.	A net of opportunities: IT and Telecommunications, Policy Report to the Danish Parliament	Denmark	2000	Action Plan
20.	Education, training and research in the Information Society. A National Strategy for 2000-2004	Finland	1999	Action Plan
21.	Education and research 1999-2004 - Development Plan	Finland	1999	Action Plan
22.	Information Strategy for Education and Research 2000-2004 – Implementation Plan	Finland	2000	Report
23.	Finnish comments on the communication of the European Commission “Towards a European Research Area”	Finland	N/a	Communication
24.	Finnish comments on the communication of the Commission "MAKING A REALITY ...	Finland	2000	Communication
25.	Research and technology programme activities in Finland 106/2001	Finland	2001	Report
26.	Status of Research and Technological Development	France	2000	Report
27.	Dix orientations prioritaires pour la Recherche et la Technologie (available only in French)	France	2000	Action Plan
28.	Développer l'effort public de recherche (available only in French)	France	2000	Report
29.	Building a European area for knowledge, science and innovation	France	2000	Action Plan
30.	Multimedia in the university (available only in German)	Germany	2000	Report
31.	Federal report on research 2000 (only in German)	Germany	2000	Report
32.	1996 White Paper on Science, Technology a. Innovation	Ireland	1996	Green Paper
33.	Position Paper on Sixth Framework Programme	Ireland	2001	Communication
34.	National Program of Research (available only in Italian)	Italy	2000	Action Plan
35.	Italy’s contribution in the debate on the future of European research (available only in Italian)	Italy	2001	Communication
36.	Science Budget 2000	Netherlands	1999	Report
37.	Beleidsagenda HOOP 2000 (available only in Dutch)	Netherlands	1999	Communication
38.	Boeiend, Betrouwbaar en Belangrijk, Nota Wetenschap- en Techniekcommunicatie (available only in Dutch)	Netherlands	2000	Action Plan
39.	National Plan of Scientific Research, Development and Technological Innovation for the period 2000 to 2003 (only partially available in English	Spain	1999	Action Plan
40.	Research and Renewal: the Government's research policy bill	Sweden	2001	Report
41.	Goals for the Swiss research policy in the period 2000 to 2003 (available only in German)	Switzerland	1997	Report

No	Title of document (Date)	Region	Pub. date	Type of doc.
42.	Message on the support of education, research and technology in the years 2000 to 2003 (available only in German)	Switzerland	1998	Action Plan
43.	Excellence and Opportunity: a science and innovation policy for the 21st century	UK	2000	Green Paper
44.	Implementation Plan on the White Paper "Excellence and Opportunity ..."	UK	2000	Action Plan
45.	UK Response to: "Making a Reality of the ERA"	UK	2000	Communication
46.	Science Research Priorities 2001-02 to 2003-04	UK	2001	Action Plan
47.	JISC Five-Year Strategy 2001-05	UK	2001	Action Plan

For bibliographical and retrieval information see annex 1 from p. 72 onward.

Source: FHSO compilation.

The assessment of the policy documents returned no consistent approach for monitoring and evaluating Internet-related R&D policy measures. Therefore, the summary below concentrates on a documentation of the most important goals and measures in relation to the Internet in R&D. In the course of developing the SIBIS indicators these will be taken into account. Thus an additional value might be created if indicators can be found which are relevant for evaluating the R&D-related *e*Europe goals as well as the Internet-related goals of other R&D policy initiatives.

3.2 Internet-related issues in European R&D policy documents

The main goal of the current European Commission's policy on scientific research and technological development is the creation of an integrated European research area (ERA). The strategy is laid down in the similarly named communication and in various documents subsequently published (see the documents [1], [3], [7], [8], [9] from table 7). It has been developed with the idea of generating multi-layered networks within the research and development system as well as across its boundaries, including other socio-economic areas and political institutions. Aiming at the integration and co-ordination of European research activities, the new European research policy definitely goes beyond the old practice of adding another Framework Programme. It also reaches further than national research policies usually do. Many desirable outcomes of this strategy can be pictured, such as tighter and better co-ordinated networks between national and European institutions in the field of R&D policy, intensified interregional and international co-operation between researchers from the European Union and between them and their colleagues from Eastern Europe, or a greater success in the implementation of research results. However, whether these results can be achieved with ERA and the means envisioned in it – or through an alternative strategy – must be discussed in a rigorous and minute policy evaluation which is without doubt valuable and necessary but not the objective of the present investigation.

The collaboration and network-oriented target level of the Commission's research policy also entails specific policy measures which are listed in the referenced documents on the European research area. With regard to the ICT the documents list:

- the creation of 'virtual centres of excellence' based on electronic networks[211]
- actions to encourage the use of electronic networks in various fields of research (such as the development of databases and access to advanced Internet services, promotion of

[211] According to one of the external reviewers of the draft version of this report, Ian Butterworth, scientists have expressed their concern about the Commission's goal of establishing "centres of excellence". This concern is based on the observation that political institutions usually lack the information and knowledge to identify and support, not to mention create, excellence.

the production of multimedia content and interactive uses, support of new forms of electronic collaboration of researchers)

- awareness-building and training of researchers on the possibilities of the ICT.

Targets of the European Commission's *e*Europe initiative (see [4]) are the upgrading of the Internet infrastructure for European researchers and students and the establishment of campus networks which are capable of supporting multimedia communications. These targets were further detailed in the *e*Europe 2002 action plan (see [5]). In the latter document the Commission formulated four actions which aim at improving research networks at different levels (European, national, local/campus) and at promoting research technologies which use the available computing and transmission powers. As an outcome of that, the pan European GÉANT project was set up to increase the transmission capacity between national research networks across Europe and to provide a platform for the interconnection with similar networks in other parts of the world (see section 2.1.1).

All in all, the Commission's measures go beyond the OECD suggestions on ICT and R&D. The latter proposes that governmental support of ICT use in research should concentrate on two areas [see document 11]:

1) resources (e.g. high-performance information and communication networks, scientific databases, digital libraries)
2) the appropriate legal and regulatory framework for the exploitation of ICT (e.g. privacy protection, infrastructure security).

These OECD proposals can be found at least partially in the Commission's *e*Europe initiative, whereas with the ERA initiative a much more active position has been adopted in respect to the promotion of ICT.

Contrary to European research policy, national research policies have defined priorities rather narrowly: In general they focus on improving the competitiveness of the national research system in order to increase the amount of basic and applied knowledge and the number of innovations available to society, thus improving its economic performance. A broad range of goals and objectives, measures and activities is listed in the national research strategy documents to put this priority into practice. Among them the upgrading of communication networks and the usage of information and communication technologies for information retrieval, computer-based research methods, co-operation between scientists and the dissemination of their research results usually play an important but not outstanding role.

The available national documents give the impression that the situation is different in two EU member states: Finland and the UK should be considered as the most advanced European countries in regard to specific ICT-oriented actions in support of R&D. The Finnish Government elaborated "A National Strategy for 2000-2004. Education, training and research in the Information Society" which deals with ICT and the information society as research objects and also covers the multiple opportunities of employing ICT for scientific research. The government also tries to realise its vision of Finland as a leading interactive knowledge society in the year 2004 by means of many different programmes, actions and instruments (see table 7). The British Government communicated its science and innovation policy in a White Paper on "Excellence and Opportunity: a science and innovation policy for the 21st century" together with an implementation plan. The documents laid down many commitments which aim at improving the usage and the effects of ICT in R&D and they also determine concrete responsibilities and activities to monitor and evaluate the progress achieved.

While some data restrictions exist, which make it difficult to assess the capabilities of the national research systems in general and the effects of research and technology policies on R&D in particular, it is virtually impossible to monitor the broad aim of the Commission's new research policy with the currently available range of research data and indicators. A short tradition of collecting and publishing data on research expenditure, research personnel and research output (especially patents) exists in the OECD member states (see section 4.2). However, there is hardly any data available on most of the issues elaborated in the

Commission's research strategy; e.g. figures on the size and the density of research-related networks, on the mobility of researchers or on the provision and usage of research infrastructure such as high-speed networks, scientific databases and digital libraries simply have not been made available up to date. Usually the data have not even been collected yet.

4 R&D data from multinational organisations, national statistics and private organisations

4.1 Overview of R&D statistics

Another important source of information on current R&D statistics – which also gives details on the computerisation of science and technology within these statistics – are database descriptions, reports, handbooks and other publications from various multinational and national organisations. The OECD in particular has been at the leading edge of developing and standardising R&D and innovation statistics for many years. But EU institutions, other multinational and national organisations such as statistical offices and science boards and some particular enterprises and researchers also contributed to the advancement of indicators on R&D. Table 9 gives an overview of the most important documents and databases in this field.

Table 9: Overview of statistical documents on the Internet for R&D

No.	Title	Region	Pub. date	Type of document
1.	Main Science and Technology Indicators Database (MSTI)	OECD countries	2001	Public database
2.	Analytical Business Enterprise Research and Development Database (ANBERD)	OECD countries	2000	Public database
3.	Science, technology and industry scoreboard 1999	OECD countries	1999 (biannual)	Benchmarking report
4.	Science, Technology and Industry Outlook 2000	OECD countries	2000 (biannual)	Benchmarking report
5.	Frascati Manual 1993	-	1994	Handbook
6.	Oslo Manual	-	1997	Handbook
7.	Community Innovation Surveys in the New Cronos Database	EU, CEEC	N/a	Public database
8.	Science, Technology and Innovation. Key Figures 2000.	EU	2000	Statistical report
9.	Statistics on Science and Technology in Europe	EU	2001	Statistical report
10.	ESIS II WWW indicators	CEEC	2001	Survey report
11.	ISI®Essential Science Indicators℠ (ESI)	World	N/a	Private database
12.	WIPO Industrial Property Annual Statistics	World	2001 (annual)	Public database
13.	Cooperative Agreements and Technology Indicators (CATI) database	World	1999	Private database
14.	Facts and Figures on Danish Research	Denmark	2000	Statistical report, benchmarking report
15.	On the Road to the Finnish Information Society II	Finland	1999	Statistical report, survey report, benchmarking report
16.	Science, technology and information society	Finland	2001	Statistical report
17.	Science and technology indicators	Germany	2001	Statistical report, benchmarking report

No.	Title	Region	Pub. date	Type of document
18.	Research and Development in Italy in the years 1998-2000	Italy	2000	Statistical report, benchmarking report
19.	Science and Technology Indicators 2000.	The Netherlands	2001	Statistical report, benchmarking report
20.	Survey on the National Resources in Science and Technology 1988-1997	Portugal	1998	Statistical report, benchmarking report
21.	Contribution to a indicator system for science and technology policy	Switzerland	1998	Statistical report, benchmarking report
22.	SET Statistics 2000	UK	2000	Statistical report, benchmarking report
23.	Science and Engineering Indicators 2000	USA	2000	Statistical report, survey report, benchmarking report

For bibliographical and retrieval information see annex 2 from p. 74 onwards.

Source: FHSO compilation.

The analysis of these documents has been focussed on approaches and indicators that reflect the usage and impact of the Internet on R&D. But specific results have proved scarce and no systematic approaches could be found. We decided therefore to focus on two issues:

- First, we briefly describe important concepts of R&D statistics and what is generally being measured, with the aim of outlining possible paths towards enriching the existing data with new, more Internet-oriented indicators.
- Second, we highlight the few instances which have provided data on the Internet for R&D, though this has usually been performed in a rather non-systematic and incidental way.

4.2 Brief summary of international R&D statistics and their methodological foundations

R&D statistics

The OECD manages the largest and most comprehensive international databases on R&D and innovation including both, input and output indicators. Other international and national organisations and statistical bodies extensively use these databases when they assemble statistical reports and benchmarking studies (of course, the OECD databases are founded on national R&D and innovation surveys, patent and public budget statistics). Main Science and Technology Indicators (MSTI) is the most important database for this study. It contains R&D input data, such as R&D expenditures and R&D personnel, and output data like patents. Useful data for calculating indicators on the inputs and outputs of R&D activities are also included in the Analytical Business Enterprise Research and Development database (ANBERD) which describes R&D in the private business sector. This database includes for example data on the R&D expenditure of business enterprises in the ICT sector.

Another broad data set also including data on the input and output sides of R&D is Eurostat's second Community Innovation Survey (CIS 2). The survey was carried out by the national statistical offices on the basis of a core questionnaire. This questionnaire asked a broad range of innovation-related questions: on the firm's engagement in product and process innovations, the resources they devoted to this, influencing and hampering factors, information sources, innovation co-operation and others. The time horizon of the questions was 1994 to 1996 and the results have been included in the New Cronos database. A follow-up survey, CIS 3, has been carried out in most countries in 2001. Again, a common core questionnaire has formed the basis.

Further international databases which are frequently used to compute science, technology and research and development statistics are:

- The database of The Cooperative Agreements and Technology Indicators (CATI) database developed at the Maastricht Economic Research Institute on Innovation and Technology (MERIT) which collects co-operation agreements that contain arrangements for transferring technology or joint research.[212] The co-operation counts are restricted to strategic technology alliances, such as joint ventures for which R&D or technology sharing is a major objective, research corporations and joint R&D pacts.

Output indicators can be found particularly in the following data sets:

- Patent databases from the World Intellectual Property Organisation (WIPO) or the European Patent Office (EPO) which contain patent applications, patents granted and sometimes both.
- The bibliographical databases of the Institute for Scientific Information (ISI) include information on the authors and citations of articles from roughly 8,500 scientific journals. This data set offers data for ranking nations (as well as authors, institutions and journals) by field of research. The most renowned indicators are the Science Citation Index (SCI), the Social Science Citation Index (SSCI) and the Arts and Humanities Citation Index (A&HCI).[213]

National publications on R&D statistics are usually based on the above mentioned databases, as they either contribute to the data collection or prefer internationally comparable data for benchmarking reasons. Highly remarkable is the US National Science Board's compendium "Science and Engineering Indicators 2000" which presents a compilation of information not only on the US but at a global level on almost 500 pages of text and over 600 pages of annex tables.[214]

Methodological issues

The work of the OECD and the OECD lead Group of National Experts on Science and Technology Indicators (NESTI) has been central to the development of the research, science, technology and innovation statistics. NESTI has developed a series of methodological manuals, of which the Frascati Manual and the Oslo Manual are of particular importance for this study.

- The Frascati Manual[215] presents a broad range of recommendations regarding the collection and interpretation of R&D data. It lays down definitions for research and development and classifies different R&D activities according to their nature (basic research, applied research, experimental development). It sets out which personnel should be considered to be working on R&D activities and which expenditures should be

[212] See http://meritbbs.unimaas.nl.

[213] See http://www.isinet.com/isi/products/citation/sci/index.html.

[214] See National Science Board (2000), op. cit.

[215] OECD (1994), op. cit.

listed among R&D expenditures. The manual also contains a chapter on procedures for R&D surveys.

- The Oslo Manual[216] provides definitions, criteria and classifications which are relevant to the study of industrial innovation. It contains a range of suggestions and recommendations on national and international innovation surveys and statistics.

Besides the two OECD publications, most publications from Eurostat, national statistical offices or other national sources define the R&D indicators used and describe their significance and suitability for analysing the R&D sector. A good example is the publication of the German Federal Statistical Office (in German) on science and technology indicators. This features detailed methodological overviews, including the indicator formulas, explanations on the delimitation of numerators and denominators and indications of sources (document [17] of table 8).

4.3 The Internet in current R&D statistics

OECD statistics on R&D still ignore the computerisation of science and research. The European Commission has at least recognised the importance of computer networks for R&D. One indication can be found in Eurostat's CIS 2 which contains a question on the relevance of computer-based information networks as a source of information for innovation. Other listed sources are the enterprise or the enterprise group, competitors, clients or customers, universities or other higher education institutes, patent disclosures, fairs, exhibitions etc. Unfortunately, these data will not be obtained in CIS 3 as this question has not been followed up. Another indication can be found in the European Survey of the Information Society (ESIS) II project (see document [10] in table 8), a follow-up project which was carried out for Central and Eastern European Countries (CEEC).[217] This survey includes data on the percentage of high schools and universities with Internet websites.

One more source which should be highlighted in the context of this report, even though the connection to our topic is somewhat weak: Statistics Finland produced its second report "On the Road to the Finnish Information Society" in 1999. The report also contains the number of scientific magazines published in Internet-compatible form (a private enterprise database is given as the source).[218] Unfortunately, the follow-up report which is scheduled for October 2001 will not present further details.[219]

This evidence leads to a disappointing conclusion: the use of computer networks for R&D and their impact are virtually non-existent in traditional R&D statistics.

216 OECD (1997), op. cit.

217 See http://www.eu-esis.org/esis2www/synthCEEC7.htm.

218 See Statistics Finland (1999): On the Road to the Finnish Information Society. Helsinki 1999.

219 According to e-mail correspondence with the responsible editor, Mrs Lea Parjo.

5 Summary of Part A

There is no doubt that the Internet has changed research and development. It has become easier to access information that otherwise would be difficult to obtain. Asynchronous communication has become easier and more frequent. R&D collaboration over large distances has been enhanced, as it has become easier to communicate and transmit information even if it is large and bulky. There are essentially three perspectives on the influence of the Internet which in the further course of the project will be considered as 'sub-topics':

- Internet-related ICT infrastructure for R&D,
- integration of the new network technologies into research activities,
- enhancement of new collaboration-based institutions for performing R&D.

5.1 Summary of available indicators and data collection methods

This section summarises the indicators which could be retrieved from the scientific literature related to the Internet for R&D, policy documents and R&D statistics.[220] Table 10 contains an overview of existing indicators and indicators under development for evaluating R&D-oriented Internet issues. The table is limited to indicators and concepts that were found in one (or more) of the different sources evaluated. *It does not contain any new indicators that are to be developed within the SIBIS project.* This task will be carried out in the second deliverable of this workpackage 2, the indicator report on the Internet for R&D. As we can infer from table 10, the majority of indicators – 19 out of 24 – are somehow related to measuring infrastructure aspects. Fewer indicators are related to the integration of the new network technologies into research activities and only one indicator may provide some information on computer networks and R&D collaborations. This imbalance may be due to the fact that R&D collaborations in general seem to be rather difficult and unsatisfactory to measure, even when well established concepts form the guiding framework (see section 2.3.2, p. 45).

Table 10: Overview of traditional and innovative indicators on the Internet for R&D

No.	Name of indicator	Sub-topic	Methods needed
1.	Core transmission capacity of national research networks	Internet-related ICT infrastructure	Survey of research network providers
2.	Core network size of national research networks	Internet-related ICT infrastructure	Survey of research network providers
3.	Market penetration of national research networks	Internet-related ICT infrastructure	Survey of research network providers
4.	Transmission capacity to Europe/United States	Internet-related ICT infrastructure	Survey of research network providers
5.	Transmission capacity to large sites	Internet-related ICT infrastructure	Survey of research network providers
6.	Traffic on national research networks per student	Internet-related ICT infrastructure	Survey of research network providers
7.	Total budgets of national research networks	Internet-related ICT infrastructure	Survey of research network providers

[220] As there were few occurrences of traditional and innovative indicators in each of the three source categories that were analysed, we will not make any distinction between traditional or innovative indicators as performed in the reports on other SIBIS topics.

Table 10

No.	Name of indicator	Sub-topic	Methods needed
8.	Total budgets of national research networks per student/capita	Internet-related ICT infrastructure	Survey of research network providers
9.	Staffing of national research networks	Internet-related ICT infrastructure	Survey of research network providers
10.	Total amount of effort available to national research networks	Internet-related ICT infrastructure	Survey of research network providers
11.	Internet penetration rate of high schools and universities (HEI)	Internet-related ICT infrastructure	Evaluation of websites
12.	Titles in digital collections	Internet-related ICT infrastructure	Survey of providers of electronic resources (e.g. digital libraries)
13.	Usage of digital libraries	Internet-related ICT infrastructure	Analysis of server log files
14.	Library staff for electronic services	Internet-related ICT infrastructure	Survey of providers of electronic resources (e.g. digital libraries)
15.	Expenditure on electronic resources in relation to total expenditures	Internet-related ICT infrastructure	Survey of providers of electronic resources (e.g. digital libraries)
16.	Perceived benefits of computer-mediated communication	Internet-related ICT infrastructure	Survey of R&D personnel
17.	Communication activity within a mailing list	Internet-related ICT infrastructure	Analysis of mailing list archives
18.	Perceived benefits of the Internet	Internet-related ICT infrastructure/ Integration of the new network technologies into research activities	Survey of R&D personnel
19.	Frequency of computer-mediated communication	Internet-related ICT infrastructure/ Integration of the new network technologies into research activities	Survey of R&D personnel
20.	Web impact factor (Web-IF) of a country	Integration of the new network technologies into research activities	Evaluation of websites (webometrics)
21.	Frequency of information retrieval from the Internet	Integration of the new network technologies into research activities	Survey of R&D personnel, analysis of client-side log files
22.	Computer networks as information sources for innovation	Integration of the new network technologies into research activities	Survey of business enterprises
23.	E-journals citation index	Integration of the new network technologies into research activities	Bibliometrics
24.	Web reference indicator	Computer networks and R&D collaborations	Evaluation of websites (webometrics)

For more detailed information on the indicators see annex 3 from p. 76 onwards.

Source: FHSO compilation.

Various methodological concepts have been found for collecting and processing data for any of the three sub-topics:

- Surveys (of R&D personnel, business enterprises, providers of electronic resources) by regular mail or via the Internet,
- the evaluation of WWW pages,
- the analysis of log files,
- the analysis of archives (e.g. of mailing lists),
- bibliometrics.

These methods might either be used as substitutes to obtain one piece of information on a more sound and secure basis. They might also be applied to obtain complementary information for describing more facets of a subject. Each method has advantages and disadvantages: surveys are easy to construct and they might render exactly the information which is needed. But they are costly to carry out and it can be difficult to construct a representative sample and have the questions answered by an overstressed set of researchers and R&D managers. The information on WWW pages is available day and night and is easily accessible. But it might be outdated, faulty and the coverage of pages retrieved by means of search engines might be biased and not representative. Server log files record all traffic on a server but the available tools for analysing these files are calibrated to record different features. Frequently servers might not even record requests, if a web page has already been cached locally. These are only some examples of methodological problems which will have to be dealt with when developing the SIBIS indicators within the further course of the project.

In order to ensure a structured analysis with a similar scheme for all topics of the SIBIS project, the indicators are described in tables. These are included in annex 3 (see pp. 76-92). There, each indicator is defined and the methods that are needed to compute it are listed (e.g. survey, web page analysis etc.). Furthermore, some additional information is given on potential measurement problems and inaccuracies. Policy relevance and factors affecting the future relevance and the possibility of assembling time series are also mentioned. Information on the availability of time series and the country coverage can only be presented if the indicator has already been calculated and data has been collected which is usually not the case. More often, additional data and computational steps will be necessary to aggregate 'raw' data to the country level.

5.2 Topic-related summary: Issues from the literature on the Internet and R&D

Sub-topic A: Internet-related ICT infrastructure for R&D

A large body of literature on the Internet-related ICT infrastructure for R&D has been evaluated in this report. Research networks, the first infrastructure element described, exist on different spatial levels. TERENA, the Trans-European Research and Education Networking Association, has recently initiated the collection of data and formation of indicators on national networks. It offers indicators on transmission capacities, budgets, staffing and traffic on the networks which form a good basis for further considerations (see indicators [1]-[10] in annex 3).

Furthermore, various on-line information sources and computer-mediated communication tools were analysed. Their market penetration differs markedly: A few, such as e-mail, have become omnipresent tools for researchers. Some are mainly used in specific research disciplines such as numerical databases. Other applications are still in the market introduction phase such as on-line meeting tools. The same applies to grid computing, a new concept of distributed research work which is based on the capacities of ICT infrastructures. Consequently, the possibilities of evaluating these tools and concepts also differ: When usage is not very common, it is impossible or at least difficult to collect quantitative data, establish benchmarks or make comparisons at national level. But also when usage becomes widespread, certain data might loose their value: the simple distinctions between ‘e-mail users’ and ‘e-mail non-users’ or between ‘on-line’ and ‘offline-only’ journals are hardly relevant anymore, as e-mail usage rates and the on-line availability of journals peak. This has significant implications for the construction of time series and data collection, inasmuch as the scale has to be considered closely. Binary scales might be rather short lived.

Nevertheless, some preliminary efforts at quantifying Internet applications in R&D have been undertaken and we found some studies which even attempted to assess the effects of Internet applications on research success.[221] These studies and most of the data collection on Internet infrastructure for research were carried out at the micro-level, i.e. the individual researcher, the individual research paper, the individual research organisation etc. as units of observation. Even with digital libraries, for which the most advanced indicator concepts were found, only the micro-level perspective has been taken (see indicators [12]-[15] in annex 3). The concepts were developed to answer the basic question: What indicators are necessary to measure the performance of electronic library services?[222] Of course, this micro-level approach is very useful for comparing performance, to attribute funds and even to set policy goals. But it has rarely led to indicators which are appropriate for the comparison of national research systems. Usually for that purpose additional steps are compulsory, if only to aggregate the data of the observation units. More often than not these additional steps are difficult, for example as rights to privacy would have to be violated, or because it is hard to establish a representative sample, or even as the data collection is very costly in relation to the expected value.

Sub-topic B: Integration of the new network technologies into research activities

The Internet is being used by researchers mainly for collecting data and retrieving scientific information from different virtual storage spaces and for presenting, disseminating and discussing research results. Internet-based data analysis has not yet become commonplace, but new developments in the field of grid computing will increase the possibilities in this field.

The literature that deals with the usage of network infrastructure, on-line resources and Internet tools in R&D mainly discusses their advantages and disadvantages for different steps of the R&D process, e.g. the benefits of on-line questionnaires compared to regular methods; or the time and cost savings and distribution effects of e-publishing compared to print publishing. Again, typically a micro-level perspective has been employed and macro-level research questions have been disregarded (e.g. in which countries is advanced on-line research carried out, where do the authors of e-journal articles come from?). Nevertheless, this should not be considered as a problem, it is simply a fact.

Still, SIBIS can also benefit from the literature that examines how the Internet is being integrated into different steps of the R&D process. The benefit comes from a methodological point of view: Internet-based data collection might constitute a valuable method for gathering data and constructing indicators on different issues of the Internet for R&D.

Sub-topic C: Computer networks and R&D collaborations

The new research policy of the European Commission has been laid down in the documents on a European Research Area (ERA). Increasing *collaboration among R&D institutes* from different EU and CEEC countries in general as well as creating top-level virtual R&D institutes, labelled ‘virtual centres of excellence’, are important goals of ERA. The empirical evidence that worldwide research collaboration has increased is strong. Some authors also

---

221 See e.g. the following studies: Walsh, J.P.; Maloney, N. G. (2001), op. cit. – Lawrence, S. (2001), op. cit. – Kaminer, N.; Braunstein, Y.M. (1998), op. cit. – Cohen, J. (1996), op. cit. – Hesse, B.W. et al. (1993), op. cit.

222 See International Coalition of Library Consortia (1998), op. cit. – Brophy, P. et al. (2000), op. cit. – Shim, W. et al. (2000), op. cit.

attribute this fact to the diffusion of capable ICT infrastructure and the Internet applications discussed above. The operationalisation of R&D collaboration has been based either on few available output indicators (bibliometric data, patents) or on count data of R&D partnerships. Neither of the two methodologies permits an adequate measurement of the different concepts of R&D collaboration which have been developed in social sciences theoretical literature.

The novel concepts of collaboratories and virtual teams, two new forms of network-supported long-distance collaborations, have only been described in case studies so far. How to quantify their significance or their geographical spread still remains unsolved.

The Indicator Report will further discuss the strengths and weaknesses of the indicators which have been found in the literature and advance some of them with the objective of applying them to an international comparison of the usage and impact of the Internet on R&D. The next report will also work out its own proposals for new indicators which have not yet been computed and discuss the strengths and weaknesses of these indicators in a similar manner.

6 Bibliography for PART A

The bibliography does not include the policy documents and statistical documents analysed in section 3 and 4. Bibliographical and retrieval information on these documents can be found in annexes 1 and 2.

Aiken, R. A. (2000): New Frontiers for Research Networks in the 21st Century, in: Wouters, P.; Schröder, P. (eds.): Access to publicly financed research. The Global Research Village III, Amsterdam 2000. Background papers. Amsterdam.

Almind, T.; Ingwersen, P. (1997): Informetric analyses on the World Wide Web: methodological approaches to 'webometrics', in: Journal of Documentation, vol. 53, no. 4, pp. 404-426.

Anderson, K.; Sack, J.; Krauss, L.; O'Keefe, L. (2001): Publishing Online-Only Peer-Reviewed Biomedical Literature: Three Years of Citation, Author Perception, and Usage Experience, in: Journal of Electronic Publishing, vol. 6, no. 3 (http://www.press.umich.edu/jep/06-03/anderson.html).

Armstrong, C. J.; Lonsdale, R. E.; Stoker, D. A.; Urquhart, C. J. (2000): JUSTEIS - JISC Usage Surveys: Trends in Electronic Information Services. (http://www.dil.aber.ac.uk/dils/research/justeis/cyc1rep0.htm).

Aubert, J.E.; Bayar, V. (1999): Maximising the Benefits of Information Technology for Science: Overview and Major Issues, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, pp. 7-28.

Axmann, H.-P.; Payr, S. (1999): A Global Research Infrastructure for the Global Research Village: The European Perspective, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, pp. 29-47.

Bainbridge, W. S. (1999): Information infrastructure issues in the social sciences, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, pp. 123-135.

Baker, M.; Buyya, R.; Laforenza, D. (2000): The Grid: international efforts in global computing. Presentation at the SSGRR International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, L'Aquila, July 31 - August 06 2000. (http://www.ssgrr.it/en/ssgrr2000/papers/268.pdf).

Bar-Ilan, J. (1999): Search engine results over time – A case study on search engine stability, in: Cybermetrics, vol. 2/3, no. 1. (http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html).

Bauer, K. (2000): Who Goes There? Measuring Library Web Site Usage, in: Online, January 2000. (http://www.onlineinc.com/onlinemag/OL2000/bauer1.html).

Becker, H. J.; Klaproth, F.; Lepschy, P. (2000): Survey of end user surveys. Renardus Deliverable 6.6. (http://www.renardus.org/deliverables/d6_6/D6_6_final.pdf).

Beheshti, J. (2000): Looking for the needle in a haystack: retrieving information from the web. Presentation at the SSGRR International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, L'Aquila, July 31 - August 06, 2000. (http://www.ssgrr.it/en/ssgrr2000/papers/206.pdf).

Bingham, C. (1998): Peer review on the Internet: a better class of conversation, in: The Lancet, vol. 351 (suppl I), pp. 10-14. (http://thelancet.com/journal/vol351/isss1/full/llan.351.s1.reviews.13114.1).

Björk, B.-C.; Turk, Z. (2000): How scientists retrieve publications: an empirical study of how the Internet is overtaking paper media, in: The Journal of Electronic Publishing, vol. 6, no. 2. (http://www.press.umich.edu/jep/06-02/bjork.html).

Bosnjak, M.; Bandilla, W.; Schneid, M.; Lorch, G.; Batinic, B.; Werner, A.; Stiegler, A. (1998): On-line Forschung im deutschsprachigen Raum. Erste Ergebnisse einer Umfrage unter Mitgliedern der 'German Internet Research' Mailingliste. (http://www.or.zuma-mannheim.de/inhalt/projekte/or_expert/girl98_1.pdf).

Boudourides, M.A. (1999): Complexity of science and technology data and their communication in networks of social understanding. SOEIS Project, Task 2.1 deliverable, June 1999 (http://hyperion.math.upatras.gr/webometrics)

Broder, A.; Kumar, R.; Maghoul, F.; Raghavan, P.; Rajagopalan, S.; Stata, R.; Tomkins, A.; Wiener, J. (2000): Graph structure in the web. (http://www9.org/w9cdrom/160/160.html)

Brophy, P.; Clarke, Z.; Brinkley, M.; Mundt, S.; Poll, R. (2000): EQUINOX - Library Performance Measurement and Quality Management System: Performance Indicators for Electronic Library Services. (http://equinox.dcu.ie/reports/pilist.html).

Burton, M. C.; Walther, J. B. (2001): The Value of Web Log Data in Use-Based Design and Testing, in: Journal of Computer-Mediated Communication, vol. 6, no. 3. (http://www.ascusc.org/jcmc/vol6/issue3/burton.html).

Butler, D. (2000): Souped-up search engines, in: Nature, vol. 405, 11 May 2000, pp. 112-115.

Butterworth, I. (1998): The impact of electronic publishing on the academic community an international workshop org. by the Academia Europaea and the Wenner-Gren Foundation. London [etc.]: Portland Press.

Cairncross, F. (2001): The Death of Distance, 2nd edition, Boston, Mass.: Harvard Univ. School Publications.

Chen, C.; Newman, J.; Newman, R.; Rada, R. (1998): How did university departments interweave the Web: A study of connectivity and underlying factors, in: Interacting with computers, vol. 10, pp. 353-373.

Clemons, E. K.; Row, M. C. (1992): Information Technology and Industrial Cooperation: The Changing Economics of Coordination and Ownership, in: Journal of Management Information Systems, vol. 9, no. 2, pp. 9-28.

Cohen, J. (1996): Computer mediated communication and publication productivity among faculty, in: Internet Research: Electronic Networking Applications and Policy, vol. 6, no. 2/3, pp. 41-63.

Cowan, R.; David, P. A.; Foray, D. (2000): The explicit economics of knowledge codification and tacitness, in: Industrial and corporate change, vol. 9, no. 2, pp. 211-253.

Cybermetrics. International Journal of Scientometrics, Informetrics and Bibliometrics (http://www.cindoc.csic.es/cybermetrics/cybermetrics.html).

David, P.; Foray, D. (1995): Accessing and expanding the science and technology knowledge base, in: STI-review, no. 16, pp. 13-68.

Day, J.; Bartle, C. (1998): The Internet as an Electronic Information Service: Its Impact on Academic Staff in Higher Education. Proceedings IRISS '98 International Conference: 25-27 March 1998, Bristol, UK. (http://sosig.ac.uk/iriss/papers/paper06.htm)

Dodgson, M. (1996): Technological collaboration and innovation, in: Dodgson, M.; Rothwell, R. (eds.): The handbook of industrial innovation, Cheltenham and Brookfields, pp. 285-292.

European Commission (2000a): eEurope: An Information Society for all. Communication on a Commission Initiative for the Special European Council of Lisbon, 23 and 24 March 2000.

European Commission (2000b): eEurope 2002: An Information Society for all. Action plan prepared by the Council and the European Commission for the Feira European Council, 19 and 20 June 2000.

European Commission (2000c): Towards a European research area. Communication from the Commission COM 2000 (6). Brussels, 18 January 2000.

European Research Consortium for Informatics and Mathematics (ERCIM): Grids: e-science to e-business, ERCIM News, No. 45, April 2001. (http://www.ercim.org/publication/Ercim_News/enw45/)

European Technology Assessment Network (ETAN) (1999): Transforming European science through information and communication technologies: challenges and opportunities of the digital age. Final version.

Etzkowitz, H.; Leydesdorff, L. (eds.) (1997): Universities in the global knowledge economy: A triple helix of university-industry-government relations. London.

Finholt, T.A. (2001): Collaboratories, in: B. Cronin (Ed.): Annual Review of Information Science and Technology, vol. 36. (also available at: http://intel.si.umich.edu/crew/technical_reports_alphabetical.htm)

Finholt, T.A.; Rocco, E.; Bree, D.; Jain, N.; Herbsleb, J.D. (1998): NotMeeting: A field trial of NetMeeting in a geographically distributed organization, in: SIGGROUP BULLETIN, vol. 20, no. 1, pp. 66-69. (http://www.bell-labs.com/org/11359/colab_prod/).

Foray, D. (1999): Building the Virtual 'House of Salomon': Digital collaboration technologies, the organisation of scientific work and the economics of knowledge access. Report of the ESF-IIASA-NSF Workshop - 3 to 5 December 1999 - At the International Institute for Applied System Analysis, Laxenburg, Austria, p. 9. (http://www.esf.org/policy/pdf/iiasa.pdf)

Foray, D.; Steinmueller, E. (1999): Collective invention and European policies "Colline" 1997-1999. (http://www.dauphine.fr/imri/COLLINE/exesummary.html).

Foster, I. (2000): Internet Computing and the Emerging Grid, in: nature web matters, 07. December 2000. (http://www.nature.com/nature/webmatters/grid/grid.html)

Foster, I; Kesselman, C.; Tuecke, S. (2001): The Anatomy of the Grid. Enabling Scalable Virtual Organisations, to appear in: International journal of Supercomputer Applications. (http://www.globus.org/research/papers/anatomy.pdf)

Fox, E.A. (1999): Networked Digital Library of Theses and Dissertations, in: Nature web matters, 12 Aug 1999 (http://www.nature.com/nature/webmatters/library/library.html).

Foyn, F. (2000): Community Innovation Survey 1997/98 - final results. Eurostat (Ed.): Statistics in Focus: Theme 9 Research and Development, no. 2/2000.

Garvey, W.D. (1977): Communication: the essence of science. Facilitating information exchange among librarians, scientists, engineers and students. Pergamon Press: Oxford, New York, Toronto, Sydney, Paris, Frankfurt.

Gibbons, M.; Limoges, C.; Nowotny, H.; Schwartzman, S.; Scott, P.; Trow, M. (1994): The New Production of Knowledge. Sage: London; Thousand Oaks; New Delhi.

Ginsparg, P. (1994): First steps toward electronic research communication, in: Computers in Physics, vol. 8, no. 4, pp. 390-396. (also available from: http://arxiv.org/blurb).

Godin, B.; Gingras, Y. (2000): The Place of Universities in the System of Knowledge Production, in: Research Policy, vol. 29, no.2, pp. 273-278.

Grenzmann, C.; Marquardt, R.; Wudtke, J. (1999): Forschung und Entwicklung in der Wirtschaft 1997 - 1999. Bericht über die FuE-Erhebungen 1997 und 1998. Essen 1999.

Gresham, J. (1994): From invisible college to cyberspace college: computer conferencing and the transformation of informal scholarly communication networks. (http://www.helsinki.fi/science/optek/1994/n4/gresham.txt).

Guthrie, K. (2000): Revitalizing Older Published Literature: Preliminary Lessons from the Use of JSTOR. Paper for the conference "The Economics and Usage of Digital Library Collections", 23-24 March 2000, Ann Arbor, Michigan. (http://www.si.umich.edu/PEAK-2000/guthrie.pdf).

Hagedoorn, J.; Link, A. N.; Vonortas, N. S. (2000): Research partnerships, in: Research Policy, vol. 29, pp. 567-586.

Hagedoorn, J. (1996): Trends and patterns in strategic technology partnering since the early seventies, in: Review of Industrial Organisation, vol. 11, pp. 601-616.

Harter, S. P.; Kim, H. J. (1997): Electronic Journals and Scholarly Communication: A Citation and Reference Study, in: The Journal of Electronic Publishing, vol. 3, no. 2. (http://www.press.umich.edu/jep/archive/harter.html).

Harter, S. P. (1996): The Impact of Electronic Journals on Scholarly Communication: A Citation Analysis, in: The Public-Access Computer Systems Review, vol. 7, no. 5. (http://info.lib.uh.edu/pr/v7/n5/hart7n5.html).

Herbsleb, J.D.; Mockus, A.; Finholt, T.A.; Grinter, R.E. (2000): Distance, Dependencies, and Delay in a Global Collaboration, in: Proceedings of CSCW 2000, Philadelphia, PA, Dec. 2-7, 2000 (http://www.bell-labs.com/org/11359/colab_prod/).

Hesse, B. W.; Sproull, L. S.; Kiesler, S. B.; Walsh, J. P. (1993): Returns to science: Computer networks in oceanography, in: Communications of the ACM, vol. 36, no. 8, pp. 90-101.

Ingwersen, P. (1998): The calculation of web impact factors, in: Journal of Documentation, vol. 54, no. 2, pp. 236-243.

Interagency Working Group (IWG) on Information Technology Research and Development (IT R&D) of the National Science and Technology Council (2001): Information Technology: The 21st Century Revolution. (http://www.itrd.gov/pubs/blue01/).

International Coalition of Library Consortia (1998): Guidelines for Statistical Measures of Usage of Web-based Indexed, Abstracted, and Full-Text Resources. (http://www.library.yale.edu/consortia/webstats.html).

Journal of Computer-Mediated Communication (http://www.ascusc.org/jcmc).

Kaminer, N.; Braunstein, Y.M. (1998): Bibliometric analysis of the impact of Internet use on scholarly productivity, in: Journal of the American Society for Information Science, vol. 49, no. 8, pp. 720-730.

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

Kling, R.; McKim, G. (2000): Not just a matter of time: Field differences and the Shaping of Electronic Media in Supporting Scientific Communication, in: Journal of the American Society for Information Science, vol. 51, no. 14, pp. 1306-1320. (also available at: http://xxx.lanl.gov/ftp/cs/papers/9909/9909008.pdf).

Kling, R.; McKim, G. (1998): The Shaping of Electronic Media in Supporting Scientific Communication: The Contribution of Social Informatics. Paper presented at the "European Science and Technology Forum: Electronic Communication and Research in Europe" Darmstadt/Seeheim, 15 to 17 April 1998. (http://www.slis.indiana.edu/kling/pubs/seeheim.htm).

Koku, E.; Nazer, N.; Wellman, B. (2000): Netting Scholars: Online and Offline, in: American Behavioral Scientist, vol. 43. (also available at: http://www.chass.utoronto.ca/~wellman/ publications/nettingscholars/scholnet-abs9a.pdf)

Krantz, J. H.; Dalal, R. (2000). Validity of Web-based psychological research, in: M. H. Birnbaum (Ed.): Psychological Experiments on the Internet. San Diego, CA.

Lawrence, S. (2001): Free online availability substantially increases a paper's impact, in: Nature, vol. 411, no. 6837, p. 521. (also available as ‘Online or invisible?’ at: http://www.neci.nec.com/~lawrence/papers.html).

Lawrence, S.; Giles, C. L. (1999): Accessibility of information on the web, in: Nature, vol. 400, 8 July 1999, pp. 107-109.

Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997): Internet use by faculty members in various disciplines: a comparative case study, in: Journal of the American Society for Information Science, vol. 48, no. 6, pp. 508-518.

Leiner, B. (1998): The Scope of the Digital Library. (http://www.dlib.org/metrics/public/papers/dig-lib-scope.html).

Lenares, D. (1999): Faculty use of electronic journals at research institutions, in: Proceedings of the ACRL Ninth National Conference. (http://www.ala.org/acrl/lenares.pdf).

Leydesdorff, L. (2001): The challenge of scientometrics. The development, measurement, and self-organization of scientific communications. 2nd edition.

Leydesdorff, L.; Curran, M.: (2000): Mapping university-industry-government relations on the Internet: the construction of indicators for a knowledge-based economy, in: Cybermetrics, vol. 4, no. 1, paper 2 (http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html).

Lide, D. R. (1994): The impact of information technology on the access to science, in: Expanding Access to Science and Technology: The Role of Information Technologies, edited by Ines Wesley-Tanaskovic, Jacques Tocatlian, and Kenneth H. Roberts. Proceedings of the Second International Symposium on the Frontiers of Science and

Technology, Kyoto, Japan, 12-14 May 1992. (http://www.unu.edu/unupress/unupbooks/uu07ee/uu07ee05.htm).

Liebscher, P.; Abels, E. G.; Denman, D. W. (1997): Factors that influence the use of electronic networks by science and engineering faculty at small institutions. Part II: Preliminary use indicators, in: Journal of the American Society for Information Science, vol. 48, no. 6, p. 496-507.

Lubanski, A.; Matthew, L. (1998). Socio-economic Impact of the Internet in the Academic Research Environment. Proceedings IRISS '98 International Conference: 25-27 March 1998, Bristol, UK. (http://sosig.ac.uk/iriss/papers/paper18.htm.

Luther, J. (2000): White Paper on Electronic Journal Usage Statistics, in: Journal of Electronic Publishing, vol. 6, no. 3. (http://www.press.umich.edu/jep/06-03/luther.html).

MacElroy, B. (1999): Comparing seven forms of on-line surveying, in: Quirk’s Marketing Research Review, article no. 510. (http://www.quirks.com/articles/article_print.asp?arg_articleid=510).

Machlup, F. (1962): The production and diffusion of knowledge in the United States. Princeton.

Malone, T.W.; Rockart, J. (1992): Information technology and the new organisation, in: Proceedings of the 25th Hawaii International Conference on Systems Sciences, vol. 4, pp. 636-643.

Mark, G; Grudin, J.; Poltrock, S. E. (1999): Meeting at the Desktop: An Empirical Study of Virtually Collocated Teams, in: Proceedings of ECSCW'99, The 6th European Conference on Computer Supported Cooperative Work, 12-16 September 1999, Copenhagen, Denmark, pp. 159-178. (http://research.microsoft.com/users/jgrudin/)

Matzat, U. (2001): Academic Communication and Internet Discussion Groups. Their Spread, Use and 'Survival' within Academic Communities. (http://www.ppsw.rug.nl/matzat/short-project-description.html)

Matzat, U. (1998): Informal Academic Communication and Scientific Usage of Internet Discussion Groups. Proceedings IRISS '98 International Conference: 25-27 March 1998, Bristol, UK. (http://sosig.ac.uk/iriss/papers/paper19.htm)

Merten, K.: Kommunikation. Eine Begriffs- und Prozessanalyse. Opladen 1977.

Mettrop, W., Nieuwenhuysen, P. (2000): The reliability of Internet search engines: Fluctuations in Document Accessibility, in: Proceedings of National Online Meeting 2000, May 16-18, 2000, New York, edited by Martha E. Williams, Information Today Inc. Medford, N. J., pp. 271- 281.

Mitra, A.; Hazen, M.D.; LaFrance, B.; Rogan, R.G. (1999): Faculty Use and Non-Use of Electronic Mail: Attitudes, Expectations and Profiles, in: Journal of Computer-Mediated Communication, vol. 4, no. 3 (http://www.ascusc.org/jcmc/vol4/issue3/mitra.html).

Muzart, G. (1999): Description of national innovation surveys carried out, or foreseen, in 1997-99 in OECD non-CIS-2 participants and NESTI observer countries. OECD STI working papers 19999/1, Paris.

National Science Board (2000): Science & engineering indicators 2000. Arlington, Va. (http://www.nsf.gov/sbe/srs/seind00/start.htm).

Nature: Web debate: Future e-access to the primary literature (http://www.nature.com/nature/debates/e-access/).

Nelson, R. R.; Winter, S. G. (1982): An evolutionary theory of economic change. Cambridge and London.

Noam, E. M. (1995): Electronics and the Dim Future of the University, in: Science, vol. 270, pp. 247-249.

Nonaka, I. (1994): A dynamic theory of organizational knowledge creation, in: Organization science, vol. 5, pp. 14-37.

OECD (1999): Science, technology and industry scoreboard 1999. Benchmarking Knowledge-based Economies. Paris.

OECD (1998): The Global research village: How information and communication technologies affect the science system. Paris. (http://www.oecd.org/dsti/sti/s_t/scs/prod/ global.pdf)

OECD (1994): The Measurement of Scientific and Technological Activities: Proposed Standard Practice for Surveys of Research and Experimental Development - Frascati Manual 1993. 5th Edition. Paris.

Olson, G.M.; Olson, J.S. (2002): Distance matters, in: John M. Carroll (Ed.): Human-Computer Interaction in the New Millenium. (forthcoming), pp. 139-179.

Place, E. (2000): Recent Results from the DESIRE project, in: Scimitar, vol. 16 (http://www.iihe.ac.be/scimitar/J0100/yuri_d2-bristol-ejpaper.html).

Porat, M. U. (1977): The information economy: definition and measurement. OT special publication 77-12 (1).

PricewaterhouseCoopers (1999): Impact of the Telematics for Libraries Programme under the Fourth Framework Programme. (ftp://ftp.cordis.lu/pub/ist/docs/digicult/impact.pdf).

Rallet, A.; Torre, A. (1999): Is Geographical Proximity Necessary in the Innovation Networks in the Era of the Global Economy, in: GeoJournal, vol. 49, no. 4, pp. 373-380.

Reips, U.-D. (2000): Was ist Online-Forschung?, in: Internet-Umfragen Newsletter, no. 18. (http://www.or.zuma-mannheim.de/inhalt/Informationsquellen/newsletter/um18.htm).

ResearchIndex (http://citeseer.nj.nec.com/cs).

Rojo, A. (1995): Participation in scholarly electronic forums. (http://www.digitaltempo.com/e-forums/Oframes.html).

Rous, B. (2001): Usage statistics for online literature, in: Professional/Scholarly Publishing Bulletin, vol. 2, no. 1. (http://www.pspcentral.org/bulletins/spring2001_bulletin_pdf.pdf).

Rousseau, R. (1999): Daily time series of common single word searches in AltaVista and NorthernLight, in: Cybermetrics, vol. 2/3, no. 1. (http://www.cindoc.csic.es/cybermetrics/articles/ v2i1p2.html).

Rusch-Feja, D.; Siebeky, U. (1999): Evaluation of Usage and Acceptance of Electronic Journals. Results of an Electronic Survey of Max Planck Society Researchers including Usage Statistics from Elsevier, Springer and Academic Press (Full Report), in: D-Lib Magazine, vol. 5, no. 10. (http://www.dlib.org/dlib/october99/rusch-feja/10rusch-feja-full-report.html).

Sanderson, D. (1996). Cooperative and collaborative mediated research, in T. M. Harrison and T. D. Stephen (eds): Computer Networking and Scholarly Communication in the 21st Century University, New York, pp. 95-114.

Sandstrom, T. A.; Watson, V. R. (2001): Collaborative Expeditions through Remote Scientific Data via the Internet, SSGRR 2001 International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, L'Aquila, Aug 06 - Aug 12 2001. (http://www.ssgrr.it/en/ssgrr2001/papers/ Velvin%R.%Watson.pdf)

Segal, B. (2000): Grid Computing: The European Data Grid Project. Presentation to the IEEE Nuclear Science Symposium and Medical Imaging Conference, Lyon, 15-20 October 2000. (http://web.datagrid.cnr.it/pls/portal30/docs/1442.DOC).

Shapiro; C.; Varian, H. (1999): Information rules. A strategic guide to the network economy. Boston, Mass.

Shim, W.; McClure, C.R.; Bertot, J.C.; Sweet, J.T.; Maffré de Lastens, J.-M.; Dagli, A.; Fraser, B.T. (2000): ARL E-metrics project: developing statistics and performance measures to describe electronic information services and resources for ARL libraries. (http://www.arl.org/stats/newmeas/emetrics/phaseone.pdf).

Smarr, L. (2000): Infrastructure for Science Portals, in: IEEE Internet Computing, January/February 2000, pp. 71-73. (http://www.computer.org/internet/v4n1/smarr.htm).

Smith, D.; Katz, S. (2000): Collaborative Approaches to Research. Fundamental Review of Research Policy and Funding. Final Report. The Higher Education Funding Council for England (HEFCE).

Smith, R. (1998): The Future of Peer Review, in: OECD: The Global Research Village: How Information and communication Technologies Affect the Science System, Paris, pp. 255-265. (http://www.oecd.org/dsti/sti/s_t/scs/prod/global.pdf)

Sudweeks, F.; Allbritton, M. (1996): Working together apart: Communication and Collaboration in a Networked Group. (http://www.it.murdoch.edu.au/~sudweeks/papers/acis96.html), also published in: C. D. Keen, C. Urquhart and J. Lamp (eds), Proceedings of the 7th Australasian Conference of Information Systems (ACIS96), Vol. 2, Department of Computer Science, University of Tasmania, pp. 701-712.

Tague-Sutcliffe, J. (1992): An introduction to informetrics, in: Information Processing & Management, vol. 28, no. 1, pp. 1-3.

Tenopir, C.; King, D.W. (2000): Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers. Washington, D.C.: Special Libraries Association.

Trans-European Research and Education Networking Association (TERENA) (2001): Terena Compendium of National Research and Educational Networks in Europe. (http://www.terena.nl/compendium/ToC.html).

The Global Research Village III (2000): Access to publicly financed research. Background papers. Amsterdam. (http://www.minocw.nl/english/conferentie/deel4.pdf)

The Journal of Electronic Publishing (http://www.press.umich.edu/jep/index.html).

Thelwall, M. (2001): The responsiveness of search engine indexes, in: Cybermetrics, vol. 5, no. 1. (http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html).

Van Alstyne, M.; Brynjolfsson, E. (1996): Could the Internet Balkanize Science?, in: Science, vol. 274, no. 5292, pp. 1479-1480.

Vietsch, K. (2000): Research Networking in Europe, in: Scimitar Electronic Journal, vol. 17. (http://www.iihe.ac.be/scimitar/J1000/rns.html).

Waldorp, M.M. (1995): On-Line Archives Let Biologists Interrogate the Genome, in: Science 269, 8 September.

Walsh, J. P.; Bayma, T. (1996): The Virtual College: Computer-Mediated Communication and Scientific Work, in: The Information Society, vol. 12, pp. 343-363.

Walsh, J.P.; Maloney, N. G. (2001): Computer network use, collaboration structures and productivity, in: P. Hinds and S. Kiesler (ed.): Distributed work. Cambridge, Mass.

Walsh, J.P.; Roselle, A. (1999): Computer Networks and the Virtual College, in: STI Review No. 24: Special Issue on The Global Research Village. Paris, pp. 49-77.

Wellman, B. (2001): Computer networks as social networks, in: Science, vol. 293, 14. Sep. 2001, pp. 2031-2034.

Wierba, E.E; Finholt, T.A.; Steves, M. (2001): "What Have You Done for Me Lately?" A Case Study of Barriers to Collaborative Tool Adoption in a Manufacturing Engineering Setting. (http://intel.si.umich.edu/crew/Technical%20reports/Wierba_Finholt__What_have_you_done_for_me_lately_05_31_01.pdf)

Wormell, Irene (1998): Informetric Analysis of the international impact of scientific journals: How 'international' are the international journals?, in: Journal of Documentation, vol. 54, no. 5, pp. 584-605.

Zelman, A.; Leydesdorff, L. (2000): Threaded Email Messages in Self-Organization and Science & Technology Studies Oriented Mailing Lists, in: Scientometrics, vol. 48, no. 3, pp. 361-380.

7 Annex to Part A

7.1 Policy documents

Austrian Federal Government (2000): Declaration of the Federal Government on Current Issues in Research and Technology Policy. (http://www.bmwf.gv.at/4fte/materialien/regerkl-en.htm).

Austrian Federal Ministries for Education, Science and Culture and for Transport, Innovation and Technology (2000): Forschungsbericht 2000. [Research report 2000]. (http://www.bmwf.gv.at/4fte/materialien/fober_2000/fb2000.pdf).

Austrian Federal Ministry for Education, Science and Culture (2000): Austrian Memorandum on the Commission Communication "Towards a European Research Area" (http://www.bmwf.gv.at/eu/fueprogramme/memo-en.htm).

Austrian Federal Ministry for Science and Transport (1999): Grünbuch zur österreichischen Forschungspolitik [Green Paper on Austrian Research Policy] (http://www.bmwf.gv.at/ 4fte/materialien/gruenbuch/kurzf_engl.htm).

British Department of Trade and Industry (2000): Excellence and Opportunity: a science and innovation policy for the 21st century. (http://www2.dti.gov.uk/ost/aboutost/dtiwhite/index.html).

British Department of Trade and Industry (2000): Implementation Plan on the White Paper “Excellence and Opportunity: a science and innovation policy for the 21st century”. (http://www2.dti.gov.uk/ost/whatsnew/01.htm).

British Joint Information Systems Committee (2001): Joint Information Systems Committee (JISC) Five-Year Strategy 2001-05. (http://www.jisc.ac.uk/pub01/strat_01_05/exec.html).

British Office of Science and Technology (2001): Science research priorities 2001-02 to 2003-04. (http://www2.dti.gov.uk/ost/ostbusiness/srif_strategy_doc623feb.pdf).

British Office of Science and Technology; Department of Trade and Industry (2000): UK Response to: “Making a Reality of the European Research Area: Guidelines for EU research activities (2002-2006)”. (http://www2.dti.gov.uk/ostinternational/eu_rtd_d/src/fp6/era.doc).

Committee for EU RTD Affairs in Finland (2000): Finnish comments on the communication of the Commission "Making a reality of the European research area: guidelines for EU research activities (2002-2006)”. (ftp://ftp.cordis.lu/pub/finland/docs/finnish_comments_021100.pdf).

Danish Committee for Promotion of IT Research and Education (1999): Promotion of IT Research and Education. (http://www.fsk.dk/fsk/publ/1998/styrke/uk/word.doc).

Danish Council for Research Policy (2000): Danmarks Forskningsråds Årsrapport 1999. [Danish Council for Research Policy Annual Report 1999]. (http://www.fsk.dk/fsk/publ/ 2000/dfaarsrapport/inde0015.htm#0001).

Danish Ministry of Information Technology and Research (2000): A net of opportunities: IT and Telecommunications. Policy Report to the Danish Parliament. (http://www.fsk.dk/fsk/publ/ 2001/anetofopportunities).

Dutch Ministry of Education, Culture and Science (1999): Beleidsagenda HOOP 2000. (http://www.minocw.nl/onderwijs/ho/agenda2000/index.htm).

Dutch Ministry of Education, Culture and Science (1999): Science Budget 2000. (http://www.minocw.nl/english/pdf/budget2k.pdf).

Dutch Ministry of Education, Culture and Science (2000): Boeiend, Betrouwbaar en Belangrijk, Nota Wetenschap- en Techniekcommunicatie. (http://www.minocw.nl/wetenschap/ wtc.doc).

Europe’s Research Council (2000): Resolution Future of European research. 2272nd Council meeting, Press release 9411/00. Strasbourg, 15 June 2000. (http://ue.eu.int/newsroom/ LoadDoc.cfm?MAX=1&DOC=!!!&BID=101&DID=61869&GRP=2577&LANG=1#TOP).

European Commission (2000): Innovation in a knowledge-driven economy. Communication from the European Commission COM(2000) 567 final. Brussels, 20.09.2000. (Http://europa.eu.int/eur-lex/en/com/cnc/2000/com2000_0567en01.pdf).

European Commission (2000): Making a reality of the European research area: guidelines for EU research activities (2002-2006). Communication from the European Commission COM(2000) 612 final. Brussels, 4 October 2000. (Http://europa.eu.int/comm/research/area/ com2000-612-en.pdf).

European Commission (2000): Towards a European research area. Communication from the Commission COM 2000 (6). Brussels, 18 January 2000. (http://europa.eu.int/comm/research/ area/com2000-6-en.pdf).

European Commission (2000): eEurope: An Information Society for all. Communication on a Commission Initiative for the Special European Council of Lisbon, 23 and 24 March 2000.

European Commission (2000): eEurope 2002: An Information Society for all. Action plan prepared by the Council and the European Commission for the Feira European Council, 19 and 20 June 2000.

European Commission (2001): A European Research Area for Infrastructures. Working document of the Commission services SEC (2001) 356. Brussels, 27.02.2001. (ftp://ftp.cordis.lu/pub/improving/docs/infrastructures_sec_2001_356.pdf).

European Commission (2001): Faster Internet – Member States' actions. March 2001. (http://europa.eu.int/information_society/eeurope/action_plan/research/fast/member_states/targets_2001/index_en.htm).

European Commission (2001): First Report on Progress towards the European Research and Innovation Area. Commission Staff Working Paper, Report for the Stockholm Council 23/24 March 2001. Brussels, 16.03.2001. (http://europa.eu.int/comm/stockholm_council/pdf/ resinnov_en.pdf).

European Commission (2001): Proposal for a decision of the European Parliament and the European Council concerning the multiannual framework programme 2002-2006 of the European Community for research, technological development and demonstration activities aimed at contributing towards the creation of the European research area. Communication of the European Commission COM (2001)94 final. Brussels, 21.02.2001. (http://europa.eu.int/ comm/research/pdf/com-2001-94-en.pdf).

European Science Foundation; Academia Europaea (2000): The Need for High Bandwidth Computer-based Networking in Europe. Joint statement of the Academia Europaea and the ESF. (http://www.esf.org/ftp/pdf/SciencePolicy/ESPB7.pdf).

Finnish Government: Finnish comments on the communication of the European Commission “Towards a European Research Area”. (http://www.cordis.lu/finland/en/news-4.htm).

Finnish Ministry of Education (1999): Education and research 1999-2004 - Development Plan. (http://www.minedu.fi/julkaisut/KESU2004/eng/engKESUsis.html).

Finnish Ministry of Education (1999): Education, training and research in the Information Society. A National Strategy for 2000-2004. (http://www.minedu.fi/julkaisut/information/ englishU/index.html).

Finnish Ministry of Education (2000): Information Strategy for Education and Research 2000-2004 – Implementation Plan. (http://www.minedu.fi/julkaisut/pdf/tietostrategia/ toimeenpanosuunnitelmaENG.pdf).

French Interministerial Committee on the Information Society (2000): Développer l'effort public de recherche. [Developing the public effort on research]. (http://www.recherche.gouv.fr/ recherche/politic/cisi).

French Minister of Research (2000): Building a European area for knowledge, science and innovation. (http://dbs.cordis.lu/cordis-cgi/autoftp?FTP=/documents_r5/natdir0000002/s_1559005_ 20001102_170704_GL001379en.pdf&ORFN=GL001379en.pdf).

French Minister of Research (2000): Une nouvelle etape pour la recherche: Dix orientations prioritaires pour la recherche et la technologie. (http://www.recherche.gouv.fr/discours/2000/ priorite/conf0405.rtf).

French Ministry of Research (2000): Status of Research and Technological Development. (ftp://trf.education.gouv.fr/pub/edutel/rt/brochure/plfre00.pdf).

German Federal Ministry on Education and Research (2000): Bundesbericht Forschung 2000 [Federal report on research 2000]. (http://www.bmbf.de/digipubl_1929.html).

German Joint Commission of the Federation and the Länder on education planning and the promotion of research (2000): Multimedia in der Hochschule. [Multimedia in the university]. (http://www.blk-bonn.de/papers/heft85.pdf).

Irish Minister for Commerce, Science and Technology (1996): The 1996 White Paper on Science, Technology and Innovation. (ftp://ftp.cordis.lu/pub/ireland/docs/whitepa.pdf)

Irish Office of Science & Technology (2001): Position Paper on Sixth Framework Programme. (ftp://ftp.cordis.lu/pub/ireland/docs/fp6.pdf).

Italian Minister of the University and Scientific Research and Technology (2000): Programma Nazionale della Ricerca. [National Program of Research]. (http://www.murst.it/ricerca/frameset.html).

Italian Minister of the University and Scientific Research and Technology (2001): Contributo Italiano al Dibattito sul Futuro della Ricerca Europea. [Italy's contribution in the debate on the future of European research]. (http://www.murst.it/ricerca/frameset.html).

Johnston, D. J. (2000): Access to Global Research as a Political Issue. Opening statement at the 3rd Global Research Village Conference Amsterdam, The Netherlands 6-8 December 2000. (http://www.minocw.nl/english/conferentie/report.pdf)

National Technology Agency (Tekes) (2001): Research and technology programme activities in Finland 106/2001. (http://www.tekes.fi/julkaisut/Research_and_technology.pdf)

Spanish Inter-ministerial Commission of Science and Technology (1999): Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica para el período 2000-2003. [National Plan of Scientific Research, Development and Technological Innovation for the period 2000 to 2003]. (http://www.mcyt.es/sepct/PLAN_I+D/pnididocu.htm).

Swedish federal government, Department of Innovation Policy Studies (2000): Research and Renewal: the Government's research policy bill. (http://www.cordis.lu/sweden/res_bill.htm).

Swiss Federal Office for Education and Science; Swiss Federal Council (1998): Botschaft über die Förderung von Bildung, Forschung und Technologie in den Jahren 2000-2003. [Message on the support of education, research and technology in the years 2000 to 2003]. (http://www.admin.ch/bbw/bbtd/bbtdindex.html).

Swiss Science Council (1997): Ziele für die schweizerische Forschungspolitik Periode 2000-2003. Vorschläge des SWR an den Bundesrat. [Goals for the Swiss research policy in the period 2000 to 2003. Proposals from the SSC to the Federal Council]. (http://www.admin.ch/bbw/bbtd/bbtdindex.html).

7.2 Statistical documents

British Office of Science and Technology (2000): SET statistics 2000. (http://www.dti.gov.uk/ost/setstats).

Centrum voor Wetenschaps- en Technologie-Studies (CWTS); Maastricht Economic Research Institute on Innovation and Technology (MERIT) (2001): Netherlands Observatory of Science and Technology (NOWT). Science and Technology Indicators 2000. (http://meritbbs.unimaas.nl/nowt).

European Commission (2001, latest issue): European Survey of the Information Society (ESIS) II WWW indicators. (http://www.eu-esis.org/esis2www/synthCEEC7.htm).

European Commission Research DG; Eurostat (2000): Towards a European Research Area. Science, Technology and Innovation. Key Figures 2000. Brussels. (http://europa.eu.int/comm/ research/pdf/keyfiguresihp.pdf).

Eurostat (2001): Statistics on Science and Technology in Europe. Data 1985-1999. Luxembourg.

Eurostat: Community Innovation Surveys in the New Cronos Database.

German Federal Statistical Office (2000): Wissenschafts-und Technologieindikatoren. Wiesbaden.

Institute for Scientific Information (ISI) (http://www.isinet.com/isi/products/citation/sci/index.html)

Italian National Statistical Institute (2000): La ricerca e sviluppo in Italia nel periodo 1998-2000. [Research and development in Italy in the period 1998-2000]. (www.istat.it/Anotizie/ Aaltrein/statinbrev/ricesvil/index.html).

Maastricht Economic Research Institute on Innovation and Technology (MERIT): Cooperative Agreements and Technology Indicators (CATI) database. (http://meritbbs.unimaas.nl).

Ministry of and Research Information Technology (2000): Facts and Figures on Danish Research. (http://www.fsk.dk/fsk/publ/2000/facts/factsandfigures.pdf)

National Science Board (2000): Science & engineering indicators 2000. Arlington, Va. (http://www.nsf.gov/sbe/srs/seind00/start.htm).

OECD (1994): The Measurement of Scientific and Technological Activities: Proposed Standard Practice for Surveys of Research and Experimental Development - Frascati Manual 1993. 5th Edition. Paris.

OECD (1997): Proposed Guidelines for Collecting and Interpreting Technological Innovation Data: Oslo Manual. 2nd Edition. Paris. (http://www.oecd.org/publications/e-book/9297031e.pdf)

OECD (1999): Science, Technology and Industry Scoreboard 1999. Benchmarking Knowledge-based Economies. Paris.

OECD (2000): Research and Development Expenditure in Industry 1977-98. Paris.

OECD (2000): Science, Technology and Industry Outlook 2000. Paris.

OECD (2000, latest issue): Analytical Business Enterprise Research and Development data-base (ANBERD). Paris.

OECD (2001, latest issue): Main Science and Technology Indicators database (MSTI). February 2000. Paris.

Portuguese Observatory for Science and Technology (1998): Survey on the National Resources in Science and Technology 1988-1997 (Higher Education; Government and Private Non-Profit). (www.oct.mct.pt/en/statistics/official/indicators/index.htm).

Statistics Finland (1999): On the Road to the Finnish Information Society. Helsinki 1999.

Statistics Finland (2001, latest update): Science, technology and information society. (http://www.stat.fi/tk/yr/tttiede_en.html).

Swiss Science Council (1998): Beitrag zu einem Indikatorensystem für die Wissenschafts- und Technologiepolitik. Fakten und Bewertungen für die Ziele der schweizerischen Forschungspolitik 2000 – 2003. [Contribution to an Indicator System for Science and Technology Policy. Facts and Benchmarks for the Goals of Swiss Research Policy 2000 – 2003]. (http://www.cest.ch/www-archive/1999/am/fub_99_2/synthese_anf_dt.pdf).

World Intellectual Property Organisation (WIPO) (2001, latest issue): WIPO Industrial Property Annual Statistics. (http://www.wipo.int/ipstats/en).

7.3 Existing indicators on the Internet for R&D

Internet-related ICT infrastructure

Name of indicator	**[1]: Core transmission capacity of NRNs**
Definition	Core or maximum transmission capacity within a national research network
Methods needed	Survey of providers of research networks
Notes	The indicator covers only the supply side and not the demand side.
Sources	http://www.terena.nl/compendium/corecap.html, NRN (see table on page 13)
Countries covered	EU, Switzerland, Central and Eastern European Countries
Time series available	–
Policy relevance	The *e*Europe initiative lists the upgrade of NRNs and the establishment of high-speed Internet access in universities as two of its actions.
Future value	The indicator will be important as long as data transfer uses stationary networks.
Links to other indicators	See also the indicators on core network size of NRNs [2], market penetration [3], the transmission capacity to Europe/US [4], the transmission capacity to large sites [5] and traffic per student [6]

Source: FHSO compilation.

Name of indicator	**[2]: Core network size of NRNs**
Definition	Core network size of national research networks: core transmission capacity multiplied by length of the network (in Mb/s x km)
Methods needed	Survey of providers of research networks
Notes	The core network size can be considered as a measure of the absolute network size. It was included in the TERENA questionnaire as an experiment, to see if it would produce useful results. The indicator covers only the supply side and not the demand side.
Sources	http://www.terena.nl/compendium/size.html
Countries covered	EU, Switzerland, Central and Eastern European Countries
Time series available	–
Policy relevance	The eEurope initiative lists the upgrade of NRNs and the establishment of high-speed Internet access in universities as two of its actions.
Future value	The indicator will be important as long as data transfer uses stationary networks.
Links to other indicators	See also the indicator on core transmission capacity [1]

Source: FHSO compilation.

Name of indicator	[3]: Market penetration of NRNs
Definition	Market penetration of national research networks.
Methods needed	Survey of providers of research networks
Notes	The indicator was not developed intentionally. It results from a misunderstanding on the part of some NRN administrators due to an ambiguously phrased survey question from Terena regarding the NRN client structure. It might be difficult to collect valid data, e.g. in respect to the percentage of the research institutes of private businesses which are connected.
Sources	http://www.terena.nl/compendium/instit.html
Countries covered	–
Time series available	–
Policy relevance	The eEurope initiative lists the upgrade of NRNs and the establishment of high-speed Internet access in universities as two of its actions. The present indicator returns the percentage of universities and other research organisations that are connected.
Future value	Depends on the future market structure for this kind of service.
Links to other indicators	See also the indicator on core transmission capacity [1].

Source: FHSO compilation.

Name of indicator	[4]: Transmission capacity to Europe/US
Definition	Transmission capacity to Europe/US
Methods needed	Survey of providers of research networks
Notes	The indicator covers only the supply side and not the demand side.
Sources	http://www.terena.nl/compendium/cap.html
Countries covered	EU, Switzerland, Central and Eastern European Countries
Time series available	–
Policy relevance	The eEurope initiative lists the establishment of high-speed Internet access in universities as one of its actions. The transmission capacity to Europe/US determines the speed of international data transfer.
Future value	The indicator will be important as long as data transfer uses stationary networks.
Links to other indicators	See also the indicators on the core transmission capacity of NRNs [1] and the transmission capacity to large sites [5]

Source: FHSO compilation.

Name of indicator	[5]: Transmission capacity to large sites
Definition	Transmission capacity to large sites
Methods needed	Survey of providers of research networks
Notes	The providers were asked to provide information on the typical transmission capacity to a large site. Some countries returned a range of transmission capacities which were used for calculating a mean value. The indicator covers only the supply side and not the demand side.
Sources	http://www.terena.nl/compendium/xscap.html
Countries covered	EU, Switzerland, Central and Eastern European Countries
Time series available	–
Policy relevance	The eEurope initiative lists the upgrade of NRNs as one of its actions. The transmission capacity to large sites monitors the connectivity of important research sites within a national research system.
Future value	The indicator will be important as long as data transfer uses stationary networks.
Links to other indicators	See also the indicators on the core transmission capacity of NRNs [1] and the transmission capacity to Europe/US [4]

Source: FHSO compilation.

Name of indicator	[6]: Traffic on NRNs per student
Definition	Data transmitted on the NRN per student and per year (in Gigabytes)
Methods needed	Survey of providers of research networks
Notes	The traffic per student can be considered as a measure of the relative network size. A country with a fewer number of inhabitants and a fewer number of students can be as well-served by a smaller-sized network as a large country with a larger network. The indicator covers especially the demand side and to a limited extent also the capacity.
Sources	http://www.terena.nl/compendium/trafstud.html
Countries covered	EU, Switzerland, Central and Eastern European Countries
Time series available	–
Policy relevance	The eEurope initiative lists the upgrade of NRNs and the establishment of high-speed Internet access in universities as two of its actions.
Future value	The indicator will be important as long as data transfer uses stationary networks.
Links to other indicators	See also the indicator on the core transmission capacity [1].

Source: FHSO compilation.

Name of indicator	[7]:Total budgets of NRNs
Definition	Total budget of national research networks in million €
Methods needed	Survey of providers of research networks
Notes	Budget sizes depend on the size of an NRN (e.g. in some countries NRNs also deliver services to primary and secondary education), the range of services it delivers and the amount of services that have been outsourced.
Sources	http://www.terena.nl/compendium/totbud.html
Countries covered	EU, Switzerland, Central and Eastern European Countries
Time series available	–
Policy relevance	The eEurope initiative lists the upgrade of NRNs and the establishment of high-speed Internet access in universities as two of its actions.
Future value	The value of the indicator depends on the problems mentioned above.
Links to other indicators	See also the indicator on total budgets per student/capita [8].

Source: FHSO compilation.

Name of indicator	[8]:Total budgets of NRNs per student
Definition	Total budget of national research networks per university student in €
Methods needed	Survey of providers of research networks
Notes	The indicator has also been calculated as a per capita rate. Budget size depends on the size of an NRN (e.g. in some countries NRNs also deliver services to primary and secondary education), the range of services it delivers and the amount of services that have been outsourced.
Sources	http://www.terena.nl/compendium/percap.html
Countries covered	EU, Switzerland, Central and Eastern European Countries
Time series available	–
Policy relevance	The eEurope initiative lists the upgrade of NRNs and the establishment of high-speed Internet access in universities as two of its actions.
Future value	The value of the indicator depends on the problems mentioned above.
Links to other indicators	See also the indicator on total budgets of NRNs [7].

Source: FHSO compilation.

Name of indicator	[9]: Staffing of NRNs
Definition	Total staff size in full-time equivalents
Methods needed	Survey of providers of research networks
Notes	The indicator can also be calculated for the technical staff only. Staff size depends on the size of an NRN, the range of services it delivers and the amount of services that have been outsourced.
Sources	http://www.terena.nl/compendium/staff.html
Countries covered	EU, Switzerland, Central and Eastern European Countries
Time series available	–
Policy relevance	The eEurope initiative lists the upgrade of NRNs and the establishment of high-speed Internet access in universities as two of its actions.
Future value	The value of the indicator depends on the problems mentioned above.
Links to other indicators	See also the indicator on total amount of effort available [10].

Source: FHSO compilation.

Name of indicator	[10]: Total amount of effort available to NRNs
Definition	Total amount of effort available to NRNs in person-months per year
Methods needed	Survey of providers of research networks
Notes	Because indicators on the number of staff available to an NRN are subject to some organisational differences between NRNs (see above indicator [9]) the calculation of the entire input of work effort might be a better indicator. However, it also depends on the size of an NRN.
Sources	http://www.terena.nl/compendium/staff.html
Countries covered	EU, Switzerland, Central and Eastern European Countries
Time series available	–
Policy relevance	The eEurope initiative lists the upgrade of NRNs and the establishment of high-speed Internet access in universities as two of its actions.
Future value	The value of the indicator depends on the problems mentioned above.
Links to other indicators	See also the indicator on staffing of NRNs [9].

Source: FHSO compilation.

Name of indicator	[11]: Internet penetration rate of high schools and universities (HEI)
Definition	Percentage of high schools and universities with Internet websites.
Methods needed	Analysis of websites
Notes	Among other national institutions (Primary and secondary schools, national ministries, regional and local authorities, hospitals & clinics, museums, libraries) the Internet presence of high schools and universities was evaluated. Comparisons are possible between the different institutions within a country and between the countries.
Sources	ESIS II (http://www.eu-esis.org/esis2www/synthCEEC7.htm)
Countries covered	Central and Eastern European Countries (CEEC)
Time series available	Not available.
Policy relevance	Potential indicator to assess the use of electronic networks which should be encouraged according to the new European Research Policy laid down in ERA (see chapter 3.2).
Future value	The indicator showed relatively high Internet penetration rates in HEI for some CEEC (100 % in Bosnia, Hungary, Lithuania, Slovakia). This leads to the conclusion that in most Western European countries the indicator might reach 100 % and therefore not show any variance.
Links to other indicators	–

Source: FHSO compilation.

Name of indicator	[12]: Titles in digital collections
Definition	The number of all titles, or separate types such as e-journals, e-books, other electronic resources, of a digital collection in relation to its target group
Methods needed	Survey of digital libraries
Notes	–
Sources	Shim, W. et al. (2000), op. cit.
Countries covered	–
Time series available	–
Policy relevance	The indicator monitors one element of the electronic resource base of scientific research, an important target of public R&D policy according to the OECD.
Future value	The indicator is straightforward and does not depend on very sophisticated data. Hence it is virtually the only one which can be computed using data which are regularly collected and monitored by digital collection providers.
Links to other indicators	–

Source: FHSO compilation.

Name of indicator	[13]: Usage of digital libraries
Definition	The numbers of sessions (logins) of searches and menu selections, of documents/items examined and of rejected sessions because of requests exceeding the simultaneous user limit are possible indicators of the usage of a digital library
Methods needed	Analysis of server log files or survey
Notes	The calculation of the data is based on an analysis of the log files of digital library servers. This analysis still requires answers to many open technical issues, a common set of definitions and standards and more experience regarding human behaviour when searching for information at digital library sources.
Sources	International Coalition of Library Consortia (1998), op. cit. – Brophy, P. et al. (2000) op. cit. – Shim, W. et al. (2000), op. cit.
Countries covered	–
Time series available	–
Policy relevance	Potential indicator to assess the use of electronic networks which should be encouraged according to the new European Research Policy laid down in ERA (see chapter 3.2).
Future value	The long-time value of server log-based indicators depends on the future development of protocol languages and content. A standardisation of protocol content would support the compilation of comparable indicators.
Links to other indicators	–

Source: FHSO compilation.

Name of indicator	[14]: Library staff for electronic services
Definition	Library staff involved in developing, managing and providing ELS and user training as a percentage of all staff
Methods needed	Survey of providers of electronic resources (e.g. digital libraries)
Notes	A problem might result from the different types of services provided by digital libraries: while some might restrict themselves to the pure provision of information objects, others also provide training to their users. A categorisation of the services and the assessment of staff data for the different categories might solve this problem and lead to comparable data.
Sources	Brophy, P. et al. (2000) op. cit. – Shim, W. et al. (2000), op. cit.
Countries covered	–
Time series available	–
Policy relevance	The indicator monitors one element of the electronic resource base of scientific research, an important target of public R&D policy according to the OECD.
Future value	–
Links to other indicators	–

Source: FHSO compilation.

Name of indicator	**[15]: Expenditure on electronic resources in relation to total expenditure**
Definition	Expenditure on electronic resources such as e-journal subscriptions, e-books, licenses for network-based data sets etc. in relation to total expenditure.
Methods needed	Survey of providers of electronic resources (e.g. digital libraries)
Notes	The indicator has been developed to assess the furnishing of electronic library services.
Sources	Brophy, P. et al. (2000) op. cit. – Shim, W. et al. (2000), op. cit.
Countries covered	–
Time series available	–
Policy relevance	The indicator monitors one element of the electronic resource base of scientific research, an important target of public R&D policy according to the OECD.
Future value	Changes of the pricing schemes for electronic media and financial restrictions on the buyers' side (sinking library budgets) might introduce biases in time series.
Links to other indicators	–

Source: FHSO compilation.

Name of indicator	**[16]: Perceived benefits of computer-mediated communication**
Definition	Indicator of the perceived benefits of different services of computer-mediated communication such as e-mail, mailing lists and tools for on-line meetings.
Methods needed	Survey of R&D personnel
Notes	The benefits might either be assessed very generally (large-small) or more specifically (better integration into information flows, higher productivity etc.).
Sources	Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997), op. cit., p. 512. - Lubanski, A.; Matthew, L. (1998), op. cit. - Walsh, J.P.; Maloney, N. G. (2001), op. cit. - Rojo, A. (1995), op. cit. - Finholt, T.A. et al. (1998), op. cit. - Mark, G; Grudin, J.; Poltrock, S. E. (1999), op. cit.
Countries covered	–
Time series available	–
Policy relevance	Potential indicator to assess the benefits of electronic networks.
Future value	The benefits of specific indicators depend again on the benefits of specific services themselves.
Links to other indicators	Frequency of computer-mediated communication [19]

Source: FHSO compilation.

Name of indicator	[17]: Communication activity within a mailing list
Definition	Number of messages within a mailing list during a certain period of time
Methods needed	Analysis of mailing list archives
Notes	Access to the archives of a mailing list is required to compute this indicator. In some cases information from the owner/manager of the list is necessary. To make the indicator relevant for country comparisons at least two pieces of information would be compulsory: A list of important and representative mailing lists in R&D, Information about the home countries of the senders of messages within these lists.
Sources	Boudourides, M.A. (1999), op. cit.
Countries covered	–
Time series available	–
Policy relevance	The indicator as it is calculated bears no direct relevance to R&D policy. It could constitute an indicator to assess the use of electronic networks (which should be encouraged according to the new European Research Policy laid down in ERA, see chapter 3.2), if the above mentioned additional information could be added.
Future value	As with the frequency of other forms of computer-mediated communication, the value of the indicator depends on the future value of the service itself: when mailing lists are no longer used as a communication channel, the indicator will be useless.
Links to other indicators	Frequency of computer-mediated communication [19]

Source: FHSO compilation.

Name of indicator	**[18]: Perceived benefits of the Internet**
Definition	Indicator on the perceived benefits of the Internet in general, or of specific sources (databases, digital libraries), services (ftp, gopher etc.) or for specific tasks (data collection, analysis, dissemination of results) in particular
Methods needed	Survey of R&D personnel
Notes	Various indicators of Internet service benefits have been constructed on the basis of survey questions.
Sources	Lubanski, A.; Matthew, L. (1998), op. cit. - Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997), op. cit.
Countries covered	–
Time series available	–
Policy relevance	Potential indicator to assess the benefits of electronic networks.
Future value	The value of specific indicators depends on the benefits of the specific services themselves; e.g. to calculate an indicator for the benefits of ftp-servers might be a waste of time if ftp-servers become unimportant for transmitting files.
Links to other indicators	Frequency of information retrieval via the Internet [21]

Source: FHSO compilation.

Name of indicator	**[19]: Frequency of computer-mediated communication**
Definition	Frequency of the usage of different services of computer-mediated communication such as e-mail, mailing lists and tools for on-line meetings during a specified period of time.
Methods needed	Survey of R&D personnel
Notes	The binary information 'usage: yes-no' leads to a reduction of information and might reduce the variance of the sample to zero, e.g. when everybody uses e-mail the frequency distribution will be '100 % yes', '0 % no'.
Sources	Day, J.; Bartle, C. (1998), op. cit.
Countries covered	–
Time series available	–
Policy relevance	Potential indicator to assess the use of electronic networks which should be encouraged according to the new European Research Policy laid down in ERA (see chapter 3.2).
Future value	The value of specific indicators depends on the value of the specific services themselves; e.g. calculating an indicator for the usage of on-line meeting tools might be a waste of time if these tools have only been used in a few selected cases so far.
Links to other indicators	Perceived benefits of computer-mediated communication [16], communication activity within a mailing list [17]

Source: FHSO compilation.

Integration of the new network technologies into research activities

Name of indicator	[20]: Web impact factor of a country
Definition	Web impact factor (Web IF) of a country: the number of web pages which link to a country divided by the number of pages found in its domain
Methods needed	Evaluation of websites (webometrics)
Notes	By differentiating between pages which link to other pages within a country and pages which link to them from abroad, self-link and external-link Web-IFs can be computed. A global Web-IF serves as a benchmark. Measurement problems might result depending on how representative a search engine is and the stability of its coverage.
Sources	Ingwersen (1998), op. cit.
Countries covered	Selected countries and domains
Time series available	Calculated for 1997, no time series available
Policy relevance	The indicator as it is calculated bears no direct relevance to R&D policy, though this relevance could be enhanced by calculating Web-IFs for national science systems. Then it would give clues as to the importance of the Internet in general (as a means of information generation and retrieval and social communication etc.) for national science systems.
Future value	The indicator will have significance as long as scientific content is displayed on the web and the principle of linking websites exists. Some problems in respect to time series construction might result from the inconsistency of search engines over longer time periods.
Links to other indicators	See also the web reference indicator [24]

Source: FHSO compilation.

Name of indicator	[21]: Frequency of information retrieval from Internet sources
Definition	Frequency of retrieving information from web pages in general or from specific sources (numerical databases, full-text databases) and by means of specific services (ftp, gopher etc.) during a specified period of time.
Methods needed	Survey of R&D personnel, analysis of client-side log files
Notes	Various indicators of Internet service usage have been constructed on the basis of survey questions. One analysis (Kaminer and Braunstein 1998) made use of both methods, survey and log file analysis, to measure the use of Internet applications.
Sources	Lubanski, A.; Matthew, L. (1998), op. cit. – Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997), op. cit. - Kaminer, N.; Braunstein, Y.M. (1998), op. cit. – Brophy, P. et al. (2000), op. cit. – Shim, W. et al. (2000), op. cit.
Countries covered	–
Time series available	–
Policy relevance	Potential indicator to assess the use of electronic networks which should be encouraged according to the new European Research Policy laid down in ERA (see chapter 3.2).
Future value	The value of specific indicators depends on the value of the specific services themselves; e.g. calculating an indicator for the usage of ftp-servers might be a waste of time if ftp-servers become unimportant for transmitting files.
Links to other indicators	See also the indicator of the perceived benefits of the Internet [18]

Source: FHSO compilation.

Name of indicator	[22]: Computer networks as information sources for innovation
Definition	Relevance of computer-based information networks as a source of information for innovation in relation to other sources
Methods needed	Survey of business enterprises
Notes	Four answering options were provided in CIS 2 to assess the relevance of different sources of information ('unimportant', 'slightly', 'moderately' or 'very important'). Computer-based information networks are one of the possible answers. It can be ranked in comparison to other sources such as sources within the enterprise or the enterprise group, competitors, clients or customers, universities or other higher education institutes, patent disclosures, fairs, exhibitions etc.
Sources	New Cronos Database, Eurostat
Countries covered	EU member states
Time series available	No, only aggregated data for the time period 1994-96.
Policy relevance	Potential indicator to assess the benefits of electronic networks.
Future value	Basically the indicator could be valuable in the future. But newer data will not be available as the question was omitted from the core questionnaire of the current Community Innovation Survey CIS 3.
Links to other indicators	–

Source: FHSO compilation.

Name of indicator	[23]: E-journals citation index
Definition	Citations of e-journals in relation to citations of non-electronic journals/all journals
Methods needed	Bibliometrics
Notes	This indicator has been computed to assess the impact of electronic journals on their fields of research.
Sources	Harter, S. P. (1996), op. cit.; data sources: Institute of Scientific Information (Science Citation Index, Social Science Citation Index, Arts and Humanities Citation Index)
Countries covered	–
Time series available	–
Policy relevance	The indicator as it has been calculated bears no direct relevance either to eEurope or to ERA. But this relevance could be included by aggregating e-journal citation indexes over the authors' research fields and home countries. Then it would indicate the extent to which e-journals as a specific form of electronic media are being used.
Future value	Due to identification problems (the diffusion of electronic versions of printed journals makes a distinction between 'on-line only' and 'off-line only' journals impossible) the future value of this indicator will probably be low.
Links to other indicators	–

Source: FHSO compilation.

Internet and R&D collaborations

Name of indicator	[24]: Web reference indicator
Definition	A web reference indicator in the broad sense measures the relation between different web pages by means of html-references (links) or regular textual references.
Methods needed	Evaluation of websites (webometrics)
Notes	Indicator construction depends on the available features of search engines. Due to its versatility the AltaVista search engine has been used frequently, although different studies have revealed problems concerning how representative the WWW is and its unstable coverage.
Sources	Leydesdorff, L.; Curran, M. (2000), op. cit.
Countries covered	–
Time series available	–
Policy relevance	Web reference indicators could be highly relevant for analysing the relations between different organisations and the density and spread of collaboration networks in R&D. Thus they could reflect the participation of individual organisations (countries) to collaborative research within the European Research Area and with external partners.
Future value	The indicator will have significance as long as scientific content is displayed on the web and web pages contain references. Some problems in respect to time series construction might result from the inconsistency of search engines and changing principles of reference construction over longer time periods.
Links to other indicators	See also the web impact factor [20]

Source: FHSO compilation.

PART B (D 2.2)

8 Overview over Part B

The present report follows on from the topic report, deliverable 2.1, which documented the scientific, political and statistical literature available on the Internet and related to research and development (R&D). This report discusses potential indicators which could serve to assess the utilisation and the effects of the Internet on R&D and lays down the basis of an indicator system suitable for carrying out a comparison and benchmarking exercise of European countries.

The indicator system covers constructs (conceptions of reality with varying degrees of abstraction and latency) in all three sub-topics of “the Internet for R&D”: “Internet-related ICT infrastructure for R&D activities”, “integration of the new network technologies into R&D activities” and “computer networks and R&D collaborations” (see table). The largest number of indicators has been developed for the sub-topic “Internet-related ICT infrastructure for research activities”. However, most indicators produce more than one item of information; e.g. an indicator on the effects of information retrieval from and via the Internet includes as response categories: time budgets, contacts, productivity and quality of work results.

Only a few separate indicators could be found for the importance and effects of computer networks regarding R&D collaborations. This may be due to general difficulties of operationalising and evaluating R&D collaborations. Nevertheless, the two indicators developed provide information on both the quantitative importance of new, network-based forms of R&D collaboration and their impact on the inputs and outputs of collaborative R&D.

Table: Indicators on the Internet for R&D

No.	Construct	Name of indicator	Sub-topic	Suggested method
(1)	Expenditure on ICT infrastructure	R&D expenditure on ICT	Infra-structure	Decision ma-ker survey
(2)	Expenditure on ICT infrastructure	R&D expenditure on different types of ICT	Infra-structure	Decision ma-ker survey
(3)	Electronic library services	Number of titles in digital collections	Infra-structure	Survey of di-gital collecti-on providers
(4)	Electronic library services	Staff providing electronic library services	Infra-structure	Survey of di-gital collecti-on providers
(5)	Researchers’ websites	Information displayed on a researcher’s web page(s)	Infra-structure	Survey of researchers
(6)	Researchers’ websites	Effects of researchers’ web page(s) (on time budget, communication, contacts and recognition)	Infra-structure	Survey of researchers
(7)	E-mail	E-mail communication for R&D purposes	Infra-structure	Survey of researchers
(8)	E-mail	Effects of e-mail use for R&D purposes (on information, contacts, collaborations, productivity, quality of work)	Infra-structure	Survey of researchers
(9)	Computer skills of R&D personnel	Computer skills of R&D personnel	Infra-structure	Survey of researchers
(10)	Specialised computer staff	Computer staff providing services to R&D	Infra-structure	Decision ma-ker survey

No.	Construct	Name of indicator	Sub-topic	Suggested method
(11)	Specialised computer staff	Unfilled vacancies in private businesses for computer staff providing services to R&D	Infra-structure	Decision ma-ker survey
(12)	Computer skills of R&D personnel	Effects of computer skills on R&D	Infra-structure	Survey of researchers
(13)	Digital library and peer site usage	Frequency of information retrieval from electronic sources	Research processes	Survey of researchers
(14)	Digital library and peer site usage	Documents/items from electronic sources	Research processes	Survey of researchers
(15)	Software usage	Frequency of software usage	Research processes	Survey of researchers
(16)	Information retrieval	Effects of information retrieval from and via the Internet (on time budgets, productivity, quality of work, contacts)	Research processes	Survey of researchers
(17)	E-publishing	Amount of work published in electronic media	Research processes	Survey of researchers
(18)	Quality control	Review activities for e-journals	Research processes	Survey of researchers
(19)	E-publishing	Impact of publications in electronic media (on size of readership, time to publication)	Research processes	Survey of researchers
(20)	R&D collaboration	Participation in long-distance R&D collaborations	R&D collaboration	Survey of de-cision makers/ researchers
(21)	R&D collaboration	Impact of computer networks on R&D collaborations (on communication, data & information transfer, project management and duration, quality of work, productivity)	R&D collaboration	Survey of de-cision makers/ researchers

Source: FHSO compilation

The report has ruled out some of the available methods of data collection for the purpose of this study. The level of methodological development of web content analysis and log file analysis has not advanced sufficiently to be able to use them effectively for answering the kind of macro-analytical questions we pose. We also dispensed with bibliometrics, a common method for assessing the output data of academics, as the empirical material available for bibliometrical analyses is not sufficiently detailed to account for Internet issues (e.g. whether the on-line or printed version of a journal was cited, whether and to what extent a research collaboration that produced a co-authored article used the Internet). The proposed indicators rely exclusively on surveys as the method of choice, with different categories of respondents as target groups.

A somewhat small number of indicators is appropriate for the planned SIBIS decision maker survey (see indicators (1), (2), (10), (11), (20), (21)), whereas the largest number of indicators is appropriate for a survey of researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants). For some questions it seems advisable to restrict the sample to academic researchers, as their information disclosing behaviour should be more autonomous and less influenced by information policies of a higher hierarchical level. This applies for example to the content of a researcher's web presentation. A pilot survey of researchers might take this into account and limit the sample to researchers from the academic sector. The indicators suitable for this survey do not prejudice the use of a traditional survey form (personal, phone, written) over a novel, Internet-based form (e-mailed or on-line questionnaire). However, as one of the targets of this survey is to assess the utilisation of various Internet tools, an on-line only survey might suffer from a sample selection bias. Hence, a combination of on-line and off-line methods seems most appropriate.

A third group of indicators can only be assessed if ancillary service organisations are willing to contribute information about the (computer-based) services they provide to R&D (see indicators (3) and (4) and section 11.1.2 on research networks). Important organisations in this field are the national research networks and the providers of electronic library services. However, instead of carrying out our own pilot surveys, it is a lot more effective to join forces with expert organisations in these fields which carry out indicator development and data collection as part of their regular tasks.

The next phase of research on the topic Internet for R&D will be quality checks with external experts to ensure that the developed indicators are both meaningful and feasible for an empirical investigation. These quality checks will be carried out in parallel with the further transformation of the indicators into survey questions, which has been performed in this report in a rapid, prototyping manner. Furthermore, preparatory work must be performed and a choice made as to how the majority of the indicators can be tested, as this cannot be integrated into the planned SIBIS surveys. The work on the topic “the Internet for R&D” will continue with these tasks and will document and present the results within the following SIBIS deliverables.

9 Introduction

9.1 The need for an indicator system "the Internet for R&D"

Buzzwords such as "Information Society", "Digital Economy" or "Cyberspace" indicate the trend of increasing integration of information and communication technologies into the lives of people all around the world. Though this trend has been described on probably thousands of book pages, we still lack detailed knowledge regarding its features and effects in many areas. This explains why MIT researchers recently suggested a research program that should deal with the role of organisations in the development of the digital economy, with the social transformations within and across internetworking organisations and with the use and effects of Internet technologies.[223] We are not yet able to fully understand or even see the consequences of the information revolution as too many areas are affected by the new technologies, many technological and social developments are too new and the pace of change is too rapid. The fundamental problem is that we often lack valid concepts which tell us where we have to look and what we have to measure in order to assess the size and effects of ICT developments, not to mention the empirical data itself. This applies to all of the different subsystems of society and economy and hence, with regard to the SIBIS project, analysing nine of these subsystems can only be a modest contribution to what has to become a broader effort in various fields of science.

In the topic report on "the Internet for R&D", one of the nine topics of SIBIS, we identified and evaluated the relevant available literature from three main sources:

- Scientific literature published in journals and books, published and unpublished conference papers and Internet documents,
- Policy documents from multinational governmental institutions: primarily the European Commission, international organisations such as research associations and national institutions (governments, associations in the field of R&D),
- Documents from statistical offices, other organisations generating data such as the OECD and private enterprises.

On the basis of this literature we drew three conclusions:

1) The Internet does indeed affect R&D in many ways.
2) There is hardly any data available for measuring these effects.
3) The European Commission and national governments have exhibited political commitment to promote the usage of Internet technologies in science.

Ad 1) In deliverable 2.1 we described the Internet in relation to R&D from different perspectives. Firstly we studied the new Internet-related infrastructure for science which has been developed and put into operation or is still in the process of being developed. For some tools we even found studies which undertook preliminary evaluations of their effects on R&D. The second approach assessed the technology from a process-oriented point of view and reviewed the known effects of ICT on the gathering of scientific data and information, data analysis and technology development and the publication and dissemination of research results. The third perspective acknowledged that the new ICT have not only created new inputs for R&D and provided new methods of performing existing tasks, but they have also created the possibility of giving R&D a new organisational setting. The most important feature of this new setting is that it is independent of location and therefore enables specialised scientists from all over the world to join and work on a common research project and thus create a new, virtual research organisation.

[223] Orlikowski, W. J.; Iacono, C. S. (2000): The truth is not out there: an enacted view of the "Digital Economy", in: Brynjolfsson, E. and Kahin, B. (eds): Understanding the digital economy. Data, tools, and research. Cambridge and London, pp. 352-380.

This has laid out the structure for the further work on this SIBIS topic. It would certainly do no justice to the scope and complexity of the topic and not produce any reliable results if we were to reduce the analysis to a general assessment along the lines of "How did the Internet affect your research?", or if we only selected a few sub-topics which are currently hotly debated. The Internet technologies are too new and current developments are too rapid to irrefutably stipulate which tools will establish themselves on a general level, which activities and steps within research work will benefit most, or which new organisations will turn out to be stable over a certain time frame and therefore appropriate for performing scientific work. Though we might not be able to develop indicators for each sub-topic with the same comprehensiveness, due to our constraints of time and limited expertise, we will at least try to indicate the directions in which research could proceed.

Ad 2) The literature review of deliverable 2.1 found very few indicators or data appropriate for monitoring the new technological developments and their effects on science. Different conferences, workshops and studies on the topic have also repeatedly highlighted this. Statements such as the following are frequently found in the literature:

> *"This is perhaps the most important recommendation stemming from this workshop: there is a pressing need to increase efforts and resources to undertake in-depth empirical studies on the innovative uses of Internet in science and to carry out European-wide surveys on this issue. Such studies are the only way to generate a sufficient amount of data and information necessary to evaluate the impact of new, high capacity electronic communication facilities upon the organization, distribution and conduct of collaboration on fundamental research problems."*[224]

> *"There was general agreement at the Conference on the broad tasks that governments will have to fulfil in order to maximise the benefits of ICT for science: …*
> *Develop internationally comparable indicators for a quantitative assessment of ongoing developments and performance, including return on investments in this rapidly evolving area. Diversity in the way each discipline, country and region views and is able to organise infrastructure, content and access seems to be the dominant characteristic in the evolution of ICT use in science, and needs to be measured."*[225]

> *"At the moment, we don't even have much data about usage patterns online. This is especially regrettable since these patterns appear to be in the midst of substantial changes."*[226]

This shortage of concepts, indicators and data means that in the topic area "the Internet for R&D" SIBIS must develop its own concepts and indicators which are valid and reliable and must collect the necessary data wherever possible. The project largely performs pilot work that must be tested, broadened in its scope and investigated in more depth within future projects and studies. Hence, the development of indicators and especially the multitude of proposals and ideas in the present report should be considered as a first contribution to the process of advancing R&D statistics that better take into account information and communication technologies in general and the Internet in particular.

---

[224] From the summary of an European Science Foundation conference, Foray, D. (1999): Building the Virtual 'House of Salomon': Digital collaboration technologies, the organisation of scientific work and the economics of knowledge access. Report of the ESF-IIASA-NSF Workshop - 3 to 5 December 1999 - at the International Institute for Applied System Analysis, Laxenburg, Austria, p. 9. (http://www.esf.org/policy/pdf/iiasa.pdf)

[225] From the summary of a OECD conference, Aubert, J.E.; Bayar, V. (1999): Maximising the Benefits of Information Technology for Science: Overview and Major Issues, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, p. 26.

[226] The renowned mathematician and ICT researcher Andrew Odlyzko, University of Minnesota, on the usage of electronic publications in academia, see Odlyzko, A. (2001): The rapid evolution of scholarly communication. (http://www.dtc.umn.edu/~odlyzko/doc/rapid.evolution.pdf).

Ad 3) The European Commission has promoted a new research policy in its publications on the European Research Area (ERA). The common thread of many goals and measures described in these documents is the creation of multi-layered networks within the R&D system as well as across its boundaries, including other socio-economic areas and political institutions. These networks among researchers, research institutes, protagonists of specific research topics, stakeholders of R&D in different organisations etc. have to be paralleled by modern and high-capacity communication networks. The European Commission acknowledges this fact and consequently promotes within its *e*Europe initiative the enhancement of Research and Education Networks for data transmission and the development of novel collaboration-oriented computer systems (“Grids”). The ERA communications cover the usage side of ICT and encourage the development and implementation of further computer-based tools for science, the training of researchers on the possibilities of ICT and the use of computer networks to connect the best researchers in Europe to form "virtual centres of excellence". Many European countries have developed new concepts of research policy over the last three years and some, primarily Finland and the U.K., expect ICT to contribute to the development of science.

The consequence of this orientation of research policy towards ICT is that it becomes even more necessary to develop indicators and collect data that are capable of measuring the broad spectrum of effects that the ICT have on R&D. Only then will it be possible to identify gaps (e.g. in the infrastructure, the skills) and work out policy measures to close them, substantiate priorities and develop an overall research policy that optimally reaps the benefits of computer networks.

The present report continues the development of indicators to measure Internet usage in R&D as well as the effects of this. It synthesises the findings from Part A and develops proposals for further indicators covering the sub-topics identified in that deliverable.

9.2 Structure of Part B

The structure of the subsequent sections of Part B is as follows: in section 10 we will define and explain important terms and concepts which are related to the construction of an indicator system in general, and the construction of the indicator system “the Internet for R&D” in particular. Of course, we rely to a large extent on the logic that was elaborated in Part A (deliverable 2.1), and employ a matrix structure composed of topics/sub-topics and dimensions.

In section 11 we elaborate on this indicator system. Each sub-section of section 11 contains a brief overview that serves to repeat important definitions, terms and results of the literature review of Part A. Then a description of possible indicators is given, firstly within the dimension of quantity and secondly within the dimension of impact (see p. 100 on the terms). *It must be emphasised, however, that this deliverable 2.2 is an interim report and that the indicator system should not be considered exhaustive. On the contrary, if new possibilities of measuring important developments become apparent, these will be integrated as far as possible.* The indicator descriptions also contain discussions as to whether the indicators meet the necessary criteria, which are outlined in section 10, for inclusion within the SIBIS indicators. For these SIBIS indicators we provide details and possibilities for empirical assessment in tables which are reproduced in the annex.

Section 12 provides a summary of the results of Part B and points the way forward to the next workpackages, especially the future empirical work on the Internet for R&D.

10 The structure of the indicator system

10.1 Constructs and indicators

The indicators are central to our indicator system. Their function is to transfer latent and not-observable ideas, hypotheses and conceptions of reality into quantifiable units, in other words: it operationalises them. When the conceptions of reality are abstract and latent rather than concrete and observable they are called "constructs". As Nunnally and Bernstein put it: “Such a variable is literally something that scientists “construct” (put together from their own imaginations) and which does not exist as an observable dimension of behavior.”[227] Indicators thus can be considered as measures of constructs. Indicators and constructs can be linked in three different ways:[228]

1) *Effect indicators* are observable results (effects) of the constructs. The construct is broader in meaning than the individual indicators, which only reflect parts of it. Examples of indicators and constructs are "co-authored publications" or "co-invented patents" as indicators for "R&D collaboration".
2) *Components* are indicators which are transformed by simple linear combinations. They do not contain a link to constructs and are largely transformed for convenience, e.g. to reduce the amount of data presented or to facilitate comparison. Such transformations could be, for example, calculating the fraction of "co-authored publications" per scientist or adding "co-authored publications" and "co-inventions" to "total results of co-operative work".
3) *Causal indicators* imply a causal effect of the indicator on the construct (that is in the other direction than in bullet 1). Changes in the observed indicators may consequently lead to changes in the constructs. Examples are the numbers of "co-authored publications", "co-invented patents" and other achievements (causal indicators) which may be used to measure the "recognition" or "merit" of a scientist. The constructs "merit" and "recognition" do not exist independently of their causes.

Evaluation research has developed a list of the desirable features of indicators which should render them useful for the measurement of constructs:[229]

- *Validity*: The indicator should measure what it is intended to measure. Validity is based on theoretical reasoning, that is on the arguments substantiating why hidden facts come to light in an indicator and explaining its suitability for measurement.[230] But validity must be tested and proved empirically, e.g. by correlating different indicators which aim to measure the same construct or by analysing an indicator’s predictive value (comparing prediction and actual development).
- *Reliability* refers to the necessity that an indicator produces the same results whenever it is implemented to measure the same construct. Reliability is not inherent to the indicator, it also depends on the context and diligence of data collection. Reliability requires, for example, that the method of data collection should not bias the results (e.g. produce answers which the interviewer obviously wants to hear).
- An indicator should also have *direction*: it should be unambiguous what is beneficial and what is detrimental. The direction is based on normative decisions which are also part of the theoretical foundation of an indicator.

---

227 Nunnally, J. S.; Bernstein, I. H. (1994): Psychometric Theory. 3rd edition, New York et al., p. 85.

228 See Nunnally, J. S.; Bernstein, I. H. (1994), op. cit., p. 449.

229 See Weiss, C. H. (1998): Evaluation. 2nd edition, Upper Saddle River, pp. 144-150.

230 Orlikowski and Iacono give a good example of what happens when an indicator is not valid: an evaluation study of Internet usage at home detected the disturbing finding that the social and psychological well-being of the Internet users was reduced. The measure for this was “Internet use”, i.e. the number of hours the subjects were connected to the Net. But, according to Orlikowski and Iacono, there was no account of what the people were actually doing on the Internet, and the reduction in well-being may well be attributable to specific uses and not the Internet in general, see Orlikowski, W. J.; Iacono, C. S. (2000): op. cit., p. 362.

- *Sensitivity to differences*: if all answers to a question are agglomerated in one of the different answering options, these are obviously not suited to assess the differences of the construct over the sample of measurement objects.
- An indicator should also be *accessible*: the data required to produce the indicator should be available and it should be possible to check its validity in order to assure its value for measurement.

We use these characteristics within this report to decide whether an indicator is suitable for inclusion in the SIBIS indicators.

Another consideration which is specific to this topic will also help us to select the indicators: the Internet for R&D includes, as we have shown extensively in deliverable 2.1, many different disciplines which perform research activities in universities, public research institutes or private firms. But the aim of the SIBIS project is to produce indicators which are suitable for reflecting the situation in the entire national research systems across the European Union and Switzerland. Therefore we have to abstract as much as possible from the specifics of an individual discipline. Mono-disciplinary (sub-)indicators which only contain information on one or a few research disciplines are not particularly helpful. For example most of the Internet-based data collection tools are rather mono-disciplinary, as different research disciplines make use of different data: social sciences mainly use data on human activities, natural sciences on stocks and flows in nature and engineering on technical processes. On the other hand, e-mail and other forms of computer-mediated communication are multi-disciplinary. They are used at least to some extent in all research disciplines as the communicative value of this communication system is not particularly dependent on the content that is being communicated. Of course, this does not mean that each discipline uses e-mail to the same extent. We will concentrate on multi-disciplinary indicators and exclude mono-disciplinary ones, even if this leads to gaps in the coverage of the indicator system. But it seems preferable to have some gaps rather than to construct an indicator system that is not really representative of the national R&D systems.

10.2 Achieving comparability and reducing complexity

While the above-mentioned indicator characteristics help us to select the valuable ones from less valuable ones, and to exclude those that would eventually lead to erroneous results and conclusions, they do not help us to condense and sort the large volume of data which an empirical implementation of the indicators produces. For some purposes it might not be necessary, or may even be misleading, to reduce the data. For example, if the goal is to design policy measures that promote computer-mediated communication among researchers, all available data on different ways of communicating via a computer network, the barriers against using it or the benefits and costs that accrue will prove useful. But if the intention is to compare which national research system has the best resources for computer-mediated communication, harnesses these resources better than other research systems or attributes more importance to communication than to information retrieval via the computer, a large amount of data may obscure the underlying picture. In this case some sort of simple information reduction or the establishment of the link between indicators and constructs may be helpful.

The theoretically possible links between indicators and constructs listed above (effect indicators, components, causal indicators) also have consequences for the methods that should be used. In the case of effect indicators, the most common method is *factor analysis* which may be both exploratory or confirmatory, depending on the previous knowledge of the constructs and their effects. Given the constraints of time and space we cannot describe this method in detail now,[231] but we will document our approach at later stages of the project in the event that we use factor analysis.

[231] See on factor analysis e.g. Nunnally, J. S.; Bernstein, I. H. (1994), op. cit. – Harman, H.H. (1976): Modern factor analysis, 3rd edition, Chicago.

For components, factor analysis is also applicable, but it may be sufficient to simply carry out algebraic transformations. Usually we have to do this to achieve comparability. For example we could count the number of e-mails sent by two samples of researchers from Switzerland and Germany and come up with two numbers. But are they really significant? The number of e-mails sent depends certainly on the sample size, and the communication cultures might play a role: in a smaller country researchers might rely to a larger extent on personal communication as travel distances are shorter, whereas in a larger country the use of telecommunications may be more intensive. Of course, we have not yet tested these assumptions, and this example only serves to illustrate that we need to develop comparable indicators. A standard procedure for making indicators comparable is to calculate fractions using a denominator that controls for the important factor of data variation. This procedure is also called "normalisation". Tenopir and King (2000) propose five different types of comparable indicators:[232]

- performance measures, i.e. quantitative input in relation to a quantitative output indicator
- effectiveness measures, i.e. output characteristics (quantity and quality) in relation to demand
- cost-effectiveness measures which reflect the monetary input necessary to provide a service in relation to the demand for this service
- impact measures relating usage/demand and effects, or usage/demand and "environmental characteristics" such as the size of target group
- efficiency or cost-benefit relationships

Vedung uses similar evaluation criteria for the public sector and lists various possibilities to achieve some level of comparison: the past performance, intra- and/or international comparisons, benchmarks, goals-, expectations- or interest-related comparisons, definition of least, average or optimum levels achievable.[233]

It is not possible to define a general rule for ensuring comparability for all our different indicators and sub-indicators. In accordance with the purpose of the entire project, we will look for international comparability. This makes it essential to check for identical definitions in the different countries in order to avoid any systematic measurement error. If different concepts and definitions are employed, as is often the case for services organised and provided at the national level, we either have to consider modifications or even re-assess the data, based on identical definitions.

Algebraic transformations of indicators not only serve to establish comparability, they are also sometimes useful for reducing the amount of information. This is commonly known as the *computation of indexes*. Precisely because it is an easy task to calculate an index from a variety of indicators, this should be carried out only with great care. Usually, neither the choice of the method of combination (addition, subtraction, multiplication, division or non-linear combinations) nor the choice of the weighting is trivial. Both choices must be based on a clear and theory-based understanding, e.g. if and to what extent the individual indicators are complements or substitutes.

Causal relationships between indicators or between indicators and constructs have to be ascertained by appropriate methods such as *regression analysis*. A problem may emerge when the proper operationalisation of a construct is difficult. However, progress in the field of regression analysis within the last decades has reduced this problem somewhat, e.g. the

[232] See Tenopir, C.; King, D.W. (2000): Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers. Washington, D.C.: Special Libraries Association, pp. 106-111.
Other proposals for indicator systems seem somewhat less structured. McClure and Lopata (1996) mix primary (e.g. the amount or extent to which a service is provided, the benefit or result of a service or activity) and derived indicators (e.g. the service output in relation to costs), see McClure, C. R.; Lopata, C. (1996): Assessing the Academic Networked Environment: Strategies and Options. (http://istweb.syr.edu/~mcclure/network/toc.html).

[233] He cites Dror (1968) as the original source of this list, see Vedung, E. (1999): Evaluation im öffentlichen Sektor. Wien, Köln and Graz, p. 224.

development of models with binary dependent variables such as logit and probit models have reduced the pressure to find dependent variables with a normal distribution.

We will not yet make any proposals for data reduction within this report. As we have hopefully made clear in the previous section, the use of an indicator as an effect, component or causal indicator is dependent on the desired results. Each may be possible with the very same indicator and data available. Further processing of the data should be carried out if the "hidden" characteristics of constructs can be revealed, or if it is necessary to reduce the amount of data. We have to establish the indicators first (with the constructs in mind, of course) which will be the objective of the subsequent section 3.

In order to keep track of the variety of indicators that we will present in section 3 we have to employ some structure or rules of grouping and categorising the indicators. We will use two structures which are described in the following two sections 2.3 and 2.4.

10.3 Topics and sub-topics

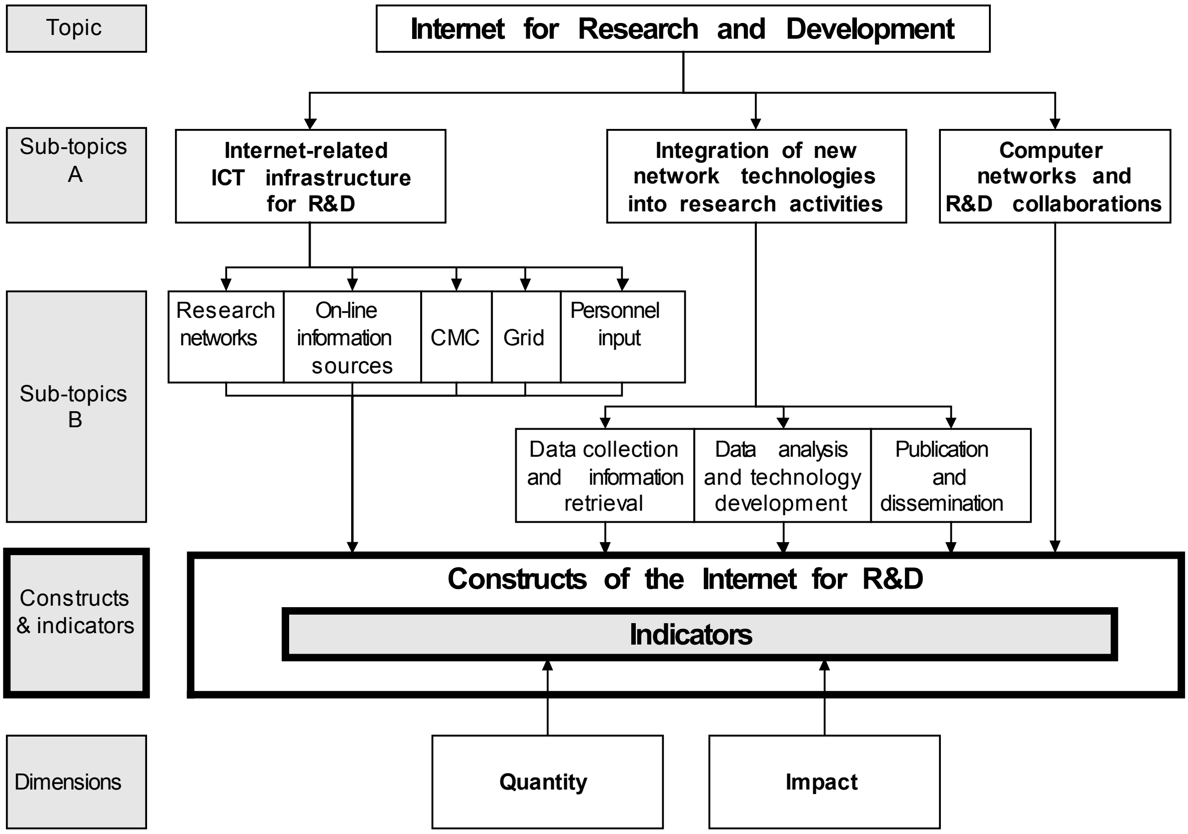
The present report develops a system of indicators which can serve to evaluate Internet usage for R&D and the effects of this. For this purpose we will keep the structure that was elaborated in the literature review of the topic report (deliverable 2.1) on the Internet for R&D. This structure is basically hierarchical, i.e. each level is more narrow and precise – and therefore more appropriate for empirical measurement – than the previous level. Consequently, the highest level "**sub-topics A**" contains (see figure 1):

- “Internet-related ICT infrastructure for R&D activities”
- “Integration of the new network technologies into R&D activities”
- “Computer networks and R&D collaborations”

These can be differentiated further into the next level "**sub-topics B**" which still cover rather broad issues. They have some things in common, e.g. each one of the computer programmes, methods or digital objects discussed under the heading of "Internet-related ICT infrastructure for R&D" can be considered as an element of infrastructure (employing an admittedly broad concept of infrastructure). But most of the elements serve different purposes which might be to transmit data, to store information or to enable communication. Therefore they are grouped into different sub-topics B. While in this example we used the main function of each infrastructure element to group it, for other sub-topics A we had to use different schemes (i.e. steps of a research process). These sub-topics B, though they are more precise than sub-topics A, are still very general and in some instances it has been necessary to introduce a further level "**sub-topics C"** (not included in figure 1). Again, an adequate way to group the indicators had to be found each time depending on the specific characteristics of the sub-topic.

We omit from this report a detailed discussion of two sub-topics: grid technologies from the infrastructure part and data analysis and technology development from the process-oriented perspective. Both are highly dynamic and to some extent interrelated: grid technologies might introduce major changes to scientific data analysis and create the possibility of not only transferring data but also modifying and analysing them from different locations at the same time. This highly dynamic nature still extends to the basic constructs (e.g. what are grid services and applications?) and therefore an operationalisation would be very difficult and data collection would only produce meaningful data for a very short period of time. But, as we already stated above, though it seems advisable to exclude grids and data analysis issues for the time being, if realities change within the project duration, we are willing to reconsider this decision.

Figure 1: The structure of the indicator system "the Internet for R&D"



Source: FHSO compilation.

10.4 Dimensions

In addition to the "top down" approach of pooling the indicators for different sub-topics, we also employed another approach to structure the presentation of indicators. It is presented as "bottom up" in figure 1, *though it is not really "bottom up" as is commonly understood but is rather of a matrix type*. This approach serves to categorise the sub-indicators differently, according to an indicator-related logic and not a topic-related logic as the sub-topic levels. Usually indicators can be found which reflect different **dimensions** of a construct. We will differentiate two dimensions within this report:

1) Issues related to the supply and usage of the constructs are taken into account by means of indicators in the dimension "quantity". This dimension encompasses indicators that measure for example how many "titles" are provided in digital collections, how many computer personnel are employed, or how many e-mails are sent or received during a certain period of time. The units of measurement are usually physical units and seldom monetary units such as expenditure on Internet-related infrastructure.

2) The "impact" dimension takes into account the fact that the effects of producing and using Internet services on R&D can be measured and evaluated directly. Impact can be multifaceted depending on the spectator's view: for example, a research manager may be interested in the budgetary effects of the purchase and use of a specific computer programme, while a researcher is more interested in its effects on time saving, output quality and other needs, and the computer staff might care particularly about its installation and maintenance/update features. As a principle we will take the researcher's view. The

distinction between physical and monetary units of measurement is more relevant in the impact dimension than in the quantity dimension:

- Physical effects can be either additional inputs into R&D activities (e.g. additional data and information due to improved resource access via the Internet) or productivity increases and a qualitative improvement of the output (e.g. more research articles per time period, higher quality of the articles).
- Monetary values can sometimes be used instead of physical units, e.g. the effects of easier long-distance data transmission on research budgets or the additional revenues generated through higher quality research papers and patents. However, a monetarisation of the physical effects is often impossible or only possible by making many assumptions and accepting large errors.[234] Hence, we will have to concentrate on the concept of physical units.

Of course, the distinction between the quantity and impact dimensions is purely analytical. There is a relationship between both dimensions, as quantitative issues reveal a lot about cost-benefit ratios: usage levels would be low, if little benefits were accompanied by large costs. Therefore at least in some cases the utilisation also provides useful information on the value of a resource or technique.

We now advance our indicator system "the Internet for R&D" beginning with the Internet-related infrastructure and specifically R&D expenditure on ICT.

[234] This problem is frequently encountered regarding the effects of ICT. For example the effects of ICT on the quality of consumable outputs have been found difficult to measure, see Brynjolfsson, E.; Hitt, L. M. (2000): Beyond Computation: Information technology, organizational transformation and business performance, in: Journal of Economic Perspectives, vol. 14, no. 4, p. 41.

11 The system of indicators for "the Internet for R&D"

11.1 Internet-related ICT infrastructure for research activities

11.1.1 General assessment of ICT infrastructure for research activities using expenditure figures

Overview of indicators on ICT expenditure for research activities

Various programs and initiatives at international and national levels have the objective of supporting and promoting the use of ICT for research activities. The motivation for these efforts were the expected high returns regarding R&D productivity, the quality of R&D results and the capacities of research to tackle computation-intensive research questions. A general assessment of the ICT infrastructure for research activities is only possible by means of a general unit of measurement which converts the various physical units (data transmission capacities in Megabit per second, "volumes" in a digital library, patent files in a networked patent data base, computer staff assisting in research work etc.) into a single unit. Monetary data are predestined to fulfil this purpose and hence the ICT expenditure for R&D can serve as a general indicator for assessing the "ICT-intensity" of an organisation's research activities (firm, university, government agency etc.).

As we are not dealing with the impact of ICT in general on R&D, but specifically with the Internet, we could limit the analysis to the latter. However, it may be difficult to differentiate exactly between Internet-related and non Internet-related expenditure. Hence, it seems advisable to assess both and investigate their relationship. Another more serious drawback for a statistical assessment is the rapid growth in capacity and price reductions of ICT equipment which makes it very difficult to calculate meaningful time series and capital stock figures from past investments. Notwithstanding these drawbacks, we will include two indicators on ICT expenditure for R&D and consider them as supplementary to the other, more detailed indicators on ICT and the Internet.

Table 1: Indicators on expenditure on Internet-related ICT infrastructure for research activities

No.	Construct	Name of indicator	Dimension	Suggested method
(1)	Expenditure on ICT infrastructure	R&D expenditure on ICT	Quantity	DMS
(2)	Expenditure on ICT infrastructure	R&D expenditure on different types of ICT	Quantity	DMS

Source: FHSO compilation

The quantity dimension "ICT expenditure for research activities"

The OECD has elaborated a proposal for the standardised measurement of R&D expenditure in the Frascati Manual.[235] The manual differentiates between intramural and extramural, and current and capital expenditure. It does not provide any further distinction and does not deal explicitly with R&D expenditure on ICT. Hence we must develop our own understanding of the term. Among the R&D expenditure for ICT we include expenditure on:

(a) Equipment: hardware and software licences

[235] See OECD (1994): The Measurement of Scientific and Technological Activities: Proposed Standard Practice for Surveys of Research and Experimental Development - Frascati Manual 1993. 5th Edition. Paris, pp. 91-104.

(b) Labour: internal support staff

(c) Services: external computer services (including hard- and software-related support)

In principle, expenditure on ICT equipment, labour and services can be assessed in private businesses, in universities and in government funded R&D agencies. Whether a R&D unit is private or public does not preclude the assessment of expenditure data. However, for the private business sector the expenditure could be normalised as a fraction of total turnover, in analogy to the "R&D intensity" which depicts the total intramural R&D expenditure of a company in relation to its turnover. This is not possible for the public R&D sector because it may not be able to come up with turnover data. Furthermore, this kind of normalisation may produce an upward bias towards entities whose sole purpose is R&D. Hence, we may obtain better comparability if we calculate a fraction of R&D expenditure on ICT and total "intramural R&D expenditure" – in OECD terms – or total budgets of universities and public R&D agencies (see indicator (1)). Besides the aggregated figure we may also obtain differentiated data for each of the following: ICT equipment, labour and services (see indicator (2)).

11.1.2 Research networks

Research networks (RNs) are physical networks which permit the transmission of data between research institutions. They are the basic prerequisite for performing research that involves the transmission of large volumes of data. Usually RNs are organised on the local (campus), national and supra-national level with ports to the next level/other networks. In principle we could aim to assess their services, usage and effects at each spatial level. This would require a broad collection of data, e.g. from all the different local/campus research networks, the development of appropriate methods to compute the data for the national level, as this is our spatial level of research, and the development of methods to integrate the data from the different spatial levels. This is outside of the scope of SIBIS which must deal with a broad set of different concepts, whilst at the same time focusing on one – the national – spatial level.[236] Therefore we will concentrate on national research networks (NRNs).

Indicators on NRNs have been developed by the Trans-European Research and Education Networking Association (TERENA) in collaboration with some of its members. In a pilot survey among the NRNs, these were tested and the results made available to the public.[237] In deliverable 2.1 we provided a detailed discussion and critique of these indicators (see section 2.1.1 of deliverable 2.1 on the Internet for R&D). The TERENA-led group is currently undertaking a revision of the indicators and the next survey is planned for February 2002. We do not see great promise in the independent and competitive development and assessment of indicators as TERENA and the NRNs both have a wealth of expertise and the proprietary data which are necessary to carry out this undertaking. Instead, we made our comments on the indicators and ideas for potential improvements available to TERENA and SWITCH, the Swiss national research network that also participates in the mentioned group. Additionally, we agreed with TERENA to provide comments on the new questionnaire. The latter is currently (December 2001 - January 2002) being discussed and assembled, and hence could not be included in this report. Consequently, this report and the empirical work of SIBIS in workpackage 3 will not contain any indicators on RNs. However, in workpackage 5 we will include the new data from the next phase of the TERENA survey in our benchmarking exercise and provide a thorough discussion of its results.

[236] Especially the evaluation of service levels and impact at the local level might be a worthwhile undertaking from a national point of view. It can serve to identify breaches and bottlenecks which could be targeted by national investment programmes for RN.

[237] See http://www.terena.nl/compendium/ToC.html.

11.1.3 On-line information sources

Overview of indicators on information sources on-line

On-line information sources were discussed in deliverable 2.1 as yet another element of the Internet infrastructure which is useful for R&D. We differentiated between three on-line information sources: textual databases, numerical databases and the WWW in general.

- The concepts of measuring *textual databases* are well advanced. Different indicators have been proposed in the scientific literature (see deliverable 2.1) but mainly from a micro-level point of view and with the objective of assessing the performance of electronic library services (ELS). Subsequently we will only consider those indicators that are in principal suited to be aggregated to the macro-level in order to compare the quality of ELS provision in different countries. As potential data sources, server log files and questionnaire-based surveys have been singled out. Server log files do not constitute a valid, reliable and easily accessible source of information to date and therefore we will only include survey-based indicators on textual databases among the SIBIS indicators. However, if the quality of log files and log file analysis tools improve in the future, it should be discussed whether information from log files can be collected and disclosed for scientific purposes.
- *Numerical databases* were among the first information sources stored on electronic storage devices and could therefore be exploited through access via computer networks. Though the scientific literature states many examples and provides some case studies on important and expanding databases, to our knowledge no reports have been assembled so far which provide indicators and comparisons of database production and use on a country level.

Table 2: Indicators on on-line information sources

No.	Construct	Name of indicator	Dimension	Suggested method
(3)	Electronic library services	Number of titles in digital collections	Quantity	Survey of di-gital collecti-on providers
(4)	Electronic library services	Staff providing electronic library services	Quantity	Survey of di-gital collecti-on providers
(5)	Researchers' websites	Information displayed on a researcher's web page(s)	Quantity	Survey of researchers
(6)	Researchers' websites	Effects of researchers' web page(s) (on time budget, communication, contacts and recognition)	Impact	Survey of researchers

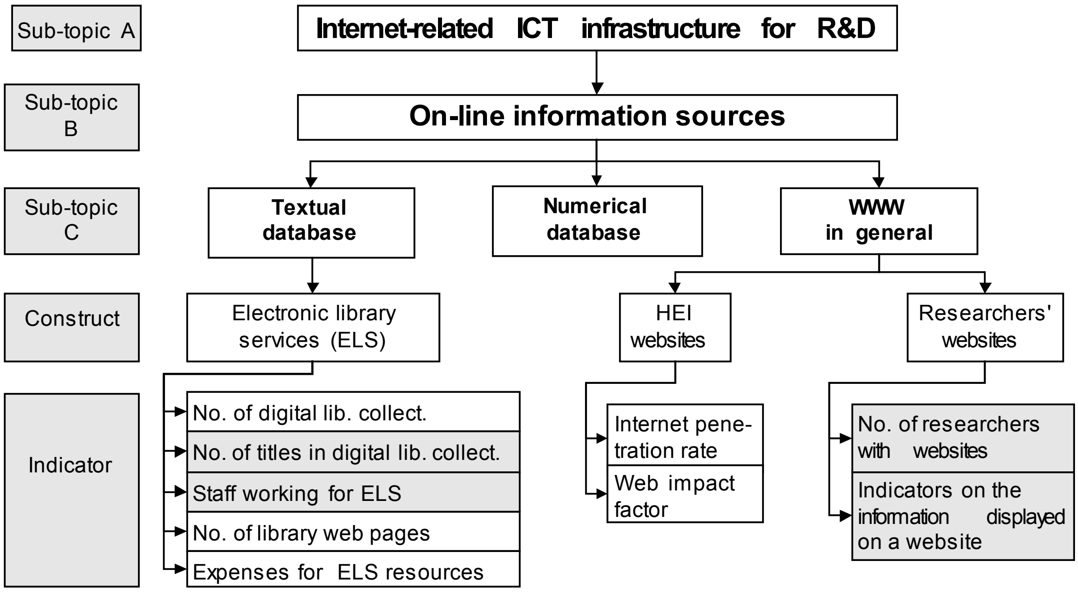
Source: FHSO compilation

- The *World Wide Web* is less structured than any regular database but it contains a lot more information of a broader nature which is often accessible without any user fees being levied. Hence, it has become a vital information source for many researchers. The development of indicators useful for monitoring to what extent the Internet is used as a platform for providing and retrieving information in R&D has just begun. Very few approaches for measuring this exist in the available literature. The improvement of new methods such as web content analysis and log file analysis (see deliverable 2.1, pp. 36-39) should provide large benefits for this aspect in particular.

The quantity dimension "on-line information sources"

According to the structure outlined above we differentiate between quantity indicators for textual databases, numerical databases and the WWW in general (see figure 2). We will discuss these indicators below.

Figure 2: Indicators on the quantity dimension "on-line information sources"



Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

The number of digital collections would seem to be an obvious and accessible figure for comparing the pervasiveness of textual databases between countries. But, unfortunately, this is not true. As both formerly off-line only libraries and newly founded on-line collections may have installed access to their information collections via computer networks, the first problem is the definition of what a digital collection is. Furthermore, the number of digital collections is valuable particularly from the point of view of library and information access management. It is less valuable from the research-oriented perspective, as the number of digital library collections is not a very good measure of the amount of information available.

A better indicator for this is the number of objects in a digital collection. This figure is more appropriate from a researcher's point of view, as each title in a database constitutes a possible information source which might be accessed via the Net. It may also be useful to include some weighting to take into account the time and effort required to access a title; on-line titles would then receive the highest rank, off-line titles which can only be accessed by having them sent by mail would receive the lowest rank. Again, we could find many reasons to criticise this indicator, as the number of objects does not give an indication of their informational value which may be better expressed by their size and quality. But, bearing in mind that an indicator is only useful if it is accessible, we have to dispense with an overly sophisticated operationalisation. Commonly accepted definitions do not exist for either the size or the quality of digital titles, nor can they be measured with an acceptable effort. Surveys in the US indicate that the number of electronic database titles is actually the only information that is collected by libraries on a regular basis.[238] We therefore stick to the number of titles in digital collections and propose to normalise it by calculating its relation to the customer base, that is the number of researchers having access to it (see indicator (3)).

---

[238] See Shim, W. et al. (2000): ARL E-metrics project: developing statistics and performance measures to describe electronic information services and resources for ARL libraries, p. 18. (http://www.arl.org/stats/newmeas/emetrics/ phaseone.pdf).

Alternatively we could also use the entire target population which is probably easier to identify.

Cost figures have been proposed as normalised figures, either in relation to total expenditure or in relation to some usage measure such as user sessions, documents examined etc.[239] They could also be calculated at the national level, but this doesn't seem to be a useful approach for our purpose: the cost-effectiveness of ELS is not what we are looking at. Instead it is rather the expenses for digital resources per scientific user which would be of interest. This indicator is related to the amount of electronic resources (see indicator (3)). But it might be biased by differing price structures among European countries,[240] and therefore purely the number of digital resources seems to be the preferable indicator.

Staff figures were proposed by the Equinox project as another indicator for the provision of digital library services.[241] The objective of this indicator is stated on the Equinox website: "To assess the human resources the library puts into its electronic library services, in order to indicate the library's efforts to develop and provide its services, user training and prepare for future requirements."[242] To normalise the indicator, Equinox proposes calculating the fraction of staff providing electronic services in relation to all staff. Also the staff at external institutions responsible for providing the service should be included. While the latter accords with our proposal on computer network staff (see p. 116), the former does not seem to be a useful normalisation for our purposes. As we employ a different perspective and do not look for the performance data related to digital library services but rather for the availability of these services to R&D on the national level, the relation between the total personnel effort and the size of the target population or the number of researchers seems to be more appropriate (see above the discussion on indicator (3)). This information requires the surveying of libraries and digital libraries (see indicator (4)). One problem might result from different types of services provided by digital libraries: while some might restrict themselves to the pure provision of information objects, others also provide their users with training. A categorisation of the services and the assessment of staff data for the different categories should solve this problem and lead to comparable data.

Another indicator on the information supply of ELS was developed in the ARL E-metrics project: the number of library web pages in service. "This information can provide not only a trend-line in terms of the amount of information in the local context but also some crude measure of staff productivity."[243] However, the authors state some problems relating to the comparability of this indicator, as a web page in one library can be quite different from a page in another library. From our point of view it is even more important that we cannot be sure whether the number of (local, national) library web pages is really a valid indicator for what we want to measure, namely the information available from on-line textual databases. Web pages contain bibliographical information which is necessary and important for R&D, but which is also rather standardised and accessible at different local, national or foreign sources. For obtaining bibliographical information the source that promises the best "quality of response" will always be chosen, i.e. the largest number of titles, the most extensive information provided with each title or the easiest handling and access.[244] Consequently the number of web pages of the "local" electronic library service is not really relevant for searching bibliographical information. It might be relevant from another point of view, as it more or less reflects the number of accessible (electronic and non-electronic) information objects. But this was already part of indicator (3) which contains a more precise and therefore

---

239 See Brophy, P. et al. (2000): EQUINOX - Library Performance Measurement and Quality Management System: Performance Indicators for Electronic Library Services. (http://equinox.dcu.ie/reports/pilist.html).

240 Being typical information goods, digital resources are produced with high costs for the first unit and rather low costs for each additional unit. This permits economies of scale, i.e. the producer with the larger output can charge a lower price as his average production costs decrease. Therefore it is plausible to expect lower relative prices for digital resources in countries with large demand and production.

241 See Brophy, P. et al. (2000), op. cit.

242 Ibid. (http://equinox.dcu.ie/reports/method.html#Appendix%201).

243 See Shim, W. et al. (2000), op. cit., pp. 18-19.

244 See for example in regard to mathematical databases: Bourguignon, J.-P.; European Mathematical Society (1999): The future of mathematical databases, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, p. 112.

superior method of collecting it. Therefore we will not include the number of ELS web pages among the SIBIS indicators.

Though *numerical databases* constitute an important input for many research disciplines we will not consider them further in the course of our analysis. We exclude numerical databases because it seems impossible to collect representative and meaningful information on their producers, content, users and usage at national levels. The Gale Directory of Databases (GDDB), a proprietary database of databases which is available from the Gale Group, provides information on the producers of more than 13,000 databases (covering bibliographical, full text and numerical databases).[245] But we cannot be sure that this information is comprehensive and really representative at the country level. Gale is a US-based company and there might be a bias in its sample to American databases or English language databases in general.[246] Moreover, though databases are used over the entire range of research disciplines, the utilisation is especially intensive in the physical sciences, geosciences, and biosciences (see deliverable 2.1, p. 21). An assessment of database usage levels, e.g. by means of a survey, would therefore encounter great problems associated with constructing a representative sample of researchers.

WWW in general: the European Survey of the Information Society (ESIS) II project for Central and Eastern European Countries (CEEC) collected data on the percentage of high schools and universities with Internet websites.[247] This indicator may be useful in an environment with low Internet penetration rates and few organisations connected to the Net. For Western Europe we do not expect much benefit from this indicator as effectively 100% of all higher education and research institutions should be connected. The next step would be to evaluate the Internet presentations and assess different service levels. But as many R&D institutions also fulfil other functions (universities: education; private companies: production) it would be very difficult to single out R&D-oriented web pages or to judge general purpose pages from a rather narrow perspective.

Carrying out this evaluation by means of the new technologies themselves would in many ways be an innovative solution to the problem. The calculation of web impact factors that has already been described in deliverable 2.1 (p. 37) uses the hyperlinks on other web pages to assess the value of a page.[248] The usefulness of a web page is confirmed by many links pointing to it. This method also has some deficits: there is in general a time lag between a newly established or redesigned and improved web page and the reflection of this on the users' web pages. This lag might lead to misjudgements. Furthermore, the current naming of hosts makes it difficult to pick out research institutions in most European countries. This relates to the problem of limited search options on search engines and the instability and unreliability of search results (see deliverable 2.1, p. 23). Therefore we exclude both the Internet penetration rate and the web impact factor from the SIBIS indicators, but we recommend re-considering the latter as soon as new developments have improved the representation of the Net on search engines and the search facilities.

Another option for gauging the utilisation of the Web as an information resource would be to ask the researchers instead of the research institutions. Among researchers we still expect differences regarding the usage of the WWW to communicate expertise, current projects, publications and other relevant information. So it should therefore be possible to assess the Internet penetration rate among researchers and the amount of information that is displayed on their web pages. For the latter a categorisation is necessary that takes into account different types of information that may be included in a researcher's web presentation, such

245 See http://www.galegroup.com/pdf/facts/gdod.pdf.

246 An analysis of the Cuadra Directory, the predecessor of GDDB, returned three English speaking countries, the USA, England and Canada, as the three top producer countries in the early 90s, see: Dusoulier, N. (1994): Databases and data banks, in: Wesley-Tanaskovic, I.; Tocatlian, J.; Roberts, K.H. (eds.): Expanding access to science and technology. The role of information technology. Proceedings of the Second International Symposium on the Frontiers of Science and Technology held in Kyoto, Japan, 12-14 May 1992. (available at: http://www.unu.edu/unupress/unupbooks/uu07ee/ uu07ee0a.htm).

247 See http://www.eu-esis.org/esis2www/synthCEEC7.htm.

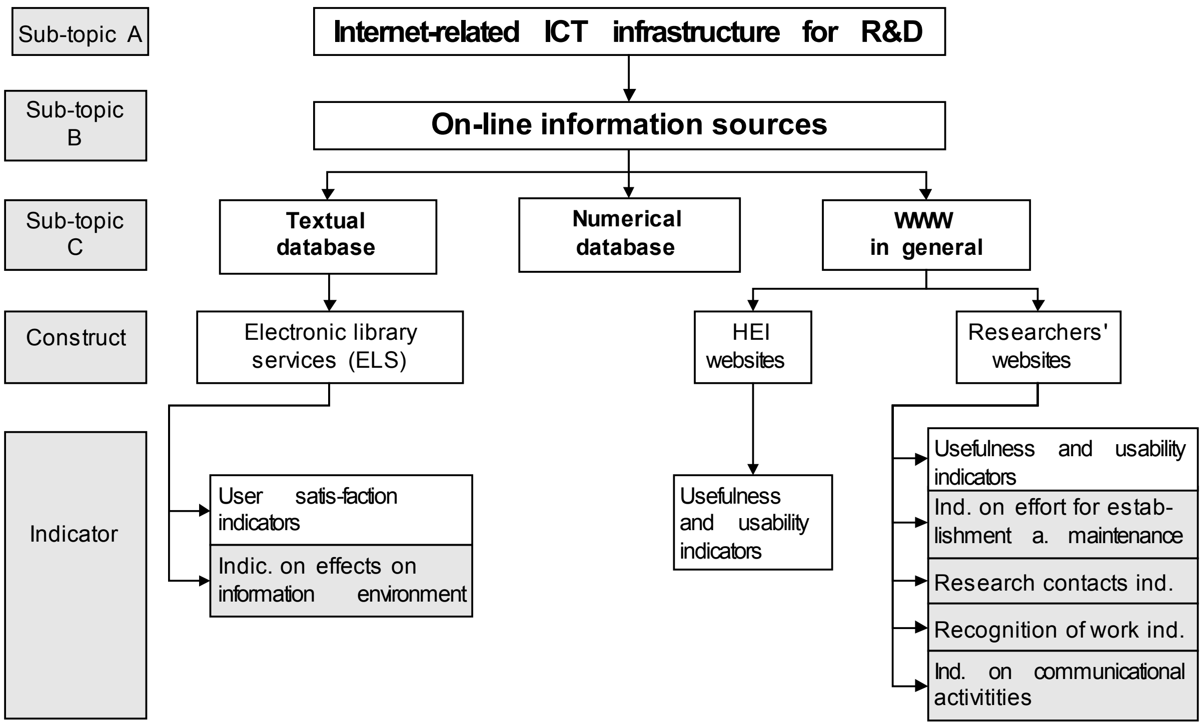
248 See for an introduction and example: Ingwersen, P. (1998): The calculation of web impact factors, in: Journal of Documentation, vol. 54, no. 2, pp. 236-243.

as biographical information, subjects of expertise and interest, current activities, publication lists and entire publications. This information could be gathered by means of closed questions in a questionnaire. Thus we could avoid the problem of the multi-functionality of web pages by focusing the questions on research-related content. We have elaborated an example of this indicator in the annex as indicator (5). The dissemination of research interests and research results should be a general objective in academia and we do not expect any bias in a purely academic sample. Regarding private businesses this might be different, as researchers might (have to) refrain from publishing too much information because of proprietary reasons and company policies. It would therefore be best to limit a pilot assessment to the academic sector.

The impact dimension "on-line information sources"

The impact dimension "indicators on on-line information sources" is located at the intersection of the infrastructure-oriented perspective, which relates to this section, and the task- or R&D process-oriented perspective which will be discussed in section 11.2 below. Hence, we differentiate here only between the different on-line information sources and investigate how their benefits for R&D could be measured, whereas in section 11.2.1 we will look at on-line information sources in general and explore how their usage for different steps of retrieving information for R&D purposes could be evaluated. Though in principle we make a distinction between physical and monetary effects, as described in section 10.4 above, we must concentrate on physical effects: it is often impossible to construct monetary indicators or their assessment may not lead to meaningful results.

Figure 3: Indicators on the impact dimension "on-line information sources"



Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

Recent projects on the assessment of electronic library services have indicator lists which contain user satisfaction indicators as the basic indicators on the impact dimension.[249] Brophy et al. propose to assess user satisfaction of ELS by means of the methodology elaborated in

[249] See Brophy, P. et al. (2000), op. cit. - Shim, W. et al. (2000), op. cit., p. 22.

ISO 11620 on library performance indicators.[250] They point out that due to different user needs and expectations, the comparability of the results might be limited.[251] The micro-level approach with the objective of developing indicators for the assessment and comparison of individual ELS is not suited to our purpose. From our point of view it is not relevant whether a researcher is satisfied with a specific ELS but whether all the ELSs available meet his specific needs. Hence, instead of a user satisfaction assessment we must carry out an impact assessment which investigates the effects of ELS on the researchers' information environment. We integrate this into section 11.2.1 which also deals with the effects of the Internet on information retrieval in R&D (see indicator (16)).

We discussed above the Internet penetration rate, i.e. the connectivity of higher education and research institutions to the Web, though we excluded it from the SIBIS indicators as we did not expect any particularly pertinent information. Regarding the effects of the Internet presentations of HEI, we should expect to encounter some distinctions caused e.g. by the quality and the content of the presentations. However, it will be very difficult to measure these effects, and we can think of only three viable methods:

(1) The first would be direct user assessment, i.e. including the question "How useful are our web pages for you?" on the website of a research institution in the hope that it will trigger some usable responses.

(2) The second possible method consists of an evaluation of the website along pre-established criteria.[252]

(3) The last option would be an "automated assessment", using log files and taking e.g. the viewing time, the download frequency or other parameters as indications of the usability and usefulness. The problems of this method have been discussed at various places of this and the previous deliverable.

Methods (1) and (3) require the co-operation of R&D institutions as they are based on the analysis of proprietary data which leaves us with the problem of data accessibility. Method (2) is very costly to implement, because an evaluation scheme which is sufficiently general to be applicable to different university systems must be developed. Moreover, great efforts must be made to construct a sample that is large enough and sufficiently representative for an evaluation. Other methods, such as questioning the research institutions themselves do not seem to be promising.

Things are somewhat different regarding the effects of researchers' own web page(s) on their time budgets, communicational activities, recognition, contacts and collaboration. Each of these might be affected by the maintenance of a web presentation and the reactions prompted by it and we can expect that researchers are at least aware of some of these effects (see indicator (6)). We could in principle also employ the two methods described above and take the "customers'" reactions as an indication of the usefulness and usability of a researcher's web page(s).

The data obtained from physical indicators could be used to construct some monetary indicators and calculate monetary effects, such as additional research funds obtained due to an open information policy or losses suffered due to violations of intellectual property rights as too much information has been published. But the data necessary for this exceeds by far what is currently available or accessible under optimistic assumptions and therefore we will not pursue this lead any further.

---

250 See International Organization for Standardization (1998): Information and documentation – library performance indicators.

251 See Brophy, P. et al. (2000), op. cit.

252 This method is for example carried out to evaluate the quality of HEI course material presented on the Web in Chen, C. et al. (1998): How did university departments interweave the Web: A study of connectivity and underlying factors, in: Interacting with computers, vol. 10, pp. 353-373.

11.1.4 Tools for Computer-Mediated Communication (CMC)

Overview of indicators on CMC

Computer-mediated communication has been defined as intended human communication between two or more individuals in which the receiver has been personally addressed by the sender through the use of central computers (see deliverable 2.1, p. 26). We differentiated in deliverable 2.1 between three different tools for CMC: e-mail, mailing lists and on-line meeting tools. For e-mail, the most widely diffused of the three tools, some quantitative analyses exist which evaluate the impact of e-mail on research productivity. Usually data collection in this field has been carried out by means of surveys. In regard to the importance and impact of mailing lists and on-line meeting tools, the existing information is rather cursory, fragmentary and based on case studies. Therefore we limit our approach to indicator formation on e-mail (see table 3). However, some information on mailing lists will also be obtained, as they use the same transmission technology.

Table 3: Indicators on computer-mediated communication

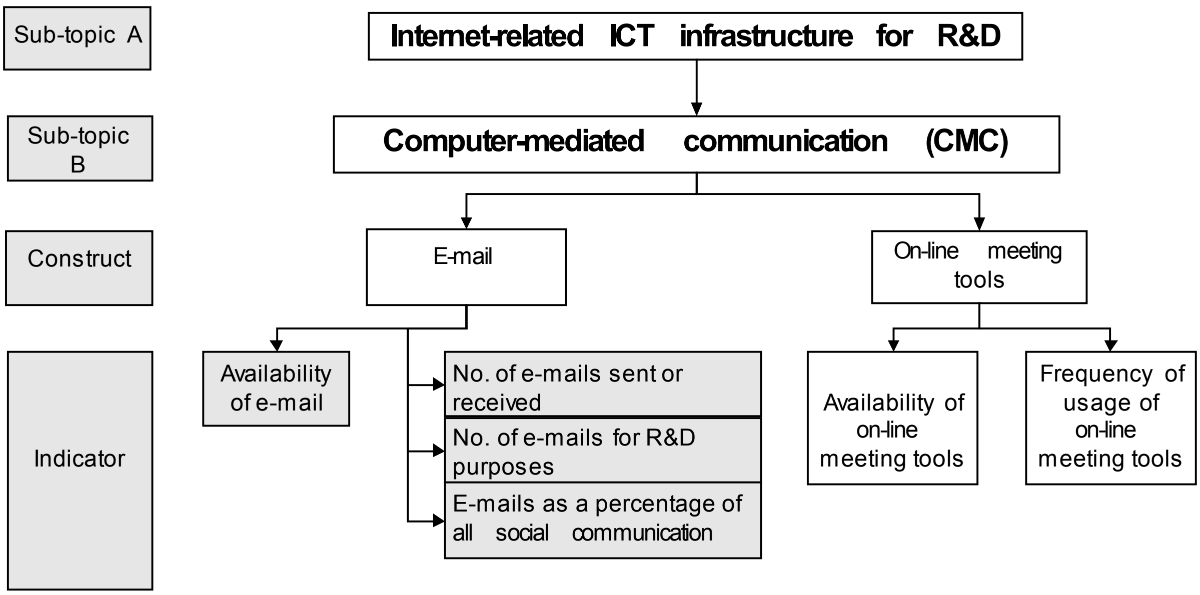
No.	Construct	Name of indicator	Dimension	Suggested method
(7)	E-mail	E-mail communication for R&D purposes	Quantity	Survey of researchers
(8)	E-mail	Effects of e-mail use for R&D purposes (on information, contacts, collaborations, productivity, quality of work)	Impact	Survey of researchers

Source: FHSO compilation

The quantity dimension "tools for CMC"

The structure of possible indicators on tools for computer-mediated communication is shown in figure 4. It includes availability and usage indicators for e-mail and on-line meeting tools.

Figure 4: Indicators on the quantity dimension "tools for CMC"



Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

As e-mail has developed into an indispensable communication channel among researchers in all disciplines, it represents a good application for assessing the amount of computer-mediated communication within a country's R&D system. The assessment of the availability of e-mail services functions as a filtering criterion. However, the simple distinction between users and non-users would probably not provide much insight: the information value of the indicator would be low, as researchers who send and receive dozens of e-mails per day and others who receive one or two or even less would be counted equally. In the worst case (from a methodological point of view), with everybody using e-mail to at least some extent, the nominal variable "usage of e-mail: yes-no" might have no variance at all. Therefore it is preferable to collect a ratio measure such as the number of e-mails in a certain time period. It might be worthwhile to differentiate between e-mails sent and received as this could indicate different communication profiles.

Mailing lists which have been discussed separately in the SIBIS deliverable 2.1 can be considered as a special type of e-mail. If we also take into account the fact that university scholars use e-mail not only for research-motivated communication but also for teaching or administration issues,[253] it seems to be highly appropriate to categorise e-mails according to their objectives. Then the data obtained will be more specific and more reliable for assessing the effects of e-mail on R&D, university education or other activities. An alternative could be a distinction between different types of senders and receivers, e.g. peers, students, professional associations, sponsors etc. But the latter categorisation seems to be inferior to the former, as it constitutes only a measure of the function of an e-mail which is also directly measurable.

A bias might be introduced into communication data by researchers' productivity, their reputation and position within their organisations. These factors influence the amount of communication and the structure: e.g. the director of a research department would probably both send and receive lots of e-mails on administrative issues, whereas a research assistant might receive some informational e-mails from his (her) administration but (s)he probably does not send many. One way to cope with that bias would be to control for the position of an e-mail user in his or her organisation. Another more precise but also more costly way would be the inclusion of other means of communication (personal meetings, phone, fax, regular mail etc.) when investigating the importance of e-mail.

The necessary assessment of the objectives of an e-mail and its usage compared to other means of communication rule out server log files as data sources. The most promising method should be a survey question that asks researchers about their communication habits (see indicator (7)). It has been shown that the number of self-reported e-mail messages correlates fairly well with the actual usage data.[254]

In principle, data on the quantity dimension "on-line meeting tools" would have to be assessed in a similar way as e-mail data. But, as we have already stated in the deliverable 2.1 (pp. 28-29), on-line meeting tools are not yet widely diffused. With their specific features which facilitate synchronous communication and collaborative work they are also of greater importance in those fields of research that benefit from the close interaction of researchers. They are of lesser importance in other fields where desk research and the subsequent distribution and publication of the results are the norm. Therefore an assessment of the availability of on-line meeting tools by means of a survey among researchers will probably not provide any valid results at this point in time. We have added on-line meeting tools in figure 4, but we will not consider them any further among the SIBIS indicators.

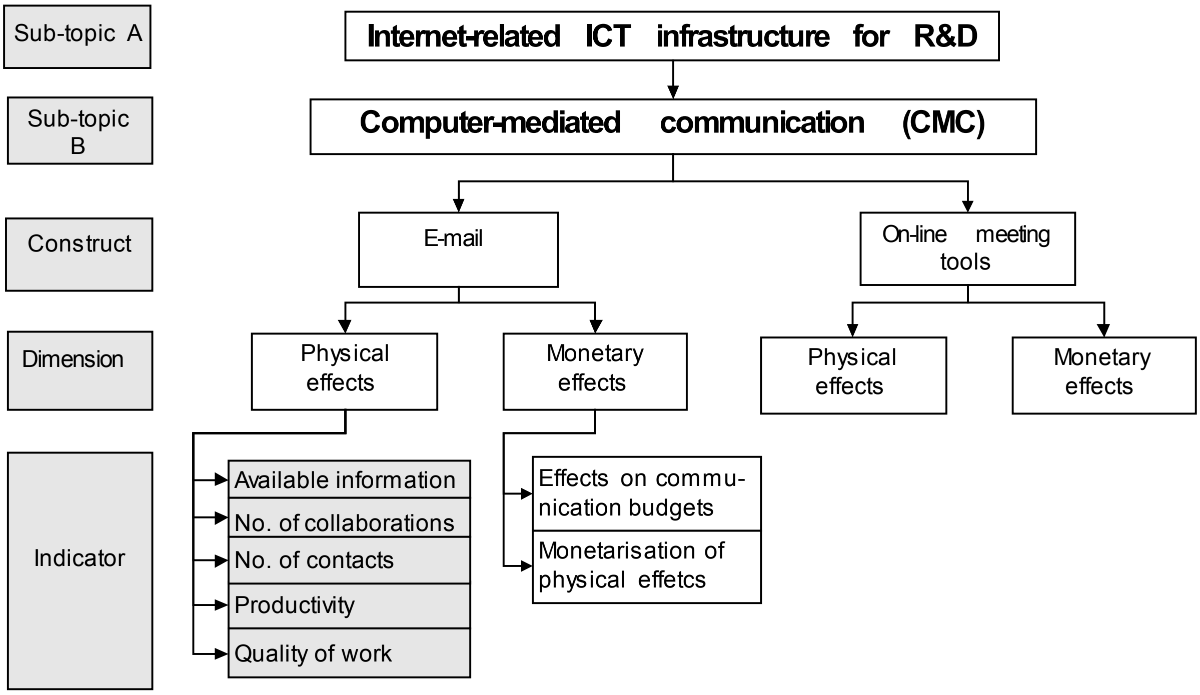
253 See Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997): Internet use by faculty members in various disciplines: a comparative case study, in: Journal of the American Society for Information Science, vol. 48, no. 6, p. 512.

254 See Hesse, B. W. et al. (1993): Returns to science: Computer networks in oceanography, in: Communications of the ACM, vol. 36, no. 8, p. 93.

The impact dimension "CMC"

The present section deals with the impact dimension "CMC". Within the scientific literature on this topic, a clear focus has been laid on e-mail as the most common communication method. We will restrict ourselves to the costs and benefits of e-mail, for more or less the same reasons for which we excluded on-line meeting tools from the quantitative indicators (see above). However, we include it in the overview of indicators on the impact dimension "tools for CMC" (see figure 5).

Figure 5: Indicators on the impact dimension "tools for CMC"



Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

Communication and information science have discussed the effects of computer-mediated communication on R&D, particularly from the viewpoint of its impact on the information base, amount of collaboration, research productivity and quality of results (see deliverable 2.1, pp. 27 and 88). It should be possible to collect valid data on the perceived effects of e-mail by means of self-assessment (see indicator (8)). Different analyses have undertaken this exercise and in general they have produced valid results.[255]

It is also feasible to relate usage data on CMC to "objective" data on research output or other impact dimensions such as the levels of information, collaboration or the size of the professional network. As we noted in deliverable 2.1 this kind of analysis has been performed at various times.[256] Nevertheless, it suffers from a causality problem: we do not know whether larger output causes more CMC or vice versa. This type of analysis would also require additional data collection, e.g. on the researchers' output (articles, presentations, patents etc.) and careful consideration of impact lags (increased CMC in period t may affect output in t+1, t+2, t+3). Due to these data requirements and the problems with interpreting the data, we will not consider this alternative further.

[255] See Walsh, J.P.; Maloney, N. G. (2001): Computer network use, collaboration structures and productivity, in: P. Hinds and S. Kiesler (ed.): Distributed work. Cambridge, Mass. – Cohen, J. (1996): Computer mediated communication and publication productivity among faculty, in: Internet Research: Electronic Networking Applications and Policy, vol. 6, no. 2/3, pp. 55.

[256] See Walsh, J.P.; Maloney, N. G. (2001), op. cit. - Cohen, J. (1996), op. cit. - Hesse, B. W. et al. (1993), op. cit.

In principle the monetary effects of CMC could also be measured.[257] This is very difficult for an individual researcher as (s)he would have to know the costs and benefits of CMC messages in relation to other forms of communication. But research managers, project managers and the cost accounting units of universities and firms should be in possession of the necessary information. However, this is still no simple undertaking: as hard- and software are also used for other purposes, it is difficult to compute the cost shares attributable to CMC and it is even more difficult to quantify benefits such as the "increased available information" or "the increased collaboration". Reducing the assessment to simple budgetary effects, such as the decrease in telephone or regular mail expenses due to CMC, however, would definitely underestimate the impact. Due to these difficulties a monetary impact assessment cannot be considered optimal.

11.1.5 Computer skills for R&D

Overview of indicators on computer skills for R&D

A broad variety of computer skills are necessary to use research networks, network applications and the available ICT hard- and software properly for R&D projects. A shortage of these skills might lead to delays in R&D projects, an inefficient organisation (trying to substitute for the missing resources) or sub-optimal results. Various initiatives have stressed the necessity of improving computer skills within science,[258] but no attempt to measure these has been documented. Therefore it is vital for SIBIS to develop "humanware" indicators which monitor the availability of computer skills for R&D, in addition to the hardware and software indicators described above.

As the SIBIS project specifically analyses Internet-related issues for R&D we could be satisfied with the Internet-related skills of researchers. But it might be difficult to separate general computer skills from the more specialised Internet-related skills. Both are necessary for using the Internet efficiently and for performing successful R&D. Therefore, we will focus the analysis in this section on computer skills in general and indicators for these (see table 4). However, during collection of empirical data and pretesting of the developed indicators, the accessibility of more specific data could be checked.

---

257 See e.g. the rough assessment of costs for Internet services to R&D in Lubanski, A.; Matthew, L. (1998). Socio-economic Impact of the Internet in the Academic Research Environment. Proceedings IRISS '98 International Conference: 25-27 March 1998, Bristol, UK. (http://sosig.ac.uk/iriss/papers/paper18.htm).

258 See Aubert, J.E.; Bayar, V. (1999): Maximising the Benefits of Information Technology for Science: Overview and Major Issues, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, pp. 24-25. – European Technology Assessment Network (1999): Transforming European science through information and communication technologies: challenges and opportunities of the digital age. Final version, p. 57. – European Commission (2000a): Towards a European research area. Communication from the Commission COM 2000 (6). Brussels, 18 January 2000, p. 11. (http://europa.eu.int/comm/research/ area/com2000-6-en.pdf). – European Commission (2000b): Making a reality of the European research area: guidelines for EU research activities (2002-2006). Communication from the European Commission COM(2000) 612 final. Brussels, 4 October 2000, p. 17. (Http://europa.eu.int/comm/research/area/ com2000-612-en.pdf).

Table 4: Indicators on computer skills for R&D

No.	Construct	Name of indicator	Dimension	Suggested method
(9)	Computer skills of R&D personnel	Computer skills of R&D personnel	Quantity	Survey of researchers
(10)	Specialised computer staff	Computer staff providing services to R&D	Quantity	DMS
(11)	Specialised computer staff	Unfilled vacancies in private businesses for computer staff providing services to R&D	Quantity	DMS
(12)	Computer skills of R&D personnel	Effects of computer skills on R&D	Impact	Survey of researchers

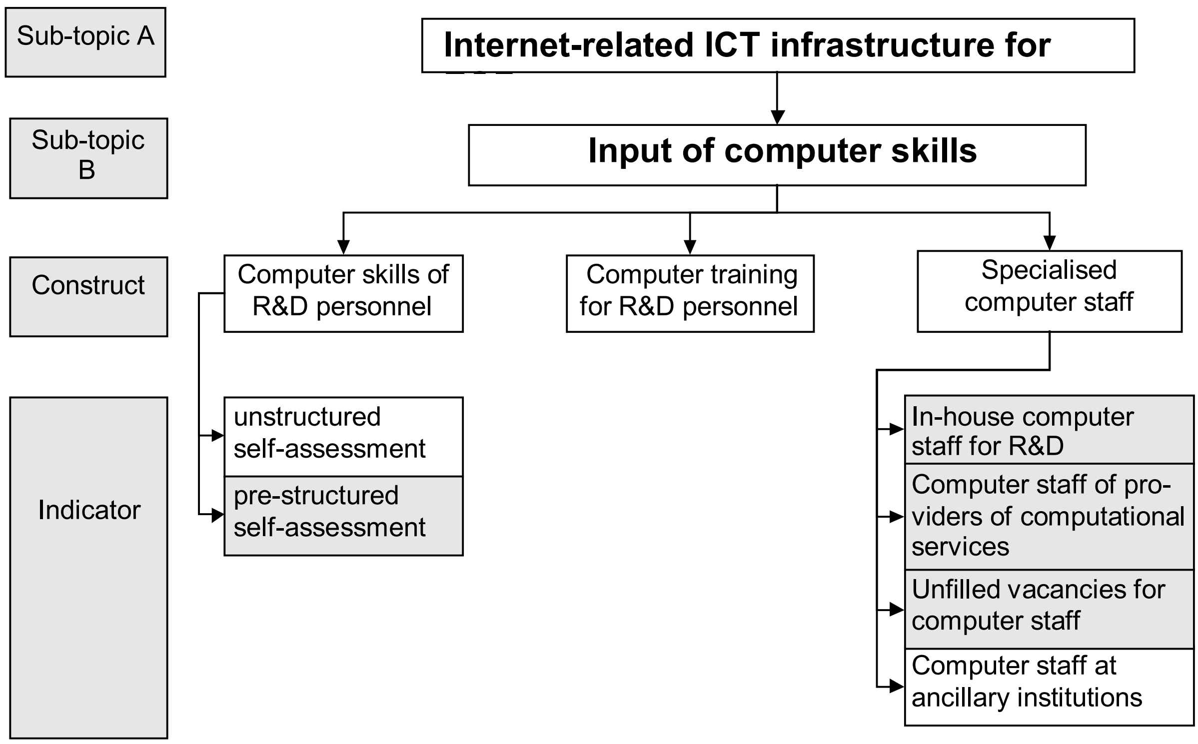
Source: FHSO compilation

The quantity dimension " computer skills for R&D"

We discuss three different approaches for measuring the computer skills of R&D personnel available for R&D activities (see figure 6):

e) Assessment of the computer skills levels of R&D personnel
f) Evaluation of the training activities on computers for R&D personnel
g) Assessment of the specialised computer staff and the computer skills bought from external providers for R&D

Figure 6: Indicators on the quantity dimension "computer skills for R&D"



Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

Ad a) A standardised scheme for computer skills does not exist. A straightforward approach would be the conceptualisation of different skills elements and levels and their subsequent assessment within a survey. This could be carried out in a specific survey of researchers which permits the assessment of specific R&D-related skills, or in a general labour force survey. The latter would have to use broader categories but creates the possibility of comparing the skills level of different occupations and groups among the workforce. We will not deal with the broader concept here, as the SIBIS topic on work, skills and employment discusses it in more detail (see the respective deliverables 2.1 and 2.2 on "Work, skills & employment").

The measurement of computer skills among R&D personnel could take as a starting point the different functions of computers within research. Word processing, spread sheet analysis, database management, statistical analyses, presentation and graphic applications, administration and time management, communication and Internet research – each of these tasks involves computers to some extent and requires the knowledge of specific software and applications. The easiest method, an unstructured self-assessment ("Mark on the scale your knowledge of your word processing software 1=excellent, 5=very poor") suffers from reliability problems: the better you know a software application, the better you know your limits. Therefore this sort of self-assessment might not produce reliable results.

An alternative would be to describe certain features of computer applications and ask the respondents if they have mastered them. However, this pre-structured self-assessment would have to deal with the problem that it is not the mastering of a specific software application which has to be measured, but the mastering of groups of software applications available for certain tasks and functions. But even the software within these groups has different features and specialisations (e.g. the statistics program SPSS offers a broad range of services, while Eviews, another program, offers fewer applications but goes into a lot more depth) and it might be difficult to formulate general skills levels. Also, in the course of time the capabilities of computer programmes change and the indicator has to be defined anew each time it is used. Another problem might result from the fact that researchers do not necessarily carry out all the computer work themselves as they have assistants who support them. So if somebody has not mastered or doesn't even use a group of software applications, this might be for different reasons: (s)he doesn't need it for his (her) work, (s)he has somebody else doing this kind of work or (s)he doesn't know how to do it on the computer and consequently does it the "old" way instead.

A pre-structured self-assessment still seems to be the only possible method of collecting reliable data on researchers' computer skills. The table on indicator (9) provides an example of how this could be done.

Ad b) Another possible approach could be to quantify the volume of training on the use of ICT that scientists and researchers receive. Though universities generally provide the basic skills, there might be a skills gap, especially with older researchers or due to the further development of ICT. An indicator could specify the efforts undertaken within the different national research systems to close an ICT skills gap in R&D. Of course, it would not provide any information on the initial ICT skills of R&D personnel and it also depends on the quality of schooling and the provision of ICT skills in other educational institutions (i.e. a low demand for ICT vocational training could also be due to a high level of ICT training at secondary and higher education institutions). Because of these uncertainties regarding its value, we will not include it among the SIBIS indicators. It might be worthwhile including it in other (pilot) surveys that describe and evaluate the national education systems to a better extent.

Ad c) The number of specialised computer staff could be used as a measure for the computer skills available for R&D activities. We use the term "computer staff" to summarise staff that

- manage the computers, networks and digital resources,
- manage Internet access and the presentation,
- carry out information searches and computations and
- provide user training

on behalf of academic institutions and R&D departments.

According to OECD classifications, it makes sense to differentiate between the staff of R&D computer departments and central computer departments: whereas the former should be included in counts of R&D personnel, the latter should not.[259] This might introduce a bias to R&D personnel/computer staff counts depending on the structure of an organisation. To avoid this bias it is preferable to include three different types of computing staff: staff within R&D departments, within central computer departments providing services for R&D and at external service providers.

In general, a quantification of the staff responsible for computer services in universities, research institutes, private business R&D departments and other R&D-related organisations seems difficult to achieve. In universities in particular, computer services are provided by various units: a large part is managed by central computer departments, but other central departments such as press offices and libraries may also contribute. Some departments employ additional technical staff and others rely heavily on their own abilities or on students. Data collection seems to be less difficult in private businesses and research institutes: they are either smaller or have R&D or ICT departments which monitor the number of internal and external computer services provided. Therefore an appropriate way to collect comparable data on computer staff for R&D in the business sector seems to be the surveying of chief executives, heads of R&D departments or ICT departments (see indicators (10) and (11)).

In the future, questions regarding the computer staff providing services to R&D could become part of the national R&D surveys in the private business sector. These are carried out irregularly in the OECD member states and other countries worldwide. Universities and public research institutes could be added subsequently to make the sample representative for the entire national R&D systems.

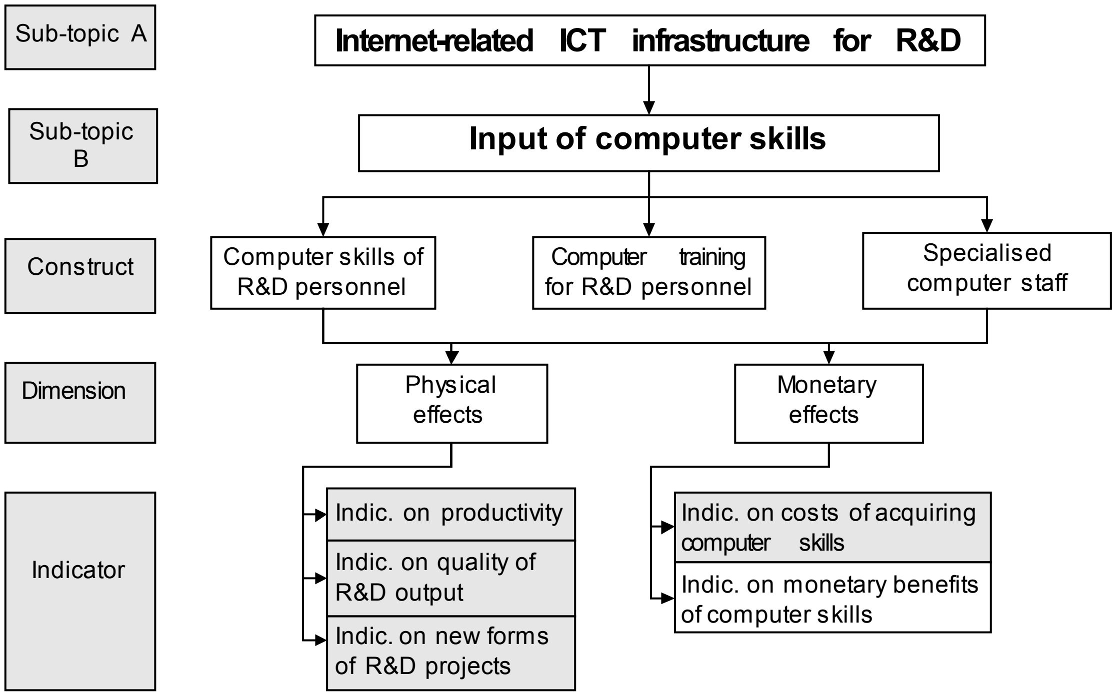
Further computer staff data from ancillary organisations might be included on the basis of availability: TERENA collects data on the staff sizes of, and the total amount of personnel effort available to, the national research networks (see deliverable 2.1, p. 16). Digital library working groups discuss the collection of staff size data as an indicator of their patron-oriented resources and services (see p. 106).

The impact dimension “computer skills for R&D”

We can confidently assume that computer skills both entail costs and benefits for research activities. Costs manifest themselves for example as learning costs: using a software application requires prior investment in learning its capabilities and specific features. Benefits are manifold and might entail higher productivity and improved quality of work. We can again differentiate between physical effects and monetary effects (see figure 7).

[259] See OECD (1994), op. cit., p. 37.

Figure 7: Indicators on the impact dimension “computer skills for R&D”



Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

The physical effects of the exploitation of computer skills for R&D are basically the same as for the physical infrastructure discussed in this section: a change of productivity and of the quality of R&D results. Productivity could be assessed by means of calculating performance measures, e.g. research papers per researcher or amount of work necessary for writing one paper, as a function of the amount of computer skills employed. Quality could be assessed similarly, by taking into account a quality indicator such as the percentage of peer reviewed research papers. Another option could be to ask researchers within a survey whether the availability of computer skills (their own, from specialised staff) has affected their research output. One problem eventually leads to a bias in both approaches: some research projects, e.g. large and computation-intensive ones, might only have become possible because computer staff with sufficient expertise were available. However, this is not necessarily reflected in the output of an R&D project, and hence we might run the risk of obtaining invalid results. In a survey this problem may be better accounted for than in performance calculations.

The monetary costs and benefits of computer skills within R&D activities are difficult to measure. The benefits are mainly returns to human capital: a rising level of computer skills increases the level of human capital employed, resulting in higher wages and a substitution of human capital for other production factors. While some costs, such as learning costs that accrue for learning a new software application or a new technique of computer-based analysis, may be interpreted and measured as labour costs, for a total assessment this perspective is too narrow. Some of the costs resulting from the acquisition of computer skills are opportunity costs, costs incurred for not employing the working time and effort for alternative uses. Furthermore, it is difficult to measure these opportunity costs as they are revenues which would have resulted from some other, hypothetical use of the resources. Additionally, it is very difficult to evaluate funding for R&D projects which have only become possible due to computer skills if we employ a long-range perspective. All in all, an entire monetarisation of the effects of computer skills on R&D seems overly ambitious and we will have to be satisfied with an evaluation of the physical effects.

Indicator (12) formulates a proposal as to how the effects of computer skills might be assessed in a survey of researchers.

The sub-topic of computer skills was the last one we subsumed among the infrastructure sub-topics. The next section takes a different approach, looking at the Internet from a process-oriented perspective. The main question is: how can we measure the extent to which the Internet is used in different phases of R&D work and what are the resulting effects? As we have to cover all research disciplines and the public as well as the private sector, we must employ a rather coarse definition of R&D work which is not particularly differentiated and distinguishes only between three phases.

11.2 Integration of the new network technologies into research activities

11.2.1 Data collection and information retrieval

Overview of indicators on data collection and information retrieval

In deliverable 2.1 we defined "information" as structured data that needs some additional knowledge to be understood and exploited properly. The collection of data and information both utilise computer networks, but in a different manner and they rely on different methods.

- Network-based data collection employs a range of different methods. While an increasing number of studies use these methods, we could not find any meta-analyses which assess their utilisation among research communities.
- Scientific information is stored on specific Internet-accessible databases and the WWW in general, and various tools exist for retrieval. The utilisation of both sources and tools (software) has been assessed in pilot studies.

In general, the utilisation of the Internet for, and its impact on, information retrieval has been documented significantly better than for data collection. This may be due to the different nature of the methods and especially their suitability for use across different research disciplines. We consequently will not develop any indicators on data collection as their usefulness is doubtful.

Table 5: Indicators on data collection and information retrieval

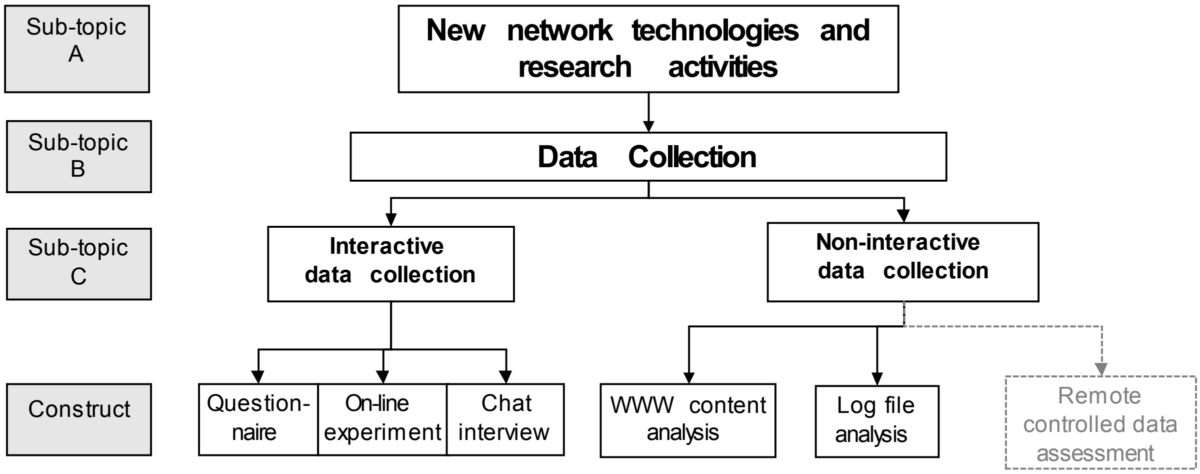
No.	Construct	Name of indicator	Dimension	Suggested method
(13)	Digital library and peer site usage	Frequency of information retrieval from electronic sources	Quantity	Survey of researchers
(14)	Digital library and peer site usage	Documents/items from electronic sources	Quantity	Survey of researchers
(15)	Software usage	Frequency of software usage	Quantity	Survey of researchers
(16)	Information retrieval	Effects of information retrieval from and via the Internet (on time budgets, productivity, quality of work, contacts)	Impact	Survey of researchers

Source: FHSO compilation

The quantity dimension "data collection and information retrieval"

In deliverable 2.1 we distinguished between interactive and non-interactive methods of data collection (pp. 35-39). Interactive data collection takes place by means of questionnaires, interviews and experiments carried out via the computer, whereas non-interactive data collection uses information objects on computer servers (web pages, log files) for data collection (see figure 8). Automated data collection carried out by remotely controlled instruments might be considered as a borderline case: data is not collected from or via a computer network, but without this network other procedures for data collection and subsequent transmission to analytical devices would have to be found. Both the network and the instrumentation are therefore essential for implementing this sort of data collection.

Figure 8: Indicators on the quantity dimension "data collection"



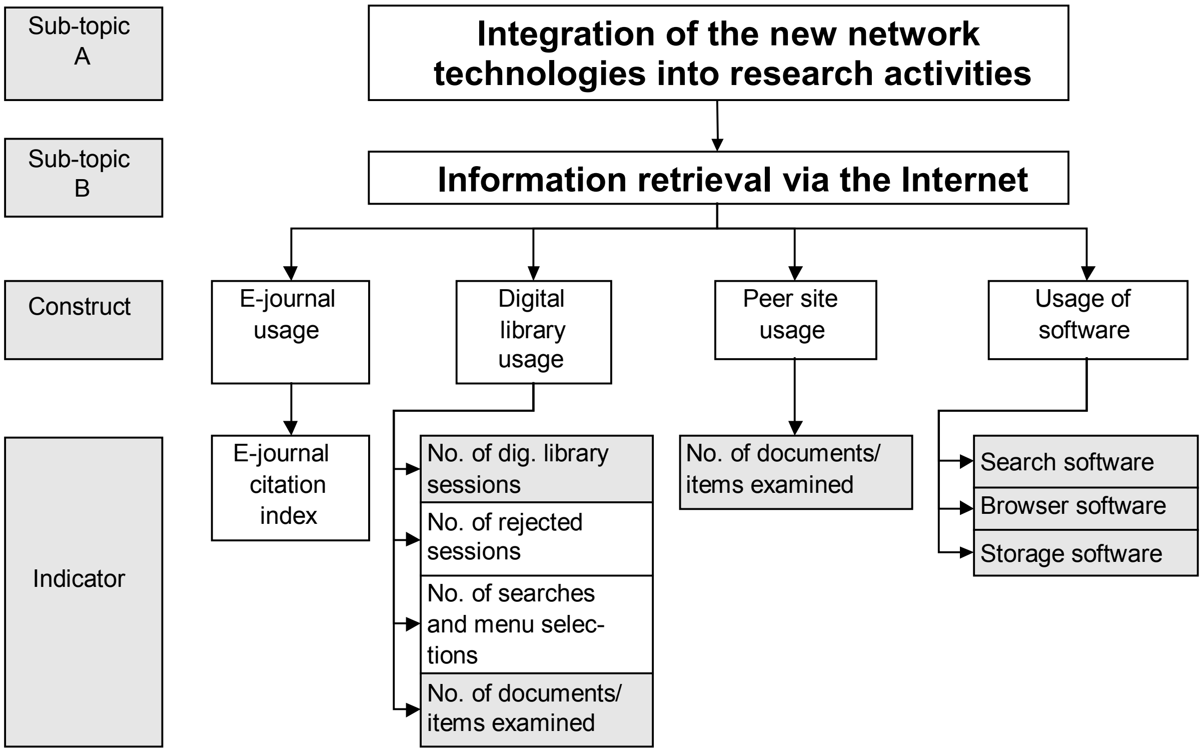
Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

However, we will not consider among the SIBIS indicators any of the data collection methods identified in deliverable 2.1. The reason for this lies in their specificity: though none is strictly limited to only one research discipline, they have particular features which make them suitable only for a certain sector within science. On-line surveys, experiments or interviews are usually carried out by researchers from the social sciences. The same applies to non-interactive data collection from Web servers with above average usage in information and communication science. Data collection via instruments and subsequent transmission via the Net are more important in disciplines that need large amounts of data from places that are difficult to reach, e.g. climate-related data in meteorology, earth-related data in geophysics or spatial data in astrophysics. Of course, we still might construct indicators for each of the different methods, but it would be impossible to gather data that can be regarded as representative for the entire national research system. Furthermore, the standing of a country within a research discipline will also influence the data and we might even expect a correlation between the extent to which the national scientific community within a specific discipline uses novel and innovative computer-based methods and its importance and recognition among the international scientific community in this discipline. In general, representative data may be collected for research disciplines, but not for entire national research systems.

Things are somewhat easier and more promising for information retrieval (see figure 9).

Figure 9: Indicators on the quantity dimension "information retrieval"



Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

The construction of *e-journal citation indexes* was usually motivated by the impact assessment of electronic journals in comparison to "regular" off-line journals.[260] The empirical basis for this were citation data such as the Science Citation Index from the Institute for Scientific Information. But the latter distinction has become more and more problematic as electronic versions of printed journals have been made available by the majority of publishers. And it is impossible to deduce from a citation if the author has used the electronic or the printed version of a publication. Pure citation data without additional information from the author are therefore in many cases not a valid indicator for distinguishing between on-line and off-line sources. Consequently we will not use this indicator among the SIBIS indicators.

Another approach would be to analyse the extent to which *digital sources* are used. The different studies on digital library metrics proposed four indicators which aim to assess the usage of electronic library services:[261]

a) the number of digital library sessions (logins)
b) the number of rejected sessions
c) the number of searches and menu selections and
d) the number of documents/items examined

Two of these (a and d) may also be employed on the macro-level of national research systems for the purpose of evaluating the informational value of the Internet for R&D. Indicators b and c are less suited due to methodical reasons: the mentioned studies consider

---

260 See Harter, S. P. (1996): The Impact of Electronic Journals on Scholarly Communication: A Citation Analysis, in: The Public-Access Computer Systems Review, vol. 7, no. 5. (http://info.lib.uh.edu/pr/v7/n5/hart7n5.html)

261 See International Coalition of Library Consortia (1998): Guidelines for Statistical Measures of Usage of Web-based Indexed, Abstracted, and Full-Text Resources. (http://www.library.yale.edu/consortia/webstats.html). – Brophy, P. et al. (2000), op. cit. – Shim, W. et al. (2000), op. cit.

log file analysis as the method of data generation. We have dispensed with this method for the time being for various reasons[262] and prefer to rely on a survey of researchers. However, indicators b and c seem to be too specific to be included in a survey that is based on the self-assessment of researchers. But indicators a and d should be suitable for this approach. One problem might result from the varying requirements for information search and retrieval or the differences in reading behaviour between disciplines or research domains. Therefore, as in section 11.1.4 on CMC, we have decided to include the usage of all possible sources of information, computer-based as well as non computer-based, to provide for a control mechanism for the general information search and analysis behaviour.[263] Though this leads to some additional questioning effort, it creates the possibility of combining indicators on different electronic sources. Indicators (13) and (14) take this into account and integrate into the questions both digital libraries and peers' Internet presentations as important electronic information sources.

Besides looking at the electronic sources of information it is also possible to investigate the *usage of tools* which are employed for locating and gathering R&D-related information. This has been a common objective of investigations attempting to assess the effects of the Internet in academia.[264] As already stated in deliverable 2.1, there are different levels at which the process of information retrieval might start: locating possible sources of required information, browsing and evaluating accessible information objects or saving and transmitting them for further, off-line uses. Different computer-based tools and methods exist for each step, in addition to "traditional" methods. To secure the maximum level of comparability we must again gather data on all of the different options and stages, from information search through to retrieval and transmission (see indicator (15)).

The impact dimension "data collection and information retrieval"

We restrict our discussion of the impact dimension to information retrieval and omit data collection issues because the same limitation applies as in the quantity dimension: data collection methods are rather mono-disciplinary, and we cannot expect to find any indicators that give an accurate picture of the entire research systems.

As in other sections (e.g. on-line information sources in 11.1.3 and computer-mediated communication in 11.1.4) we propose to evaluate the effects of Internet-based information retrieval by means of the self-assessment of researchers. We do not have to differentiate between on-line information sources, as this has already been covered in section 11.1.3 (see p. 108). Instead, we assess the impact on different features of R&D output, such as productivity, quality and the contact network of a researcher (see indicator (16)).

We will now continue with the sub-topic "publication and dissemination of research results" and omit "data analysis and technology development" from the indicator system, as we consider the latter as highly dynamic and not accessible to a stable and reliable assessment.

262 Log file analysis is still implemented mainly for gauging the demand on a server. Multiple problems related to technological features, a lack of standardisation and the human usage of the Internet make it impossible to use the recorded log data for other purposes (such as assessing the usage of Internet applications in different countries), see deliverable 2.1.

263 A similar approach has been carried out by Björk, B.-C.; Turk, Z. (2000): How scientists retrieve publications: an empirical study of how the Internet is overtaking paper media, in: The Journal of Electronic Publishing, vol. 6, no. 2. (http://www.press.umich.edu/jep/06-02/bjork.html).

264 See Lubanski, A.; Matthew, L. (1998), op. cit. – Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997), op. cit. – Kaminer, N.; Braunstein, Y.M. (1998): Bibliometric analysis of the impact of Internet use on scholarly productivity, in: Journal of the American Society for Information Science, vol. 49, no. 8, pp. 720-730.

11.2.2 Publication and dissemination of research results

Overview of indicators on the publication and dissemination of research results

We have shown in deliverable 2.1 (pp. 43-44) that in the last decade various new forms of electronic scientific publications have appeared which to some extent both complement and replace traditional scientific publishing. Besides the simple distribution of research papers such as e-mail attachments and postings on WWW pages, e-publishing has become the most important innovation in presenting and disseminating research results. While acceptance from the readers has been rapid, acceptance from the authors has been somewhat slower, due to concerns regarding the reduced reputation and perceived lower impact of e-journals and other e-publications.

The acceleration of communication and the new ways of giving comments and critique have also changed the review and quality control part of scientific publishing in both electronic and non-electronic media. The scientific discussion has been focused very much on the advantages and disadvantages of different review procedures.

The major source for gathering empirical information has been the survey-based questionnaire. Table 6 gives an overview of indicators which we consider worthy of collection in a more detailed analysis of the effects of the Net on the publication and dissemination of research results.

Table 6: Indicators on the publication and dissemination of research results

No.	Construct	Name of indicator	Dimension	Suggested method
(17)	E-publishing	Amount of work published in electronic media	Quantity	Survey of researchers
(18)	Quality control	Review activities for e-journals	Quantity	Survey of researchers
(19)	E-publishing	Impact of publications in electronic media	Impact	Survey of researchers

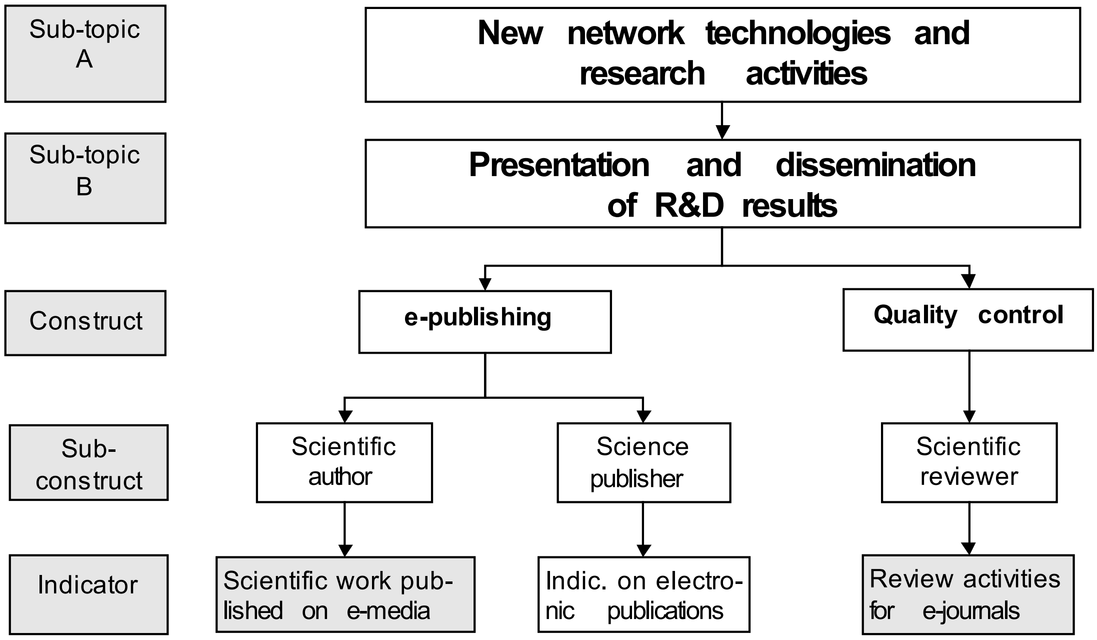
Source: FHSO compilation

Quantity indicators on the publication and dissemination of research results

It is possible to assess the significance of electronic media for the publication and dissemination of research results at two sources: the scientist and the publisher (see figure 10).

A straightforward approach to assessing the significance of electronic publications from the perspective of the scientific author is to ask for the amount of work published in electronic media. The posting of research papers on a scholar's web page(s) has already been part of indicator (5) but nevertheless it seems worthwhile to make a differentiation between different electronic media. Indicator (17) contains a survey question on the frequency of publication of research results in different scientific media. Besides providing insight into the importance of e-journals from the authors' point of view, this indicator is also useful for computing performance indicators or carrying out causal analyses to evaluate the effects of R&D inputs.

Figure 10: Indicators on the quantity dimension "the publication and dissemination of research results"



Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

Scientists are not only involved as producers in the publishing process but also as reviewers, editors and advisers. The role of reviewer in particular has been hotly debated since the advent of the new, computer-based forms of publication. An alternative to assessing the scientists' affinity for, and experience with, electronic media could be to ask about their experiences with reviewing electronic publications. We should be aware that we have already elaborated on questions concerning the perception of electronic publications from the reader's and the author's perspective. Consequently the willingness to respond may be jeopardised if another set of questions from the reviewer's perspective is included in the same questionnaire. However, we have formulated a possible indicator that takes this reviewer perspective and this is included in the annex (indicator (18)).

The publishers of scientific publications are the second source for assessing the importance of electronic publications for disseminating scientific knowledge. They should be in possession of figures which document how their publications develop on a country level and, above all, what amount of electronic content is produced and sold. However, we have to be aware that these data are proprietary data and that some, such as the access to e-journals and other electronic media, may be considered as sensitive information. Not every publisher may be willing to provide it on a voluntary basis fearing that libraries may cancel journal subscriptions due to low usage rates. Furthermore, a common understanding of usable data for electronic resources must still be developed.[265] Another problem could be the large effort necessary to collect meaningful data from the publishers: it seems to be very difficult to draw a representative sample and a complete survey could only be handled by the national statistical offices. For these reasons we do not take publishers into consideration as a potential and easily accessible source for indicators and data within this report.

[265] See Luther, J. (2000): White Paper on Electronic Journal Usage Statistics, in: Journal of Electronic Publishing, vol. 6, no. 3. (http://www.press.umich.edu/jep/06-03/luther.html).

Impact indicators on the publication and dissemination of research results

The effects of the electronic dissemination of research output lie primarily in a larger or smaller readership and the reduced delay between submitting research results and having them published.

Both an increase or a decrease in the readership are theoretically possible and the effect realised depends on many factors, such as the reading and publication behaviour in a discipline, the reputation of electronic and non-electronic media etc. It is possible to ask for the perceived impact of electronic publications compared to other publications within a survey. Such impact may include the number of citations or received e-mails after publication. Yet it has been shown that responses to such questions might not be reliable if the scientists lack the information necessary to make valid comparisons.[266] Another option for impact assessment would be to evaluate the inclusion of e-publications in CVs or their acceptance in tenure deliberations. However, the survey cited above has returned no variance at all for the two latter indicators. Also, due to differing university and career systems in Europe, tenure information is not available on a comparable basis.

While in many cases it may not be of importance whether research results are published after 60 or 180 days, as they are not time-sensitive or as drafts and preliminary versions were distributed and presented at conferences beforehand, in some cases it may be crucial: e.g. when parallel research has been carried out and only the first receives the merit and obtains access to revenues from property rights. Nevertheless, a faster publication process might speed up the research process in general, as peers who do not belong to a "virtual college", the group of researchers that informally exchange research results on a certain topic area, might gain the opportunity to provide comments and perform control analyses earlier.

We include an indicator (see indicator (19)) on the effects of electronic publishing among the SIBIS indicators. However, due to the problems mentioned we do not give it top priority.

11.3 Computer networks and R&D collaborations

Overview of indicators on computer networks and R&D collaborations

It has been shown that R&D collaborations in general and international R&D collaborations in particular have increased over the last few years (see section 2.3 of deliverable 2.1). The diffusion of the new information and communication technologies is assumed to be both a result of the increased demand for collaboration (and communication) and one of the causes for the growth of collaborative R&D. It is unquestionable that the Internet has lowered collaboration costs through improving the processing of information and communication. Various concepts of R&D collaborations exist, which differentiate collaborations depending e.g. on the participants' status (public-private versus private-private), the degree of formalisation (formal versus informal) or the position along the production chain (horizontal versus vertical). A new concept of R&D collaborations has recently appeared. It singles out collaborations which are supposed to be critically dependent on ICT such as so-called collaboratories or virtual teams. The European Commission has put special emphasis on the promotion of these new forms of collaboration, including the creation of "virtual centres of excellence", among the objectives laid down in its ERA communication.[267] The Commission

[266] See Anderson, K.; Sack, J.; Krauss, L.; O'Keefe, L. (2001): Publishing Online-Only Peer-Reviewed Biomedical Literature: Three Years of Citation, Author Perception, and Usage Experience, in: Journal of Electronic Publishing, vol. 6, no. 3 (http://www.press.umich.edu/jep/06-03/anderson.html).

[267] See European Commission (2000a), op. cit., p. 10.

also decided to support the development and deployment of grid technologies, which are basically collaboration technologies, as part of its *e*Europe initiative.[268]

As we have also shown in deliverable 2.1 (pp. 49-52), valid data on research collaborations are scarce and often not sufficiently sophisticated to properly evaluate even the older and well-developed concepts. Therefore it would add value if the SIBIS project managed to develop indicators for collaborative R&D. Of special interest should be indicators which assess the impact of the Internet on R&D collaborations and the spread of new, network-based forms of collaboration. This is certainly not an easy undertaking, as the definitions of collaboratories and virtual teams are difficult to operationalise, and it is even more difficult to separate them from traditional forms of R&D collaboration.

Table 7: Indicators on R&D collaborations

No.	Construct	Name of indicator	Dimension	Suggested method
(20)	R&D collaboration	Participation in long-distance R&D collaborations	Quantity	DMS/Survey of resear-chers
(21)	R&D collaboration	Impact of computer networks on R&D collaborations (on communication, data & information transfer, project management and duration, quality of work, productivity)	Impact	DMS/Survey of resear-chers

Source: FHSO compilation

The quantity dimension "indicators on computer networks and R&D collaborations"

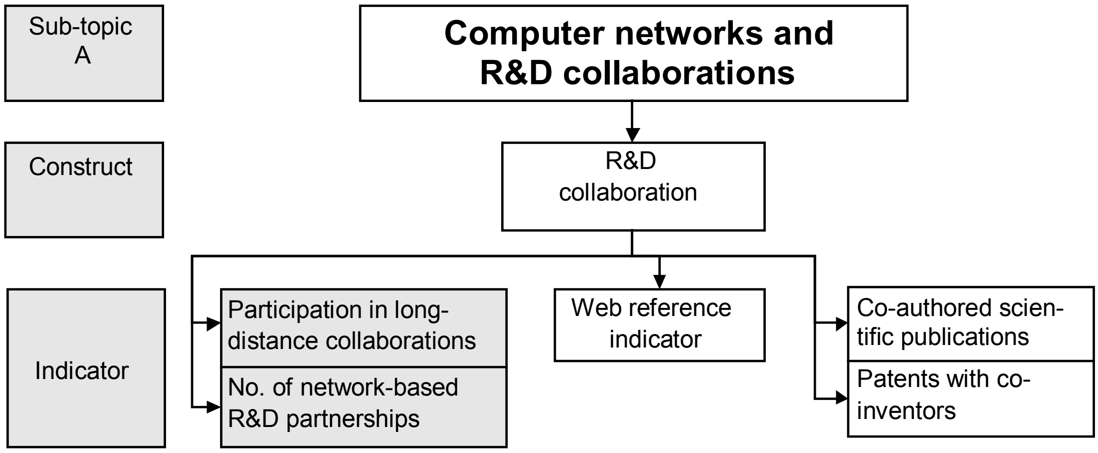
Figure 11 shows how the quantity dimension "R&D collaborations" could be measured.

One possible approach to quantify the extent to which private enterprises are involved in R&D collaborations and the role that computer networks play for these collaborations is to ask decision makers in the field of R&D (heads of R&D departments, in smaller firms the chief executives). This sort of question has been posed in innovation-related surveys and it has generally produced useful results.[269] We include two questions for a decision maker survey among the R&D collaboration indicators. The first, indicator (20) (see annex), indicates the amount of collaboration and serves as a baseline for the second indicator (21) which covers both the quantity and the impact dimensions of ICT for R&D collaborations. In principle both questions are also suitable for a survey in the public R&D sector, i.e. universities and research institutes. The list of answers to the question of indicator (20) would have to be reduced and slightly changed; an example is also included in the table.

[268] See European Commission (2000c): *e*Europe 2002: An Information Society for all. Action plan prepared by the Council and the European Commission for the Feira European Council, 19 and 20 June 2000, p. 9.

[269] See Foyn, F. (2000): Community Innovation Survey 1997/98 - final results. Eurostat (Ed.): Statistics in Focus: Theme 9 Research and Development, no. 2/2000.

Figure 11: Indicators on the quantity dimension “computer networks in R&D”



Note: Only the indicators in the grey boxes will be used as SIBIS indicators.

Source: FHSO compilation.

An alternative method of quantifying the number of R&D partnerships would be to collect collaboration-related information from the web pages of organisations which are part of the national R&D systems. This method has been used to evaluate the relations between universities, firms and governmental institutions (Triple Helix).[270] To assess the Triple Helix relations, complex queries were launched and those web pages which contained relevant references within their links or texts were retrieved. The searches were carried out with the AltaVista search engine. The cited study concluded that it was possible to measure Triple Helix relations via the Internet and that the WWW can be considered as a less codified counterpart of traditional scientometric sources (e.g. scientific journals). However, for a number of reasons this approach doesn’t seem to be feasible for the purpose of our analysis: firstly, even if we could retrieve all the references to other agents within the national R&D systems and outside of them from a website, these could hardly be classified a priori as indications of R&D collaborations. They might simply constitute a service to the reader pointing to other interesting sites on a certain topic such as "link-lists" which usually serve this purpose. Research institutions frequently also carry out other tasks (such as higher education) and the web content consequently does not only reflect R&D activities. We do not know in advance if a reference belongs to research or other functions; for example bibliographies or literature lists for university courses also increasingly contain hyperlinks. Second, there are some technical problems which would render the results of an empirical analysis at least uncertain: the coverage of search engines is limited, none of them manages to cover the entire WWW (see deliverable 2.1, p. 23). Therefore we could only analyse a segment of an R&D system without knowing how representative this segment is. Additionally, the results of searches by search engines were found to be inconsistent[271] and often not up-to-date.[272] The only way to solve these drawbacks would be to manually check each retrieved web page to ascertain that it contains valid and up-to-date information on an R&D collaboration. This might be feasible for small institutions but it certainly is not for entire national R&D systems. Under these circumstances we do not consider it appropriate to pursue further the idea of assessing R&D collaborations via a website analysis.

---

[270] See Leydesdorff, L.; Curran, M. (2000): Mapping university-industry-government relations on the Internet: the construction of indicators for a knowledge-based economy, in: Cybermetrics, vol. 4, no. 1, paper 2 (http://www.cindoc.csic.es/ cybermetrics/articles/v4i1p2.html).

[271] See Rousseau, R. (1999): Daily time series of common single word searches in AltaVista and NorthernLight, in: Cybermetrics, vol. 2/3, paper 2 (http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html). – Snyder, H.; Rosenbaum, H. (1999): Can search engines be used as tools for web-link analysis? A critical view, in: Journal of Documentation, vol. 55, no. 4, pp. 375-384. – Thelwall, M. (2000): Web impact factors and search engine coverage, in: Journal of Documentation, vol. 56, no. 2, pp. 185-189. – Thelwall, M. (2001): The Responsiveness of Search Engine Indexes, in: Cybermetrics, vol. 5, paper 1 (http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html).

[272] See Lawrence, S.; Giles, C. L. (1999): Accessibility of information on the web, in: Nature, vol. 400, 8 July 1999, pp. 107-109.

A third alternative to assess R&D collaborations has also been pointed out in deliverable 2.1: the analysis of data which contain information on the output of R&D activities such as co-authored scientific articles or co-invented patents. These data are generally available at the country level, but they are not sufficiently detailed to measure the impact of the Internet on these collaborations. Collaborations which have become possible due to the Net and others which hardly use it might be equally likely to result in a co-authored research paper. Therefore, we also omit these output indicators from the list of SIBIS indicators. They may come into play again for validating SIBIS research results at a later point in time.

The impact dimension "indicators on computer networks and R&D collaborations"

The impact of computer networks on R&D collaborations can again be assessed in physical and in monetary units. The direct physical impact of computer networks on R&D collaborations are related to the amount and characteristics of communication, interaction, data and information transfer among the partners. Overall communication may increase leading to the improved co-ordination of R&D projects (see indicator (21)). ICT may also replace personal communication, e.g. if partners "meet" at a video conference, and thereby reduce travel expenses. The direct effects may additionally trigger a range of indirect effects: more or improved communication and information exchange can enhance the management of R&D projects, reduce the project duration, improve the quality of collaborative output and the productivity of R&D partnerships. All these physical effects undoubtedly affect the costs and returns of collaborative R&D projects and contribute to their financial performance.

While the costs can be calculated using the budgets of collaborative R&D projects, it is extremely difficult to calculate a monetary equivalent of the benefits achieved or an additional return on investment for computer network use within collaborative R&D. The output of collaborative (as well as non-collaborative) R&D projects consists of publications, conference presentations, patents etc. Some output forms such as patents may have an immediate monetary value, others such as publications and conference presentations usually do not. But even when the existence of a monetary value can be ascertained, it is almost impossible to calculate a priori its exact size: patents may lead to innovations which produce returns on not pre-determined markets for an indeterminate period of time. Hence, we will not try to carry out an assessment of the monetary effects of computer networks on R&D collaborations.

12 Summary of Part B and conclusions

Part B of this report continues the development work to produce an indicator system to measure Internet usage in R&D as well as its effects. It synthesises the findings from deliverable 2.1 (Part A) and develops proposals for further indicators.

The indicator system that was developed covers constructs (conceptions of reality with varying degrees of abstraction and latency) in all three sub-topics of the "Internet for R&D" and in the two dimensions "quantity" and "impact" (see table 8). The largest number of indicators has been developed for the different sub-topics of the Internet-related ICT infrastructure for research activities. However, the figures reproduced in table 8 understate the amount of data which the indicator system contains: most indicators, especially those in the dimension "impact", provide more than one item of information; e.g. an indicator on the effects of information retrieval from and via the Internet includes as response categories: time budgets, contacts, productivity and quality of work results.

Table 8: Overview of indicators on the Internet for R&D

	Dimension		**Suggested method of data collection**		
	Quantity	Impact	Decision maker survey	Survey of researchers	Survey of ancillary organisations
Internet-related ICT infrastructure for R&D activities	11	3	6	5	2
Integration of the new network technologies into R&D activities	5	2	-	7	-
Computer networks and R&D collaborations	1	1	2	(2)	-

Source: FHSO compilation.

The quantity dimension "Internet-related ICT infrastructure for R&D activities" contains two indicators on R&D expenditure on ICT, five indicators on computer skills available for R&D, three on on-line information sources and one on computer-mediated communication. Indicators on research networks have been excluded from this report as an expert group is currently developing new indicators and planning a survey for early 2002 (see section 11.1.2). Grid technologies, a new and important application which uses the Internet for R&D, have not been included, as they are too new and immature to make a stable and reliable measurement possible.

Regarding the "integration of Internet technologies into R&D activities", it was possible to construct some accessible indicators on information retrieval from the Net and the publication and dissemination of research results via the Internet.

Only a few separate indicators on the importance and effects of "computer networks and R&D collaborations" were included. This imbalance is partly due to general difficulties in operationalising and evaluating specific types of R&D collaborations. Nevertheless, the two indicators developed provide information on both the quantitative importance of new, network-based forms of R&D collaboration and their impact on the input and output of collaborative R&D.

This report has not discussed in detail the options for gathering empirical data as this would have overburdened the text. Additionally there is a relationship between indicators and empirical methods, and usually different forms of an indicator may be assessed by means of different techniques. However, we ruled out some of the available methods for the purpose of this study as their level of methodological development has not advanced sufficiently to be able to use them effectively for answering the kind of macro-analytical questions we pose. This applies to web content analysis and log file analysis. We also dispensed with bibliometrics, as the empirical material obtained by a bibliometrical analysis is not sufficiently detailed to account for Internet issues (e.g. whether the on-line or printed version of a journal was cited, whether and to what extent a research collaboration that produced a co-authored article used the Internet). The proposed indicators rely exclusively on surveys as the method of choice, with different categories of respondents as target groups.

It is appropriate to ask a somewhat small number of questions related to indicators in a decision maker survey as is planned by SIBIS (see table 8 above). However, questions on R&D expenditure on ICT, computer staff available for R&D activities, and R&D collaborations must be addressed to a certain group of decision makers in order to obtain useful responses. For staff and expenditure this group should be the heads of IT departments, for R&D collaborations probably only the heads of R&D departments can provide the requested information (within university institutes it could also be the individual researcher).

The largest number of indicators is appropriate for a survey of researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants). For some questions it seems advisable to restrict the sample to academic researchers, as their information disclosing behaviour should be more autonomous and less influenced by information policies of a higher hierarchical level. This applies for example to the content of a researcher’s web presentation. A pilot survey of researchers might take this into account and limit the sample to researchers from the academic sector. The indicators suitable for this survey do not prejudice the usage of a traditional survey form (personal, phone, written) over a novel, Internet-based form (e-mailed or on-line questionnaire). However, as one of the targets of this survey would be to assess the utilisation of various Internet tools, an on-line only survey might suffer from a sample selection bias. Hence, a combination of on- and off-line methods seems most appropriate.

A third group of indicators can only be assessed if ancillary service organisations are willing to contribute information on the (computer-based) services they provide to R&D. Important organisations in this field are the national research networks and the providers of electronic library services. However, instead of carrying out our own pilot surveys, it would be a lot more effective to join forces with organisations in these fields which carry out indicator development and data collection as these belong to their regular tasks.

The next stage of research on the Internet for R&D should include the following:

- The indicators developed within this report have been transformed into survey questions in a rapid, prototyping manner. These survey questions must be discussed further with topic and methodological experts, finalised and assembled into questionnaires.
- The different survey alternatives ("traditional" vs. on-line) must be evaluated and selection of the appropriate method and sampling must be made.
- Compilation of address files
- Surveys

The work on the topic “the Internet for R&D” will continue with these tasks and will document and present the results within the following SIBIS deliverables.

13 Bibliography for Part B

Anderson, K.; Sack, J.; Krauss, L.; O'Keefe, L. (2001): Publishing Online-Only Peer-Reviewed Biomedical Literature: Three Years of Citation, Author Perception, and Usage Experience, in: Journal of Electronic Publishing, vol. 6, no. 3 (http://www.press.umich.edu/jep/06-03/anderson.html).

Aubert, J.E.; Bayar, V. (1999): Maximising the Benefits of Information Technology for Science: Overview and Major Issues, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, pp. 7-28.

Björk, B.-C.; Turk, Z. (2000): How scientists retrieve publications: an empirical study of how the Internet is overtaking paper media, in: The Journal of Electronic Publishing, vol. 6, no. 2. (http://www.press.umich.edu/jep/06-02/bjork.html).

Bourguignon, J.-P.; European Mathematical Society (1999): The future of mathematical databases, in: STI Review No. 24: Special Issue on "The Global Research Village". Paris, pp. 107-121.

Brophy, P.; Clarke, Z.; Brinkley, M.; Mundt, S.; Poll, R. (2000): EQUINOX - Library Performance Measurement and Quality Management System: Performance Indicators for Electronic Library Services. (http://equinox.dcu.ie/reports/pilist.html).

Brynjolfsson, E.; Hitt, L. M. (2000): Beyond Computation: Information technology, organizational transformation and business performance, in: Journal of Economic Perspectives, vol. 14, no. 4, pp. 23-48.

Chen, C.; Newman, J.; Newman, R.; Rada, R. (1998): How did university departments interweave the Web: A study of connectivity and underlying factors, in: Interacting with computers, vol. 10, pp. 353-373.

Cohen, J. (1996): Computer mediated communication and publication productivity among faculty, in: Internet Research: Electronic Networking Applications and Policy, vol. 6, no. 2/3, pp. 41-63.

Dusoulier, N. (1994): Databases and data banks, in: Ines Wesley-Tanaskovic, Jacques Tocatlian, and Kenneth H. Roberts (eds): Expanding access to science and technology. The role of information technology. Proceedings of the Second International Symposium on the Frontiers of Science and Technology Held in Kyoto, Japan, 12-14 May 1992. (available at: http://www.unu.edu/unupress/unupbooks/uu07ee/uu07ee0a.htm).

European Commission (2000a): Towards a European research area. Communication from the Commission COM 2000 (6). Brussels, 18 January 2000. (http://europa.eu.int/comm/research/ area/com2000-6-en.pdf).

European Commission (2000b): Making a reality of the European research area: guidelines for EU research activities (2002-2006). Communication from the European Commission COM(2000) 612 final. Brussels, 4 October 2000. (Http://europa.eu.int/comm/research/area/ com2000-612-en.pdf).

European Commission (2000c): *e*Europe 2002: An Information Society for all. Action plan prepared by the Council and the European Commission for the Feira European Council, 19 and 20 June 2000.

Foray, D. (1999): Building the Virtual "House of Salomon": Digital collaboration technologies, the organisation of scientific work and the economics of knowledge access. Report of the ESF-IIASA-NSF Workshop - 3 to 5 December 1999 - at the International Institute for Applied System Analysis, Laxenburg, Austria, p. 9. (http://www.esf.org/policy/pdf/iiasa.pdf)

Foyn, F. (2000): Community Innovation Survey 1997/98 - final results. Eurostat (Ed.): Statistics in Focus: Theme 9 Research and Development, no. 2/2000.

Harman, H.H. (1976): Modern factor analysis, 3rd edition, Chicago.

Harter, S. P. (1996): The Impact of Electronic Journals on Scholarly Communication: A Citation Analysis, in: The Public-Access Computer Systems Review, vol. 7, no. 5. (http://info.lib.uh.edu/pr/v7/n5/hart7n5.html)

Hesse, B. W.; Sproull, L. S.; Kiesler, S. B.; Walsh, J. P. (1993): Returns to science: Computer networks in oceanography, in: Communications of the ACM, vol. 36, no. 8, pp. 90-101.

Ingwersen, P. (1998): The calculation of web impact factors, in: Journal of Documentation, vol. 54, no. 2, pp. 236-243.

International Coalition of Library Consortia (1998): Guidelines for Statistical Measures of Usage of Web-based Indexed, Abstracted, and Full-Text Resources. (http://www.library.yale.edu/consortia/webstats.html).

Lawrence, S.; Giles, C. L. (1999): Accessibility of information on the web, in: Nature, vol. 400, 8 July 1999, pp. 107-109.

Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997): Internet use by faculty members in various disciplines: a comparative case study, in: Journal of the American Society for Information Science, vol. 48, no. 6, pp. 508-518.

Lubanski, A.; Matthew, L. (1998). Socio-economic Impact of the Internet in the Academic Research Environment. Proceedings IRISS '98 International Conference: 25-27 March 1998, Bristol, UK. (http://sosig.ac.uk/iriss/papers/paper18.htm).

Luther, J. (2000): White Paper on Electronic Journal Usage Statistics, in: Journal of Electronic Publishing, vol. 6, no. 3. (http://www.press.umich.edu/jep/06-03/luther.html).

McClure, C. R.; Lopata, C. (1996): Assessing the Academic Networked Environment: Strategies and Options. (http://istweb.syr.edu/~mcclure/network/toc.html).

National Science Board (2000): Science & engineering indicators 2000. Arlington, Va.

Nunnally, J. S.; Bernstein, I. H. (1994): Psychometric Theory. 3rd edition, New York et al.

OECD (1994): The Measurement of Scientific and Technological Activities: Proposed Standard Practice for Surveys of Research and Experimental Development - Frascati Manual 1993. 5th Edition. Paris.

Odlyzko, A. (2001): The rapid evolution of scholarly communication. (http://www.dtc.umn.edu/~odlyzko/doc/rapid.evolution.pdf).

Orlikowski, W. J.; Iacono, C. S. (2000): The truth is not out there: an enacted view of the “Digital Economy”, in: Brynjolfsson, E. and Kahin, B. (eds): Understanding the digital economy. Data, tools, and research. Cambridge and London, pp. 352-380.

Rousseau, R. (1999): Daily time series of common single word searches in AltaVista and NorthernLight, in: Cybermetrics, vol. 2/3, paper 2 (http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html).

Shim, W.; McClure, C.R.; Bertot, J.C.; Sweet, J.T.; Maffré de Lastens, J.-M.; Dagli, A.; Fraser, B.T. (2000): ARL E-metrics project: developing statistics and performance measures to describe electronic information services and resources for ARL libraries. (http://www.arl.org/stats/newmeas/emetrics/phaseone.pdf).

Snyder, H.; Rosenbaum, H. (1999): Can search engines be used as tools for web-link analysis? A critical view, in: Journal of Documentation, vol. 55, no. 4, pp. 375-384.

Tenopir, C.; King, D.W. (2000): Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers. Washington, D.C.: Special Libraries Association.

Thelwall, M. (2000): Web impact factors and search engine coverage, in: Journal of Documentation, vol. 56, no. 2, pp. 185-189.

Thelwall, M. (2001): The Responsiveness of Search Engine Indexes, in: Cybermetrics, vol. 5, paper 1 (http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html).

Vedung, E. (1999): Evaluation im öffentlichen Sektor. Wien, Köln and Graz.

Walsh, J.P.; Maloney, N. G. (2001): Computer network use, collaboration structures and productivity, in: P. Hinds and S. Kiesler (ed.): Distributed work. Cambridge, Mass.

Weiss, C. H. (1998): Evaluation. 2nd edition, Upper Saddle River.

14 Annex

14.1 Indicator tables

Name of indicator	**(1): R&D expenditure on ICT**
Definition	R&D expenditure on ICT as a fraction of total intramural R&D *expenditure*
Notes	Definitions of R&D and ICT expenditure must be given. The OECD includes among R&D expenditure the labour costs of R&D personnel and other current and capital costs for R&D equipment. Valued Added Taxes and the depreciation of assets are *excluded*. Expenditure can be classified as ICT-related when it is for • hardware and software licences • internal support staff • external computer services (including hard- and software-related support)
Methods needed	Decision maker survey
SIBIS survey: Q and group to be asked	**Target group: DMS** (chief executives, heads of R&D departments or ICT departments) *Q1: What was the total size of R&D expenditure of your organisation last year?* *Q2: What percentage of the R&D expenditure was for ICT* An additional question may be added for further clarity: *(Q3: What percentage of R&D expenditure was Internet-related?)*
Indicator sources	FHSO compilation
Policy relevance	The indicator gives a general picture of the extent to which R&D activities use ICT. It is suited to an overview and provides data that can be generalised more easily than specific indicators on individual parts and tools of the ICT infrastructure.

Source: FHSO compilation.

Name of indicator	**(2): R&D expenditure on different types of ICT**
Definition	R&D expenditure on different types of ICT as a percentage of total R&D expenditure on ICT
Notes	ICT equipment expenditure is expenditure on hardware and software licences. ICT labour expenditure is for internal support staff, while external services expenditure is for external computer support services (e.g. external mainframe computing services).
Methods needed	Decision maker survey
SIBIS survey: Q and group to be asked	**Target group: DMS** (chief executives, heads of R&D departments or ICT departments) *Q1: What percentage of total R&D expenditure in 2001 was for ICT equipment, labour and external services?* An additional question may be added for further clarity: *(Q2: What percentage of R&D expenditure in 2001 was for Internet-related equipment, labour and external services?)*
Indicator sources	FHSO compilation.
Policy relevance	The indicator provides more detailed information on how R&D expenditure on ICT is structured. It is suited to analysing whether particular types of expenditure (equipment, staff, services) lead to a greater R&D success rate (when information on this is also obtained).

Source: FHSO compilation.

Name of indicator	(3): Number of titles in digital collections
Definition	Number of titles in digital collections per target population/researchers
Notes	It seems advisable to distinguish between different types of titles and to define them properly in order to make sure that comparable categories are collected. Also the terms "target population" and "researchers" must be defined. Some definitions can be found in the literature: “Electronic library resources: every document in electronic form which needs special equipment to be used. [NOTE: electronic resources include digital documents, electronic serials, databases, patents in electronic form and networked audio-visual documents.] ISO/DIS 2789” (http://equinox.dcu.ie/reports/pilist.html#pis) “Target population: groups of actual and potential users appropriate to an individual library as the object of a specific service or as the primary users of specific materials. ISO 11620 [NOTE: the target population may be the population to be served by the library, a specific group within that population, or some other group that the library is aiming to serve. The target population must be defined by the library in each instance and carefully recorded to facilitate benchmarking.]” (http://equinox.dcu.ie/reports/pilist.html#pis)
Methods needed	Survey
SIBIS survey: Q and group to be asked	**Target group:** heads of digital collection providers *Q0 (filter): Do you provide access to your collections by means of the Internet or other computer networks?* *Q1: How many separate titles did you provide for electronic access on 31 December 2001? How is the access possible? (the question should be put in a table, answering categories for the second part must be explored)* • Books (including networked CD-ROMs) • Journals • Video resources • Audio resources • Other (graphics, maps etc.) • Total *Q2: What was the size of your target population on 31 December 2001? (If there are different target populations for different sources please try to provide an average figure that is valid for the entire service).*
Indicator sources	Shim, W. et al. (2000), op. cit.
Policy relevance	The indicator serves to measure an important aspect of the Internet-related infrastructure available for research activities: the more resources available on-line, the larger the reduction in transaction costs and the greater the benefit for R&D. A short supply of on-line resources should consequently lead to compensatory policy measures.

Source: FHSO compilation.

Name of indicator	(4): Staff providing electronic library services
Definition	Staff of digital libraries and external staff providing electronic library services in relation to the target population.
Notes	A definition of "electronic library services (ELS)" and the "staff providing ELS" should be included in a questionnaire to ensure that the assembled data are comparable. Definitions can be found in the literature: ELS: “A service which is either supplied from local servers or accessible via networks. [NOTE: electronic library services comprise the OPAC, the library website, electronic resources, electronic document delivery and internet access offered via the library.] ISO/DIS 2789” (http://equinox.dcu.ie/reports/pilist.html#pis) Staff providing ELS: number of library staff providing, maintaining and developing ELS and training users (http://equinox.dcu.ie/reports/method.html#Appendix%201)
Methods needed	Survey
SIBIS survey: Q and group to be asked	**Target group:** heads of digital collection providers *Q1: How many members of staff in your organisation were responsible for providing digital library services on 31 December 2001? (in full-time equivalents).* *Q2: How many additional members of staff in your organisation would be necessary to substitute for the services related to the digital collections which were obtained from external service providers on 31 December 2001? (in full-time equivalents).* Additional questions necessary for calculating the indicator: *Q3, Q4: Q0 and Q2 of indicator (3)*
Indicator sources	Brophy, P. et al. (2000), op. cit.; Shim, W. et al. (2000), op. cit.
Policy relevance	The indicator serves to measure an important aspect of the Internet-related infrastructure available for research activities: understaffing may limit the benefits of ELS if it results in insufficient consulting with, or training of, the users.

Source: FHSO compilation.

Name of indicator	(5): Information displayed on a researcher's web page(s)
Definition	Information displayed on a researcher's web page(s)
Notes	The incentives to disseminate and communicate research expertise and results via the Net differ between the academic and the private business sector. To avoid a bias in the sample, the two groups should be distinguished.
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q0 (filter): Do you have an individual WWW presentation of your professional activities and expertise?* *Q1: What does your WWW presentation include?* • Biographical information • Description or listing of the fields of interest and expertise • Past or current research projects • Publication list, list of patents or other documentation of research output • Working papers, full text articles, other forms of research output or hyperlinks that point to any of the former • Addresses of, or hyperlinks to, collaborators • Other (please specify)
Indicator sources	FHSO Data to be collected in a pilot survey of FHSO.
Policy relevance	Indicator on the importance of the Internet as a media for information dissemination in R&D. In the short-term it may also be a valuable indicator on the awareness of researchers on the communicative possibilities of the Internet.

Source: FHSO compilation.

Name of indicator	**(6): Effects of researchers' web page(s)**
Definition	Effects of researchers' web page(s) per researcher
Notes	-
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q1: Please indicate the extent to which you agree with the following statements on the effects of your WWW presentation? [agree/disagree on a five-point scale]* • To establish and maintain my Web presentation requires a lot of effort and work time. • The Web presentation has increased the amount of inquiries and mail I receive. • The Web presentation does not particularly benefit my research. • The Web presentation has increased the visibility of my work in general. • The Web presentation has increased the visibility of my work among peers and in my scientific community. • The Web presentation has led to new contacts and valuable input (e.g. information, collaborative partners) for my research. • Other (please specify) Additional questions necessary for calculating the indicator: *Q0: Q0 of indicator (5)*
Indicator sources	FHSO Data to be collected in a pilot survey of FHSO.
Policy relevance	Indicator on the effects of the WWW on a researcher's time, communicational activities, contacts and recognition.

Source: FHSO compilation.

Name of indicator	(7): E-mail communication for R&D purposes
Definition	E-mails sent and e-mails received for R&D purposes as a percentage of total social communication for R&D purposes (alternatively differentiated according to the user's position in his (her) organisation)
Notes	The question should be in table format and include various explanations.
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q0: Do you have access to e-mail services? (filter)* *Q1a: Please describe your communication pattern during an average day within the last month (if one instance of media use has various purposes, please try to differentiate it and give the breakdown):* • How many e-mails do you send/receive for the following purposes ... • How many phone calls do you make/get for the following purposes ... • How many letters or other written communication (fax, written messages etc.) do you send/receive ... • How many personal meetings do you attend ...(please try to be as comprehensive as possible and include formal and informal meetings) The list of purposes should include: teaching, R&D, general administration, mailing lists, private purposes, spam, others. An alternative to Q1a could be to assess the user's position within his (her) organisation to account for differences in the amount of overall communication: *Q1b: How many e-mails did you send and receive during an average week within the last three months for the following purposes ... ? (if one instance of media use has various purposes, please try to differentiate it and give the breakdown)* Additional question necessary when using Q1b: *Q2: What is your position in your organisation? – Head of organisation, head of sub-unit, researcher or research fellow, research assistant, technical staff, others*
Indicator sources	Lazinger, S. S.; Bar-Ilan, J.; Peritz, B. C. (1997), op. cit. Data to be collected in a pilot survey of FHSO.
Policy relevance	E-mail was one of the breakthrough applications of the Internet, as it offers significant advantages over some older forms of communication. The extent to which researchers use it may be considered as an indicator for the acceptance of new ICT in R&D which is one of the ERA goals.

Source: FHSO compilation.

Name of indicator	(8): Effects of e-mail use for R&D purposes
Definition	Effects of e-mail use for R&D purposes
Notes	-
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q1: How has e-mail affected your work? [agree/disagree on a five-point scale]* • It has increased the amount of information available. • It has facilitated collaborations or increased their number. • It has broadened my scientific network and led to more contacts. • It has improved my productivity (e.g. more publications, more patents etc.). • It has improved the quality of my work. Additional questions necessary for calculating the indicator: *Q0: Q0 of indicator (7)*
Indicator sources	Walsh, J.P.; Maloney, N. G. (2001), op. cit. – Cohen, J. (1996), op. cit. Data to be collected in a pilot survey of FHSO.
Policy relevance	E-mail was one of the breakthrough applications of the Internet, as it offers significant advantages over some older forms of communication. Showing the benefits of using e-mail may be a powerful argument to convince late adopters.

Source: FHSO compilation.

Name of indicator	**(9): Computer skills of R&D personnel**
Definition	Mean value of skills in various fields of computer use for R&D purposes
Notes	
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) Questions should be presented in table format. *Q0: Do you need the following type of computer programs for your work? (filter)* *Word processing:* • Yes, I need and I use them. • Yes, I need them but somebody else carries out the work for me. • Yes, I need them but I do not use them. • No, I do not need this type of program. *Q1: Please indicate your skill level for the following types of computer programs* *Word processing:* • I know how to enter and edit texts. • I know how to enter and edit long and complicated texts (with columns, sections), create and use styles, templates, tables and graphs. • I know how to carry out automated mailing and use macros in texts.
Indicator sources	Data to be collected in a pilot survey of FHSO.
Policy relevance	General computer skills indicator, relevant for OECD and EU (ERA) targets regarding the usage of computers within R&D.

Source: FHSO compilation.

Name of indicator	(10): Computer staff providing services to R&D
Definition	Computer staff providing services to R&D in relation to total R&D personnel/total personnel (in full-time equivalents at a reference date).
Notes	Definitions of "R&D personnel", "computer staff", "computer services" and "R&D personnel/staff" must be included in a questionnaire.
Methods needed	Survey
SIBIS survey: Q and group to be asked	Target group: **DMS** (chief executives, heads of R&D departments or ICT departments) *Q0 (filter): Do you carry out R&D in your organisation?* *Q1: How do you secure the specialised computer staff needed for R&D activities?* *Answering options (in employees, full-time equivalents):* • We have computer staff in R&D unit(s). • We have computer staff in IT departments who provide services to R&D. • We buy IT services for R&D from external service providers. Additional questions necessary for calculating the indicator: *Q2: What is the total R&D staff size of your organisation? (in full-time equivalents)*
Indicator sources	SIBIS survey
Policy relevance	General indicator for "Internet for R&D". Indicates the extent to which R&D projects have access to the skills needed to use ICT efficiently. Provides additional information on the percentage of computer services obtained from external sources, possibly an indicator for evaluating the efficiency of make-or-buy decisions in respect to computing services in R&D.

Source: FHSO compilation.

Name of indicator	**(11): Unfilled vacancies in private businesses for computer staff providing services to R&D**
Definition	Unfilled vacancies in private businesses for computer staff providing services to R&D in relation to total computer staff providing services to R&D/total personnel (in full-time equivalents at a reference date).
Notes	Definitions of "computer staff" and "computer services" must be included in a questionnaire. The indicator is only useful if indicator (10) is also obtained.
Methods needed	Survey
SIBIS survey: Q and group to be asked	Target group: **DMS** (chief executives, heads of R&D departments or ICT departments) *Q0 (filter): Did you have unfilled vacancies for computer staff that should provide services to R&D projects on 31 December 2001?* *Q1: Please specify the number of unfilled vacancies for computer staff needed to provide services for R&D projects in your reporting unit (in full-time equivalents on 31 December 2001).* Additional questions necessary for calculating the indicator: *Q2, 3: Q1 and Q2 of indicator (10)*
Indicator sources	SIBIS survey
Policy relevance	General indicator for “Internet for R&D”. Specifies computer personnel shortages which might adversely affect R&D projects.

Source: FHSO compilation.

Name of indicator	(12): Effects of computer skills on R&D
Definition	Effects of computer skills on R&D
Notes	-
Methods needed	Survey
SIBIS survey: Q and group to be asked	**Target group:** researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q1: With the following question we want to find out how computer skills affect your R&D activities. Please indicate the extent to which you agree with the following statements on the effects of computer skills in R&D. [agree/disagree on a five-point scale]* • The computer skills available for my work (my own, of specialised staff) have increased productivity. • The available computer skills have improved the quality of my research results. • The available computer skills have led to new forms of research and new types of project which would not be feasible without these skills. • A lot of time and money was required to acquire the computer skills necessary for carrying out my research projects.
Indicator sources	SIBIS survey
Policy relevance	General indicator for "Internet for R&D". Specifies how computer skills affect R&D and how a skills shortage might adversely affect R&D projects.

Source: FHSO compilation.

Name of indicator	(13): Frequency of information retrieval from electronic sources
Definition	Frequency of information retrieval from electronic sources compared to information retrieval from other sources per researcher
Notes	The term "information" should be defined to make sure that the differences between "data" and "information" are respected.
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q1: Do you have access to electronic sources for retrieving information? (yes, no, I don't know)* *Q2: How frequently do you typically use one of the following sources for searching for and retrieving information for R&D purposes? [answering options: never, less than once a month, one or two times per month, one or two times per week, daily]* • Libraries • Local colleagues, assistants and superiors • Conferences • Specific Intranet sources (excluding the local library) • Websites of libraries • Electronic journals, working paper and article databases (excluding lists on peers' websites!) • Peers' websites • Other Internet sources (professional associations, catalogues, dictionaries etc.) • Other (please specify)
Indicator sources	Lubanski, A.; Matthew, L. (1998), op. cit. – International Coalition of Library Consortia (1998), op. cit. – Brophy, P. et al. (2000) op. cit. – Shim, W. et al. (2000), op. cit. – Björk, B.-C.; Turk, Z. (2000), op. cit. Data to be collected in a pilot survey of FHSO.
Policy relevance	Potential indicator to assess the use of electronic networks (and information sources) which should be encouraged according to the new European Research Policy laid down in the ERA.

Source: FHSO compilation.

<table>
<tr><th>Name of indicator</th><th>(14): Documents/items from electronic sources</th></tr>
<tr><td>Definition</td><td>Number of documents/items from electronic sources in relation to documents/items from other sources</td></tr>
<tr><td>Notes</td><td>-</td></tr>
<tr><td>Methods needed</td><td>Survey of researchers</td></tr>
<tr><td>SIBIS survey: Q and group to be asked</td><td>Target group: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants)
Q1: How many documents/items do you typically examine per week which were originally obtained from one of the listed sources?
• Libraries (personally, by mail, but excluding electronic transmission)
• Electronic journals, working paper and article databases
• Electronic transmission from libraries (e.g. download from the website, received by e-mail)
• Electronic transmission from colleagues, assistants and superiors (e.g. download from the website, received by e-mail)
• Colleagues, assistants and superiors (personally, by mail, but excluding electronic transmission)
• Specific Intranet sources (excluding the local library)
• Other Internet sources (professional associations, catalogues, dictionaries etc.)
• Other (please specify)</td></tr>
<tr><td>Indicator sources</td><td>FHSO
Data to be collected in a pilot survey of FHSO.</td></tr>
<tr><td>Policy relevance</td><td>Potential indicator to assess the use of electronic networks (and information sources) which should be encouraged according to the new European Research Policy laid down in the ERA.</td></tr>
</table>

Source: FHSO compilation.

Name of indicator	**(15): Frequency of software usage**
Definition	Frequency of software usage for computer-based information search and retrieval compared to other methods per researcher
Notes	-
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q1: How frequently do you **yourself** typically use one of the following methods for searching for information for R&D purposes? [answering options: never, less than once a month, one or two times per month, one or two times per week, daily]* • I delegate information search and retrieval to assistants. • I ask colleagues, assistants, superiors or other people personally or over the phone. • I use an Internet search engine (e.g. AltaVista, Google etc.) or a commercial search software. • I browse various known and established sources (e.g. the "Favorites") on the computer. • I leaf through printed books, journals, working papers or other printed sources. • Other *Q2: How frequently do you **yourself** typically use one of the following methods for evaluating information for R&D purposes? [answering options see Q1]* • I delegate information evaluation to assistants. • I leaf through printed books, journals, working papers etc. or read their abstracts. • I read key words, abstracts or full texts on the computer monitor. • I listen to presentations at seminars, workshops or conferences. • Other *Q3: How frequently do you **yourself** typically use one of the following methods for storing information for subsequent closer examination? [answering options see Q1]* • I delegate information storing to assistants. • I buy printed material or order it from the library to have it delivered via regular mail. • I make copies of printed material. • I save electronic files which I received by mail, found on the Internet or other on-line information sources. • Other
Indicator sources	Lubanski, A.; Matthew, L. (1998), op. cit. – Lazinger, S.S.; Bar-Ilan, J.; Peritz, B.C. (1997), op. cit. - Kaminer, N.; Braunstein, Y.M. (1998), op. cit. Data to be collected in a pilot survey of FHSO.
Policy relevance	Potential indicator to assess the use of electronic networks (and information sources) which should be encouraged according to the new European Research Policy laid down in the ERA.

Name of indicator	(16): Effects of information retrieval from and via the Internet
Definition	Effects of information retrieval from and via the Internet
Notes	-
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q1: Please indicate the extent to which you agree with the following statements on the effects of Internet-based information sources. [agree/disagree on a five-point scale]* • I require less time to perform information searches due to the Internet. • The information available via the Internet has increased my productivity (e.g. measured as the amount of work I get done in a certain period of time). • The information available via the Internet has improved the quality of my work (e.g. as I obtain more information, more concepts and opinions etc.). • Information gathered on the Internet has helped me to establish new contacts for my research. • Other … Additional questions necessary for calculating the indicator: *Q0: Some sort of filter to determine whether the respondent uses the Internet seems advisable, the answering options from indicator (13) could possibly serve this purpose.*
Indicator sources	FHSO Data to be collected in a pilot survey of FHSO.
Policy relevance	Indicator on the effects of the Internet, interpreted as a source of information, on productivity and other relevant measures; potentially an indicator for assessing the usefulness of the available electronic information sources.

Source: FHSO compilation.

Name of indicator	(17): Amount of work published in electronic media
Definition	Amount of work published in electronic media as a percentage of total number of publications
Notes	The results of R&D might differ systematically between the academic and the private business sector. To avoid a bias in the sample, the two groups should be distinguished. The indicator also provides information on the R&D output of researchers which can be helpful for evaluating the effects of some Internet tools and the information search behaviour (e.g. by means of calculating performance indicators, regression analysis etc.).
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q0 (filter): Have you made the results of any of your work available to a larger readership within the last three years?* *Q1: Which media have you used and to what extent? (In case of "double use", e.g. in a journal and on the website, please make two entries) [the question should be in table format, answering options yes/no and frequency of media use]* • Scientific journals (excluding e-journals) • E-journals (on-line only) • Monographs • Chapters in books • Conference presentations • Patent publications • Numerical databases • Reports with limited distribution • Working and discussion papers • Postings on your website • Other (please specify)
Indicator sources	FHSO Data to be collected in a pilot survey of FHSO.
Policy relevance	Output indicator and indicator on the importance of electronic publishing forms and the Internet as a media for information dissemination in R&D. The ERA also aims to promote the production of multimedia content which is not a *new* medium but rather a combination of different media in one.

Source: FHSO compilation.

Name of indicator	(18): Review activities for e-journals
Definition	Review activities for e-journals as a percentage of total review activities
Notes	
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q0 (filter): Have you served as a reviewer (ex-ante or ex-post) to scientific publications within the last three years?* *Q1: For which media have you reviewed scientific publications? [the question should be in table format, answering options yes/no and frequency of reviews]* • Scientific journals (excluding e-journals) • E-journals (on-line only) • Monographs • Chapters in books • Conference presentations • Reports with limited distribution • Working and discussion papers • Other (please specify)
Indicator sources	FHSO Data to be collected in a pilot survey of FHSO.
Policy relevance	Output indicator and indicator on the importance of electronic publishing forms and the Internet as a media for information dissemination in R&D.

Source: FHSO compilation.

Name of indicator	(19): Impact of publications in electronic media
Definition	Perceived impact of publications in electronic media compared to their impact in traditional media
Notes	
Methods needed	Survey of researchers
SIBIS survey: Q and group to be asked	**Target group**: researchers on different hierarchical levels (heads of R&D departments and university institutes, professors, research assistants) *Q1: Please indicate the extent to which you agree with the following statements on experiences with publications in electronic media* ***compared to publications in traditional media****? [agree/disagree on a five-point scale]* • The publication was made available to its most important target group more quickly. • It triggered more feedback (requests, comments etc.). • It was cited more often. • Other (please specify)
Indicator sources	FHSO Data to be collected in a pilot survey of FHSO.
Policy relevance	Indicator on the effects of the Internet as a medium for disseminating knowledge.

Source: FHSO compilation.

Name of indicator	**(20): Participation in long-distance R&D collaborations**
Definition	Participation in long-distance R&D collaborations in relation to short-distance R&D collaborations/all R&D collaborations
Notes	Definitions of "R&D", "R&D collaborations" must be included in a questionnaire.
Methods needed	Decision Maker Survey and/or survey of academic researchers
SIBIS survey: Q and group to be asked	**Target group 1:** chief executives, heads of R&D departments *Q0 (filter): Is your company involved in R&D collaborations?* *Q1: Have you participated in R&D collaborations with any of the following collaborators? Please also specify their location (local, national, other EU countries, US, other countries worldwide).* • Affiliated companies • Competitors, firms from the same industry • Customers • Suppliers • Consultants, market researchers • Universities, other HEI and research institutes • Government institutions • Other **Target group 2:** heads of university departments and research institutes, researchers *Q0 (filter): Is your department/institute involved in R&D collaborations?* *Q1: Have you participated in R&D collaborations with any of the following collaborators? Please also specify their location (local, national, other EU countries, US, other countries worldwide).* • Private companies • Universities, other HEI and research institutes • Government institutions • Other
Indicator sources	SIBIS survey
Policy relevance	General indicator on the importance of different types of R&D collaborations with different types of partners. It is important to obtain a general picture of the collaboration activities and to calculate correlations with the more specific indicator (21). It could also be a possible variable to evaluate the effects of ICT on companies, e.g. as the endogenous variable in cross-correlations with ICT variables (investment, physical ICT indicators).

Source: FHSO compilation.

Name of indicator	**(21): Impact of computer networks on R&D collaborations**
Definition	Impact of computer networks on R&D collaborations
Notes	Definitions of "computer networks", "R&D", "R&D collaborations" must be included in a questionnaire.
Methods needed	Decision Maker Survey and/or survey of academic researchers
SIBIS survey: Q and group to be asked	Target group: chief executives, heads of R&D departments, researchers *Q1: Have computer networks affected the performance and management of your R&D collaborations? [agree/disagree on a five-point scale]* • They have made new R&D collaborations possible. • They have made new R&D collaborations possible that are reliant on them and could not function without them. • They have led to more long-distance collaboration. • They have had a positive effect on existing R&D collaborations. • They have led to more communication and interaction between the collaborators. • They have reduced the amount of personal communication among the collaborators. • They have increased data and information transfer among the collaborators. • They have improved the monitoring and evaluation of collaborative R&D projects. • They have accelerated collaborative R&D projects. • They have improved the quality of collaborative R&D output. • They have made collaborative R&D more productive (i.e. we get the same output with less effort and expense or we get more output with the same input). • They have reduced the costs of collaborative R&D projects. Additional questions necessary for calculating the indicator: *Q2, Q3: Q0, Q1 of indicator (20)*
Indicator sources	SIBIS survey
Policy relevance	The usage of electronic networks among researchers should be encouraged according to the new European Research Policy laid down in the ERA. The Commission also aims to support the development and the deployment of grid technologies which are basically collaboration technologies. The indicator might serve to evaluate the impact of computer network use on collaborations.

Source: FHSO compilation.