

THE
> PUBLIC
DOMAIN
OF
DIGITAL
RESEARCH
DATA

THE
> PUBLIC
DOMAIN
OF
DIGITAL
RESEARCH
DATA

Paul Wouters and Peter Schröder
Series Editors

Policies on Digital Research Data: An International Survey

June 2002

> Paul Wouters

Networked Research and Digital Information (Nerdi)
NIWI-KNAW

ACKNOWLEDGMENTS

This research project has been funded by the Ministry of Education, Culture and Sciences (OC&W). I would like to thank Peter Schröder, Jacky Bax and Emiel Broesterhuizen (Ministry OC&W), Paul Uhlir (NAS), Tony Mayer (ESF), Peter Arzberger (UCSD), Lisette Bros, Helga van Gelder, Gaspard de Jong (NIWI-KNAW), and Anne Beaulieu and Andrea Scharnhorst (Nerdi) for their comments on earlier drafts. Helga van Gelder helped collect the data. Repke de Vries (now at the Royal Library of the Netherlands) installed the software. I am indebted to Colin Reddy for his editorial assistance.

NIWI-KNAW, Joan Muyskenweg 25, PO Box 95110, 1090 HC Amsterdam, NL
email paul.wouters@niwi.knaw.nl

Copyright © 2002 NIWI-KNAW, Amsterdam
ISBN 90 6472 183 1

graphic design Magenta Ontwerpers, Amsterdam/Bussum

Contents

Preface by Enric Banda 7

Introduction 9

Part I

Policies on Data-Sharing. A Preliminary Assessment of the Current State of the Art
by an Email Survey 13

Part II

Access to and Sharing of Research Data – The Policy Context. A Web Scan of Principles
and Regulations in the US 25

Appendices

Part I

Appendix 1: The Questionnaire 47

Appendix 2: Response Rate to the Mini-Survey 51

Part II

Appendix 3: Methods 53

Appendix 4: Excerpts from the FOIA and the Bayh-Dole Act 55

Appendix 5: Members of the OECD Follow-Up Group on Issues of Access to Data
from Public Funding 57

Bibliography 63

Preface by Enric Banda

7

The importance of access to scientific data as one of the key tools in modern research cannot be over-emphasised. New ICT methodologies now allow both the construction and storage of very large amounts of data and their interrogation and use, often in real time and remotely. The European Science Foundation considers that the development of policies concerned with access to scientific databases must be based on international comparison and cooperation and the adoption of best practice. The acquisition of data, their storage and accessibility has become a very significant cost in research. At the same time, issues of trust in the science system have also become of concern.

The ESF welcomed the OECD initiative to address the issue of access to publicly financed data within the context of “Global Research Village” and was pleased to be able to act as a partner and thus provide a conduit for its Member Organisations (the national research agencies in 27 European countries) to be involved in this activity.

This report by Paul Wouters is a very important review of current policies and provides the essential building block for the refinement of international best practice for improving access to data and for encouraging and intensifying international research collaboration.

Enric Banda
Secretary General, ESF

Data sharing will become an issue
in all scientific and
scholarly fields.

Introduction

9

Providing access to scientific data is fast becoming a crucial aspect of science policy at the national and international level (National Research Council 1997). The need for increased levels of data processing are related to a number of developments: the application of information and communication technologies (ICT) in research; the development of new, often interdisciplinary, research questions; and the increased social and economic role of science, social science and the humanities. At the same time, a prudent use of state of the art information and communication technologies may help create new methods of providing access to scientific data in a timely and cost-effective way on a truly global scale.

The application of ICTs to promote access to publicly financed research was the main topic of the Third Global Research Village Conference (GRV III), held in December 2000. The GRV III conference in Amsterdam concluded “governments and research organisations should pay more attention to the conditions for access to data, information and knowledge” (Franken 2000). The sharing of information was seen as one of the key conditions for the development of scientific knowledge. A special session of the conference was devoted to policy issues related to the promotion of data sharing among researchers. This session concluded “governments and funding agencies should demand, in dealing with proposals for the funding of research infrastructures, that applications include an ICT paragraph addressing the question of the sharing of data and tools, including the software, and the sharing of instruments”. The OECD/CSTP was asked to produce a short report and a Web resource on the best practices of international sharing and data, tools (including software) and instruments. Moreover, the conference concluded that it would be useful to develop “a set of principles” for the (international) access to and dissemination of data, information and knowledge. One of the key recommendations from the conference was to form a Working Group on current practices and underlying principles for gaining access to research data (Franken 2000).

The two studies in this report aim to contribute to the work of this group of experts, the *OECD Follow-up Group on Issues of Access to Data from Public Funding* (see Appendix 5), by providing an assessment of the present state of affairs with respect to the access to, and sharing of, research data. The first study zooms in on non-US countries, on the basis of an email survey among members of the European Science Foundation and national research organisations in Australia, Canada and Japan. This email survey is complementary to the second study. This is a Web scan, which provides an overview of the policy principles with respect to the access to, and sharing of, research data in the United States.

The importance of to scientific data cannot be over-emphasised.

10

At the international level, data sharing is still in its infancy as a policy issue. However, most research organisations expect that the access to, and sharing of, research data will become a pertinent issue in the next few years. This is the main outcome of the email survey of members of the European Science Foundation and national research organisations in Australia, Canada, Japan and Europe. The contrast with the results of the second study is striking. Public availability and accessibility of research data is a basic policy principle of the US organisations in this Web scan. This includes the availability of research data for sharing among researchers.

The existence of the federal laws governing the data handling processes (Privacy Act, Freedom of Information Act and the Bayh-Dole Act) are the principal cause of the difference between the US and Europe. These laws can be understood in the framework of a political tradition in the US in which public access to government data is seen as crucial. They have created a regulatory context to which research organisations seem to have adapted by developing explicit principles, policies and guidelines. Outside of the US this is not (yet) the case.

Given this relevance of clear policy principles, the next question is how they compare with actual data-sharing practices. This is the topic of a set of case studies, which are now being undertaken within the framework of the Working Group on current practices and underlying principles for gaining access to research data. A number of players are crucial in the practice of data-sharing: funding agencies, data repositories and archives, dedicated Web sites with data, and not least the researchers themselves. Their interaction determines to what extent data are actually being shared among researchers and between researchers and non-expert audiences. The case studies aim to draw lessons from present data-sharing practices, illustrate the issues that are most pressing, locate best practices and exemplary models, find out which additional policies or funding mechanisms may be needed, and identify the main barriers and obstacles for heightened data-sharing. Which types of tools and regulation are most conducive to data sharing, and which effects increased data sharing may have on the research process, will also be addressed in the case studies. One can expect that these effects will vary by scientific field and probably also by the type of data involved. Data sharing is not always uncontroversial in the scientific community. In some specialties, the duty to make research data publicly available seems to clash with established traditions and routines (or lack thereof).

This raises the additional question of the transaction costs of rules set by funding agencies. Moreover, the application of general principles of data sharing in research contract conditions requires

specialist knowledge of the types of data involved and of the various stages in the research process. This is usually acquired in some form of cooperation or communication with the researchers in question. In other words, the application of the general principles and guidelines is based on, and produces, configurations of trust relationships and practical provisions. Data sharing is not only a technical issue, but also a complex social process in which researchers have to balance different pressures and tensions. Basically, two different modes of data sharing can be distinguished: peer-to-peer forms of data sharing and repository-based data sharing. In the first mode, researchers communicate directly with each other. In the second mode, there is a distance between the supplier of data and the user in which the rules of the specific data repository determine the conditions of data sharing. In both modes, the existence, or lack, of trust between the data supplier and the data user is crucial, though in different configurations. One of the case studies focuses on the systematic study of these configurations of trust relationships in data sharing. The other case studies will result in best practice models for data sharing. Together with the study of economic and legal aspects of data-sharing they will hopefully provide us with more knowledge about the basic social mechanisms shaping the access to and sharing of research data and help identify the most important barriers to an increased level of use of existing scientific knowledge and data.

Cutting edge research
is becoming data-driven.

Part I Policies on Data-Sharing. A Preliminary Assessment of the Current State of the Art by an Email Survey

BACKGROUND

13

Increasingly, cutting edge research is becoming data-driven in a larger number of disciplines than in the recent past. The creation of new scientific knowledge needs more and more data as *input* for novel research. At the same time, science is also *producing* an exponentially rising amount of data. These data are often not only relevant for the data-producing communities but also for researchers in other fields, for industry, and for non-profit organisations and institutions.

This “tidal wave” of data threatens to engulf the existing data infrastructure in science. No longer can the acquisition, generation, production, and archiving of data be organised on a case-by-case basis. Economically as well as organisationally, guaranteeing access to the relevant data will become a major concern in science policy.

In the near future, the challenge posed by the production of data will clearly exceed the level of the individual researcher or research group. The issues relating to the gaining of access to public research data are moving to centre stage in science policy making. This raises the question of to what extent these issues have been addressed in the science policy area. What is the current state of the art in the access to, and sharing of, data in science policy in non-US countries? To what extent have research organisations and institutions developed explicit principles, guidelines and regulations to actively promote the access to, and sharing of, publicly funded research data? This is the topic of the present study.

By conducting an email survey on data-sharing of member organisations of the European Science Foundation (ESF) and of relevant research organisations in Australia, Canada, and Japan we have tried to acquire an overview of the current policies and practice among these research organisations. As will become clear in this report, the results of this mini-survey give a clear indication that policies relating to the access to, and sharing of, research data are still a relatively unexplored domain for many important organisations. The survey has also produced a snapshot of the expectations that are currently held by experts of these organisations.

The "tidal wave" the existing data infrastructure

QUESTIONS

The questions posed aimed at acquiring a quick overview of the current state of affairs with respect to data issues and identifying those issues that were deemed most important (see Appendix 1 for the full questionnaire and the accompanying letters). Firstly, the organisations were asked to indicate whether the access to, and sharing of, data was addressed by government regulation, and if so by what level of policy making (under discussion, topic in policy documents, or addressed in legislation). Secondly, the question posed was whether the organisation itself had developed explicit policies on data issues. Thirdly, whether the participants expected that data sharing would become an important issue in the next three years. The remainder of the questionnaire was aimed at filling in the details. Amongst other topics, we wanted to know in which fields the participants expected data sharing issues to be the most relevant (both now and in the future), as well as what sort of problems they expected (technical, legal, economic or standard-setting issues).

A draft version of the questionnaire was developed in cooperation with the Dutch Ministry of Education and discussed with the ESF. A questionnaire with 10 questions was then posted on the Web site of NIWI-KNAW (data-sharing.niwi.knaw.nl). In total, 53 institutional addresses obtained from the European Science Foundation were approached by both email and regular mail with an accompanying letter, a letter from the Dutch Ministry of Education and Sciences explaining the survey, and a letter from the ESF asking for cooperation. The organisations were asked to fill in the Web form. Additionally, the three national research organisations in Japan, Australia and Canada were approached. Responses were obtained through the Web site, via regular mail, and by email. Non-respondents were reminded of the survey and asked to participate. The Web forms were automatically processed with Perseus Survey Solutions software. The documents received by email and regular mail were processed manually.

of data threatens to engulf in science.

RESULTS - GENERAL OVERVIEW

15

Response In total 31 answers were obtained from 29 different institutions¹ (50 % of the addressees, which is less than expected). The responses are from 21 different countries. This is 78 % of the countries involved in the survey. We have not been able to obtain answers from 6 countries (see Table 1).

Response

Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Estonia, France, Germany, Hungary, Iceland, Ireland, Italy, the Netherlands, Norway, Poland, Slovenia, Spain, Sweden, Turkey, United Kingdom

Non-response

Finland, Japan, Greece, Portugal, Slovakia, Switzerland

Table 1 Overview of response by country

The organisations addressed The institutions represent different types. Four categories can be distinguished:

- national research organisations and funding agencies
- scientific academies and societies
- research institutions
- governmental bodies

The boundaries between the different categories are not always clear-cut. For example, the relationships between research organisations and ministries may vary from country to country. The same holds for the other types. Scientific academies do not always have the same functions. In Eastern Europe, they tend to combine the role of learned society with that of national research organisation

¹ The Slovenian Academy of Sciences sent in two forms, one filled in by the medical section, the other by the central bureau. The Spanish CSIC also sent in two forms but since these were substantially identical, we have treated these as one form. The Norwegian Research Council and Academy of Sciences responded together in one form.

running a network of research institutes. This is different from academies of science for which the learned society is the main role.

The national research organisations responded more than average, whereas the reverse holds for the academies and societies. As a result, the national research organisations and funding agencies are over represented in the survey response; the academies are underrepresented (see Table 2).

	Total	Responding
Funding Agencies / Research Councils	27 (50%)	19 (66%)
Academies/Societies	18 (33%)	6 (21%)
Research Institutions	8 (15%)	3 (10%)
Ministries	1 (2%)	1 (3%)
<i>Total</i>	54 (100%)	29 (100%)

Table 2 Response by type of institution

Current state of affairs at the national level In slightly more than half of the countries (12 out of 21) from which we derived answers, data sharing is becoming an issue of science policy. In these countries, data sharing is presently under discussion, subject of policy documents, or part of the national legislation according to respondents from these countries. In 8 countries, this is not the case. Only in 2 countries, France and Poland, is data-sharing subject of national legislation. In 6 countries data sharing is part of policy documents but not of legislation. This is the case in: Australia, Canada, Hungary, Iceland, Netherlands and Norway. In 4 countries the issues are under discussion: Estonia, Germany, Italy and Slovenia. The remaining 9 countries are not developing policies on access to and sharing of research data according to the respondents (see Table 3).

It should be noted that in all countries some form of legislation pertaining to data does exist. For example, in the form of privacy-protection, rules on the use of clinical data, and protection of intellectual property rights (which may affect “embedded data”). However, the state of affairs, in different countries, pertaining to more advanced science policy focussing on the promotion of access to

Data sharing is most urgent in the life sciences.

Current state of affairs	Countries
Legislation	France, Poland
Part of policy documents	Australia, Canada, Hungary, Iceland, Netherlands, Norway
In discussion	Estonia, Germany, Italy and Slovenia
No policy in development	Austria, Belgium, Denmark, Ireland, Spain, Sweden, Turkey, UK and Czech Republic

Table 3 Overview of current state of affairs in national data policies

and sharing of research data is rather diverse. For example, in Iceland, national GIS-based databases on Icelandic nature are being developed which run against some major institutional and standard-setting problems. In most countries, this type of initiative is not even under consideration according to the respondents. The historical development of the political system is sometimes an important factor. In Hungary, for example, researchers were obliged by law to supply data on any research topic. Since the political turnover, research institutes have largely ignored this law, resulting in the creation of a new national data and technological information centre in Hungary. Within one country, the situation may be different in different institutions and fields. In Norway, all data from publicly funded research projects in the social sciences are stored and distributed through the Norwegian Social Science Data service, a branch of the Research Council. These data are freely available to students and researchers. However, no such system exists for the natural sciences and technology in Norway.

Current organisational policies Ten institutions have developed some form of policy on issues of the access to, and sharing of, research data. Most institutions (17) have not (see Table 4)².

2 Two institutions did not fill in this question.

	Data policy developed	No data policy
Funding agencies / Research councils	Australia, Canada, Iceland, Netherlands, Norway	Belgium, Denmark, Germany, Italy, Spain, Slovenia, Sweden, Turkey, UK
Academies / societies	Hungary, Norway, Slovenia	Austria, Estonia, Ireland, Slovenia (Med.), Czech Republic
Research institutions	France, Italy	France

Table 4 Organisational data policy by type of institution

Although a majority of the respondents have not developed data-sharing policies so far, a small majority does expect to develop policies on data-sharing in the near future: 9 out of 17. Seven organisations do not have this expectation: the Austrian Academy of Science, the Royal Irish Academy, Information and Innovation Systems at INRA (France), FWO (Belgium), the research councils EPSRC and NERC (UK), the Slovenian Research Council, the Swedish Research Council and the Czech network of universities and the academy CESNET.

POLICY ACTIONS AND VARIATION BY SCIENTIFIC FIELD

The specific forms of data sharing may vary by scientific discipline or field. It is therefore relevant to know in which fields the research organisations and academies expect that issues of data sharing will become most pressing. According to the respondents in this survey, the access to, and sharing of, research data will be an issue in all scientific and scholarly disciplines. The respondents did, however, identify a field in which data sharing is most urgent: the life sciences. The humanities, on the other hand, are least expected to be confronted with issues of data sharing. In the classical experimental sciences such as chemistry and physics, some respondents indicated that data sharing might not be such an urgent problem because existing practices and databases may usually be sufficient to provide for the data needed. This may, however, be quite different in new, multidisciplinary, fields (such as materials science and nano-technology) and in fields, which use large data generating instruments (such as high energy physics and astronomy).

We also inquired about the type of activities, which were undertaken by organisations with a policy on data sharing, broken down by field. The answers show no relationship between the type of policy action (from non-binding recommendations to legislation) and the scientific field. This means that if organisations are involved in, for example, the formulation of recommendations, they tend to develop this for all fields for which they bear responsibility. Asked about the type of policy action they expected for the future, “development and implementation of regulation” was the most frequently mentioned, followed by the formulation of “non-binding recommendations”. Two respondents only expected legislation in countries where it does not yet exist.

An important issue in data sharing is also the identification of the nature of barriers and problems that may prevent the further development of data-sharing practices in the sciences and humanities. The respondents were asked to identify which type of problem they expected to encounter in the future development of their policies on the access to, and sharing of, research data. This resulted in the following rank order (see Table 5).

Type of problem	Number of responses
legal problems (among others privacy)	9
technical problems	9
standards	8
institutional barriers	3
prohibitive cost	3

Table 5 Types of problems expected in data-sharing policies

Lastly, we inquired about the nature of the activities developed under the guidance of the research councils and academies. This should give some insight in the type of expertise that is, and will be, developed by the respondents. Selling data is definitely not popular among the respondents: only 3 organisations are active in this respect. The funding and/or management of data archives and depositories is presently, and probably also in the near future, the most practised type of activity that is included in the policies of the respondents (see Table 6).

Type of activity	Number of respondents
Funding/managing data archives	12
Co-operation with governmental data collecting agencies	11
Co-operation with national archives	9 (plus 1 which is itself an archive)
Selling/buying data from commercial firms	3

Table 6 Type of activities in data-sharing policies

IS THERE A RELATION BETWEEN NATIONAL AND ORGANISATIONAL DATA POLICIES?

The survey results give a clear indication that there is a statistically significant relationship between the existence of policies on issues of data sharing and the existence of national policies on these issues.

On the basis of the questionnaire, it is possible to construct four different types of data-sharing configurations. These are:

- Type A Respondents which have a policy on data-sharing in a country where data-sharing is an issue at the national level
- Type B Respondents which have a policy on data-sharing in a country where data-sharing is not an issue at the national level
- Type C Respondents which do not have a policy on data-sharing in a country where data-sharing is an issue at the national level
- Type D Respondents which do not have a policy on data-sharing in a country where data-sharing is not an issue at the national level

This typology is basically a table showing two dimensions: national policies and organisational policies (see Table 7).

the forefront of data sharing policies.

	National policies: yes	National policies: no
Organisational policies: yes	10	0
Organisational policies: no	3	14

Table 7 Correlation between national and organisational policies on data sharing

This relationship is statistically significant at the one-promille level, which means that the probability that this relationship is due to chance is less than one in a thousand³. The total number of observations is small, but this also holds for the whole population of institutions and countries. The level of non-response does not affect the correlation between data-sharing policies at the level of the nation and the level of the institution⁴.

The correlation is also substantially significant because it is not self-evident that initiatives in science policy at the national (or international) level lead to related changes in research organisations and funding agencies. Science policy is a political domain and hence relatively independent of the domain of scientific institutions. If novel themes like data-sharing do indeed “carry over” from the political domain to the institutional (which is suggested by the correlation), it may underline the practical relevance of formulating policy principles and guidelines at the national and international level in policy documents.

CONCLUSIONS AND DISCUSSION

Data sharing is still in its infancy as a policy issue in non-US countries. Most respondents have not yet developed explicit policies and guidelines on data sharing. This is confirmed by the interest of respondents in being informed about activities of the OECD/CSTP Working Group on Data Sharing in the future. Only 16 of the 29 respondents wish to be kept informed. Nevertheless, the majority of

³ Chi Square = 17.04 with 1 degree of freedom, hence $p < < 0.001$.

⁴ The effect of the non-response has been calculated on the basis of the known distribution of non-responding institutions over countries. In each possible configuration, the correlation turned out to be statistically significant.

research councils and academies expect that the access to, and sharing of, research data will become an important issue in the next three years. This is underlined by the fact that the respondents to this email survey tend to prioritise more consequential forms of policy initiatives (such as the formulation of regulation) above less consequential forms (such as non-binding recommendations).

The respondents expect that data sharing will become an issue in all scientific and scholarly fields. The life sciences have, however, been identified as the field in which guidelines on data sharing may be most urgent. The main problems respondents expect with respect to data sharing are technical difficulties and descriptive standards, legal restrictions and institutional barriers. Considerations of financial costs are not deemed so important.

Selling and buying data is not a major activity of the respondents. This may point to an intriguing paradox in the future. Although the life sciences are mentioned as the area where data sharing is most urgent, the respondents do not expect to be very active in selling data to, or buying data from, commercial firms. As is well known, the life sciences have become commercialised in many ways, also with respect to data handling. This may become a matter for further consideration if the trend of commercialisation affects access to research data.

Given the spread of existing national policies and policy documents on data sharing over different countries, it seems worthwhile to study the nature of these policies more in-depth and compare them in more detail with existing regulation in other countries. This may be of more relevance, as those research organisations, which expect to undertake future action, tend to emphasize binding regulation as their priority. Identifying key problems in the development of this type of regulation may therefore be useful.

There is a clear relationship between the national and organisational level of policies with respect to the access to, and sharing of, research data. This is first indicated by the statistical correlation found in this survey between the existence of a policy on data sharing at the national level and the existence of these policies at the institutional level. This may point to the intimate relationship between national science policy and national research organisations. It may also be related to the relative novelty of the issues of data sharing. New themes may perhaps “carry over” relatively easily, which would point to an agenda-setting role of national science policy. Secondly, the relationship is indicated by the difference of the results of this email survey and the findings of the Web survey

Data sharing is still in its infancy as a policy issue in non-US countries.

of data-sharing policies and principles in the US (see part II). In the US, there exists both a political tradition in which public access to data is seen as crucial and a set of federal laws that regulate how research organisations and institutions should provide access to research data and facilitate the sharing of research data. This has created a regulatory context to which research organisations seem to have adapted by developing explicit principles, policies and guidelines. Outside of the US this is not (yet) the case.

Scientific databases are
"non-linear accelerators
of research".

Part II Access to and Sharing of Research Data - The Policy Context. A Web Scan of Principles and Regulations in the US

BACKGROUND

25

The United States is probably the largest data producer in the world. Government agencies, scientific institutions, and commercial companies generate enormous amounts of data on a daily basis. Due to digitisation, data producing capabilities are also increasing exponentially. “There is barely a sector of the economy that is not significantly engaged in the creation and exploitation of digital databases, and there are many – such as insurance, banking, or direct marketing – that are completely database dependent” (National Research Council 1999). Scientific and scholarly research is no exception to this general trend. Increasingly, the creation of new knowledge is dependent upon gaining instant access to research data as well as the capacity to store massive amounts of generated data in a fast and reliable way. Scientific databases are proving to be “non-linear accelerators of research” (Cerf 1999). In some scientific fields a tradition of data sharing has evolved through the daily operation of large scientific instruments, e.g. high-energy physics (CERN), or networks of observatories, e.g. radio astronomy (Schillizzi 2000). In other fields, however, large-scale data sharing has been confronted with technical and social barriers, e.g. brain research (Jennings 2000; OHBM 2001) and genetics (Stokstad 2002).

This has led research funding agencies and scientific societies to start developing explicit policies and regulations to promote the economic use of large-scale research instruments or networks of instruments. US institutions seem to be at the forefront of this new domain of science policy. This is partly due to the dominant role of American researchers in a number of fields, especially in natural and life sciences (it is less so in the social sciences and humanities). It is also related to the political tradition in the US in which open access to government data for all citizens is seen as one of the corner stones of democracy and the constitutional state. As a consequence, data generated with public money (including scientific data) were freely available to all. However, in the last five years the status quo has been challenged by new economic, technological, and legal developments concerning (digital) databases. Digital technologies play a paradoxical role in this development. They may enable a radically heightened scale of data sharing as well as allowing for an increased level of control over data by its owner or provider. Since shared access to data seems to have become more important than ever for the creation of scientific knowledge, analysis of the contradictory tensions surrounding practices of data sharing seems pertinent for policy. As will become clear from this study and its comparison with the state of affairs in European data-sharing policies, the political and legal context does affect the ways in which institutions organise access to and sharing of

research data. The question of whether clear policy principles and guidelines have been formulated at the international and national level does matter. However, this does not mean that the relationship between policies and rules and the practice of data sharing amongst scientists is straightforward (Hilgartner 1998). For the individual researcher or research group, the policy and regulatory context provides a set of additional pressures which he needs to reconcile with other pressures in his research practice, such as the complexity of the research tasks themselves, pressure from peers and local institutional structures. Shaping the institutional contexts of research practices is probably one of the most effective ways of influencing the way research is being executed. For example, by the creation of legal boundaries for research, the imposition of conditions under which research is being funded, and the creation of infrastructures which can be used by researchers. In the United States all three dimensions have been implicated in attempts to promote access to, and sharing of, research data.

THE POLICY AND TECHNOLOGICAL CONTEXT OF ACCESS TO RESEARCH DATA

The Web documents providing the policies and regulations on shared access to data reflect these pressures on the ways that research is being performed. The following organisations have been included in this study (see Appendix 3 for this study's methodology):

- National Research Council NRC www.nas.edu/nrc
- National Science Foundation NSF www.nsf.gov
- National Institutes of Health NIH www.nih.gov
- National Aeronautics and Space Agency NASA www.nasa.gov
- American Association for the Advancement of Science AAAS www.aaas.org
- National Archives NARA www.nara.gov
- National Endowment for the Humanities NEH www.neh.gov
- Inter-University Consortium for Political and Social Research ICPSR www.icpsr.umich.edu
- Organisation for Human Brain Mapping (OBHM) www.humanbrainmapping.org
- Global Change Data and Information System (GCDIS) www.globalchange.gov
- Committee on Data for Science and Technology (CODATA) www.codata.org

do not expect active in selling data.

The results were also compared with documents from the European Science Foundation ESF www.esf.org. As most documents referred to ongoing debates about legal initiatives and (partly conflicting) legislation, additional documentation on these debates was also collected and included in the analysis.

The present state of regulation with respect to the access to and sharing of research data has mainly been shaped by two different federal laws in the US: the Freedom of Information Act, and the Bayh-Dole Act (see Appendix 4):

- In 1999, the Freedom of Information Act (FOIA) was extended to explicitly include research data. A provision was inserted in the Omnibus Appropriations Bill (Public Law 105-277) to change federal regulations in order to allow broader access to federally funded research data. The provision meant that *all* federally funded research data could be accessed through the mechanisms laid out in the Freedom of Information Act. The scientific community was opposed to the proposal, arguing that it threatened to undermine the integrity of the research process. Nevertheless, Congress adopted the extension of the FOIA, although the White House Office of Management and Budget limited the scope of the amendment in implementing its provisions in regulations. Scientific institutions that are also federal agencies (such as the National Institutes of Health) have since developed principles and policies to deal with requests for information under the FOIA.
- The Bayh-Dole Act of 1981 is aimed at the commercialisation of research results by granting patent rights to universities for inventions developed with federal funds. This includes exclusive licensing. Its reach has since been broadened, and the Act seems to have led to a substantial increase in the number of patents filed by universities, research institutes and individual researchers. The Bayh-Dole Act may have impeded the sharing of data involved in the preparation of patent applications. A patent, on the other hand, is a form of publication and does not itself limit the use of the underlying data.

Other legal frameworks shaping shared access to research data in the US are:

- The Privacy Act of 1974, which provides certain safeguards for the use of information, maintained in a database, about individuals. These safeguards include the right of individuals to

The costs of data handling and storage keep falling.

determine what personal information is maintained in Federal agencies' files (hard copy or electronic) and how it is used, to have access to such records, and to correct, amend, or request deletion of information in their records that is inaccurate, irrelevant, or outdated.

- The “fair use” exception in copyright law, which enables scientists to use copyrighted material freely in many cases and under certain conditions. The exception is rooted in the constitutional right of free speech under the First Amendment. It enables the use of all factual data in a copyright protected database as long as the creative elements in the database are not being reproduced. However, new forms of database protection may threaten exemption of copyright under the fair use exception.
- Software protection under patent law, which has been implemented since a law case in 1986. The US Patent Office changed its policy in the 1990s and it is now possible to patent algorithms. As a consequence, software falls both under patent law and under copyright protection. The algorithm and related advances in software technology are protected by patent law (as the idea). The final product is protected by copyright (the expression of the idea).
- Anticircumvention rules in the new US copyright law (the Digital Millennium Copyright Act) may, in the near future, threaten the possibilities for scientists to use digital data that is protected by encryption or other technical means (Samuelson 2001). The DMCA specifically forbids the bypassing of technical measures imposed by copyright owners to limit access to their works. It also outlaws the manufacture or distribution of technologies designed to circumvent such technical measures. Finally, it makes the removal of copyright management information, such as digital watermarks, illegal. Since all digital data can be protected with this type of encoding, the anticircumvention rules may have an impact on access to research data in more areas than computer science alone. The combination of detailed technological control over the use of data and information, together with the DMCA, may have severe downstream consequences for the reuse and redistribution of research data. However, the extent to which this will happen is unclear.

The regulation of shared access to data is not only shaped by legal frameworks and federal laws, but also by the technological and economic context of the information and data. Scientific data have predominantly become digital data distributed through the internet and stored in digital media.

Hence data have the same economic characteristics as information goods in general (Varian 1998). Data generation is very expensive, but its distribution or copying is cheap. Moreover, due to digitisation, the costs of data handling and storage keep falling. Many scientific data are generated by a sole source, or in a unique situation, which creates a natural monopoly for the data producer. Data are now usually stored in digital databases, often with (protected) access interfaces over the internet. The digitisation has led to a blurring of the boundaries between data and more aggregated forms of information. This may already happen at the level of scientific instruments when some form of processing of the raw data takes place even before the researcher sees them. As a consequence it is often difficult, or even impossible, to isolate data from its informational context. Often this does not even make sense for the user. Processed data are generally easier to interpret and use than raw data, which may be completely meaningless outside of the context of their generation. This may lead to a paradox with respect to data sharing if data processing was based on certain field-specific assumptions and discipline-specific standards. In those situations, the processing of data may make them less easy to use outside of their original disciplinary context whilst at the same time making them easier to interpret. This is one of the reasons why setting standards for data formats in order to promote the re-use and sharing of data can be such a daunting task, especially in interdisciplinary or hybrid contexts. In these contexts, the economic mechanisms and institutional incentives favouring data sharing are often also lacking.

These economic and technical characteristics of scientific data have been the subject of different, and often conflicting, legal regimes and initiatives. Traditionally, data have been free and not subject to copyright rules or of exclusive property rights. The increased economic role of data through digitisation triggered attempts to introduce new forms of data protection, some of which may significantly influence data sharing in scientific research. This relates to key characteristics of digital information and data:

- Digitisation has greatly enhanced the ease of copying and distribution of large amounts of data. This has been perceived by parts of the database industry as threatening, especially by the music industry and producers of various directories. As a result, in the 1990s, a lobby emerged to increase the legal protection of databases (see below). It should be noted, however, that most databases are protected by the copyright covering creative elements of a database. The facts themselves are not protected (even if collecting them was labour-intensive) but the organisation of the databases, the arrangement of the information, and the

Many scientific data are generated or in a unique situation,

30

coordination of the database are. Patents may also protect some elements of a database. Most databases that are used by scientists are either in the public domain (like all databases of the US federal government) or are covered by copyright law.

- The digital environment has greatly enhanced the possibilities to prevent unauthorised use of data with technological and legal means. Encryption technologies enable a database producer to limit access to the database. As most digital databases are highly dynamic entities, and their value depends on the frequency with which databases are being updated, the nature of the economics of databases has been transformed. Users do not buy the database itself anymore, but increasingly licence the rights of access to the database. This has important downstream consequences because the license (or a private contract) may impose important constraints on the use of even the most factual of data. This is especially important for science since much scientific research involves the merging of data from a large number of different sources and their redistribution in a new compilation and transformed format. Contracts, coupled with technological constraints, can put severe limits on this type of data use. Whether database owners will have an interest in impeding scientific research in this way, and if so to what extent, is presently an open question.

The discussion in the U.S. on data sharing has also been influenced by European legislation, which was adopted in response to pressure from parts of the database industry. This debate has not (yet) led to new rules with respect to access to research data. It has, however, stimulated representatives of both the scientific community and the federal government to restate their basic principles on access to scientific information and data (see below). This debate hinges upon the economic impact of digitisation:

- In 1996, the European Union adopted a strong form of legal protection of databases in its Directive on the Legal Protection of Databases. Since then, the directive has been incorporated into national law in the member states and in a number of affiliated states. The main difference from copyright law is that “the sweat of the brow” of the database producer is protected, not just the creative elements. If the investment of the database producer is substantial, the producer has the right to prevent the extraction or reuse of any substantial part of the database. This right pertains to downloading, copying, printing or reproduction in any form (Hugenholtz 2001). The right holds for 15 years from the date of completion of the data-

by a sole source, which creates a natural monopoly.

base. A substantial update also renews the right. This means that dynamic databases enjoy a virtually unlimited protection under the new database law. Even a mere “substantial” verification of the database might give the producer extension of his right. The exceptions are again far more limited than is the case in copyright law. Most traditional ways to use copyrighted materials are prevented in the new database law, such as journalistic freedom, quotation rights, privileges for libraries, the free use of government information. This also holds for data. The right to use data is far more limited than under copyright law and centres around the notion of “illustration” in teaching and research. It is not yet clear to what extent the implementation in national laws will lead to a strict or more liberal interpretation of the law by courts. The strongest impact of the database laws on scientific research is expected in those cases where the publication of merged and transformed data is crucial and where researchers form the sole market for the database. European database law does not contain provisions mandating compulsory licensing at marginal costs to individual researchers or research institutes (David 2001-003).

- In the US, a comparable debate started in 1991 after the Supreme Court ruled that databases were not protected under “sweat of the brow” terms and copyright protection was limited to its creative elements. The European directive subsequently fuelled this debate. Successive Congresses have considered the introduction of comparable regulation (draft database bills HR 3531, HR 2652, HR 354 and HR 1858). One reason for this is that the European directive contains a reciprocity provision which limits the legal protection to database producers from those countries that have similar tight database laws. The scientific community, in common with many other interest groups, has strongly opposed attempts to emulate the European directive in the US and elsewhere, since it would severely limit access to and use of data for research. A key point in the debate is whether database producers should enjoy a novel property right (as is the case in Europe) or rather protection against unfair competition comparable to already existing laws against misappropriation. The precise formulation of exceptions for scientific research is also a key point in the debate.
- The debate on databases may be especially important because the role of the federal government in the production and funding of scientific databases seems to be changing. Private-public partnerships now play a more important role. Private companies are becoming more important in the dissemination of government data, and a number of data-producing

activities have been outsourced by federal agencies, partly to cut the budget. This development may lead to new database legislation having a bigger impact on the sharing of data in scientific research.

BASIC POLICY PRINCIPLES ON ACCESS TO AND SHARING OF RESEARCH DATA

Under U.S. federal government law and policy, publicly funded information, including research data, should be in the public domain. This is the basic principle informing most data-sharing rules included in this study. It is laid down in the guidelines published by the National Institutes of Health (NIH 2001): “Most grant-related information submitted to NIH by the applicant or grantee in the application or in the post award phase is considered public information and is subject to possible release to individuals or organizations outside NIH. The statutes and policies that require this information to be made public are intended to foster an open system of Government and accountability for governmental programs and expenditures, and, in the case of research, to provide information about federally funded activities.” Only certain types of information that may be considered proprietary or private information may be withheld from the public. This means that NIH will generally release the following types of records in response to an FOIA request:

- Funded applications;
- Pending and funded non-competing continuations;
- Grant progress reports;
- Final reports of any audit, survey, review, or evaluation of grantee performance that have been transmitted to the grantee.

Other types of information will generally be kept confidential. These include, amongst others, pending competing grant applications; unfunded new and competing applications; financial information regarding a person; information pertaining to an individual; pre-decisional opinions; evaluative portions of site visit reports and peer review summary statements; trade secrets; information which, if released, would adversely affect the competitive position of the person or organization; and patent or other valuable commercial rights. As will be clear, the exceptions are mostly based on the Privacy Act and on the Bayh-Dole Act.

Research data may be included in either category of research information. In the NIH Grants Policy Statement “data” is defined as “recorded information, regardless of the form or media on which it may be recorded, and includes writings, films, sound recordings, pictorial reproductions, drawings, designs, or other graphic representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing or computer programs (software), statistical records, and other research data.” NIH has developed project/programme specific guidelines for access to research data. “Whenever possible, data should be deposited in public databases and materials in public repositories. Where appropriate repositories do not exist or are unable to accept the data or materials, investigators should accommodate requests to the extent possible.”

Recently, NIH announced the further extension of its policy regarding sharing research resources through a new draft statement on data sharing (NIH announcement 1 March 2002). The new statement will expect and support the “timely release and sharing of final research data from NIH-supported studies for use by other researchers”. Investigators submitting an NIH application will be required to include a plan for data sharing or to state why data sharing is not possible. The statement focuses on “final research data”. NIH defines this as follows: “recorded factual material commonly accepted in the scientific community as necessary to validate research findings”. Final research data will, therefore, not include: “laboratory notebooks, partial data sets, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects such as gels or laboratory specimens” (NIH FAQ on Data Sharing, March 1, 2002).

Public access to research data is also the basic principle of the National Science Foundation. “NSF advocates and encourages open scientific communication” (NSF Grant Proposal Guide, V, H, 1-1-2002). NSF expects significant findings from supported research and educational activities to be promptly submitted for publication with authorship that accurately reflects the contributions of those involved. “It expects PIs to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages grantees to share software and inventions, once appropriate protection for them has been secured, and otherwise act to make the innovations they embody widely useful and usable.”

NASA “shall provide for the widest practicable and appropriate dissemination” of the STI resulting from its research effort, “while precluding the inappropriate dissemination of sensitive information”.

NASA disseminates scientific information “in a manner consistent with U.S. laws and regulations, Federal information policy, intellectual property rights, technology transfer protection requirements, and budgetary and technological limitations”. In this, NASA follows the principle of “non-discriminatory access so that all users within the same data use category will be treated equally”. NASA will allow preferential treatment for U.S. government users and affiliates only where expressly permitted by law. Archiving is seen as part of NASA’s responsibility. NASA has developed an elaborate set of rules covering the publication of technical reports and technical manuals in its *Guidelines for Documentation, Approval, and Dissemination of NASA STI* (valid until September 2002). Technical publications usually include extensive data or theoretical analysis, but they may also be compilations of significant scientific and technical data.

In 1999, the US government stated its basic policy principles before the House of Representatives (Pincus 1999) discussing the Collections of Information Antipiracy Act (H.R. 354). These include:

- databases generated with Government funding generally should not be placed under exclusive control, *de jure* or *de facto*, of private parties;
- any database misappropriation regime should provide exceptions analogous to “fair use” principles of copyright law; in particular, any effects on non-commercial research should be *de minimis*.

These principles are based on “weighing the need to protect database creators against the potential impact on scientific research in particular, and the dissemination of information within the society generally”. Therefore, database protection should leave room for transformative use of data. Facts should also be excluded from protection: “The Copyright Clause and the Copyright Act permit protection only of an author’s expression, and do not authorize protection of facts. This comports with the First Amendment principles.” Government information should be publicly available because it is a valuable national resource. “It provides the public with knowledge of the government, society and economy – past, present and future. It is a means to ensure the accountability of government, to manage the government’s operations, to maintain the healthy performance of the economy, and is itself a commodity in the marketplace”. Pincus explicitly included universities in the

Data producing capabilities are increasing exponentially.

governmental domain. “We believe that public universities should fall within a broad definition of government institutions which generate collections of information. Instead of trying to draw a distinction between public universities and other government institutions, it might be more appropriate to concentrate on the distinction between public research and privately funded research at public institutions”. The US government also believes that databases produced with substantial government funding should be treated like databases of government-generated data (unless a contrary provision has been included in the contract or grant).

The National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine and the American Association for the Advancement of Science gave a joint statement in the same House discussion on database protection (Lederberg 1999). “Thus, freedom of inquiry, the open availability of scientific data, and the open publication of results are cornerstones of our research system that US law and tradition have long upheld”. Hence, full and open access to data is the basic principle for many scientific institutions in the U.S. Lederberg, citing the Bits of Power report (NAS 1997), defined full and open as follows: “by full and open we mean that data and information derived from publicly funded research are made available with as few restrictions as possible, on a non-discriminatory basis, for no more than the cost of reproduction and dissemination”.

Data from the private sector should be made available on a “fair and equitable” basis. This means that “if commercial content providers receive enhanced protections in their databases, that preferential terms of access to and use of those data by researchers, educators, libraries, and other public-interest entities, firmly rooted in our Constitution and legal tradition, are retained and, when necessary, adapted to the digital and online environment”.

In November 2000, CODATA formulated six “principles for science in the internet era” to support “full and open access to data needed for research and education”. These principles are:

- science is an investment in the public interest
- scientific advances rely on full and open access to data
- a market model for access to data is unsuitable for research and education⁵
- publication of data is essential to scientific research and the dissemination of knowledge

5 Although the text of this principle seems to include all forms of research (including private research), the context of the document indicates that what is meant here is first and foremost public research.

The digitisation has led to a blurring of the boundaries between data and information.

36

- the interests of database owners must be balanced with society’s need for open exchange of ideas
- legislators should take into account the impact intellectual property laws may have on research and education

The US Global Change Research Program initiated a Data and Information Working Group to develop interagency data management in 1987 (DWIG 2001). The program has had “full and open access” as policy guidance for federally obtained data since its inception (DWIG 1999). This means “data and information should be available without restriction, on a non-discriminatory basis, for no more than the cost of reproduction and distribution” (DWIG 1998). Where possible, access to data should be provided through the World Wide Web to keep the costs as low as possible and to allow distribution to be as wide as possible.

The National Endowment for the Humanities has been encouraging and supporting humanities research and scholarship involving computer technologies since the early 1970s. Although the term data sharing as such is not used often, a large number of NEH funded projects are in fact forms of data sharing, e.g. the creation of large repositories and databases of digitised information. The same holds for projects in the area of preserving human and cultural heritage. NEH also addresses data sharing by funding projects aimed at developing standards for creating and preserving digital data for research.

The National Archives, for which making data accessible is the very reason of its existence, states “increased data-sharing” as one of the goals for the improvement of its data administration. The Inter-University Consortium for Political and Social Research, ICPSR, is an organisation of member institutions working together to acquire and preserve social science data, to provide open and equitable access to these data, and to promote effective data use. The ICPSR promotes and facilitates research and instruction in the social sciences and related areas by “acquiring, developing, archiving, and disseminating data and documentation for instruction and research and by conducting related instructional programs”.

Two different motivations for promoting data sharing emerge in this study: public policy considerations and the needs of scientific research itself.

In the first category, the following motivations can be distinguished:

- the principle that the various forms of data collected with public funds belong in the public domain
- researchers have a special obligation to scientific openness and accountability when the research is publicly funded
- the obligation to abide by the law, especially the Freedom of Information Act
- to improve U.S. competitiveness

In the second category, motivations are:

- the advancement of science
- the widespread and timely distribution of tools for further discovery
- verification and refinement of research findings
- the replication and secondary analyses of valuable (and costly) data sets to address new, and quite possibly unforeseen, research questions
- to reduce unnecessary duplication of research
- reduction of the need for new data collection and social surveys.
- economies of scale
- to improve the productivity and cost-effectiveness of research
- the need for large data sets to answer research questions that cannot otherwise be addressed
- the application of cutting-edge technologies to data sets by multidisciplinary research teams
- when research tools are used only within one or a small number of institutions, there is a great risk that fruitful avenues of research will be neglected
- providing access to data for new but talented researchers
- to improve training for graduate and undergraduate students

All organisations used motivations from both categories, although the emphasis does vary. The US government, NIH and NASA tend to emphasize first of all the public policy considerations. The NSF, AAAS, and the NAS/NRC tend to start with stressing the importance of science for society and the role of shared access to research data in the creation of new scientific knowledge. All organisations explicitly acknowledge the political and legal paradigms in the US that have “full and open access” to data and information as a basic tenet.

THE BAYH-DOLE ACT AND DATA SHARING

All organisations try to balance the need for sharing data with the recognition of intellectual property rights on inventions (data themselves are not protected under copyright or patent laws). In the US, this means that research organisations need to satisfy the conditions of both the Freedom of Information Act and the Bayh-Dole Act. The NSF allows grantees to retain principal legal rights to intellectual property developed under NSF grants to provide incentives for development and dissemination of inventions, software and publications that can enhance their usefulness, accessibility and upkeep. “Such incentives do not, however, reduce the responsibility that investigators and organizations have as members of the scientific and engineering community to make results, data and collections available to other researchers.”

The NIH expects recipients of funds to “maximize the use of their research findings by making them available to the research community and the public, and through their timely transfer to industry for commercialisation”. The right of researchers to retain title to inventions made with NIH funds comes with the corresponding obligations to promote utilization, commercialisation, and public availability of these inventions. The Bayh-Dole Act encourages researchers to patent and license subject inventions as one means of fulfilling these obligations. However, the NIH states, “the use of patents and exclusive licenses is not the only, nor in some cases the most appropriate, means of implementing the Act. Where the subject invention is useful primarily as a research tool, inappropriate licensing practices are likely to thwart rather than promote utilization, commercialisation and public availability of the invention.” The NIH stipulates that researchers should analyse whether further research, development and private investment are needed to realize this primary usefulness. “If it is not, the goals of the Bayh-Dole Act can be met through publication, deposit in an appropriate databank or repository, widespread non-exclusive licensing or any other number of dis-

semination techniques. Restrictive licensing of such an invention, such as to a for-profit sponsor for exclusive internal use, is antithetical to the goals of the Bayh-Dole Act.” On the other hand, where private sector involvement is desirable to assist with maintenance, reproduction, and/or distribution of the tool, or because further research and development is needed to realise the invention’s usefulness as a research tool, “licenses should be crafted to fit the circumstances, with the goal of ensuring widespread and appropriate distribution of the final tool product”. The NIH explicitly includes the possibility of exclusive licensing. The NIH also considers the burden of patenting and licensing. Researchers are asked to take “every reasonable step” to streamline the process of transferring their own research tools freely to other academic research institutions “using either no formal agreement, a cover letter, the Simple Letter Agreement of the Uniform Biological Materials Transfer Agreement (UBMTA), or the UBMTA itself”.

DATA-SHARING AS A CONDITION FOR RESEARCH FUNDING

The funding organisations covered in this Web scan increasingly require explicit data-sharing plans as a condition for research funding. These plans should cover how and where these materials will be stored at reasonable cost, and how access will be provided to other researchers, generally at their cost. Since 2001, NSF has asked researchers to explicitly include, if appropriate, “plans for preservation, documentation, and sharing of data, samples, physical collections and other related research products” (NSF 2001).

In the case of x-ray crystallographers the NIH has a policy that requires the placement of coordinate data into a data bank at the time of publication. The NIH and DOE genome programs require all applicants “expecting to generate significant amounts of genome data and materials” to describe in their application how and when they plan to make such data and materials available to the community. “These plans in each application will be reviewed in the course of peer review and by staff to assure they are reasonable and in conformity with program philosophy.” If a grant is made, the applicant’s sharing plans will become a condition of the award, and compliance will be reviewed before continuation is provided. Progress reports will be asked to address the issue. NASA also stipulates that data-sharing plans should be part of research plans. For example, all NASA’s Earth System Enterprise missions, projects, grant proposals “shall include data management plans”. For

each cooperative activity with industry, domestic or foreign, NASA “shall seek agreement on all major data management and distribution issues during the project definition phase”.

Generally, the researcher or research institution obtaining the funding is held responsible for providing access to data. This means that the costs for providing access to data can be included in the research budget. The NSF has the rule that the budget “may request funds for the costs of documenting, preparing, publishing or otherwise making available to others the findings and products of the work conducted under the grant”. The NIH prefers data sets to be put into data archives, and objects into repositories. If this is not possible, the researchers should provide access “as much as possible”. For NIH grants, the awardee is not the individual investigator but the institution. The NSF has the same position as NIH, with the exception of some post-doctoral fellowships. The NIH notes that this may create problems under the Freedom of Information Act since a request to the NIH to produce data may go to a university that no longer has an employer-employee relationship with the investigator. Within NASA the departments are responsible. The organisation also assumes responsibility for archiving. In general, however, long term archiving will not be guaranteed by research groups or organisations. For this reason, the ESF is of the opinion that “national or regional discipline-based archives should be considered where there are practical or other problems in storing data at the institution where the research was conducted”.

DATA TYPES

Different types of data may create various specific problems if they are to be shared with other researchers or made available to the public at large. The following relevant issues have been identified in the documents:

- The sharing of data as research results may meet different obstacles compared to those met by the sharing of data that have been used as research resource. In a number of cases, data used as input for research may not as easily be shared as data resulting from research. This reluctance may be motivated, for example, by the fear that the release of raw input data could unblind clinical trials, lead to erroneous conclusions, undermine investigators’ investments, and jeopardize their intellectual property rights, especially in regard to non-US patents (NIH Response 1999).

Publicly funded information, including research data, should be in the public domain.

- Different types of data may require different storage facilities and access requirements. Examples are archaeological data, specimens from physical anthropology, large-scale survey data, oral interviews with scientists and other subjects, data generated by experimental research, and field records of tribal ceremonies.
- Mathematical and computer models are both tools and data. Sharing these often means that investigators must prepare fully documented and robust versions of these models.
- Objects of research such as archaeological specimens or fossil remains pose specific problems. In these instances data consist not only of the objects themselves, but also of contextual information and quantitative and qualitative descriptions of the materials. As these physical objects do not always become the property of the investigator but often belong to a host nation or cultural group, scientists may not control access to them.
- Qualitative information ranging from microfilms and other copies of very old documents, to oral interviews and video tapes, ethnographic or linguistic field notes or recordings or transcriptions, or hand written records of open-ended interviews, need special arrangements including privacy protection and specification of the time at which they will be made available.
- Quantitative social and economic data sets generally need to be placed in specialised data archives. In experimental research, individuals, be they people, animals, or objects, are subjected to preplanned conditions and their responses tabulated in some fashion. For these data, complete information on how an experiment was conducted and any unusual stimulus materials is important, so that failures to replicate will not turn out to depend on one scientist's incomplete understanding of another's procedure. In these cases, placing such data in a formal archive may be a solution.
- On the other hand, in experimental science, the data are the result of experiments. Here the need, as perceived by a number of scientific communities, is not to make the original data available, but to make available the methods used to obtain the results. If others challenge those results, they would try to replicate the experiment and would then publish their findings.

- Longitudinal data sets present a special problem as the release of data early in a long term study could affect later waves of data collection and could risk identification of subjects (for example in medical research).

LIMITS TO DATA-SHARING

At the GRV III conference, the issue of “reasonable limits” to data sharing was raised. In this scan considerations of privacy protection seem to dominate. A second important limitation mentioned is the “protection of the research process”. The NIH states that access to research data “must occur in the context of strong protections for research participants, protection of proprietary interests, freedom from harassment of researchers, and confidence that the process will further research, not harm it.”

The following limitations are mentioned in the Web documents:

- safeguard the rights of individuals and subjects
- the rights of individuals to determine what information about them is maintained
- legitimate interest of investigators, for example materials deemed to be confidential by a researcher until publication in a peer-reviewed journal
- the time needed to check the validity of results
- the integrity of collections
- data released to the public that could lead to the identification of historically and scientifically valuable archaeological sites could invite looting and destruction
- data enabling the identification of the location of rare botanical species outside the United States could lead to unwanted bio prospecting and could damage the relationship between researchers and the host community
- differences between fields
- information related to law enforcement investigations
- national security information

The following data and research resources are generally excluded from the duty to provide access to them under the Freedom of Information Act:

- draft materials such as preliminary analyses, drafts of scientific papers and plans for future research
- peer reviews
- communications among colleagues
- physical objects (e.g., laboratory samples, audio or video tapes)
- pending competing grant applications
- unfunded new and competing continuations and competing supplemental applications
- financial information regarding a person, such as salary information pertaining to project personnel
- information pertaining to an individual, the disclosure of which would constitute a clearly unwarranted invasion of personal privacy
- evaluative portions of site visit reports and peer review summary statements, including priority scores
- trade secrets and commercial, financial, and otherwise intrinsically valuable items of information that are obtained from a person or organization and are privileged or confidential
- unpublished data: “Premature access to data could unblind clinical trials, lead to erroneous conclusions, undermine investigators’ investments, and jeopardize their intellectual property rights, especially in regard to non-US patents.”

CONCLUSIONS

Public availability and accessibility of research data is a basic policy principle of the US organisations in this Web scan. The need for scientific organisations to abide by the law has necessitated an explicit and transparent set of rules and policies. This includes the availability of research data for sharing among researchers. An important motivation for making research data available is the principle that publicly funded research data (both data used as resource and data resulting from research) should be publicly available. The second set of motivations for explicit guidelines on data-sharing results from changes in the conduct of scientific research. The application of information and communication technologies and new imaging technologies has accelerated the process in which sharing data and resources are becoming crucial for research in a variety of fields. More complex multidisciplinary research questions are also important factors driving the process of increasing data sets and creating new types of large distributed data sets. Researchers themselves are

Investigators will be required to include a plan for data sharing.

44

becoming more dependent on the increased possibilities for data sharing. The need to give new researchers access to data, and the need to increase the quality of research training, give added impetus to improved regulation of access to research data.

As a result, plans for data sharing are a condition for research funding from the funding agencies in this study. Those plans are subjected to quality control and peer review, taking into account both the rules of the funding organisation and discipline-specific quality criteria. The research organisation or individual investigator is responsible for enabling access to research data. Long term archiving is an exception to this rule. This should be the responsibility of specialised data archives and repositories.

The contrast with the outcome of the email survey of ESF members and related organisations in Australia, Canada and Japan is striking. The email survey showed that data sharing is an emerging issue in science policy. Most organisations expect to develop policies on the access to, and sharing of, research data in the next few years. In the US this is already firmly established. The Web documents in this study have proved that the existence of federal laws governing the data handling processes (Privacy Act, Freedom of Information Act and the Bayh-Dole Act) are the principal cause of the difference between the US and Europe.

This study did not cover all of US academic research. Neither can the extent to which the data are made available in digital form be concluded from these policy documents. Given the wide variety of data types involved, regulation seems to be relevant for digital data, as well as analogue data and objects. The increased digitisation of research information will no doubt lead to a sustained increase of digital research data. The variety of data types necessitates not only the availability of various technical tools and standards for data-sharing but also the development of adequate institutional arrangements.

The policy documents indicate that research contracts do indeed stipulate detailed agreements on data sharing taking the specific characteristics of the research data into account. An interesting question is which experiences have been collected with these data-sharing plans and what type of tools and arrangements have proved effective.

The limits to public accessibility of data are explicitly stated in the guidelines studied. The most important limits that are deemed reasonable arise from:

- protection of the rights of persons and research subjects (including privacy protection);
- protection of intellectual property rights;
- concerns over the integrity of the research process; and
- considerations of national economic and security interests.

The precise consequences of these limits and the ways they are addressed relate to the type of data involved. The documents give the impression that the type of organisation (funding agency; research organisation; scientific society; archive) also determines the balance which is struck between conflicting needs and the way that limits to data sharing and accessibility are being imposed. This includes the exact definition of terms (e.g. what are data), the materials that are excluded from public scrutiny (e.g. under the Freedom of Information Act) and the extent to which exclusive licensing is permitted. It should be noted here that the legitimate interests of the researcher producing the data are generally seen as part of the need to protect the integrity of the research process. No organisation claims a semi-permanent privileged access to the data for the data producing investigator, given that it concerns publicly funded research.

It is nevertheless clear that the investigator is an important party in the application of the rules on data management and the development of data-sharing practices. The types of tools and regulations that are most conducive to data-sharing, as well as the effects that increased data-sharing may have on the research process itself can only be determined in case studies and comparative studies of data-sharing practices. This is also necessary to determine how the guidelines and principles covered in this Web scan are actually being applied and which experiences and best practices have been collected. Data sharing is not always uncontroversial in the scientific community. In some specialties, the duty to make research data publicly available seems to clash with established traditions and routines (or lack thereof).

This raises the additional question of the transaction costs of rules set by funding agencies in these cases. Moreover, the application of general principles of data sharing in research contract conditions requires specialist knowledge of the types of data involved and of the various stages in the research process. This is usually acquired in some form of cooperation or communication with the researchers in question. In other words, the application of the general principles and guidelines is based on, and produces, configurations of trust relationships and practical provisions. One of the speakers at a Council meeting of the National Institutes of Mental Health touched upon this in

response to the controversy in brain research on data-sharing: “Incentives for data-sharing need to be offered that offset investigator’s loss of control over their data-bases. Usually, this is some form of added scientific value. By sharing data, an investigator gains access to more data or other tools. Ultimately, there has to be a procedural framework that makes sharing sensible, efficient, thorough, and value-added. If all of those pieces are in place, fewer external or coercive forces are needed to convince investigators to share.” Best practice cases and the study of data-sharing practices are both needed to shed more light on the nature of the international framework needed for data-sharing as well as the consequences of such a framework for the production of, and access to, scientific information.

Part I

Appendix 1: The Questionnaire

QUESTIONNAIRE

please cross the right entry

1 *Are Access to and Sharing of Research Data currently subject of governmental science policy in your country?*

- being discussed ...
- formulated in policy documents ...
- established in legislation ...

2 *Does your organisation have a policy on Access to and Sharing of Research Data?*

No ...

Yes ... *go to question 5*

3 *Do you expect Access to and Sharing of Research Data to become a policy issue for your organisation within the next 3 years?*

No ... *questionnaire completed*

Yes ... *to question 4*

4 *In what fields of research do you expect access to research data to become a policy issue on the agenda of your organisation?*

	<i>yes</i>	<i>no</i>
Natural Sciences (incl. Earth Sciences, Atmospheric Research)
Engineering & Technology
Life Sciences (incl. Environmental Research, Bio diversity)

Any database misappropriation regime analogous to "fair

48

	<i>yes</i>	<i>no</i>
Social Sciences (Inc. Behavioural Sciences)
Humanities (incl. Archaeology and Linguistics)
	<i>questionnaire completed</i>	

- 5 *Does access to research data pose problems of*
- technical difficulties ...
 - descriptive standards ...
 - institutional barriers ...
 - prohibitive cost ...
 - legal restrictions (privacy, IP, Nat. Security) ...

Could you please describe briefly your major concern?

- 6 *Is Access to and Sharing of Research Data subject of*
- a non binding recommendations from your organisation
 - b formal regulation (guidelines, funding terms, professional codes) from your organisation
 - c national legislation?

	<i>recommendations</i>	<i>regulation</i>	<i>legislation</i>
Natural Sciences (incl. Earth Sciences, Atmospheric Research)
Engineering & Technology
Life Sciences (incl. Environmental, Research, Bio diversity)

should provide exceptions useⁿ principles of copyright law.

	<i>recommendations</i>	<i>regulation</i>	<i>legislation</i>
Social Sciences (Inc. Behavioural Sciences)
Humanities (incl. Archaeology and Linguistics)

7 *Could you list the names and references of the policy documents concerned and/or the Website(s) where they can be found? (If possible, please attach an electronic version of the document(s) to your answer)*

8 *Does the policy of your organisation on Access to and Sharing of Research Data include*

- a Funding and/or managing of data archives/depositories
- b Co-operation with national (governmental) archives
- c Co-operation with governmental data collecting agencies/institutes
- d Selling data to and/or buying data from commercial firms

9 *Would your organisation be interested in the (follow-up) activities (being informed, participate in a policy workshop, participate in further consultation) of the Working Group?*

No ...

Yes ...

10 *If so, could you please give the co-ordinates of the person to contact?*

Full name

Privacy protection is
the dominant "reasonable limit"
to data sharing.

Appendix 2: Response Rate to the Mini-Survey

ORGANISATIONS THAT REPLIED

51

Australian Research Council
Austrian Research Council
Biotechnology and Biological Sciences Research Council, UK
CEA/DSM (Physics Department), France
CESNET Association, network of universities and academies, Czech Republic
Consiglio Nazionale delle Ricerche, Italy
Consejo Superior de Investigaciones Científicas (CSIC), Spain (twice)
Danish Research Agency
Deutsche Forschungsgemeinschaft, Germany
Engineering and Physical Sciences Research Council, UK
Estonian Academy of Sciences
Fonds voor Wetenschappelijk Onderzoek, Vlaanderen, Belgium
Hungarian Academy of Sciences
Icelandic Research Council
INFM, Italy
Information and Innovation Systems INRA, France
Irish Research Council for the Humanities and Social Sciences
Medical Research Council, UK
National Research Council of Canada
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Netherlands
Natural Environment Research Council, UK
Norwegian Academy of Science and Letters
Research Council of Norway, PBS/STR
Royal Irish Academy
Scientific and Technical Research Council (TÜBITAK), Turkey
Slovenian Academy of Sciences and Arts
Slovenian Academy of Sciences and Arts, Section Medical Sciences
Slovenian Science Foundation
Swedish Research Council

The researcher
is held responsible
for providing access to data.

Part II

Appendix 3: Methods

The following Web sites have been searched:

- National Research Council NRC www.nas.edu/nrc
- National Science Foundation NSF www.nsf.gov
- National Institutes of Health NIH www.nih.gov
- National Aeronautics and Space Agency NASA www.nasa.gov
- American Association for the Advancement of Science AAAS www.aaas.org
- National Archives NARA www.nara.gov
- National Endowment for the Humanities NEH www.neh.gov
- Inter-University Consortium for Political and Social Research ICPSR www.icpsr.umich.edu
- European Science Foundation ESF www.esf.org
- Library of Congress www.loc.gov

Every Web site has been searched twice. First, the Web site was searched on the keywords *data*, *sharing*, and *policy*. The documents retrieved were then studied for their relevance and, if relevant, downloaded for detailed study. After document analysis, the Web sites were visited again for a follow-up search using the particularities of the scientific fields at hand and/or of the Web site of the organisation.

This turned out to be especially useful where the practice of data sharing was referred to in other terms than data sharing, or where policy statements regarding data sharing were part of documents on other topics.

The searches were restricted to *policy documents*. This means that this Web scan did not aim to capture Web documents on the *practice of data sharing*. Some documents seemed to be midway between policy and practice. For example, pilot projects were being discussed or research proposals aimed at both a scientific and a policy audience. If the emphasis was on policy, these documents were included in this Web scan.

Based on the policy principles discussed at the GRV III conference, the retrieved Web documents were studied to answer the following questions:

- Is public access to data stated as a basic policy principle?
- What is the motivation for data-sharing rules?
- Is data-sharing a condition for research funding?

- Who is responsible for providing access to data?
- Are different types of data distinguished?
- How are issues of property rights treated?
- Which limits to data sharing are recognised as reasonable?

Appendix 4: Excerpts from the FOIA and the Bayh-Dole Act

THE FREEDOM OF INFORMATION ACT

55

The Freedom of Information Act regulates the accessibility of information in the US. In 1999, a provision was inserted in the Omnibus Appropriations Bill (Public Law 105-277) to change a federal regulation in order to allow broader access to federally funded research data. The provision, as inserted by Senator Richard Shelby (R-AL), tasks the Office of Management and Budget (OMB) to change OMB Circular A-110 so that *all* federally funded research data can be accessed through the mechanisms set forth in the Freedom of Information Act. OMB subsequently filed a proposed revision in the Federal Register on 4 February 1999 and allowed for a 60-day public comment period before any further actions would be taken. OMB's proposed revision reads:

The Federal Government has the right to (1) obtain, reproduce, publish, or otherwise use the data first produced under an award, and (2) authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes. In addition, in response to a Freedom of Information Act (FOIA) request for data relating to published research findings produced under an award that were used by the Federal Government in developing policy or rules, the Federal awarding agency shall, within a reasonable time, obtain the requested data so that they can be made available to the public through the procedures established under the FOIA. If the Federal awarding agency obtains the data solely in response to a FOIA request, the agency may charge the requester a reasonable fee equalling the full incremental cost of obtaining the data.

OMB received over 9,000 responses to its proposed revision with 55 percent of the respondents favouring the changes. Representatives of scientific organisations generally argued that the proposed amendment was anathema to the character of the research process and was not the most appropriate way to regulate access to research data. While several efforts were made in the 106th Congress to prevent any changes to OMB Circular A-110, none were successful. OMB released its second proposal on August 11, 1999, in the Federal Register. The proposal took into consideration the comments received from the February 4 proposal and greatly narrowed the scope of the Shelby amendment. The final revision was filed in the Federal Register on October 8, 1999.

THE BAYH-DOLE ACT

The Bayh-Dole Act was enacted in 1980 to spur the commercialisation of research results by granting patent rights to universities for inventions developed with federal funds. This includes exclusive licensing. The principles of the Bayh-Dole Act were the result of years of intense and emotional debate. The debate included questions whether exclusive licenses would lead to monopolies and higher prices; whether taxpayers would get their fair share; whether foreign industry would benefit unduly; and whether ownership of inventions by a contractor is anti-competitive. Economic interests rather than academic science interests were the driving forces for the change in US government policy. Until the Bayh-Dole Act became effective on July 1, 1981, the federal agencies kept tight control over intellectual property rights resulting from funded research, premised largely on traditional expectations rooted in the procurement process. After the passage of the Bayh-Dole Act, as the success of the Act became quickly apparent, subsequent legislative initiatives broadened its reach further.

Appendix 5: Members of the OECD Follow-Up Group on Issues of Access to Data from Public Funding

Mr. Peter Arzberger (Chair)

57

Director, Life Sciences Initiatives
University of California, San Diego
9500 Gilman Drive, MC 0043
La Jolla, CA 92093-0043
U.S.A.
parzberg@ucsd.edu

Mr. Geoffrey Bowker, Professor

Chair, Department of Communication
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0503
U.S.A.
Bowker@ucsd.edu

Ms. Sigrun Eckelmann

Deutsche Forschungsgemeinschaft (DFG)
Wissenschaftliche Literaturversorgungs- und Informationssysteme (LIS 4)
D-53170 Bonn
Deutschland
Sigrun.eckelmann@dfg.de

Dr. Tim Hubbard

Head of Human Genome Analysis
Wellcome Trust Sanger Institute
Wellcome Trust Genome Campus, Hinxton
Cambridgeshire, CB10 1SA
United Kingdom
Th@sanger.ac.uk

Researchers are becoming more on the increased

58

Mr. Koji Kamitani

Office of IT Promotion Research Promotion Bureau
MEXT (Ministry of Education, Culture, Sports, Science and Technology)
1-3-2 Kasumigaseki, Chiyoda-ku, Tokyo, 100-8966
Japan
kami@mext.go.jp

Mr. Leif Laaksonen

Street address:

CSC – Scientific Computing Ltd
Tekniikantie 15 a D
Otaniemi, Espoo, Finland

Postal address:

CSC – Scientific Computing Ltd
PO Box 405
FIN - 02101 Espoo Finland
Leif.Laaksonen@csc.fi

Ms. Gudrun Maass

Principal Administrator
Science and Technology Policy Division
Directorate for Science, Technology and Industry
Organization of Economic Co-Operation and Development
2, rue Andre-Pascal
F-75775 Paris Cedex 16
France
T +33 (0) 1 45 24 13 95
F +33 (0) 1 44 30 62 64 (fax)
gudrun.maass@oecd.org

dependent possibilities for data sharing.

Mr. Doug McEachern

Director Social, Behavioural and Economic Sciences
Australian Research Council
GPO Box 2720
Canberra
ACT 2601
Australia
Doug.Mceachern@arc.gov.au

Dr. David Moorman

Senior Policy Advisor, Policy and Liaison Branch
Social Sciences and Humanities Research Council
PO Box 1610
Ottawa ON Canada
K1P 6G4
david.moorman@sshr.ca

Mr. Masamitsu Negishi

Professor, Humanities and Social Science Information Research
Research Information Research Division / Director of Research Division
National Institute of Informatics
negishi@nii.ac.jp

Mr. Peter Schröder (vice-chair)

Co-ordinator Information Policy
Ministry of Education, Culture and Science
Directorate Research and Science Policy
P.O. Box 25000
2700 LZ Zoetermeer
The Netherlands
p.schroder@minocw.nl

Data sharing is not always uncontroversial in the scientific community.

60

Mr. Paul Uhlir

Director, International S&T Information Programs
National Academy of Sciences/National Research Council
2101 Constitution Avenue NW
Washington, DC 20418
U.S.A.
Puhilir@nas.edu

Mr. Mitsutoshi Wada

Manager, Office of Science and Technology Information
Japan Science and Technology Corporation (JST)
5-3, 4-Bancho, Chiyoda-ku, Tokyo 102-8666 Japan
wada@tokyo.jst.go.jp

Professor Andrzej P. Wierzbicki

Director General
National Institute of Telecommunications
Szachowa 1
04-894 Warsaw, Poland
a.wierzbicki@itl.waw.pl

Mr. Jan Windmüller

Head of Section
Ministry of Science, Technology and Innovation
Bredgade 43
DK - 1260 Copenhagen K
jwi@fsk.dk

Ms. Teri Simas

Administrative Co-ordinator
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0505
U.S.A.
simast@sdsc.edu

RESEARCHERS

Ms. Anne Beaulieu

Networked Research and Digital Information (Nerdi)
P.O. Box 95110
1090 HC Amsterdam
The Netherlands
<http://www.niwi.knaw.nl/nerdi>
anne.beaulieu@niwi.knaw.nl

Ms. Kathleen Casey

University of California, San Diego
Department of Communication
9500 Gilman Drive, MC 0503
La Jolla, CA 92093-0503
U.S.A.
kcasey@ucsd.edu

Mr. Colin Reddy

Networked Research and Digital Information (Nerdi)

P.O. Box 95110

1090 HC Amsterdam

The Netherlands

<http://www.niwi.knaw.nl/nerdi>

colin.reddy@niwi.knaw.nl

Mr. Paul Wouters

Networked Research and Digital Information (Nerdi)

P.O. Box 95110

1090 HC Amsterdam

The Netherlands

<http://www.niwi.knaw.nl/nerdi>

paul.wouters@niwi.knaw.nl

OBSERVERS**Ms. Jacky Bax**

Ministry of Education, Culture and Science – The Netherlands

Ms. Kathleen Cass

CODATA Secretariat – Paris, France

Mr. Tony Mayer

European Science Foundation – France

Mr. David Schindel

National Science Foundation – France

Mr. Hugo von Linstow

Global Diversity Information Facility – Denmark

Bibliography

AAAS Letter on the Notice of Proposed Rule Making (NPRM; Feb. 4, 1999, Vol. 64, No. 23, pp. 5684-5685 of the Federal Register) to amend OMB Circular A-110 (April 1999)

Anon. (2000), Prospect of data sharing gives brain mappers a headache, *Nature*, 406, p. 445

David d'Arcy (2001), Data hosts are vital to the Internet's future, *Nua Internet Surveys*, 2001, 3 December

Bruce Alberts (July 15, 1999), Statement of Dr. Bruce Alberts, President National Academy of Sciences before the Subcommittee on Government Management, Information, and Technology, Committee on Government Reform, U.S. House of Representatives, <http://www.nas.edu/nrc>

Duncan M. Brown (1997), Understanding Urban Interactions: Summary of a Research Workshop, <http://www.nsf.gov/pubs/1998/sb981/sb981.htm>, September 30, 1997

Eric G. Campbell and others (2002), Data Withholding in Academic Genetics. Evidence from a national survey, *Journal of the American Medical Association*, 287, no. 4, pp. 473-480

John W. Carlin (October 20, 1999) Statement by John W. Carlin, Archivist of the United States, to the Subcommittee on Government Management, Information, and Technology of the Committee on Government Reform, House of Representatives, Congress of the United States, <http://www.nara.gov/nara/vision/testimon.html>

Circular A-110 (revised) Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations (1999)

Council on Governmental Relations, *The Bayh-Dole Act – A Guide to the Law and Implementing Regulations* (1999)

Robin Cowan and Elad Harison (2001), Protecting the Digital Endeavour: Prospects For Intellectual Property Rights In The Information Society, MERIT – Maastricht Economic Research Institute on Innovation and Technology, MERIT-Infonomics Research Memorandum series 2001-028

Robin Cowan and Elad Harison (2001), Intellectual Property Rights In A Knowledge-Based Economy, MERIT – Maastricht Economic Research Institute on Innovation and Technology, MERIT-Infonomics Research Memorandum series 2001-027

Paul A. David (2001), Digital Technologies, Research Collaborations and the Extension of Protection for Intellectual Property in Science: Will Building ‘Good Fences’ Really Make ‘Good Neighbors’?, MERIT – Maastricht Economic Research Institute on Innovation and Technology, MERIT-Infonomics Research Memorandum series 2001-004

Paul A. David (2001), Tragedy of the Public Knowledge ‘Commons’? Global Science, Intellectual Property and the Digital Technology Boomerang, MERIT – Maastricht Economic Research Institute on Innovation and Technology, MERIT-Infonomics Research Memorandum series 2001-003

Ed. (2000), A debate over fMRI data sharing, *Nature Neuroscience*, 3, pp. 845–846

ESF (1999), The European Social Survey (ESS) – a research instrument for the social sciences in Europe. Report

H. Franken (2000), “Conference Conclusions” in: *Access to Publicly Financed Research*, The Global Research Village III Conference, Conference Report (P. Schröder, ed.), NIWI-KNAW, Amsterdam

The Freedom of Information Act 5 U.S.C. § 552, As Amended By Public Law No. 104-231, 110 Stat. 3048 (1996)

David M. Hart (2002), The “Corporatization” of Science, *Science*, 295, no. 5554, p. 439

Stephen Hilgartner (1998), Data Access Policy in Genome Research, in: Arnold Thackray (ed.) *Private Science. Biotechnology and the rise of the molecular sciences*, pp. 202-218, University of Pennsylvania Press, Philadelphia

P. Bernt Hugenholtz (2001), *The New Database Right: Early Case Law from Europe*, Paper presented at Ninth Annual Conference on International IP Law and Policy, Fordham University School of Law, New York, 19-20 April 2001

ICSU/CODATA Ad Hoc Group on Data and Information (November 30, 2000), *Scientific Data Policy Statements*, http://www.codata.org/data_access/index.html

ICSU/CODATA Ad Hoc Group on Data and Information (November 30, 2000), *Access to Databases. Principles for Science in the Internet Era*, http://www.codata.org/data_access/index.html

Charles Jennings and Peter Aldhous (2000), *Web discussion: Should neuroscientists share their raw data?*, *Nature*, 406, 25 August 2000, <http://www.nature.com/neuro/debate/>

Donald Kennedy (2001), *Enclosing the Research Commons*, *Science*, 294, no. 5550, pp. 2249

Joshua Lederberg (18 March 1999), *Hearing on the “Collections of Information Antipiracy Act. Statement of Joshua Lederberg, President-emeritus Rockefeller University on behalf of the National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and American Association for the Advancement of Science before the Committee on the Judiciary U.S. House of Representatives*

Anne Linn (2000), *History of Database Protection: Legal Issues of Concern to the Scientific Community*, http://www.codata.org/data_access/linn.html, March 3, 2000

Stephen M. Maurer and Suzanne Scotchmer (1999), *Database Protection: Is It Broken and Should We Fix It?*, *Science*, 284, no. 5417, pp. 1129—1130

Stephen M. Maurer and P. Bernt Hugenholtz and Harlan J. Onsrud (2001), *Europe’s Database Experiment*, *Science*, 294, no. 5543, pp. 789—790

NARA, *Strategic Plan of the National Archives and Records Administration 1997-2007*

NARA *Access to Records in the National Archives and Records Administration*

NARA Regulations to its Holdings

NASA, Guidelines for Documentation, Approval, and Dissemination of NASA STI (valid until September 2002)

NASA, External Release Of NASA Software (NPD 2210.1)

NASA, Management of NASA Scientific and Technical Information (STI) (NPD 2220.5E)

NASA Procedures and Guidelines (NPG) 2200.2A, Guidelines for Documentation, Approval, and Dissemination of NASA Scientific and Technical Information

NASA (2001), NASA Earth Science Enterprise Statement on Data Management, <http://www.earth.nasa.gov/visions/data-policy.html>, 10 July 2001

LBA Science Steering Committee (1998), LBA Data and Publication Policies, http://lba-hydromet.gsfc.nasa.gov/policies/lba_data_policies.htm

NEH (2000), Report of the Humanities, Science and Technology Working Group, National Endowment for the Humanities

NIH Response to Notice of Proposed Rule Making (NPRM; Feb. 4, 1999, Vol. 64, No. 23, pp. 5684-5685 of the Federal Register) to amend OMB Circular A-110

Office of Extramural Research, National Institutes of Health (2001), NIH Grants Policy Statement 03/01, http://grants.nih.gov/grants/policy/nihgps_2001/

Office of Extramural Research (2002), National Institutes of Health. Frequently Asked Questions on Data Sharing, http://grants1.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm, March 1

NIH Principles and Guidelines for Sharing of Biomedical Research Resources (December 1999)

data-sharing need to be offered offset investigator's loss of control over their data-bases.

NIH-DOE Guidelines for access to mapping and sequencing data and material resources

Working Group on Research Tools, National Institutes of Health (1998), Report of the National Institutes of Health (NIH) Working Group on Research Tools, National Institutes of Health

NIH (1999), Principles and Guidelines for recipients of NIH research grants and contracts on obtaining and disseminating biomedical research resources: final notice, 23 December 1999

The National Human Genome Research Institute (2001), NIH-DOE Guidelines for Access to Mapping and Sequencing Data and Material Resources,
http://www.nhgri.nih.gov/Grant_Info/Funding/Statements/data_release.html,

National Advisory Mental Health Council (2000), Minutes of the 196th NAMHC Meeting,
<http://www.nimh.nih.gov/council/min900.cfm>, 15 September 2000

National Advisory Mental Health Council (1998), Minutes of the 188th NAMHC Meeting,
<http://www.nimh.nih.gov/council/min900.cfm>, February 4, 1998

NRC Committee on National Statistics (1985), *Sharing Research Data*, National Academy Press, Washington DC

Committee on Applied and Theoretical Statistics, National Academy of Sciences/National Research Council (1995), *Massive Data Sets. Proceedings of a workshop, July 7-8, 1995*,
<http://books.nap.edu/html/massdata/>, 7-8 July 1995

NAS (1999), *Global Ocean Science – Toward an Integrated Approach*, <http://www.nap.edu>

Mapping Science Committee, Board on Earth Sciences and Resources, Commission on Geosciences, Environment, and Resources, National Research Council (1997), *The future of spatial data and society: summary of a workshop*, National Academy Press,
<http://books.nap.edu/html/spa/>

National Research Council (1997), *Bits of Power. Issues in Global Access to Scientific Data*, National Academy Press, Washington DC

National Research Council (1999), *A Question of Balance. Private Rights and the Public Interest in Scientific and Technical Databases*, National Academy Press, Washington DC

Commission on Physical Sciences, Mathematics, and Applications, National Research Council (2000), *The Digital Dilemma. Intellectual Property in the Information Age*, National Academy Press, Washington DC

NSF Grant Policy Manual (1995), NSF

Addendum to the NSF Grant Proposal Guide (June 2001), NSF

NSF Social and Economic Sciences (1995), *Connecting and Collaborating: Issues for the Sciences. Report of a workshop sponsored by the NSF and held at the Walter and Judith Munk Laboratory of the Scripps Institution of Oceanography, University of California, San Diego*, <http://www.nsf.gov>, June 22-24, 1995

The Division of Behavioral and Cognitive Sciences (2001), *Data Archiving Policy*, <http://www.nsf.gov>

NSF (1999), *Realizing the Potential of Plant Genomics: From Model Systems to the Understanding of Diversity*, <http://www.nsf.gov/pubs/2001/bio011/start.htm>

The Governing Council of the Organization for Human Brain Mapping (2001), *Neuroimaging Databases*, *Science*, 292, 5522, pp. 1673–1676

Jason Owen-Smith (2002), *Intellectual Property: Between the ivory tower and the market*, *Science*, 295, no. 5561, pp.1840

Andrew J. Pincus (18 March 1999), Statement of Andrew J. Pincus, General Counsel, United States Department of Commerce, before the Subcommittee on Courts and Intellectual Property, Committee on the Judiciary U.S. House of Representatives

Pamela Samuelson (2001), Anticircumvention Rules: Threat to Science, *Science*, 293, no. 5537, pp. 2028-2031

Mark Sincell (1999), Physicists and Astronomers Prepare for a Data Flood, *Science*, 286, no. 5446, pp. 1840-1841

Erik Stokstad (2002), Data Hoarding Blocks Progress in Genetics, *Science*, 295, 5555, p. 599

U.S. Environmental Protection Agency (July 24, 1995), Information Resources Management (IRM) Policy Manual, <http://www.epa.gov/irmpoli8/>, EPA Directive Number 2100

United States Geological Survey (USGS) (15 August 2001), U.S. Geological Survey Manual, <http://www.usgs.gov/usgs-manual/>

Data and Information Working Group, U.S. Global Change Research Program (2001), <http://www.globalchange.gov>, 17 December

Thomas H. Mace (10 March 1999), DMWG Response to OMB about Suggested FOIA Changes to A-110, <http://www.globalchange.gov>,

Subcommittee on Global Change Research (June 26, 1998), Data Management for Global Change Research, <http://www.globalchange.gov>

R. Corell (October 6, 1997), DMWG “Full and Open” Definition, <http://www.globalchange.gov>

Thomas H. Mace (August 20, 1997), DMWG Policy on Data from Federal Grants, <http://www.globalchange.gov>

R. Corell (October 30, 1996), DMWG Position on Proposed World Intellectual Property Organization Action, <http://www.globalchange.gov>

D. Allan Bromley (July 2, 1991), DMWG Global Change Data Policy Statements, <http://www.globalchange.gov>

Wouters, P. and P. Schröder, Eds. (2000). *Access to Publicly Financed Research : The Global Research Village III*, NIWI-KNAW, Amsterdam

the 1990s, the number of people in the UK who are employed in the public sector has increased from 10.5 million to 12.5 million, and the number of people in the public sector who are employed in the health sector has increased from 2.5 million to 3.5 million (Department of Health 2000).

There are a number of reasons for this increase in the number of people employed in the public sector. One of the main reasons is the increasing demand for public services, particularly in the health sector. This is due to a number of factors, including an ageing population, an increasing incidence of chronic diseases, and a growing demand for health care services. Another reason for the increase in the number of people employed in the public sector is the increasing number of people who are employed in the public sector on a part-time basis.

There are a number of reasons for the increasing number of people employed in the public sector on a part-time basis. One of the main reasons is the increasing demand for public services, particularly in the health sector. This is due to a number of factors, including an ageing population, an increasing incidence of chronic diseases, and a growing demand for health care services. Another reason for the increase in the number of people employed in the public sector on a part-time basis is the increasing number of people who are employed in the public sector on a part-time basis.

There are a number of reasons for the increasing number of people employed in the public sector on a part-time basis. One of the main reasons is the increasing demand for public services, particularly in the health sector. This is due to a number of factors, including an ageing population, an increasing incidence of chronic diseases, and a growing demand for health care services. Another reason for the increase in the number of people employed in the public sector on a part-time basis is the increasing number of people who are employed in the public sector on a part-time basis.

There are a number of reasons for the increasing number of people employed in the public sector on a part-time basis. One of the main reasons is the increasing demand for public services, particularly in the health sector. This is due to a number of factors, including an ageing population, an increasing incidence of chronic diseases, and a growing demand for health care services. Another reason for the increase in the number of people employed in the public sector on a part-time basis is the increasing number of people who are employed in the public sector on a part-time basis.

There are a number of reasons for the increasing number of people employed in the public sector on a part-time basis. One of the main reasons is the increasing demand for public services, particularly in the health sector. This is due to a number of factors, including an ageing population, an increasing incidence of chronic diseases, and a growing demand for health care services. Another reason for the increase in the number of people employed in the public sector on a part-time basis is the increasing number of people who are employed in the public sector on a part-time basis.

There are a number of reasons for the increasing number of people employed in the public sector on a part-time basis. One of the main reasons is the increasing demand for public services, particularly in the health sector. This is due to a number of factors, including an ageing population, an increasing incidence of chronic diseases, and a growing demand for health care services. Another reason for the increase in the number of people employed in the public sector on a part-time basis is the increasing number of people who are employed in the public sector on a part-time basis.

There are a number of reasons for the increasing number of people employed in the public sector on a part-time basis. One of the main reasons is the increasing demand for public services, particularly in the health sector. This is due to a number of factors, including an ageing population, an increasing incidence of chronic diseases, and a growing demand for health care services. Another reason for the increase in the number of people employed in the public sector on a part-time basis is the increasing number of people who are employed in the public sector on a part-time basis.