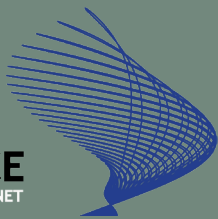


NORFACE
ERA-NET



Data and Research Infrastructure

Social Sciences

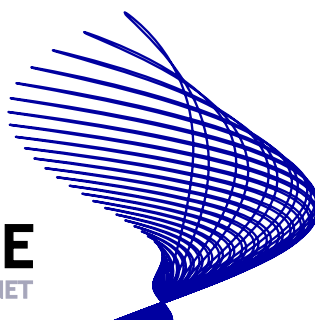
Economics

500

Data and Research Infrastructure
Recommendations to the NORFACE Partners
and
Proceedings of the NORFACE Conference on
Data & Research Infrastructure

16-17 November 2006
The Hague, The Netherlands

NORFACE
ERA-NET



ISBN 978-951-715-715-5 (Print)
ISBN 978-951-715-716-9 (pdf)

Cover design: Studio Viiva
Layout: PixPoint ky
Edita Prima Oy, Helsinki 2009

Colophon

This report has been composed by Netherlands Organisation for Scientific Research (NWO) Social Sciences Postal address P.O. Box 93461 2509 AL The Hague, The Netherlands

Co-ordinator of NORFACE

Academy of Finland
Culture and Society Research Unit
P.O. Box 99
FIN-00501 Helsinki, Finland.
Telephone +358 9 774 881
www.norface.org

The Hague, February 2007

NORFACE

Contents

Introduction	7
Recommendations	9
Recommendations for the immediate Future: the NORFACE Transnational Research Programme	9
Recommendations for the further Future: Data & Research Infrastructure Strategy	10
Proceedings of the NORFACE Conference on Data & Research Infrastructure	15
Opening by Anne Kovalainen (FI)	15
N. Bradburn (USA): Key note speech ‘Recent developments in the social sciences: growing dependency on data & research infrastructure; increase in scale and internationalization’	15
Questions and discussion	22
P. Elias (UK): Introduction to theme 1: National roadmaps	23
Questions and discussion time	31
Report from the workshops by discussion leader S. Barkçin (TR; ESFRI) and rapporteur L. Mizera (EE)	32
T. Risch (SE): Introduction to theme 2: Data access, including remote access, new ICT possibilities, international access to data.	39
Questions and discussion time	48
Report from the workshops by discussion leader R. Procter (UK) and rapporteur Sh. Duffy (IE)	49
P. Doorn (NL): Introduction to theme 3: Data networking in Europe	50
Report from the workshops by discussion leader E. Mochmann (DE) and rapporteurs M. van Leeuwen (NL) and L. Henriques (PT)	60
B. Henrichsen (NO; ESFRI): Presentation of the ESFRI Roadmap	61
D. Moorman: Key note speech ‘OECD and Research & Infrastructure’	66
Questions and discussion time	67
Plenary discussion with K. Schürer, discussion leader	68
Question and discussion time; other recommendations	68
Enclosures	70
Description of NORFACE Task: Task 4.7 – Data and Research Infrastructure	70
Conference Programme (timetable)	71
Conference Programme (content)	72
List of participants of the conference	75
Questionnaire survey	78
Important documents referred to during the conference	83
OECD Recommendation Concerning Access to Research Data from Public Funding	84

NORFACE

Introduction

This report contains recommendations with regard to the data & research infrastructure in the social sciences for the twelve European funding agencies and the associate partner in Canada co-operating in the ERA-NET NORFACE. The recommendations are from the NORFACE Conference on Data & Research Infrastructure on 16 and 17 November 2006 in The Hague, The Netherlands. The Conference was chaired by Dr. Renée van Kessel-Hagesteijn, Director NWO Social Sciences (NL), and Dr. Jan Karel Koppen, NWO Member NORFACE Network Board (NL).

This NORFACE Conference concluded the activities of the NORFACE Task on Data & Research Infrastructure (Task 4.7). NWO as Task leader acknowledges gratefully that we could not have executed this task without the support, advice and helpful suggestions from the Dialogue Partners: the United Kingdom (Suzanne Tanner, Economic and Social Research Council), Denmark (Lars Christensen, Danish Research Agency for Science, Technology and Innovation), and Ireland (Dipti Pandey and her representative Sheena Duffy, Irish Research Council for the Humanities and Social Sciences). NWO also wishes to thank the two Experts in the field of Data & Research Infrastructure, Professor Peter Elias (UK) and Dr. Peter Doorn (NL).

NWO organized two preparatory meetings, in March with the Dialogue Partners, and in June 2006 with the Dialogue Partners and the

Experts. Based on the information gathered by a questionnaire sent to all NORFACE partners plus additional information sent by the respondents and helped by suggestions from the Dialogue Partners and both Experts, the programme of the Conference in November 2006 was designed, and the speakers, discussion leaders, rapporteurs, and Conference participants were invited.

In the Conference we focussed on the 'NORFACE niche' with regard to data & research infrastructure: we looked for ways for the Funding Agencies together to support and enhance social sciences research at a European level, to promote a more easy transnational access to databases, and to take transnational research to a wider, higher level and a better quality.

The recommendations from the Conference will first be offered to the NORFACE Management Team which meet in May 2007. Subsequently the recommendations will be offered to the NORFACE Network Board, which will meet in summer 2007. After acceptance by the NORFACE Network Board the NORFACE Partners will be invited to present these NORFACE recommendations to their respective Governing Boards. We hope that the recommendations will enrich the NORFACE Transnational Research Programme as the final activity of the NORFACE Consortium, and will also have future effects on the Funding Agencies' policies in data & research infrastructure, after the end of this ERA-NET.

NORFACE

Recommendations

We want to make recommendations for the immediate future – in which the NORFACE Transnational Research Programme is a prominent feature –, and for the further future, the period both during and after the end of the ERA-NET NORFACE.

The recommendations for the immediate future are of a practical nature, and are directed towards the Transnational Research Programme. The recommendations for the further future have a strong strategic character; most of the NORFACE conference on 16+17 November 2006 had to do with strategy and many considerations arrived at there, are recommendations concerning strategy.

Recommendations for the immediate Future: the NORFACE Transnational Research Programme

The Call for proposals

Before the Call NORFACE may need to make funding available explicitly for data & research infrastructure.

We recommend that in the Call for transnational research proposals there will be reference to the data & research infrastructure. If there is data collection during a research project, the research proposal will contain information on – and budget reservation for – the following aspects of data & research infrastructure:

Archiving

The data collected during the project will be deposited in a data archive, preferably in the country of the project leader. The place of deposition may be a data archive; or it may be the research institute or university where the project will take place, in which case the project leader arranges that the material will be put into a data-repository, or a trusted digital repository which is open to researchers from all over the world.

Preparations for the deposit of the data will be visible in the project proposal.

Documentation

The data will be documented in such a way that any other researcher from anywhere in the world will be able to find them, understand them, and reuse them. This documentation or metadata often is a book of codes. The data should be documented according to the specifications of the Data Archive involved (for example: in the Netherlands this means that the Dutch Data Archiving and Networked Services DANS will be involved.). Constructing the documentation or metadata will take time; the research proposal must show a reservation of time and budget for this specific activity.

Availability

NORFACE favours Open Access to research data that is publicly funded, in accordance with the OECD guidelines and the Berlin Declaration. Provisions should be made that the privacy of respondents is guaranteed. Considerations relating to a temporary embargo is acceptable for a time period up to two years.

Data contract

NORFACE recommends that the conditions of the data deposit are specified in a “data contract” between the project coordinator and the data archive, to be made at the start of the project. In some countries data contracts exist (for example, in The Netherlands contracts are with the Dutch DANS, Data Archiving and Networked Services). The project proposal should already give indications on how permanent access to the data will be guaranteed, giving information on data formats, metadata, safety measures and digital conservation.

If appropriate, the research proposal will include a section on dissemination concerning the secondary use of the data & research infrastructure – for instance by means of a workshop.

NORFACE letters of subsidy assignment

In the letter containing the formal decision that a project will be financed, reference will be made to data & research infrastructure. The arrangements for the data collected during the project will be referred to. As an instrument to enforce the data deposit, it is recommended that the final ten percent of the budget be paid after the data have been deposited.

OECD Standards on Open Access

If the OECD Standards on Open Access to research data from public funding are endorsed by then, they will be offered together with the Call for proposals, by way of background information.

Recommendations for the further Future: Data & Research Infrastructure Strategy

In the Conference the subject of data access was discussed in terms of crossing boundaries between countries. NORFACE should discuss data access also in terms of crossing boundaries between disciplines. The social sciences could also learn from the mechanisms for international co-operation, data use, et cetera applied by other academic disciplines. CESSDA has a transborder data agreement regulating the admission to foreign data.

Illustration: There is a general strategy to break down barriers between disciplines. Cyber-infrastructure lends itself well to a programme that is there for all scientific disciplines.
(key note speech Bradburn)
For infrastructure we do not want to have any walls between the different sciences.
(introduction Risch)

Developing Data & Research Infrastructure and Costs

The NORFACE Funding Agencies should make sufficient subsidy funds available for data & research infrastructure. A rule of thumb may be

25% of the total research funding to be reserved for infrastructure. Concerning metadata: for projects where data collection is involved, metadata should be financed as well, for approximately 10% of the total project budget.

Recommendation explained: A rule of thumb at NSF was to devote about 20 to 25% of research funds to infrastructure. A recent study by the National Science Board (NSB, 2003) suggests increasing the level of support to approximately 28% to take advantage of technological advances, particularly in computing. So, in the range of 20 to 25% of research funds might be devoted to infrastructure in times of fairly measured technological advances, with an increase to the 25 to 30% range when, as is now the case, there are rapid advances that open up large new areas of research opportunities (key-note speech Bradburn). The big costs in the social sciences are the data collection. On top of data collection you need about 10% for metadata. (introduction Risch).

National Roadmaps (Conference Theme 1)

There are excellent examples of national roadmaps and of the process of creating such an important document (e.g. the UK National Data Strategy). Every country should – in its own way and with the support of its researchers and other stakeholders – develop a national roadmap. NORFACE should play a role in the exchange of roadmaps amongst Funding Agencies and sharing of processes. Developing a national roadmap encourages countries to revise their situation and needs with regard to data & research infrastructure, and offers input to Ministries for data & research infrastructure policies.

Recommendation explained: The UK roadmap illustrates that national roadmaps should be preceded by national policies or strategies: the UK roadmap did not originate in a vacuum but originated in a situation where people were already thinking about policies or strategies concerning data. The UK social sciences research infrastructures have been

ahead compared to many other European countries.

A national roadmap allows stakeholders of social sciences research, including the government, to know where researchers want to go, and maps and addresses the national imbalances in research and research infrastructure. About identifying and funding priority areas, pitfalls were seen here. The identification of priority areas may complicate the roadmap process, the defined areas may become a “straight jacket”, a list of priority areas may draw attention to what is not on the list, it is impossible to foresee topics that are felt to be urgent in the (near) future (with reference to the 9/11 aftermath) (workshop theme 1, Barkçin/Mizera).

E-science and the GRID (Conference Theme 2)

Stimulating the use of the new facilities

NORFACE should pay more attention to E-science and the GRID. This topic has been addressed little as yet. How are the Funding Agencies and governments engaged here? (plenary discussion, Schürer) All Funding Agencies should develop an E-science programme in their country; E-science is a ‘big thing’, concerns sharing data, is supporting collaboration across time and distance (at present, only the United Kingdom and The Netherlands have E-science programmes). (plenary discussion, Schürer)

Metadata

Funding Agencies should pay special attention to metadata by providing sufficient financing for this aspect of data & research infrastructure. On top of data collection an extra 10% should be subsidised for the development of metadata in any project (see also the paragraph ‘Developing Data & Research Infrastructure and Costs’ above). There is some work that NORFACE could do on co-ordinated policies for licenses, IPR and data governance. Certainly some exchange of best practices for adopting and sustaining infrastructure including training etcetera would be very useful.

About supporting further development and promoting the use of these possibilities

The NORFACE need to co-ordinate efforts. They should share experiences of barriers (different countries are at different stages of infrastructure adoption), agree on best practices across a range of activities needed to adopt and sustain infrastructure (training, infrastructure development project management). Partners should share exemplars of research enabled by new infrastructure. Partners should translate infrastructure benefits for social sciences into their respective national research agendas and thereby help to move this up their research respective agendas to create top-down leverage. (workshop theme 2, Procter/Duffy)

Access to national data resources in order to support comparative research

Licence arrangements across national boundaries are virtually non-existent. Incompatibility of national IPR regimes is a real barrier as are differences in approaches to the use of personal data. Partners need to work together to co-ordinate common policies for licensing, IPR and data governance. (workshop theme 2, Procter/Duffy)

Promoting access to international data, e.g. by Eurostat and other international bodies

There is a need for better access to EUROSTAT, to integrate academic, statistical and commercial data. (workshop theme 2, Procter/Duffy)

Funding Agencies uniting with their social scientists to achieve certain ends

Top-down leverage must be complemented by bottom-up mechanisms. Adoption will not succeed without the buy in of research community members. The NORFACE partners can encourage this by investing in training, promoting cultural change by rewarding good practices in data publication (e.g., by counting data publication as an output on a par with papers). Partners should work with research community members to define standards for data publication, make it condition of funding projects and ensure they make available the resources to fund it. (workshop theme 2, Procter/Duffy)

Data archiving (Conference Theme 3)

Work would be welcomed on the promotion of standards for data processing/documentation – sharing best practice. This could also include a study on barriers to sharing of data – Definitely NORFACE could become involved in best practice here. Following the transnational programme there could be an audit of data at national levels.

Policy on data & research infrastructure should both be developed at a national level and at a European level. Up to now the existing infrastructure is primarily national, and there is a considerable European challenge in this area.

The NORFACE Funding Agencies should encourage their researchers – or develop a policy to that end – that research data are deposited in data archives. We recommend that in grants for research during which data are gathered, a budget for archiving and documentation of data will be specified.

The workgroup on this theme – discussion leader was E. Mochmann and rapporteurs were M. van Leeuwen and L. Henriques – has developed important recommendations for promoting international data sharing, for access to administrative data and researchers data, and for working together with national statistical bureaus or offices in the NORFACE countries.

ESFRI

The NORFACE Funding Agencies should play a role in organising the social sciences in order to support the ESFRI Roadmap proposals, and they should interest the Ministries in their countries for these proposals. The NORFACE Funding Agencies are urged to engage in the ESFRI Roadmap, early and timely. Moreover, the ESFRI proposals for the social sciences and humanities can be seen as interdependent. CESSDA covers the social sciences, DARIAH the humanities, and EROHS covers both. Together they cover the overall data & research infrastructure in the social sciences and humanities and will

supply networked services to the content oriented proposals (SHARE, ESS, CLARIN) (introduction Doorn). The NORFACE Funding Agencies should see that taking the proposals in one stride will be the strongest move forward strategically.

Recommendation explained: The ESFRI roadmap proposals need to be brought forward. Institutions in the EU member states need to support the proposals, otherwise the exercise of drawing a roadmap will prove to be a failure. The social sciences are called upon to get organised – in which process the NORFACE Funding Agencies can play a role – and to approach their Ministers for Science. When a Call for proposals is launched, Ministries or Funding Agencies should apply (not the projects). Certainly support for the ESFRI proposals is expected, and needed, from the Ministries in Europe. (presentation Henrichsen). The complementarity of the six ESFRI proposals lies, for instance, in their visionary aspects, in their visibility, their practicality, and in their short and long term aims: taken together into a bundle, the proposals in the field of the social sciences and the humanities optimize their chances against proposals from strong other academic disciplines. A very important report is the document of the Social Sciences Workgroup. The NORFACE Funding Agencies should see how they would wish to engage. Engagement is the keyword here: authors of such reports and the Funding Agencies need to engage. None of the ESFRI proposals will go forward unless the Ministries are urged by the research councils. (plenary discussion Schürer).

OECD Standards

The NORFACE Funding Agencies should begin working with the (thirteen) OECD principles and guidelines on Open Access. The standards concern: Openness, Flexibility, Transparency, Legal Conformity, Protection of Intellectual Property, Formal Responsibility, Professionalism, Interoperability, Quality, Security, Efficiency, Accountability, Sustainability. (See also the

presentation by David Moorman, the plenary discussion following it, and the Enclosure 'OECD Recommendations concerning access to research data from public funding' – the authors).

*Other Issues from the Plenary Discussion
(discussion leader Schürer)*

- The NORFACE Funding Agencies should think more about investment and training for use of data & research infrastructure in general. Social science researchers do not use the infrastructure facilities enough.
- Virtually no discussion has yet been on how data can be transported across borders. Do researchers have the right, the tools, the knowledge to use data in other countries? The NORFACE Funding Agencies should consider this aspect more.
- In the European Union's Seventh Framework Programme Research and Technological Development preparatory bids will be called for concerning a data & research infrastructure activity; the instrument will probably resemble the Co-operation Actions (ERA-NETs) or COST (co-operation activities called 'Actions'). The NORFACE Funding Agencies should consider responding to the call.
- The NORFACE Funding Agencies should respond to the need of researchers for models being set for bilateral agreements on the use of data; NORFACE could develop best practice on sharing of data and bilateral agreements.
- The NORFACE Survey on data & Research Infrastructure contains interesting information from the NORFACE partners. The Survey reveals a lot about national research councils' policies on data & research infrastructures. It is recommended that the Funding Agencies learn from it.
- There is no relation visible between the NORFACE Survey and the European Framework Programmes. Researchers in EU projects need EU data. The NORFACE Funding Agencies should address this issue more, and more explicitly.
- As for the cost aspect of Open Access, the

NORFACE Funding Agencies are advised to consider an international procurement of data. Example: the British NORFACE partner ESRC buys licenses from the World Bank; while they may only be used by British researchers, the UK Data Archive is inundated by requests from abroad to use these data. A collective action with many Funding Agencies will lead to a cheaper buy and one for the whole of Europe.

- The NORFACE Funding Agencies should consider a question on the international management of data, and should investigate what the most fruitful way is towards an international management of data, what the limitations are and how we dismantle the limitations and the barriers.

Crossing Boundaries within the Social Sciences, and Boundaries Between the Social Sciences and Other Science Disciplines, including Humanities

With regard to data & research infrastructure NORFACE should not hold on to barriers within the social sciences, or between the social sciences and humanities; co-operation with the humanities should be sought. NORFACE and the social sciences in general should understand that other academic disciplines have built up a real advantage concerning data & research infrastructure.

Recommendation explained: NORFACE must be aware of the concept of 'maturity'. The science community has got a real advantage in this area – they are organised, fundamental science driven. The social sciences have a real problem, unless they can take up an activity – compare the UK Strategy. A new initiative is needed, driven by the research community and perhaps with a role for NORFACE. (introduction Elias)

NORFACE – Europe – the World

NORFACE should keep a view on developments in research infrastructures both in Europe and in the world at large.

NORFACE

Recommendation explained: ESFRI is very significant indeed, especially because it has increased the level of awareness between researchers, that there is an interdependence between them with regard to scarce resources and that they need to discuss how to allocate funds. Nevertheless, NORFACE should also look beyond Europe, because

data & research infrastructure is a global topic. NORFACE should realize that its co-operation is between Funding Agencies only in twelve European countries plus Canada – the consortium should try to find leverage for the remaining 23 European countries. There should be an international aspect to any NORFACE policy.

Proceedings of the NORFACE Conference on Data & Research Infrastructure

16 + 17 November 2006, The Hague, The Netherlands Opening by Anne Kovalainen (FI)

The conference is opened by Anne Kovalainen, chair of the NORFACE Network Board. She emphasizes and illustrates the importance of addressing the topic of data & research infrastructure in the social sciences and wishes all participants an enjoyable and successful conference.

N. Bradburn (USA): Key note speech 'Recent developments in the social sciences: growing dependency on data & research infrastructure; increase in scale and internationalization'

Norman Bradburn illustrates his key note speech by means of slides (powerpoint presentation). Both the text spoken and his slides are presented here. For his references to literature please see Enclosures at the back of this document.

Two of the most important driving forces for progress in science are new ideas and new kinds of observations that are made possible by technological innovations. We cannot predict when new ideas, especially fundamentally new ideas, will arrive on the scene, nor do we have a very good idea of how to promote their development. They seem to spring, like Athena,

from the heads of our great theorists, although in retrospect precursors and protoideas can often be identified.

On the other hand, the technologies for making the observations that are the basis for testing ideas, and in many cases, the data that give rise to new theoretical ideas and hypotheses, are more predictable. We can more easily see them coming; we can invest in their development; we can promote their use and the training necessary to enable scientists to use them effectively. Broadly speaking these technologies are what I think of under the heading of "infrastructure". I would like to address today some of what I believe are the most important recent developments in social science infrastructure and the implications of these advances for social science funding agencies.

Definition of Infrastructure



Definition of Infrastructure

- Research tools and technologies
- Analytic methods
- Web of organization and communications of scientists

NORC AT THE UNIVERSITY OF CHICAGO
A National Organization for Research and Computing



**Recent Developments in the
Social Sciences: Data and
Research Infrastructure**

Norman M. Bradburn
NORFACE Conference
The Hague
16 November, 2006

NORC AT THE UNIVERSITY OF CHICAGO
A National Organization for Research and Computing

The term "infrastructure" is not a precise term. It is defined differently by different funding agencies and differs in detail across the diverse sciences. I shall use the term to refer to the research tools and technologies, to the analytic methods such as statistical inference and mathematical modeling, and to the web of organization and communications of scientists that promote the

effective pursuit of science. In the most general sense, infrastructure is what enables scientists to conduct their research. Without it, scientists must provide for themselves or invent the means to carry out their scientific inquiry. With solid infrastructure support, scientists are able to pursue their work to the limit of their abilities.

We can distinguish between the infrastructure that is available to particular scientists in their own institutions and the general infrastructure that is available to all or most scientists in a field. It is this more general sense of infrastructure that I will be talking about today. It is what is most important in influencing the advancement of the science as distinguished from the productivity of individual scientists.

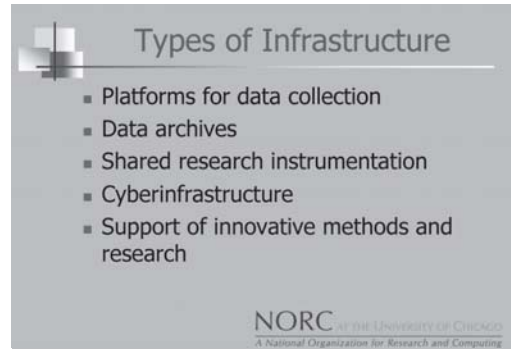
What are the principal forms of infrastructure in the social sciences? While there is no agreed upon taxonomy, a National Research Council report distinguishes 4 types of infrastructure: social, communicative, mechanical and intellectual (CBASSE, 1998) Social infrastructure refers to the resources needed to promote research collaboration within and between fields. Interdisciplinary research centers are a part of social infrastructure. Communicative infrastructure includes the Internet and other forms of electronic communication. Scientific journals and other institutionalized means of disseminating research findings are part of this kind of communicative infrastructure.

Mechanical infrastructure consists in specialized equipment such as magnetic resonance imagers and positron emission tomographers, as well as many kinds of specialized equipment that various behavioral and social scientists use in their research, such as equipment to date anthropological field specimens or that used by linguists to record and analyze languages. Intellectual infrastructure refers to large scale data bases and the methodological developments necessary for the sophisticated analysis of the data.

While this scheme gives a high level general framework within which to think about social science infrastructure, I prefer a categorization

that is more oriented toward the types of activities that are typically supported by research funding agencies.

Based on my observations of the types of infrastructure grants that we made in the Social



Science Directorate at the U. S. National Science Foundation, I divide social science infrastructure into 5 categories. These types are: 1) platforms for data collection; 2) data archives; 3) shared research instrumentation; 4) cyberinfrastructure; and 5) centers for support for innovative methods and research. These categories can be mapped on to the general framework used by the NRC. Types 1 and 2 are examples of the intellectual infrastructure; Type 3 is an example of mechanical infrastructure; type 4 is a mix of communicative and mechanical infrastructure; and Type 5 is an example of social infrastructure. I shall discuss each of these briefly.

1 Platforms for data collection



I refer here to large scale periodic data collection projects that develop comparable data over time

and space. Well known examples are the General Social Survey in the United States, the European Social Survey (ESS), the Survey of Health, Ageing and Retirement in Europe (SHARE), and the World Values Survey. The generic characteristics of these examples are that they are sample surveys of individuals living in households; they are conducted regularly using the same basic sampling frame; they usually have a core set of questions to measure social change but allow for a shifting set of topics that are studied in more depth at one point in time. They provide data that are then made publicly available and extensively used by social scientists to develop and test hypotheses about social conditions and social change.

Because of the need for methodological and temporal continuity, these platforms require long-term support if they are to realize their full value to the social sciences. Like good wine, they become better with age. Their value increases the longer the time period they cover. The General Social Survey in the U.S., which started in 1972, has measures of some variables as far back as the 1940s because it drew on some existing data series when it started. Without investments in staff, archiving, dissemination, and a governing mechanism to insure their continued relevance to the social scientific questions of the day, their value erodes and they can become outdated. Continuing funds are needed for methodological research to improve measurement, for research to improve methods of translation to assure equivalence, to standardize the classification of basic data such as occupation and education, and to compile and harmonize aggregate context variables to aid in the cross-national interpretation of the data. And of course, funds are needed to pay for the basic data collection and processing costs.

The examples just cited are part of specifically social science data programs funded, for the most part, by individual Research Councils working together, although some may be funded by government statistical or research agencies. The need for pan-European financing of such data series has been recognized by the European Commission in its financing of the ESS. The recent ESFRI Roadmap for Research

Infrastructure (ESFRI, 2006) calls for increased support for the ESS and for SHARE.

Of course, not all data series are collected specially for social scientists. Much of the data used in social science research, particularly economic research, is the product of governmental statistical agencies. While official statisticians of the OECD countries are working to harmonize their various surveys, much still needs to be done to increase the cross-national comparability of official statistics, particularly in areas such as education, where national differences in the organization of the education system make it very difficult to provide strictly comparable data. There is ample scope for the development of more social scientifically informed infrastructure that would provide, at a minimum, cross-walks between the official statistics of different European countries so that better comparative studies could be carried out. The Luxembourg Income Study has been doing this for income and labor force statistics, and, soon, for wealth data.

2 Data Archives

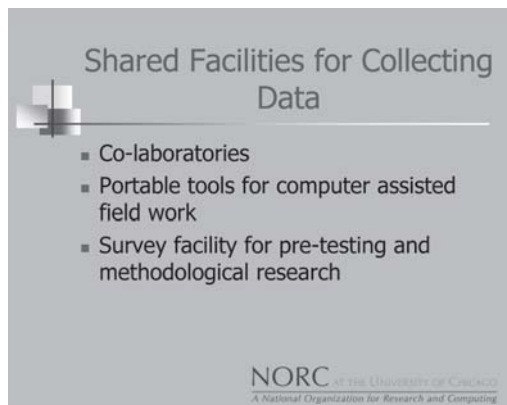


For data to be useful they need to be accessible. With the growth of data platforms as just discussed, with active data programmes in many countries and, increasingly, the availability of global data sets, the need to increase support and to upgrade web accessible data archives is great. Data archives require support for curatorial functions such as setting standards for archiving, documenting the origin, methodology and structure of the data sets, cleaning the data, enforcing confidentiality and intellectual property rights while maintaining open access to the data

and preserving and maintaining the data sets over time. The Council of European Social Science Data Archives (CESSDA) already provides many of these functions on a limited basis. The Common Language Resources and Technology Infrastructure (CLARIN) is an example of an archive of language resources and technology and reminds us that not all archives are data about people. The ESFRI Roadmap calls for substantially increased support for the CESSDA and start-up funds for CLARIN.

Not all data, however, can be made publicly available on the web at the level of disaggregation that is important for some social science research purposes. In order to protect the confidentiality of individual level data, publicly available data sets have to be aggregated and subjected to a disclosure analysis. The microdata are only made available under special circumstances or with special licenses. The trend is toward data enclaves where individual level microdata are held in a secure environment and researchers can only access the data within the confines of a secure data facility. In the U. S., a number of these enclaves have been established at universities around the country limited to access to census data, although efforts are underway to include data from other statistical agencies. In Canada, enclaves have also been established at a number of universities that give researchers access to a range of microdata collected by Statistics Canada. Such enclaves are becoming an important part of the data archival infrastructure for the social sciences.

3 Shared Facilities for Collecting Data



Shared Facilities for Collecting Data

- Co-laboratories
- Portable tools for computer assisted field work
- Survey facility for pre-testing and methodological research

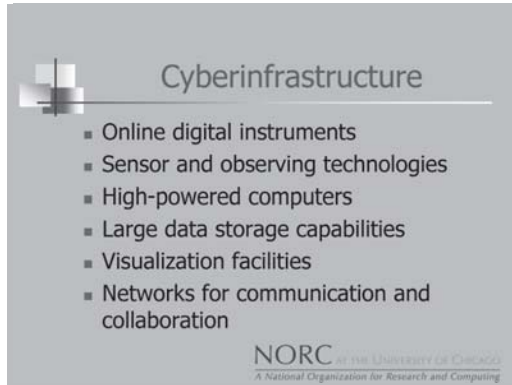
NORC AT THE UNIVERSITY OF CHICAGO
A National Organization for Research and Computing

Not all data of interest to social scientists comes from large sample surveys or large scale observational data. Increasingly laboratory experiments are being conducted in fields like economics and political science. These experiments study such phenomena as decision making under uncertainty, bargaining, and strategic behavior. To conduct these experiments requires a laboratory in which research participants can interact under controlled conditions, usually through the medium of a computer. The degree to which the results of such experiments can be generalized beyond the small scale and artificial conditions of the laboratory is unknown. In order to answer some of these generalizability questions, we need the capability to increase the scale of the experiments through shared, but distributed facilities, called appropriately “co-laboratories”. Such “co-laboratories” require secure communications links and shared research protocols so that the experiments can be conducted with larger and more diverse samples.

A second type of research infrastructure for these types of experiments frees the investigator from the confines of a laboratory by developing computer assisted tools, for example, for bargaining experiments that are portable and can be taken into the field. Prototypes of such tools have been developed and used by anthropologists in field work in remote villages as well as conducting experiments in natural settings in different countries.

A shared facility for methodological research in data collection is also a useful part of social science infrastructure. Researchers who are undertaking unique studies, perhaps of special populations or devoted to more specialized topics than are covered in the large, periodic data programs, often need to pre-test their questions or research instruments. NSF has recently funded a research facility that enables researchers to obtain data from relatively small telephone samples in a rapid time frame. This facility is for the exclusive use of social scientists who need to try out questions, get rapid feedback and retest revised questions, or who wish to conduct methodological experiments using split ballot questionnaires to test out different versions of the same questions.

4 Cyberinfrastructure



“Cyberinfrastructure” is a term applied generally to the complex of online digital instruments, emerging sensor and observing technologies, high-powered computers, extensive data storage capabilities, data mining, visualization facilities, networks for communication and collaboration and the attendant software and middleware that make this complex function (Atkins et al., 2003). Cyberinfrastructure can enable the development of more realistic models of complex social phenomena, the production and analysis of larger datasets, such as surveys, censuses, textual corpora, videotapes, cognitive neuroimaging records, and administrative data. Such computer supports can enable more complete records of human behavior, and the collection of better data through experiments and simulations on the Internet. (Berman & Brady, et al., 2005).

There is a reciprocal relationship between the development of cyberinfrastructure and the development of sophisticated research methods in the social sciences. Advances in econometrics, spatial analysis, game theory, linear programming, nonlinear dynamic modeling, linguistic annotation, simulations and agent based modeling, to name but a few, are heavily dependent on increased computing power. Cyberinfrastructure in turn provides a mechanism to use these methods and techniques more effectively, facilitating and improving both data collection and data analysis. The potentially revolutionary contribution of a vastly expanded and improved cyberinfrastructure is to make

possible these developments on a much greater scale and intensity using distributed networks and powerful computational tools.

Enhanced social science cyberinfrastructure will make possible such things as:

- Modeling life-time decision-making by individuals regarding their work, family formation, savings and retirement by following large number of people over time and formulating models that take into account the full complexity of these decisions;
- Code verbal and non-verbal cues in large numbers of video-taped interactions such as physician/patient interactions to understand proper medical diagnoses;
- Study change in urban areas over time by simultaneously geo-coding and temporally-coding land use, environmental, social interactions, institutional and other data for a large area over a long time;
- Understand the development and functioning of social networks on the Web by coding message frequency and content over time and space.

Cyberinfrastructure can help solve these problems because it provides unprecedented potential for advances in data collection and integration, computing power for analytic data, tools for data comparison, methods of data storage and communication and collaboration. (Berman & Brady, 2005).

5 Centers for Innovative Methods and Research





Finally, an important part of social science infrastructure are intellectual research and development centers where new ideas and methods can be developed. These may be physical centers where multidisciplinary teams come together to work on complex problems or, increasingly, distributed, virtual centers where the investigators work together through electronic means. Their mission is primarily to be centers of innovation that result in new ideas and methods, but they also build research capacity by training young investigators in new methods. Examples are the NSF funded in Center for Spatially Integrated Social Science, at the University of California, Santa Barbara and the ESFRI proposed European Resource Observatory for the Humanities and the Social Sciences (EROHS) which would organize the communication, coordination, documentation and sharing of information that would set the standards for research infrastructures worldwide. (ESFRI, 2006).

How are decisions made about Infrastructure

Given the large array of possible infrastructure needs and opportunities and the inevitable limitation of funds, how are Research Councils to decide on priorities? There are two general approaches that can be used which could be characterized as a top-down and a bottom-up approach.

The first, a top-down approach, is that adopted by the European Commission. A group specifically charged with the task of recommending the most

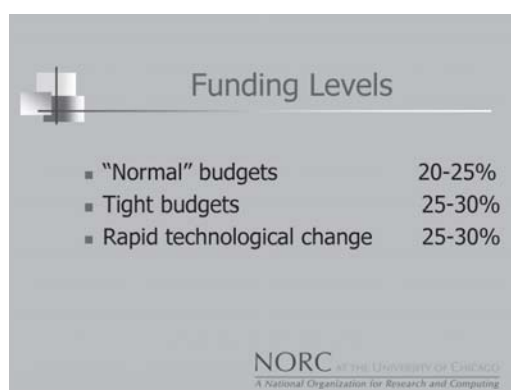
needed infrastructure projects was constituted. This group, the European Strategy Forum for Research Infrastructure (ESFRI), consulted widely with scientists in various fields and arrived at a set of recommendations for a large number of projects that could command wide-spread support among the scientific community. The results of this consultation were recently reported back to the Commission which will in turn incorporate the recommended projects in their budget plans over the next 5 or 10 years, depending on their planning horizon.

An alternative, bottom-up approach was followed by the Social, Behavioral and Economic Sciences Directorate at NSF. In this approach a general request for applications for infrastructure projects was sent out to the social and behavioral science research community with only a general description of the kinds of projects that might be funded. A large number of applications were received and reviewed using special panels of reviewers, but without any predetermined views about what was considered as most important. The proposed projects ranged across the types of infrastructure types outlined above, and varied considerably in their quality. A specially constituted review panel then recommended a set of projects to the Foundation officials, who ultimately made the funding decision based on the scientific merit of the proposals and their wider implications, limited by the amount of available funding. This special competition was repeated for several years until a portfolio of infrastructure projects was built up that met the most immediate needs. Additional special infrastructure competitions may be instituted in future years.

There is much to be said for both methods. The first does a good job in surfacing broad and well developed infrastructure ideas that can command the widest support throughout the research community. The second does a good job of surfacing innovative but perhaps riskier and less well accepted projects that may lead to transformative science, but also may well lead to dead ends. I personally favor the general competitions because I think, at least for the

social sciences, it is very difficult to predict which infrastructure projects will, in the end, be most valuable. I fear that by the time projects that have developed the broad support necessary to make it to an experts' list, they will no longer be innovative or as valuable as they would have been if a wider range of projects had been supported in a developmental stage and tested as to their usefulness.

What proportion of funding should go to infrastructure?



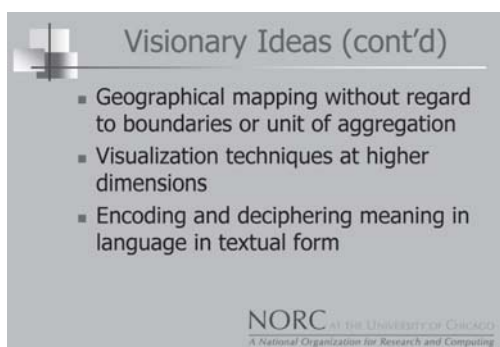
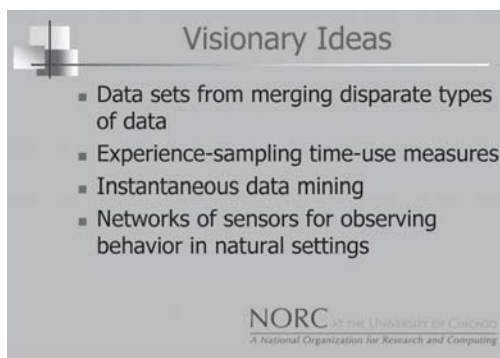
This is a difficult question for which there are no theoretical answers or agreed upon practical guidelines. To some extent it depends on the overall adequacy of research funding available and the rate of technological change. In times of generous budgets, one can have both generous infrastructure support and generous funding of research projects. Some types of infrastructure, such as data collection projects or data archives, provide basic data that many researchers with limited budgets can fruitfully use. In my view, more resources should go into infrastructure when budgets are tight than when they are more well funded.

Another factor is the rate of technological change. When there are significant and rapid advances, failure to support the infrastructure to take advantage of the new technologies can retard scientific progress. Advances in computer hardware, middleware and software together with the advent of grid computing have caused a revolution in the computing power available to social scientists. This has led in the United States

to focusing attention on cyberinfrastructure to enable social scientists to take advantage of these new and improved tools. Thus much of the excitement about infrastructure projects centers on computer tools and networks.

What does this translate into regarding funding? A rule of thumb at NSF was to devote about 20–25% of research funds to infrastructure. A recent study by the National Science Board (NSB, 2003) suggests increasing the level of support to around 28% to take advantage of technological advances, particularly in computing. So, something in the range of 20–25% of research funds might be devoted to infrastructure in times of fairly measured technological advances, with an increase to the 25–30% range when, as is now the case, there are rapid advances that open up large new areas of research opportunities.

A Vision for the Future



I have reviewed types of infrastructure that currently exist or are on the drawing boards. I would like to close on a visionary note.

Technological developments, either available now or in the near future, offer intriguing new possibilities.

For example, the creation of data sets that merge data from myriad data sources such as sample surveys, administrative records and biomarkers; experience-sampling time-use measures of location and activity; nearly instantaneous data mining; using networks of sensors to record behavior in natural settings; geographic mapping of variables without regard to boundaries or how units are aggregated; visualization techniques for higher and higher dimensions; encoding and deciphering meaning from language in textual form. For such ideas to be realized requires not only the innovative application of technology, but also new theoretical and conceptual ideas to drive the technologies. Only when the two driving forces come together will we have fundamental breakthroughs in knowledge.

I hope you will lead the way in realizing the rich possibilities that lie before us.

Questions and discussion

Question: Is there a difference between USA and Europe when we look at the co-operation of scientists in different academic fields and different funding agencies with regard to data & research infrastructure?

Answer Bradburn: In Europe I see hardly any attempts of scientists or funding agencies to agree on future developments except perhaps the astronomers. There are scientific disciplines – such as astronomy – that have to deal with industrial projects too and have big capital costs involved. In some fields the capital costs are big and industry is involved, while the operational costs are small. Social sciences are different: their

capital costs are small and their operational costs are big.

A question relating to cybernetics: Why does NSF only run one general programme for data & research infrastructure – from desk-top computers to supercomputers and extensive networks, GRID?

Answer Bradburn: There is a general strategy to break down barriers between disciplines. Cyber-infrastructure lends itself well to a programme that is there for all scientific disciplines. I would like to refer to the Atkins Committee Report on Cyber-Infrastructure [Atkins, D. E. et al. (2003). *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-ribbon Advisory Panel on Cyberinfrastructure*. Washington, D.C.: National Science Foundation].

Question: Do you know of any models for funding transcontinental projects aimed at co-eration, harmonization, et cetera?

Answer Bradburn: I consider it to be extremely hard for individual councils to pool together, such as NORFACE is doing. NSF in USA is reluctant to enter into such agreements, even with funding bodies in the United States itself. True globalization will happen through bodies such as the UN, UNESCO.

Addition: David Moorman describes an example of a transcontinental project in oceanography: there is only an informal basis for agreements concerning the division of tasks. Informal groups of oceanographers manage the project individually. Example: they agree informally that – for instance – Japan will do the deep-drilling and the USA will do the medium deep drilling. The keyword here is: informal.

P. Elias (UK): Introduction to theme 1: National roadmaps

Peter Elias illustrates his introduction of the theme of national roadmaps by means of the following slides (powerpoint presentation):

Roadmaps, strategies and research infrastructures

Promoting and fostering international research collaboration in the social sciences

Presentation by Peter Elias at the NORFACE Conference on Data and Research Infrastructure, The Hague, 16th November 2006

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Developing and improving data resources for social science research

NDS
THE NATIONAL DATA STRATEGY

Plan of the presentation

- some definitions
- developing the UK National Data Strategy
- international collaboration in the social sciences

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Developing and improving data resources for social science research

NDS
THE NATIONAL DATA STRATEGY

Some definitions

- the social sciences
- research infrastructure
- strategy
- roadmap



Developing and improving data resources for
social science research



The UK National Data Strategy

The UK National Strategy for Data Resources for the Social Sciences is a plan to develop and maintain a robust data infrastructure, ensuring that relevant and timely data are available to inform and address future research priorities in the social sciences.



Developing and improving data resources for
social science research



Why did we need a National Data Strategy?

- No clear mechanisms to link future research to data resources.
- Funding of the increasing portfolio of resources is beyond the means of single agencies.
- Duplication of efforts in some areas.
- Problems of data access, availability.



Developing and improving data resources for
social science research



The National Data Strategy - how was it developed?

- Fact-finding activities.
- Establishing the UK Data Forum.
- Identifying the key research challenges.
- Short term/medium term actions to promote and develop the strategy.
- Consultation process/revisiting priorities.



Developing and improving data resources for
social science research



Identifying key research challenges

- Ageing population
- Migration
- Globalisation
- Childhood development
- Carbon use



Developing and improving data resources for
social science research



What progress are we making in implementing the National Data Strategy?

Longitudinal data resources

- commissioning the world's largest longitudinal household study (see www.esrc.ac.uk/ukhls).
- developing proposal for a new child cohort study to start around 2012.



Developing and improving data resources for
social science research



What progress are we making in implementing the National Data Strategy?

Census and population survey data

Plans now underway to provide improvements in access to Census 2011 data and new Integrated Household Survey.



Developing and improving data resources for social science research



Progress in implementing the National Data Strategy

Administrative data

Undertaken a detailed review of the research potential of administrative data held by selected agencies, covering pensions, social security, health, businesses and vital statistics. Report detailing the scale of these resources, their potential for research, access issues and an outline proposal for better access now available.



Developing and improving data resources for social science research



Progress in implementing the National Data Strategy

Commercial data

Review of availability and research potential of commercial transactions data



Developing and improving data resources for
social science research



Progress in implementing the National Data Strategy

Linked biomedical/socio-economic data

Working with the medical research and funding communities to establish 'gateway procedures' to key linked resources (1946 cohort, 1991 birth cohort, 2002 biomedical sweep of 1958 birth cohort) and to develop biomarkers in UKHLS (new longitudinal study starting in 2008).



Developing and improving data resources for
social science research



Progress in implementing the National Data Strategy

Data access

Establishing a flexible licensing regime.

Developing remote access to secure microdata.



Developing and improving data resources for social science research



Progress in implementing the National Data Strategy

International data resources

- Working with Eurostat and European Commission to improve EU harmonised statistics
- Significant database purchases from IMF, OECD, World bank, etc. Funding ISSP and ESS
- Microdata scoping studies (China, Brazil, India, South Africa)
- Beijing conference 2007 (International Data Forum)



Developing and improving data resources for social science research



What is the strategic importance of these developments?

- to facilitate research at the boundary between the medical, biological and social sciences (genetics, mental health, lifestyle, etc.)
- to promote cross-disciplinary research on child development and social change
- to encourage major research projects on the nature and impacts of migration
- to stimulate comparative cross-national research
- to encourage international research on global problems



Developing and improving data resources for
social science research



International collaboration in the social sciences

- the role of ESFRI
- the role of NORFACE
- the role of national funding agencies
- the need for 'prototype' infrastructures
- the need for common access protocols
- the need to look beyond Europe



Developing and improving data resources for
social science research



Highlights: When talking about ‘strategy’ Peter Elias emphasizes that any policy concerning data & research infrastructure should aim at infrastructure to be ‘robust’ and ‘timely’. As social scientist you must always be engaged with a political agenda. Any National roadmap should be engaged with the political agenda in a country, which – by the way – is something different from being driven by an agenda. Cohort studies. About EUROSTAT, the European organization of some thirty statistical offices, he explains that many researchers experience this organization as some kind of a ‘brick wall’.

ESFRI is very significant indeed, especially because it has increased the level of awareness between researchers, that there is an interdependence between them with regard to scarce resources and that they need to discuss how to allocate funds. Nevertheless, NORFACE should also look beyond Europe: data & research infrastructure is a global topic. Due to the experiences with the now finished European Household Panel he stresses the need for new initiatives driven by researchers rather than by government of any office. He wonders: do the funding agencies have a role here? About the list of ESFRI for the Social Sciences and the role for NORFACE: We have to be aware of the concept of ‘maturity’. The Science community has got a real advantage in this area – they are organised, fundamental science driven. Social Sciences has a real problem, unless they can

take up an activity – compare the UK Strategy. We need a new initiative, driven by the research community. Can NORFACE have a role in this? Peter Elias is optimistic. We are starting to look beyond our national and European boundaries.

Questions and discussion time

Questions concern the importance of administrative data and the fragmentation in strategies for archiving data. Referring to the ESFRI Roadmap, one of the conference participants emphasizes the importance of the EROHS proposal: this proposal is most visionary and most important topic on the list, but it is least likely to be funded. On the contrary the ESS proposal is more visible and more likely for its funds to be continued, but it is rather practical, ‘do-able’, and perhaps more short-term. Question: How can Europe get the visionary projects funded – such as EROHS –, or how do you sell a product that people cannot see?

Answer Peter Elias: A real problem is how it will be evaluated. For EROHS to succeed, it has to be viewed as highly desirable by the research community, and have a structure. It is more of question of approach. I suggest that something visible will be put into the ‘Trojan horse’, something which people want to have and which will bring them to drawing the horse into town.

Report from the workshops by discussion leader S. Barkçin (TR; ESFRI) and rapporteur L. Mizera (EE)

Savas Barkçin illustrates his report – composed by Luule Mizera and himself – by means of the slides below:

Theme 1. Developing National Roadmaps

Flow

UK Experience
National Roadmaps
Identifying and Funding Priority
Areas
ESFRI and Role of NORFACE

The UK Experience

from a project on social sciences in the UK into a national RM

looking into many places with collections of data

need to develop projects that are flagships

national RMs should be preceded by national policies/strategies

starting points are different, the UK social sciences RIs have been quite far ahead compared to many other countries in EU

the UK Roadmap would probably not work in many other countries

need to set up “culture of resources”, a “library mentality” of research community

National Roadmaps

ESFRI played a role in encouraging countries to design national roadmaps

RM must be something more than just a wish list

RM encourages countries to revise their situation and needs, gives input to the ministries

strategy, goals and direction must precede RM: RM presumes you know where you are going and how

RM process has to be legitimized by the users (scientific community, SC, etc.)

must be based on a consensus

must engage not-so-obvious players/stakeholders as well

cost and funding design is important, but should not restrain the imagination

must take into account that RIs are public goods

coordination is needed among actors: US OSTP as an example

national RMs maybe not as useful as coordinating and implementing the ESFRI RM?

need to let the stakeholders (esp. the government) to know where you want to go

need a framework that e.g. the higher education institutions could use as a framework for research strategy

hard to predict the researchers' demands, RMs validate the funding agencies' activities if researchers are consulted

need to map and address the national imbalances in research & research infrastructures

need to bring together & coordinate "proper" research infrastructure and "proper" social science research people

IDENTIFYING AND FUNDING PRIORITY AREAS

Pitfalls

difficult to do, some RMs do not include specific priority areas

it may simply complicate RM preparing process

could become a limiting activity, a “straight jacket”

draws attention to what is not on the list

impossible to foresee everything (e.g. after 9/11 social sciences got the impetus)

country size matters: for larger countries it is more relevant

Significance and Involvement

research community must be consulted in the process

mainly project-based funding in funding agencies, therefore a need for political decision & ministries' involvement

need for research communities' involvement, evaluation & renewal process

they are good for “selling” the projects to policy-makers
(concrete examples of how social sciences can contribute to societal problems)

may be useful for the funders' internal mechanisms

a balance and combination of top-down and bottom-up approaches is needed

Caveats

limited budget, must have some criteria

needs to consider resources and market for those areas

needs to be built around already existing strengths, fields with comparative advantages

areas must be balanced political decisions (needs of different communities must be taken into account)

researchers have to use the advances in technology, interplay between research developments & underlying technologies

RI activities should not be seen as any project

long-term perspective must be adopted in taking funding decisions

in project funding mechanisms a need for new & innovative versus a need to maintain infrastructures (e.g., ESS done many times, needs revalidation)

creating data for future researchers, thus need for long-term financing

sustainability & relevance must go together

legal aspects are important

need for periodic cost review as well

the areas need to be flexible

ESFRI & NORFACE

ESFRI is useful for national purposes (raising the profile for research infrastructures in general and social sciences in particular

the ministries' (ESFRI) role very important in validation & sustainability

not all projects in ESFRI Roadmap may be materialised

it will be reviewed periodically, hence an opportunity for NORFACE

lobbying via NORFACE and/or national funding agency in the ministries

NORFACE important for maintaining cooperation, coordination of national data and improving data transfer possibilities

NORFACE could identify barriers to data use and deal with them

NORFACE could mediate national funding agencies, ESFRI, EC and research communities

NORFACE could help bringing new and innovative ideas of social science research infrastructures to the next round, in a coordinated way

NORFACE experience could be useful for ESFRI's future work

Caveats

ESFRI may become too politicized

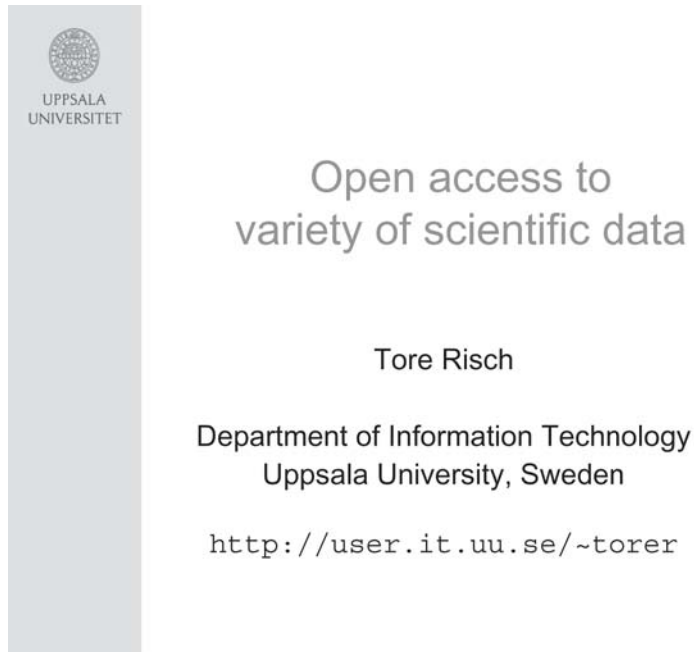
ESFRI's problem: funding agencies are missing –
NORFACE can help

The word RI may be seen as national and hence proves
to be counter-productive in trans-national or pan-
European efforts such as ESFRI

T. Risch (SE): Introduction to theme 2: Data access, including remote access, new ICT possibilities, international access to data

Tore Risch introduces himself as a member of a committee called Swedish National Infrastructure for Computing (SNIC). This committee is responsible for the strategic and scientific development and funding of high performance computing resources in Sweden.

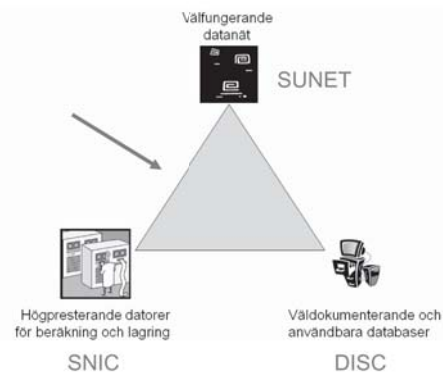
His introduction is illustrated by the slides below (powerpoint presentation):





Swedish Committee for Research Infrastructure

Astronomy and subatomic research
Molecular, cell, and materials research
The earth and its closest environment
Infrastructure for e-Science



Vision: A holistic view of research databases

Many different sciences

- Micro data in social sciences and medicine
- Linguistics and the humanities
- Climate and the environment
- Biotechnology, genomics, bioinformatics
- etc.

Open Access

Software for distributed databases; technology development

Cooperation nationally and internationally



Distributed Scientific Data Sources

Modern scientist's many data sources:

Databases:

- Patient registers
- Experimental results

Digital libraries

The web

Instruments/software producing data:

- Digital questionnaires
- Patient monitoring software
- Environmental sensors
- Simulators



Meta-data very important

Myriads of data in different forms

- ⇒ Difficult to find right information (even for systems)
- ⇒ Need meta-data, i.e. structured 'map' that describes databases

Examples of meta-data:

- Schemas in relational databases
- Both map and for efficient search!

Semantic web vision:

Define meta-data for web resources similar as database schemas

- Ontologies



Database Design

Designing meta-data
understanding data
selecting relevant data
designing database schema
adaptation to database schemas

Documentation

Availability strategy

Privacy and security strategy

Archiving strategy



Populating Databases

Data must be collected:
Data entry questionnaires
Digitalization
Instrument interfaces

and enriched:
Convert to existing database schema
Mark-up (annotation)
Statistical processing
Numerical event identification
Language analysis

and deriched:
De-identification
Data reduction (e.g. removing noise)



Database Management Systems

DBMS:

Software to manage *large volumes* of data

DBMS very useful in many applications,
including scientific work of all kinds!

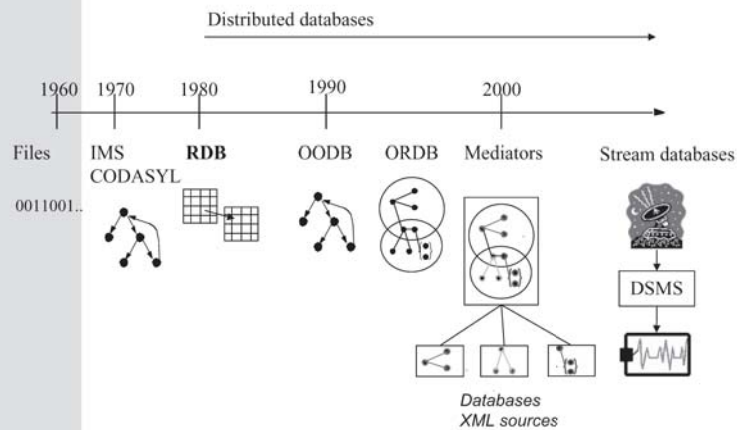
Enabling technologies:

- Efficient *search* and *update* of large datasets
- Transaction processing
- Security, authorization, integrity
- Many different data representations
(tables, statistics, arrays, XML, text, time series, images)



Tore Risch
Uppsala University, Sweden

Evolution of DBMS technology





Distributed Data Integration

Modern Information systems are distributed:

- Data represented differently in different systems
- Accessible information volume is rapidly increasing.

Standards part of the solution:

- There are many standards
=> information islands
- Different research disciplines – different standards
- New systems require new ways to structure the information.

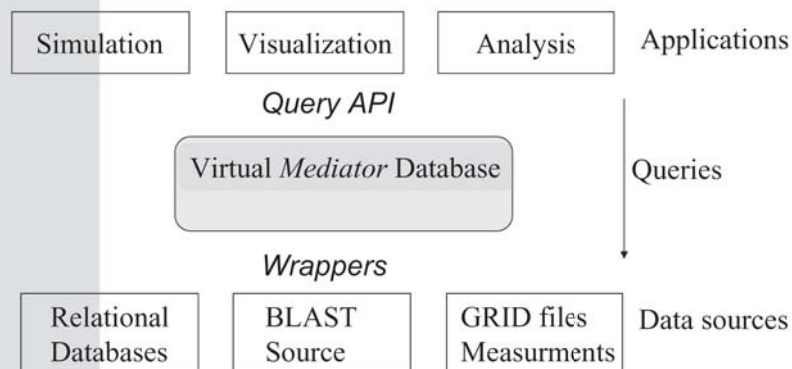
=> Methods and tools are required for integration of heterogeneous and distributed data and information



Tore Risch
Uppsala University, Sweden
<http://user.it.uu.se/~torer>

Wrappers: make data sources queryable

Mediator: combine (fuse) data from wrapped sources





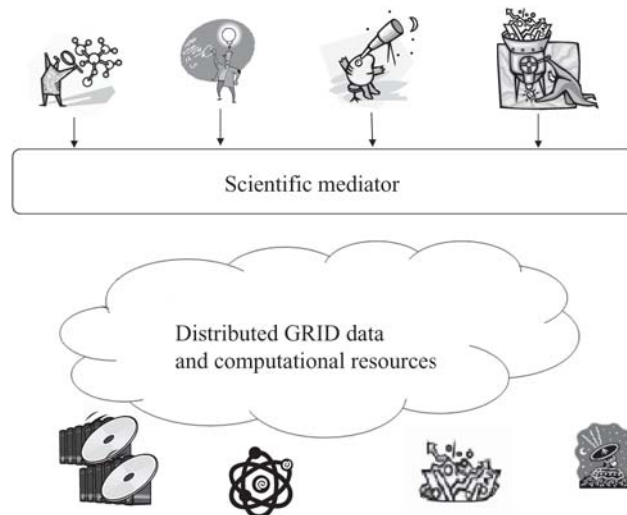
Diverse Distributed Software

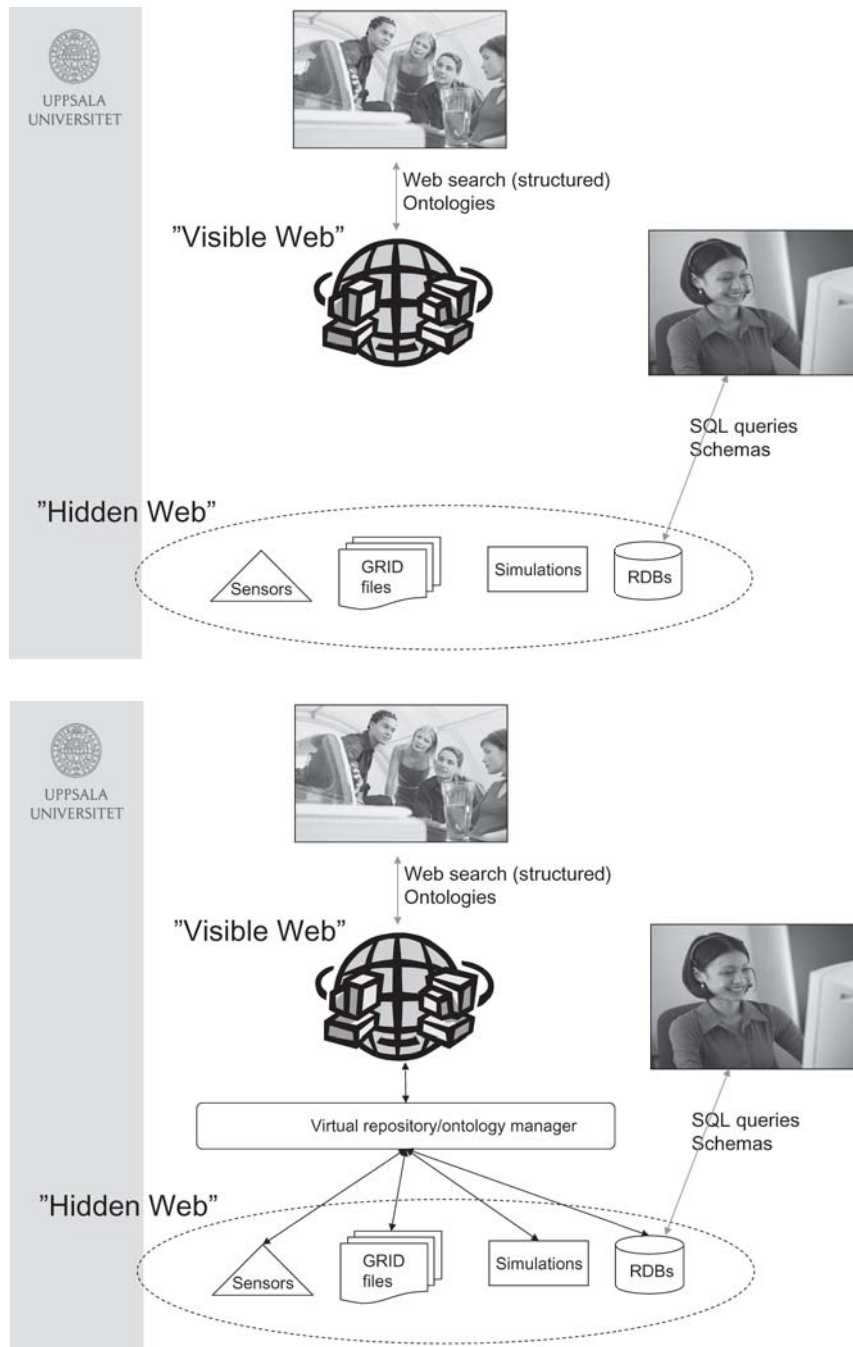
Lot's of data processing software:

- Database Management Systems
- Search engines
- Statistical processing software
- Numerical processing software
- Monitoring software
- Visualization software
- Documentation systems

Available in distributed infrastructure

- PCs
- Workstations
- The Web
- The Grid







New applications

Data comes as large data streams, e.g.

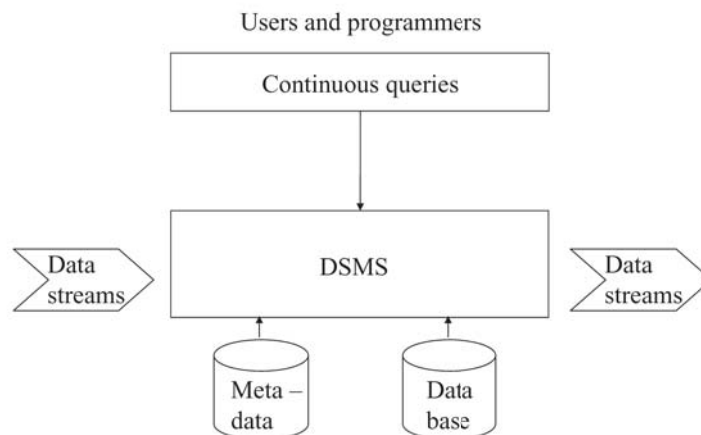
- Satellite data
- Scientific instruments
- Colliders
- Patient monitoring
- Process industry
- Traffic control

Would like to query data in *streams*



Tore Risch
Uppsala University, Sweden

What is a DSMS?



NORFACE

Highlights: In his engaging presentation Tore Risch emphasizes that for infrastructure we do not want to have any walls between the different sciences. His main message to the NORFACE partners is that meta-data is very important. They should devote enough attention to that aspect of data & research infrastructure. As Norman Bradburn did before him, he explains that – comparing the costs for infrastructures in various academic disciplines – for the social sciences it is relatively expensive to get the flow of data. In the social sciences the capital costs for data & research infrastructure are relatively low but operational costs are high.

Answer Tore Risch: This will depend on the project. I expect that the big costs in the social sciences are the data collection. You should reckon that on top of data collection you need about 10% for the metadata.

One conference participant adds that funding agencies should design procedures for their researchers to file their infrastructural material (databases) at a proper place: the data archives. Peter Doorn draws attention to data reduction. Data reduction often leads to data destruction. He pleads that this is done in a reversible way.

Questions and discussion time

Question: what implications are there related to costs?

Report from the workshops by discussion leader R. Procter (UK) and rapporteur Sh. Duffy (IE)

The workshops address the questions in the conference programme. A summary of the discussion points follows here.

How can the NORFACE partners – being funding agencies – support further development and promote the use of these possibilities in the social sciences together?

Partners need to coordinate efforts. They should share experiences of barriers (different countries are at different stages of infrastructure adoption), agree on best practices across a range of activities need to adopt and sustain infrastructure (training, infrastructure development project management). Partners should share exemplars of research enabled by new infrastructure. Partners should translate infrastructure benefits for SS into their respective national research agendas and thereby help to move this up their research respective agendas to create top-down leverage.

How can they support the wider availability of access to national data resources in order to support comparative research?

Licence arrangements across national boundaries are virtually non-existent. Incompatibility of national IPR regimes is a real barrier as are differences in approaches to the use of personal data. Partners need to work together to coordinated common policies for licensing, IPR and data governance.

How can they promote improvement of access to international data, e.g. by Eurostat and other international bodies?

Need for better access to Eurostat, to integrate academic, statistical and commercial data.

Is there a need for funding agencies to unite with their social scientists to achieve certain ends?

Top-down leverage must be complemented by bottom-up mechanisms. Adoption won't succeed without the buy in of research community members. Partners can encourage this by investing in training, promoting cultural change by rewarding good practices in data publication (e.g., by counting data publication as an output on a par with papers). Partners should work with research community members to define standards for data publication, make it condition of funding projects and ensure they make available the resources to fund it.

How do or may NORFACE and ESFRI relate concerning this topic?

ESFRI has multiple stakeholders, most of which have much bigger research budgets, so there is a real danger that SS requirements won't get heard or will not be prioritised. NORFACE should do a gap analysis to discover where ESFRI isn't addressing the SS agenda and concerns, and continue to track this as the roadmap is implemented.

P. Doorn (NL): Introduction to theme 3: Data networking in Europe

Data Archiving and Networked Services



Contents

- Some achievements of data networking in Europe, national and transnational
- New requirements for data and research infrastructures
- Towards permanent European infrastructures
- The role of the ESFRI-Roadmap as a “first stone”



Long tradition of data archiving and networking

- First data archives in the social sciences since 1960s and 1970s
- International collaboration:
 - Council of European Social Science Data Archives (CESSDA)
 - International Federation of Data Organisations (IFDO)
 - International Association for Social Science Information Service and Technology



DANS

Some examples: NSD

- Norwegian Social Science Data Services
 - Established in 1971
 - Objectives:
 - facilitate wider and more knowledge-based use of data
 - promote the preservation and sharing of data
 - ensure free and open access to information
 - Tasks:
 - collect, document, process and preserve data
 - create and integrate databases
 - disseminate data and information about available data
 - produce thematic maps and computer graphics
 - produce documentation on Norwegian current research
 - advise on issues related to privacy and research
 - produce teaching materials for high schools, colleges and universities
 - develop software for statistical analysis adapted to research needs
 - carry out research-relevant assignments



DANS

Other Examples

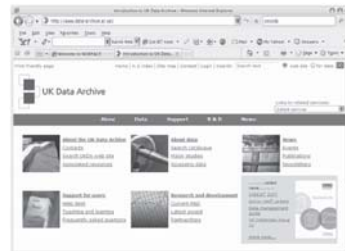
Zentralarchiv für Empirische Sozialforschung (1960):

- archives primary material (data, questionnaires, code plans) and results of empirical studies in order to prepare them for secondary analyses and to make them available to the public



UK Data Archive (1967):

- centre of expertise in data acquisition, preservation, dissemination and promotion;
- curator of the largest collection of digital data in the social sciences and humanities in the UK
- provides resource discovery and support for secondary use of quantitative and qualitative data in research, teaching and learning



DANS

Some CESSDA achievements

- Integrated Data Catalogue
- Nesstar: publishing, searching, browsing tool for data on the Web
- Madiera: multilingual, unified access portal to European social science data (2673 studies from 9 countries)
- Transborder data agreements, eg. for large European surveys (Eurobarometers, ISSP, ESS, EVS, EES, etc.)
- Common metadata standards and tools (DDI, Metadater)

DANS

Data infrastructure challenges

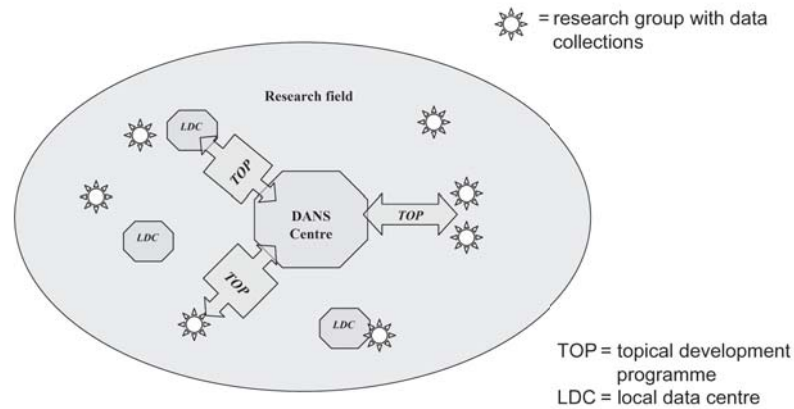
- Renewal/improvement of national data infrastructure in several countries
 - Upgrade technology and higher efficiency (eg. self-archiving)
 - Data archives no longer have a monopoly on data supply: there are now many data suppliers
 - Matching supply and demand of data, closer ties between data archives and research community
 - More transparency: single 'service desk' for researchers and data producers, cluster/network fragmented activities
 - Open access to publicly funded data (administrative or registration data, collaboration with statistical agencies)
 - Answer to larger scale and internationalisation of research (longitudinal and international surveys)



Examples of national networks for data infrastructure

- UK Data Archive part of Economic and Social Data Service
- Zentralarchiv part of GESIS - Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen
- DANS set up by NWO and the Dutch Academy to improve the data infrastructure for the humanities and social sciences
 - Brings together three previously existing data centres
 - DANS takes the form of a network with a strong centre:
 - the DANS centre is responsible for continuity, organisation and support of the data infrastructure
 - researchers play a decisive role in developing DANS by participating in decentralised "topical development programmes" (TOP's)





- In spite of achievements, existing infrastructures are primarily national
- European activities are until now funded on a project basis and carried out as voluntary activities by national centres
- Although much has been accomplished in this way, stable, truly pan-European data infrastructures for the social sciences hardly exist
- Increasing internationalisation of social science research puts new requirements for such infrastructures
- Awareness of this omission is rising: not only are the Norface partners working towards a common strategy for data and research infrastructures, but also the European Strategy Forum on Research Infrastructures (ESFRI) has just launched its first Roadmap
- 4 of the 35 proposed Roadmap-initiatives are in the domain of the social sciences



ESFRI Roadmap

- Launched on 19-10-06
- 6 proposals from SS & H



Projects (in alphabetical order per discipline)	Estimated Construction Cost (M€) *	First possible operations for users	Indicative Operational/ Deployment Cost (M€/year)
CESSDA	30	2008	6
CLARIN	108	2008	10
DARIAH	10	2008	4
EROHS	43	2008	12
ESS: European Social Survey	9	2007	9
SHARE	50	2007	< 1



CESSDA

- Now: Network across 21 countries in Europe, 15,000 data collections, serving over 20,000 researchers
- Ambitions□
 - create a European "data passport", which enables researchers and data to move virtually across national and organisational boundaries.
 - create a common platform to a fully integrated data archive infrastructure for the SSH.
 - discover datasets and related materials in a cross-national environment;
 - understand in detail the origin, methodology and structure of the underlying data collections;
 - compare and link data from different locations;
 - connect to other experts and share analyses, experiences and knowledge;
 - enforce confidentiality and intellectual property rights whilst maintaining accuracy, security and open access to data sources;
 - preserve and maintain data collections over time;
 - design new research instruments



EROHS (European Resource Observatory for the Humanities and the Social Sciences)

- Will operate both as a central and distributed facility with a strong physical hub working in close conjunction with a number of spokes across Europe
- Coordinated yet decentralised network
- Promote and ensure cooperation and integration of data, technologies and policies
- aiming at organising the communication, coordination, documentation and sharing of information
- set standards for Research Infrastructures worldwide
- open, distributed, well coordinated, well documented, on-line access to data for the humanities and the social sciences.
- Training and seminar activities organised
- Emerging infrastructures will be important stakeholders of EROHS
- Strengthen national infrastructures by linking national hubs both to each other as well as to the international hub



European Social Survey

- Set up in 2001 to monitor long term changes in social values throughout Europe, now covering 27 countries
- A long term pan-European instrument such as the ESS requires long term funding commitments
- Upgrade:
 - compiling and harmonising aggregate context variables for survey analyses;
 - experimenting with alternative (technical and traditional) methods of translation to improve equivalence;
 - investigating and mitigating longstanding problems in the collection and classification of occupation and education;
 - improving the capacity to pilot and pre-test new questions on emerging issues of public concern;
 - experimenting on a multinational basis with methods of improving response rates.





- Data infrastructure for fact-based economic and social science analyses of the on-going changes in Europe due to population ageing.
- The original 8-country survey has already being expanded to cover two new Member States;
- ideally SHARE will be expanded to all 25 Member States of the EU
- The first wave of SHARE data was collected in 2004 and the second wave is currently fielded in 2006, further waves are envisaged from 2008 onwards bi-annually.
- In the years in-between these waves, experimental modules will be tested, such as the collection of life-histories in 2007. The 24 months between the
- end of wave t and the end of wave t+1 can roughly be divided into 12 months of
- preparation and 12 months of data collection (including experimental modules).
- The SHARE data infrastructure is accessible free of charge through an archive operating as internet platform.



Humanities proposals

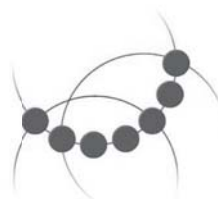
DARIAH: Digital Research Infrastructure for the Arts and Humanities

- Support digital access to all surviving humanities and cultural heritage information for Europe, and it's preservation in the long term



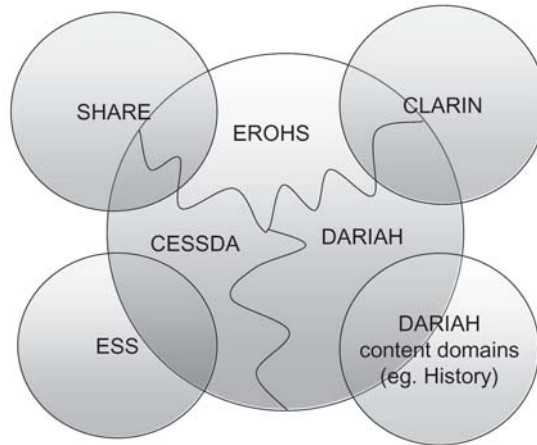
CLARIN: Common Language Resources and Technology Infrastructure

- Make language resources and technology available and useful to scholars of all disciplines, in particular the humanities and social sciences.



Interdependence of SSH Roadmap proposals

EROHS, CESSDA and DARIAH are to co-ordinate (or integrate) their functions very well and to supply networked services (data archiving, dissemination) to the content-oriented proposals (SHARE, ESS, CLARIN, etc.)



DANS

Afternoon workshop on theme 3: data networking in Europe

Discussion points:

- How may the funding agencies create a surplus value of their social sciences research by promoting international data sharing? What role can - or do - the funding agencies play here?
- How do European networked data organisations relate to initiatives concerning open access for administrative data and for researchers' data?
- How would European networked data organisations work together with national statistics bureaus or offices in the NORFACE countries?
- How do or may NORFACE relate to ESFRI concerning this topic?

Discussion leader: Ekkehard Mochmann (Zentralarchiv)

DANS

Websites

- www.gesis.org/ZA
- www.dans.knaw.nl
- www.nsd.uib.no/cessda/index.html
- www.data-archive.ac.uk/
- www.nsd.uib.no/
- www.ifdo.org/
- www.iassistdata.org/
- www.dariah.eu/
- www.share-project.org
- www.erohs.org/
- www.europeansocialsurvey.com/



Highlights: Peter Doorn is the Director of the Dutch data archive called Data Archiving and Networked Services (DANS).

He explains that the existing infrastructure is primarily national, and that there is a considerable European challenge in this area.

Referring to the ESFRI Roadmap recently published he explains his view that the Social

Sciences and Humanities proposals can be seen as interdependent. CESSDA, EROHS, DARIAH all cover a different field, but together they cover all data & research infrastructure in the social sciences and humanities and will supply networked services to the content oriented proposals (SHARE, ESS, CLARIN). An important question is: how do NORFACE and ESFRI relate to each other?

Report from the workshops by discussion leader E. Mochmann (DE) and rapporteurs M. van Leeuwen (NL) and L. Henriques (PT)

The discussions in the workshops starts with a question from Peter Doorn and then follow the questions in the conference programme.

Is co-operation between Social Sciences and Humanities needed? (question by Peter Doorn)

The discussion ended with a clear yes:

- There are comparable technical problems and similar instruments for access needed.
- They can form an alliance to lobby for SSH interests vis à vis the Sciences – but should be open for co-operation with the Sciences as well, e.g. sustainable developments / geosphere and biosphere research relating science and social science data.
- Humanities can help to contextualize Social Research (Culture/ History) (illustrated by quote from Rembrandt's Group Portraits.)

How may the funding agencies create a surplus value of their social sciences research by promoting international data sharing? What role can – or do – the funding agencies play here?

NORFACE partners can

- assess nationally the quality, coverage and range of data bases in different SSH fields (data auditing of the national data base);
- promote standards for data processing/documentation and for measurement (standard demography);
- support their national archive centres (instead of individual archiving);
- support relation between researchers and statistical offices;
- perform gap analyses.
- identifying barriers to international co-operation is important.

How do European networked data organisations relate to initiatives concerning open access for administrative data and for researchers data?

- There are different approaches to open access (e.g. green road - golden road) and different declarations (e.g. Berlin Declaration explicitly including data, and the OECD declaration re Open access to publicly funded data).

- NORFACE should support national competence centres/data infrastructures.
- NORFACE should promote ex ante / ex post harmonisation standards for data collection and integration.
- NORFACE should develop national/international data policies including national statistical offices and EUROSTAT.
- For administrative data there is a need to first develop methods and instruments to turn admin data into useful research data and then follow up with archiving.
- Also needs for qualitative research infrastructure should be addressed.

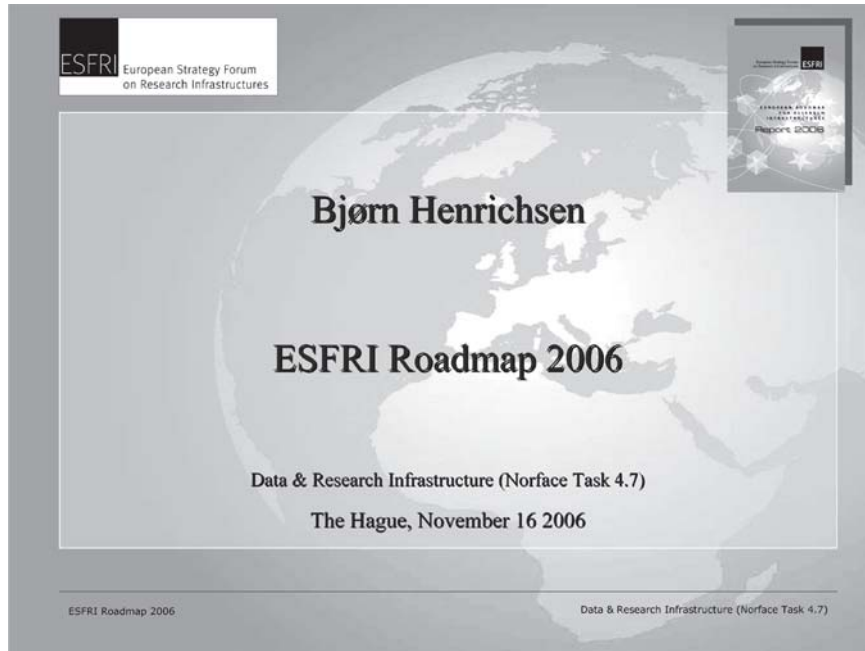
How would European networked data organisations work together with national statistical bureaus or offices in the NORFACE countries?

- There are good solutions in Norway (NSD), but the situation is difficult in most other European countries.
- There are ways to sell these ideas to agencies and politicians, claiming that these data have been publicly funded already and that research is contributing to solve societal problems (Canada).
- NORFACE could help to develop models how to transport data for bilateral co-operation (e.g. examples from Norway and Germany /France).
- Support models like GESIS Micro-data Lab to create a culture of co-operation with statistical agencies.
- Develop standards for measurement in co-operation with academic, statistical and commercial sector.

How do or may NORFACE relate to ESFRI concerning this topic?

- NORFACE could help to lift national initiatives to critical mass, so that they can profit from participating in European projects.
- NORFACE should also support access to publicly funded software.
- NORFACE should participate in the development of ESFRI – II!

B. Henrichsen (NO; ESFRI): Presentation of the ESFRI Roadmap





What is ESFRI European Strategy Forum on Research Infrastructures?

- Support a coherent approach to policy-making on research infrastructures in Europe
- Incubator for international negotiations about concrete initiatives
- Currently preparing a European road-map for new research infrastructures of pan-European interest

ESFRI Roadmap 2006 Data & Research Infrastructure (Norface Task 4.7)



The Roadmap

- The European Roadmap is the result of two years of intensive work at European Level
- Identifies 35 new Research Infrastructure (or major upgrades)
- To be realised in the next 10 to 20 years
- Wide stakeholder consultation
- It is the intention of ESFRI to update the document next year to allow it to be improved and to integrate new ideas where and when appropriate

ESFRI Roadmap 2006 Data & Research Infrastructure (Norface Task 4.7)




The Roadmap – Social Sciences

ESS European Social Survey

Main objectives:

- compiling and harmonising aggregate context variables for survey analyses
- experimenting with alternative methods of translation to improve equivalence
- investigating and mitigating longstanding problems in the collection and classification of occupation and education
- improving the capacity to pilot and pre-test new questions on emerging issues of public concern
- experimenting on a multinational basis with methods of improving response rates

ESFRI Roadmap 2006 Data & Research Infrastructure (Norface Task 4.7)




The Roadmap – Social Sciences

SHARE Survey of Health, Ageing and Retirement in Europe

Main objectives:

- collecting three waves of data in 2008, 2010 and 2012 (two waves have been collected in 2004 and 2006)
- linking the waves to understand the changes over time
- testing experimental modules such as the collection of life histories



European Strategy Forum
on Research Infrastructures

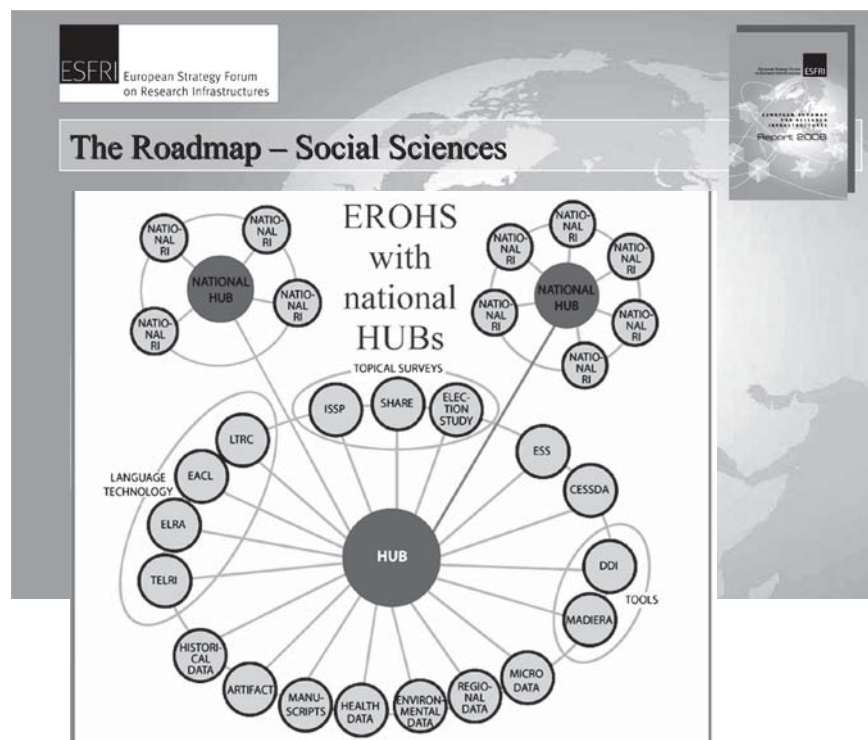


The Roadmap – Social Sciences

EROHS European Research Observatory for the Humanities and the Social Sciences

Main objectives:

- A new European-wide infrastructure
- Will operate both as a central and distributed facility
- Will be organised to promote and ensure cooperation and integration of data, technologies and policies
- In some areas it will be based on mature existing resources
- In other areas it will facilitate the development of new infrastructures based on the need






The way forward

- FP7: In the first phase, funding will only be given to projects on the ESFRI list (for the preparatory phase)
- Support expected/needed from the Ministries in Europe
- Pooling of resources from the European Commission, other international bodies as well as Ministries
- Location of services?




The way forward

- Implementation of the Roadmap
- Next year already some of the projects will be in place
- Increased role of the Research Infrastructures as a major cornerstone in the European Research Area
- ESFRI encourages constructive debate

Highlights: Being the chair of the Roadmap Working Group in the Social Sciences and Humanities Bjørn Henrichsen explains that ESFRI is neither an organization of the European Union or of the European Commission – ESFRI is a form of co-operation between research ministers and the European Union, and the secretariat is in the European Commission.

He emphasizes that the *European Roadmap for Research Infrastructures; Report 2006* is not an end product. The proposals need to be brought forward. Institutions in the EU member states need to support the proposals, otherwise the exercise of drawing a roadmap will prove to be

a failure. Bjørn Henrichsen calls upon the social sciences to get organized – in which process the NORFACE partners could play a role – and to approach their ministers for science and support the ESFRI proposals that they consider appropriate for their academic fields.

About the way forward: In the first Call you have to be on the ESFRI list; Ministries (or Funding Agencies) should apply, not the projects; support is needed / expected from the Ministries in Europe. There will be pooling of resources from the European Commission, other international organisations as well as Ministries / Funding Agencies. ESFRI encourages constructive debate.

D. Moorman: Key note speech 'OECD and Research & Infrastructure'

David Moorman introduces himself as a member of the Organisation for Economic Co-operation and Development (OECD) Committee for Science and Technology Policy (CSTP); he is also the representative of the Canadian NORFACE partner 'Social Sciences and Humanities Research Council of Canada (SSHRC)'.

His presentation is about the important document *OECD Recommendation concerning access to research data from public funding*. In the next few months this recommendation will gain a formal status within OECD, which means that the recommendation will be supported by a great number of member states. At this time of day science is not globalized yet; science still has a nationalized character. At the same time scientific databases are rapidly becoming a crucial part of the infrastructure of the global science system. The document is enclosed. Below you find the highlights from this document as presented by David Moorman.

First he explains that the recommendations are the result of a diplomatic process of setting a set of very high-level principles and guidelines.

There are two core principles:

- publicly funded research should be publicly available.
- open access is the default position; the rule is open access

To get all countries to agree on this is a major break through. NORFACE should work towards a common data policy.

The recommendation, formulated as Principles and Guidelines, is meant to apply to research data, whether already in existence or yet to be produced, that are supported by public funds for the purposes of developing publicly-accessible scientific research and knowledge.

In the context of these Principles and Guidelines, "research data" are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific

research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated. Research data from public funding is defined as the research data obtained from research conducted by government agencies or departments, or conducted using public funds provided by any level of government. Given that the nature of "public funding" of research varies significantly from one country to the next, as do existing data access policies and practices at the national, disciplinary and institutional levels, these Principles and Guidelines recognise that such differences call for a flexible approach to improved access to digital research data.

Access arrangements are defined as the regulatory, policy and procedural framework established by research institutions, research funding agencies and other partners involved, to determine the conditions of access to and use of research data.

Principles:

The recommendations refer to the following thirteen principles:

- A – Openness
- B – Flexibility
- C – Transparency
- D – Legal Conformity
- E – Protection of Intellectual Property
- F – Formal Responsibility
- G – Professionalism
- H – Interoperability
- I – Quality
- J – Security
- K – Efficiency
- L – Accountability
- M – Sustainability

[For an explanation of these principles please refer to the actual document in Enclosures – the authors].

In collaborative research programmes or projects, and especially in international scientific co-operation or in research projects based on public private partnerships where there are differences

in regulatory frameworks, the parties involved should negotiate digital research data sharing arrangements as early as possible in the life of the research project, ideally at the initial proposal stage. This will help ensure that adequate and timely consideration will be given to issues such as the allocation of resources for sharing and sustainable preservation of research data, differences in national intellectual property laws, limitations due to national security, and the protection of privacy and confidentiality.

Access arrangements also should be responsive to factors such as the characteristics of the data, their potential value for research purposes, the level of data processing (raw versus partially processed versus final), whether they are homogeneous data from a facility instrument or sensor versus heterogeneous field data collected by single researchers, data on human subjects or physical parameters, and whether the data are generated directly by a government entity or as a result of government funding. These variations in the origin or type of data should be taken into consideration when establishing data access arrangements.

Questions and discussion time

The conference participants agree that several questions need to be dealt with by NORFACE if they want to develop a common policy in data & research infrastructure. They are for instance:

- What data should be captured? All archives determine what they keep and what they throw out; the criteria is: what has most value for the future.
- What does it all cost? Costs are defined related to the same criteria: what is important for the future?
- What about intellectual property? See the OECD document on this topic.
- What about priorities?
- Can I trust your confidentiality? The debate on privacy and confidentiality will go forward continually.

- What – and when – will be the return?
- What about real long-term data-archiving?

Question: How do we convince researchers about data-archiving? Researchers are afraid to loose grip on their data.

Discussion: Kevin Schürer states that trust is a policing issue. Until a short time ago archiving was divorced from research, but that has been changed. Archiving is now embedded in research. He describes how his organization UK Data Archive keeps in contact with researchers and how the researchers welcome the message that “their data are scheduled for deposit”. In short: engagement is the keyword here.

Question: How does open access relate to costs?

Answer from David Moorman: This is a tough issue, but we must keep in mind that it is all about moving tax payers’ money and that every move of money costs something too. The main question is: who pays for what?

Question: What happens with the OECD recommendations? Will the researchers community react?

Answer from David Moorman: the heart of the matter is that the research Ministers have accepted open access as the default position. Now it is public responsibility to establish the frameworks, and researchers will comply. In the UK the NIH regulations are very well accepted.

Question: what to do with countries who do not adhere to the OECD recommendations?

Answer from David Moorman: Free-riders will always be there, but through moral persuasion most countries will be working towards a global science system together.

Plenary discussion with K. Schürer, discussion leader

The discussion starts with some introductory remarks by Kevin Schürer. He reminds the conference participants of the 'Description of work' of this conference. Also he refers to the Report questionnaires data & research infrastructures (The Hague, NWO, October 2006) which was the first product within this NORFACE Task. To his mind it is a very important report and it raises some very interesting points. The final policy document should bring those points further and add some more. The analysis started should be carried on. The timing of the report was very good. The OECD recommendations will reach a legal status in the next few months, and ESFRI just has published the European roadmap. NORFACE should use the OECD Principles as an exercise to measure how the NORFACE Partners comply to these guidelines. This is an excellent extension to the NORFACE Questionnaire.

In the discussion there were a lot of grand statements, a lot of visionary statements. A wide gap exists between some of the aspirations and the statements.

Kevin Schürer offers the following recommendations for plenary discussion.

- Virtually no discussion has yet been on how data can be transported across borders. Do researchers have the right, the tools, the knowledge to use data in other countries? NORFACE should consider this aspect more.
- The NORFACE partners should think more about investment and training.
- E-science and the GRID have been addressed little as yet. How are the research councils and governments engaged here? NORFACE should pay more attention to this topic.
- None of the ESFRI proposals will go forward unless the ministries are urged by the research councils. The NORFACE partners are urged to engage in the ESFRI roadmap very early and timely. A very important report is the document of the Social Sciences Workgroup. NORFACE should read this report line by line, and see how the partners would wish

to engage. Engagement is the keyword here: authors of such reports and the research councils need to engage. Information from the EC conference participants: in January preparatory bids will be called for concerning a data & research infrastructure activity; the instrument will probably resemble the Co-operation Actions (ERAnets) or COST (co-operation activities called 'Actions').

- Some Councils do not put money into research infrastructure, but only run projects. A recommendation to NORFACE should be that they encourage people – or develop a policy to that end – that they deposit their data in data archives.
- A bridge is needed between the funding streams for research infrastructure and the streams for research projects. NORFACE needs to consider how both streams join up.

Questions and discussion time; other recommendations

- The conference participants agree that big use of data is on government data and not on data produced by Funding Agencies.
- They also agree that researchers have a need for models being set for bilateral agreements on the use of data.
- E-science is a 'big thing', and it is also concerns sharing data. E-science is supporting collaboration across time and distance. One conference participant states that researchers do not use the infrastructure facilities enough. A recommendation to NORFACE is that all partners should develop an E-science programme in their country (there is one only in UK and NL right now).
- The NORFACE survey reveals a lot about national Funding Agencies policies on data & research infrastructures, but there is no relation visible with the European Framework Programmes. Researchers in EU/EC projects need EU data. NORFACE should address this issue more, and more explicitly.
- NORFACE should realize that its cooperation is only with twelve European countries (plus

- Canada) – the consortium should try to find leverage for the remaining 23 countries. There should be an international aspect to any NORFACE policy.
- Considering the cost aspect of Open Access: ESRC buys licenses from the World Bank. They may only be used by British researchers, but the UK Data Archive is inundated by requests from abroad to use these data. A truly international procurement of data is recommended. A collective action with many councils will lead to a cheaper buy and one for the whole of Europe.
 - It is surprising that the NORFACE survey does not contain a question on the international management for data. NORFACE should consider this subject and should investigate what the individual barriers are, what the most fruitful way is towards an international management of data; what the limitations are; how we dismantle them.
 - The subject of data access has been discussed in terms of crossing boundaries between countries. A recommendation should be that data access should also be discussed in terms of crossing boundaries between disciplines. The social sciences could also learn from the mechanisms for international cooperation, data use, et cetera applied by other academic disciplines.

Enclosures

Description of NORFACE Task: Task 4.7 – Data and Research Infrastructure

Data is the fuel of empirical research in the social sciences and scientists need scientifically driven data that is easily accessible and accurate if they are to understand, improve and critically test and validate empirical knowledge about a given subject. The task will look into the data practices and policies of research councils. The relevance for this is even more paramount with the enlargement of both the European Union and NORFACE with the pronounced uneven distribution of practices and capabilities, as some of the existing NORFACE partners have long standing traditions for handling data, whereas other partners are in the process of forming policies on this topic. A special concern is to

improve data-access, while at the same time looking at how feasible it is at an international level. Linking national and sub-national data and the issues arising with this, i.e. costs and international policies, must also be addressed. The implications of the developments in e-Science and the Grid should also be taken into account. A study will be conducted with the aim to identify best practice of the funding of research infrastructures, how and where data are archived; research councils' policies and procedures for dataaccess; policies for grant holders on data they produce; confidentiality and data protection. The Task will effectively spill into other NORFACE activities by providing a data policy for Tasks such as capacity building, the pilot and transnational programmes, as well as to European collaboration in the European roadmap (SSH) to ESFRI. Conference Programme (timetable)

Conference Programme (timetable)



DATA & RESEARCH INFRASTRUCTURE (NORFACE TASK 4.7)

Programme of the conference on 16 + 17 November 2006 in The Hague, The Netherlands

Theme: Developing a policy on data & research infrastructure by the NORFACE partners

Thursday 16 November 2006. Chair: Renée van Kessel-Hagesteijn

Room Barcelona

- | | |
|--------------------|--|
| 09.00 – 09.15 a.m. | Welcome and announcements (chair).
Opening of the conference by Anne Kovalainen (FI). |
| 09.15 – 10.00 | Key note speech on recent developments in the social sciences: growing dependency on data & research infrastructure; increase in scale and internationalization. Speaker: Norman Bradburn (USA). |
| 10.00 – 10.15 | Reflections on the key note speech |
| 10.15 – 10.45 | Theme 1: National roadmaps and Europe. Speaker: Peter Elias (UK). |
| 10.45 – 11.15 | Coffee break. Plaza. |
| 11.15 – 11.45 | Theme 2: Data access, including remote access, new ICT possibilities; international access to data. Speaker: Tore Risch (SE). |
| 11.45 – 12.15 | Theme 3: Data networking in Europe. Speaker: Peter Doorn (NL). |
| 12.15 – 12.30 | Presentation ESFRI. Speaker: Bjørn Henrichsen (NO). |

12.30 – 14.00 p.m. Lunch. NH Hotel Restaurant

Afternoon workshops. Room Barcelona and combined room Sevilla/Valencia

- | | |
|---------------|---------------------------------|
| 14.00 – 14.55 | Workshops: theme 1 and theme 2. |
| 15.00 – 15.55 | Workshops: theme 2 and theme 3. |
| 16.00 – 16.55 | Workshops: theme 3 and theme 1. |
| 17.00 – 17.15 | Drinks |

19.15 Departure for dinner.

Friday 17 November 2006. Chair: Jan Karel Koppen

Room Barcelona.

- | | |
|--------------------|---|
| 09.00 – 09.05 a.m. | Start of the second conference day and announcements (chair). |
| 09.05 – 09.45 | Key note speech on OECD and Research & Infrastructure. Speaker: David Moorman (CAN). |
| 09.45 – 10.00 | Reflections on the key note speech |
| 10.00 – 10.30 | Presentations of the recommendations by the chairpersons and/or rapporteurs from all workshops. |
| 10.30 – 10.45 | Coffee break. Plaza |
| 10.45 – 12.00 | Discussion; defining elements for a draft NORFACE policy on data & research infrastructure. Plenary discussion leader: Kevin Schürer. |
| 12.00 – 12.15 | A thank you and a farewell. |
| 12.15 – 14.00 p.m. | Lunch. <i>NH Hotel Restaurant.</i> |

Conference Programme (content)

DATA & RESEARCH INFRASTRUCTURE (NORFACE TASK 4.7)

Programme of the conference on 16 + 17 November 2006, The Hague, The Netherlands.

Theme: Developing a policy on data & research infrastructure by the NORFACE partners

CONTENT OF THE PROGRAMME

The starting document of the conference is the *Report of the NORFACE questionnaires on data & research infrastructure*.

Thursday 16 November 2006. Chair: Renée van Kessel-Hagesteijn (NWO, NL)

Room Barcelona.

09.00 – 09.15 a.m. Opening, welcome; announcements

After a welcome by NWO (chair) the conference is opened by Anne Kovalainen (FI), chair of the NORFACE Network Board.

09.15 – 10.00 **Key note speech on recent developments in the social sciences: the growing dependency on data & research infrastructure; increase in scale and internationalization. Speaker: Norman Bradburn (USA).**

This speech surpasses the three themes of the conference. Among others it will refer to the particular role of the funding agencies within the social sciences concerning data & research infrastructure. There is some time for asking questions and reflections (10.00 – 10.15 hrs).

10.15 – 10.45 **Theme 1: National roadmaps and Europe. Speaker: Peter Elias (UK).**

In this introduction the speaker will explain how to go about creating a national roadmap, based on his experience with the National Data Strategy in UK. The process will be explained in some detail. Also he will touch on the question how national roadmaps (can) relate to the ESFRI roadmap. Questions and reflections regarding the three introductions of the NORFACE themes may be offered at the end of the morning programme.

Afternoon workshop on this subject (14.00 – 14.55 hrs in room Barcelona and 16.00 – 16.55 hrs in combined room Sevilla/Valencia).

Discussion points:

- If a country is just at the start of developing a road map: what pitfalls are there to watch out for?
- How does one country connect to roadmaps in other NORFACE countries? What can they do together? How do individual country roadmaps connect to ESFRI?
- Are there any key research areas that the funding agencies could define and towards which they perhaps could direct their national subsidy instruments for data & research infrastructure (for example see the National Data Strategy of UK)? Could the NORFACE partners possibly direct parts of their budgets towards a specific topic or subject area and in that way intensify their cooperation?
- What kind of agreements could be made to realise this end?

Discussion leader: Savas Barkçin (TR; ESFRI).

Rapporteur: Luule Mizera (EE).

10.45 – 11.15 Coffee break

11.15 – 11.45 **Theme 2: Data access, including remote access, new ICT possibilities; international access to data. Speaker: Tore Risch (SE).**

This introduction will illustrate the importance of cyber-infrastructure and the GRID, E-science as exciting new possibilities for researchers.

Afternoon workshop on this subject (14.00 – 14.55 hrs in room Sevilla/Valencia and 15.00 – 15.55 hrs in room Barcelona).

Discussion points:

- How can the NORFACE partners – being funding agencies - support further development and promote the use of these possibilities in the social sciences together?
- How can they support the wider availability of access to national data resources in order to support comparative research?
- How can they promote improvement of access to international data, e.g. by Eurostat and other international bodies?
- Is there a need for funding agencies to unite with their social scientists to achieve certain ends?
- How do or may NORFACE and ESFRI relate concerning this topic?

Discussion leader: Rob Procter (UK).

Rapporteur: Sheena Duffy (IE).

11.45 – 12.15 **Theme 3: Data networking in Europe. Speaker: Peter Doorn (NL).**

The speaker will illustrate the model of Dutch Data Archiving and Networked Services (DANS), will pay attention to models as developed in other countries and reflect on recent European developments.

Afternoon workshops on this subject (15.00 – 15.55 hrs in room Sevilla/Valencia and

16.00 – 16.55 hrs in

room Barcelona)

Discussion points:

- How may the funding agencies create a surplus value of their social sciences research by promoting international data sharing? What role can - or do - the funding agencies play here?
- How do European networked data organisations relate to initiatives concerning open access for administrative data and for researchers' data?
- How would European networked data organisations work together with national statistics bureaus or offices in the NORFACE countries?
- How do or may NORFACE relate to ESFRI concerning this topic?

Discussion leader: Ekkehard Mochmann (DE).

Rapporteur: Marcus van Leeuwen (NL); Luisa Henriques (PT).

12.15 – 12.30 **The European Strategy Forum on Research Infrastructures and the recent European roadmap for research infrastructures; report 2006 will be presented by Bjørn Henrichsen (NO; ESFRI).**

Friday 17 November 2006. Chair: Jan Karel Koppen (NORFACE Network Board, NL)

Room Barcelona

09.00 – 09.05 a.m. NWO welcomes all (chair).

09.05 – 09.45 **Key note speech on OECD and Research & Infrastructure.**

Speaker: David Moorman (CAN).

The role of OECD and their work on data & research infrastructure deserves to be highlighted and could be of great importance for a NORFACE policy. OECD work done in this field needs to be taken into consideration by NORFACE. Questions and reflections at 09.45 – 10.00 hrs.

NORFACE

10.00 – 10.30 **Presentations of the recommendations from the workshops.**

The chairpersons and/or their rapporteurs report from their workshops. They explain the recommendations the participants of the workshops have arrived at.

10.30 – 10.45 Coffee break. Plaza.

10.45 – 12.00 **Discussion; defining elements for a draft NORFACE policy on data & research infrastructure.**

The discussion leader and the conference participants together formulate elements of the document on a NORFACE policy on data & research infrastructure to be presented to the NORFACE Network Board.

Discussion leader: Kevin Schürer (UK).

12.00 – 12.15 A thank you and a farewell by Jan Karel Koppen (NWO, NL).

12.15 – 14.00 Lunch. *NH Hotel Restaurant*

NORFACE – ‘New Opportunities for Research Funding Co-operation in Europe’ in the social sciences – is a partnership of funding agencies in twelve European countries plus (associate partner) Canada for the period 2005-2008. For more information about the ERAnet NORFACE please see www.norface.org.

The November conference in 2006 is organised primarily for representatives of the NORFACE funding agencies and experts in data & research infrastructure related to them.

List of participants of the conference

Participants NORFACE Conference on Data & Research Infrastructure

Ainsaar, dr M. (Mare)

Organisation: Tartu University, Department of Science and Institutional Development
Position: Head Department of Science and Institutional Development, Tartu University
E-mail: Mare.Ainsaar@ut.ee
Country: ESTONIA

Anschütz, dr S. (Susanne)

Organisation: Deutsche Forschungsgemeinschaft
E-mail: Susanne.Anschuetz@dfg.de
Country: GERMANY

Barkçin, dr. S. (Savas)

Organisation: Tübitak Position: Advisor
E-mail: savas.barkcin@tubitak.gov.tr
Country: ESFRI

Boh, dr T. (Tomaz)

Organisation: Slovenian Research Agency
Position: Researcher on the project
E-mail: tomaz.boh@arrs.si
Country: SLOVENIA

Bradburn, prof. N. (Norman)

Organisation: University of Chicago, National Opinion Research Center
Position: Senior fellow
E-mail: Bradburn-Norman@norc.uchicago.edu
Country: UNITED STATES OF AMERICA

Christensen, L.C. (Lars)

Organisation: Danish Research Agency for Science, Technology and Innovation
Position: Special advisor E-mail: lach@frsk.dk
Country: DENMARK

Dekker, R.J.P. (Ron)

Organisation: Netherlands Organisation for Scientific Research (NWO)
Position: Director operational management & institutes
E-mail: dekkerr@nwo.nl
Country: THE NETHERLANDS

Dijkstra, A. (Alice)

Organisation: Netherlands Organisation for Scientific Research (NWO)
Position: HERA project officer E-mail: dijkstra@nwo.nl
Country: THE NETHERLANDS

Doorn, dr P.K. (Peter)

Organisation: Data Archiving and Networked Services
Position: Director
E-mail: peter.doorn@dans.knaw.nl
Country: THE NETHERLANDS

Duffy, S. (Sheena)

Organisation: Irish Research Council for the Humanities and Social Sciences
Position: Acting director E-mail: sduffy@irchss.ie
Country: IRELAND

Elias, prof. P. (Peter)

Organisation: University of Warwick, Institute for Employment Research
Position: Director
E-mail: p.Elias@warwick.ac.uk
Country: UNITED KINGDOM

Ervelä-Myrreen, E. (Eili)

Organisation: Academy of Finland
Position: NORFACE coordinator
E-mail: eili.ervela-myreen@aka.fi
Country: FINLAND

Fournier, dr. J.F. (Johannes)

Organisation: Deutsche Forschungsgemeinschaft
Position: Programme director
E-mail: Johannes.Fournier@dfg.de
Country: GERMANY

Ganzeboom, prof. dr. H.B.G. (Harry)

Organisation: Free University Amsterdam, Department of Social Research Methodology
E-mail: hbg.ganzeboom@fsw.vu.nl
Country: THE NETHERLANDS

Haegeland, dr. T. (Torbjørn)

Organisation: Statistics Norway
 Position: Head of Research-Labour market and firm behaviour
 E-mail: thd@ssb.no
 Country: NORWAY

Henrichsen, B. (Bjørn)

Organisation: Norwegian Social Science Data Services
 Position: Director
 E-mail: bjorn.henrichsen@nsd.uib.no
 Country: ESFRI

Henriques, dr. L.H. (Luisa)

Organisation: Portuguese Science and Technology Foundation
 Position: Senior advisor
 E-mail: lh@fct.mctes.pt
 Country: PORTUGAL

Kämper, dr. E.S. (Eckard)

Organisation: Deutsche Forschungsgemeinschaft
 E-mail: Eckard.Kaemper@dfg.de
 Country: GERMANY

Kastrinos, dr. N.K. (Nikolaos)

Organisation: European Commission
 E-mail: nikolaos.kastrinos@ec.europa.eu
 Country: EUROPEAN COMMISSION

Kaur-Pedersen, S.K.P. (Sune)

Organisation: Danish Research Agency for Science, Technology and Innovation
 Position: Head of section
 E-mail: spe@fist.dk
 Country: DENMARK

Kessel-Hagesteijn, dr R.R. (Renée) van

Organisation: Netherlands
 Organisation for Scientific Research (NWO)
 Position: Director Social Sciences
 E-mail: kessel@nwo.nl
 Country: THE NETHERLANDS

Koppen, dr J.K. (Jan Karel)

Organisation: Netherlands
 Organisation for Scientific Research (NWO)
 Position: Director policy development & support
 E-mail: koppen@nwo.nl
 Country: THE NETHERLANDS

Kovalainen, prof. A. (Anne)

Organisation: Academy of Finland, Research Council for Culture and Society
 Position: Vice-chair + chair NORFACE Network Board
 E-mail: anne.kovalainen@tse.fi
 Country: FINLAND

Leeuwen, M.G. (Marcus) van

Organisation: Netherlands
 Organisation for Scientific Research (NWO)
 Position: Staff member
 E-mail: m.vanleeuwen@nwo.nl
 Country: THE NETHERLANDS

Mizera, dr. L.M. (Luule)

Organisation: Estonian Science Foundation
 Position: Board member
 E-mail: luule@etf.ee
 Country: ESTONIA

Mochmann, dr E. (Ekkehard)

Organisation: Central Archive for Empirical Social Research, ZAGESIS
 Position: Director E-mail: mochmann@za.uni-koeln.de
 Country: GERMANY

Moorman, dr D. (David)

Organisation: Social Sciences and Humanities Research Council of Canada
 Position: Senior policy advisor
 E-mail: david.moorman@sshr.ca
 Country: CANADA

Niemi, M. (Marko)

Organisation: Academy of Finland
 Position: Science advisor
 E-mail: marko.niemi@aka.fi
 Country: FINLAND

O’Riain, prof. S.O.R. (Seán)

Organisation: National University of Ireland
– Maynooth
Position: Professor of sociology
E-mail: sean.oriain@nuim.ie
Country: IRELAND

Pöder, I.P. (Ille)

Organisation: Estonian Science Foundation
Position: Chief specialist
E-mail: ille@etf.ee
Country: ESTONIA

Procter, prof. R.N. (Rob)

Organisation: National Centre for e-Social
Science
Position: Research director E-mail: rob.procter@
ncss.ac.uk
Country: UNITED KINGDOM

Risch, prof. J.M. Tore

Organisation: Uppsala University
Position: Professor
E-mail: Tore.Risch@it.uu.se
Country: SWEDEN

Schürer, prof. K. (Kevin)

Organisation: UK Data Archive
Position: Director
E-mail: schurer@essex.ac.uk
Country: UNITED KINGDOM

Sigfúsdóttir, I.D.S. (Inga)

Organisation: Reykjavik University
Position: Dean of the School of Health and
Education
E-mail: ingadora@ru.is
Country: ICELAND

Štebe, J. (Janez)

Organisation: Social Science Data Archives,
Slovenia
Position: Director
E-mail: Janez.stebe@fdv.uni-lj.si
Country: SLOVENIA

Strijp, dr. R.A. (Ruud)

Organisation: Netherlands
Organisation for Scientific Research (NWO)

Position: Policy officer

E-mail: strijp@nwo.nl

Country: THE NETHERLANDS

Tanner, S. (Suzanne)

Organisation: Economic and Social Research
Council
Position: Senior policy manager
E-mail: suzanne.tanner@esrc.ac.uk
Country: UNITED KINGDOM

Theofilatou, ms M. (Maria)

Organisation: European Commission, DG
Research, Research Infrastructures
Position: Project officer
E-mail: Maria.Theofilatou@ec.europa.eu
Country: EUROPEAN COMMISSION

Whelan, prof. B. (Brendan)

Organisation: Economic & Social Research
Institute
Position: Director
E-mail: brendan.whelan@esri.ie
Country: IRELAND
Organisation:

Bonenkamp, B.J. (Berry)

Organisation: Netherlands Organisation for
Scientific Research (NWO)
Position: Secretary NWO Social Sciences and
Management Team member NORFACE
E-mail: bonenkamp@nwo.nl
Country: THE NETHERLANDS

Bouma, C.M. (Margreet)

Organisation: Netherlands Organisation for
Scientific Research (NWO)
Position: NORFACE Task leader Data and
Research Infrastructure
E-mail: bouma@nwo.nl
Country: THE NETHERLANDS

Rövekamp, C.A. (Carien)

Organisation: Netherlands Organisation for
Scientific Research (NWO)
Position: Senior secretary/assistant NORFACE
conference
E-mail: rovekamp@nwo.nl
Country: THE NETHERLANDS

NORFACE

Questionnaire survey

The questionnaire was sent out to the twelve members of the NORFACE Management Team in Europe and the Associate partner in Canada in April 2006. The report was produced in October 2006 (*The Report questionnaires Data & Research Infrastructure NORFACE Task 4.7. The Hague, NWO, 2006. 45 pp.*).

Identification

Name of the NORFACE partner:
Name of respondent: ; E-mail: Telephone:

Definitions of concepts; Typologies

Does your organization use a definition of the concept of 'Data & Research Infrastructure'?

...*	Yes	Please write down the definition here: ...*
...	No	

**Please mark the correct box with an X, and write your text in the box where appropriate.*

About data resources: In the UK/ESRC document entitled 'National Strategy for Data Resources for Social Research' by prof. P. Elias (enclosed), six broad types of data resources are considered to be especially relevant for the social sciences:

Census and population surveys; Administrative data; Longitudinal surveys;
Socio-medical data; Business data; International data.

Is this typology suitable also from the point of view in your country?

...	Yes	
...	No	Please indicate what amendments are necessary from your point of view: ...

About research infrastructure: Is there any definition (of types of) research infrastructures in the social sciences in your organization?

Examples: databases, technical instruments (apparatuses), digitisations, corpora in language & linguistics, virtual laboratories, metadata linking of existing databases

...	Yes	Please give your definition of research infrastructure here: ...
...	No	

Subsidy Instruments

What subsidy instruments are being used for the funding of data & research infrastructure in the social sciences by your organization (please give the name in English, and explain in one to three sentences how they work, including the budget involved)?

Name, working, budget of subsidy instruments for data & research infrastructures: ...

Please indicate the average budgets of your subsidy instruments for data & research infrastructure in the social sciences and the average frequency per year (for the last five years):

Availability Yes/No	Budget indication for subsidies data & research infrastructure in the social sciences	Average number of projects
...	€ 110.000 – 900.000	... projects per year
...	€ 900.000 – M€ 5	... projects per year
...	M€ 5 – M€ 10	... projects per year
...	M€ 10 – M€ 20	... projects per year
...	M€ 20 – M€ 50	... projects per year
...	M€ 50 – M€ 100	... projects per year
...	More: M€	... projects per year
...	Otherwise:

Are there any examples of data & research infrastructure projects in the last two years financed by your organization with a clear, explicit *international aspect* (for instance co-ordination, co-operation, co-financing)?

...	Yes	Please give name, budget and an indication of the international aspect: ...
...	No	

Access to Data & Research Infrastructures; Training

Does your organization subsidize activities that improve access to research infrastructures for external researchers, either from your own country or abroad?

...	Yes	Please described these activities (the title of the activity and in a few words the way the improvement of access is being worked on): ...
...	No	

Does your organization subsidize research training or courses supporting the use of research infrastructure?

...	Yes	Please describe the aim of the research training/courses and the research infrastructure dealt with.: ...
...	No	

Does your organization have a policy or specific procedures concerning the access of data & research infrastructures in the social sciences?

...	Yes	Please explain the policy or specific procedures in a few words: ...
...	No	

NORFACE

Data-protection, Confidentiality, Legal Aspects

In what way is any particular interest expressed in your organization for ethical considerations concerning the sensitive nature of linked data, by legal requirements and by the resource costs of linkage?

...	Yes	Could you indicate relevant documents, persons or institutions that play a particular role in these considerations? ...
...	No	

Do you know of any particular recent developments in arrangements for research access to sensitive data in your country and/or developments relevant for your organization?

...	Yes	Please describe developments in arrangements in a few words, and perhaps mention names of relevant persons and/or institutions: ...
...	No	

Policy

For an example please refer to the JK/ESRC document 'National Strategy for Data Resources for Social Research' (enclosed).

Is there a policy or strategy document - for perhaps the next 5 to 10 years - in your country concerning data & research infrastructure?

Yes/No.

If you know of any such document in your country in the English language, could you please send a copy to NWO, leader of this NORFACE task (preferably by E-mail)?

If there is no such document – could you indicate how matters concerning policy or strategy in data & research infrastructure are being discussed or handled in your country? (Please use no more than five lines, indicating for instance persons and/or institutions in your country relevant for this topic.)

Policy or strategy for handling data & research infrastructure in my country (procedures/persons/institutions): ...

A policy or strategy document for international data: do you know of any document containing a similar plan concerning international data, resources and needs – perhaps to position the development of data resources in the context of European and global research interests and opportunities?

Yes/No. ...

If you know of any document on policy or strategy for international data & research infrastructure in your country in the English language, could you please send a copy to Task leader NWO (preferably by E-mail)?

Have any key research areas been indicated by experts in your country within which the social sciences will make important contributions over the next 5 to 10 years (areas relevant for policy development and for understanding social and economic causes and consequences)?

Example: In the ESRC document enclosed key research priorities are: the ageing population, migration, child development and 'early years' socialization, and the changing global economy.

...	Yes	Please mention the indicated areas here: ...
...	No	

Does your organization have a specific policy as to how and where data are archived, or does it contribute to any such policy in your country?

...	Yes	Please explain the policy concerning data archiving in a few words: ...
...	No	

Does your organization apply policies for grant holders on data they produce?

...	Yes	Please explain policies for grant holders in a few words: ...
...	No	

Towards a NORFACE Partners' Policy

The outcome of this NORFACE task is to be a policy on data & research infrastructure.

In what way do you think NORFACE partners could co-operate more in data & research infrastructure?

Examples: You could think of suggestions like: putting money together, develop a common policy on specific topics, improving access of databases, et cetera.

The NORFACE partners could co-operate more in data & research infrastructure ...

Do you see any possibilities for a transnational subsidizing of data & research infrastructure together with – some or all – NORFACE partners?

...	Yes	Please explain (subsidy instrument, type of project, budget, NORFACE partners, perhaps other funding organizations): ...
...	No	

Are there any new initiatives you consider to be worthwhile to be started at the NORFACE level and linked to activities of the European Commission (the European Strategy Forum on Research Infrastructures – ESFRI)?

...	Yes	Please comment on this: ...
...	No	

What elements do you consider to be essential for a NORFACE policy on data & research infrastructure?

...	Yes	Please indicate essential elements for a NORFACE policy on data & research infrastructure here: ...
...	No	

Are there any subjects that you would like to have discussed in the final workshop of this NORFACE Task 4.7?

For instance: the issues around linking national and sub-national data; implications of the developments in e-Science and the Grid, et cetera.

...	Yes	Please indicate subjects to be discussed in the final workshop of this NORFACE task here: ...
...	No	

NORFACE

Are there any persons or organizations in your country that you would like to have invited?

...	Yes	Please give names, E-mail addresses, telephone numbers, specific input to be expected of asked from them, et cetera – to a maximum of four suggestions: ..
...	No	

Comments

If you have any comments or additional remarks please give them here:

Comments or additional remarks: ...

Important documents referred to during the conference

N. Bradburn: Key Note Speech

Atkins, D. E. et al. (2003).

Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-ribbon Advisory Panel on Cyberinfrastructure. Washington, D.C.: National Science Foundation.

Berman, F. & Brady, H. E. (2005).

Workshop on Cyberinfrastructure for the Social and Behavioral Sciences, Final Report. Washington, D.C.: National Sciences Foundation

CBASSE-Commission on Behavioral and Social Sciences and Education (1998).

Investing in Research Infrastructure in the Behavioral and Social Sciences. Washington, D.C.: National Academy Press.

ESFRI-European Strategy Forum on Research Infrastructure (2006).

European Roadmap for Research Infrastructures. Luxembourg: Office for Official Publications of the European Commission.

National Science Board. (2003).

Science and Engineering Infrastructure for the 21st Century—The role of the NSF Washington, D.C.:NSB-02-190.

OECD Recommendation Concerning Access to Research Data from Public Funding

Introduction: Increasing the return on public investments in scientific research

The public science systems of OECD member countries are based on the principle of openness and the free exchange of ideas, information and knowledge. New information and communication technologies (ICTs), now in widespread use throughout all research disciplines, have greatly aided this system of free exchange and have opened up new avenues for collaboration and sharing. The progress of science, however, depends on more than just technologies. Research policies, practices, support systems and cultural values all affect the nature of new discoveries, the rate at which they are made, and the degree to which they are made accessible and used.

The power of computers and the Internet has created new fields of application for not only the results of research, but the sources of research: the base material of research data. Moreover, research data, in digital form, are increasingly being used in research endeavours beyond the original project for which they were gathered, in other research fields and in industry. Administrative data from the institutions of OECD member countries, such as employment information, are now used extensively in the social sciences, as well as in policy making. Data from public health organizations play a growing role in the advancement of life sciences. Similarly, geo-spatial data collected by many different government organizations are essential for environmental and other types of research. The list goes on.

Scientific databases are rapidly becoming a crucial part of the infrastructure of the global science system. The international Human Genome Project is but one good example of a large-scale research endeavour in which an openly accessible data repository is being used successfully by many different researchers, all over the world, for different purposes and in different contexts. Many other examples, involving research undertakings both large and small, are readily available.

Effective access to research data, in a responsible and efficient manner, is required to take full advantage of the new opportunities and benefits offered by ICTs. Accessibility to research data has become an important condition in:

- The good stewardship of the public investment in factual information;
- The creation of strong value chains of innovation; and,
- The enhancement of value from international co-operation.

More specifically, improved access to, and sharing of, data:

- Reinforces open scientific inquiry,
- Encourages diversity of analysis and opinion,
- Promotes new research,
- Makes possible the testing of new or alternative hypotheses and methods of analysis,
- Supports studies on data collection methods and measurement,
- Facilitates the education of new researchers,
- Enables the exploration of topics not envisioned by the initial investigators, and
- Permits the creation of new data sets when data from multiple sources are combined.

Sharing and open access to publicly funded research data not only helps to maximize the research potential of new digital technologies and networks, but provides greater returns from the public investment in research.

Throughout OECD member countries, continuously growing quantities of data are collected by publicly-funded researchers and research institutions. This rapidly expanding body of research data represents both a massive investment of public funds and a potential source of the knowledge needed to address the myriad challenges facing humanity.

To promote improved scientific and social return on the public investments in research data, OECD member countries have established a variety of laws, policies and practices concerning access to research data at the national level. Because the

exchange and use of research data are increasingly done on a global basis, international guidelines can help complement the approaches taken at the national level.

These *Principles and Guidelines* are meant to apply to research data that are gathered using public funds for the purposes of producing publicly-accessible knowledge. Also, the nature of “public funding” of research varies significantly from one country to the next, as do existing data access policies and practices at the national, disciplinary and institutional levels. These differences call for a flexible approach to data access and recognition that one size does not fit all. Moreover, the balance between the costs of improved access to research data and the benefits that result from such access must be judged by individual national governments and their research communities.

Whatever differences there may be between practices of, and policies on, data sharing, and whatever legitimate restrictions may be put on data access, practically all research could benefit from more systematic sharing. As the authors of the US National Research Council study, Bits of Power, pointed out:

The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research.

While publicly funded research data are subject to the default rule of openness under Principle A, this does not mean that all such data should be preserved permanently. The data archiving community appreciates the need for cost-benefit assessments and is constantly developing and refining retention protocols to ensure that those data sets with the greatest potential utility are preserved and made accessible.

The Council

Having regard to Article 5b) of the Convention on the Organisation for Economic Co-operation and Development of 14 December 1960;

Having regard to Rule 18b) of the OECD Rules of Procedure;

Recognising the wide range of benefits that arise from improving international access to, and use of, publicly funded research data, as expressed in the *Ministerial Declaration on Access to Research Data from Public Funding* of 30 March 2004 [C(2004)31/REV1];

Recognising that international exchange of data, information and knowledge is essential to the advancement of research and innovation in all Member countries;

Recognising the substantial benefits that science, the economy and society at large could gain from the opportunities that expanded use of research data have to offer;

Recognising that improved access to research data will increase the value derived from public investments in data collection, management and preservation;

Recognising that undue restrictions on access to, and use of, research data from public funding diminish the quality and efficiency of research and innovation;

Recognising that enhanced availability of research data from public funding for developing economies will enhance their participation in the global research system, thereby contributing to their social and economic development;

Recognising that rapid changes in information technology will require a careful and flexible approach to developing data access policies and that access arrangements must take into account differences in information infrastructures at the national and international levels and between research disciplines;

Recognising that the access to and use of research data from public funding may be constrained by domestic laws such as those on national security, the protection of privacy and confidentiality and intellectual property rights;

NORFACE

Recognising that delayed access to research data may be legitimate to allow for the right of first-use by researchers and for commercial exploitation;

Recognising that individual Member countries will implement these Principles and Guidelines according to national policies and practices;

Recognising that the balance between the costs of improved access to research data and the benefits that result from such access will be judged, taking into account the prevailing financial limitations, by individual Member countries;

Recognising the Member countries' declared commitment to working towards the establishment of access arrangements for research data from public funding, balancing the interests of open access to publicly funded research data with the need to restrict access in specific instances to protect social, scientific, and economic interests;

Having regard to the above-mentioned Ministerial Declaration which highlights the need to take the necessary steps to strengthen existing instruments and – where appropriate – to create within the framework of international and national law, new instruments and practices supporting

international collaboration in access to research data;

On the proposal of the Committee for Scientific and Technological Policy;

RECOMMENDS that Member countries take into consideration the Principles and Guidelines on Access to Research Data from Public Funding set out in the annex to this Recommendation and which form an integral part thereof and apply them, as appropriate for each Member country, to develop policies and good practices related to the accessibility, use and management of research data;

INSTRUCTS the Committee for Scientific and Technological Policy to review the implementation of this Recommendation as necessary;

INSTRUCTS the Committee for Scientific and Technological Policy to review the Principles and Guidelines on Access to Research Data from Public Funding, when appropriate, to take into account advances in technology and research practices, with the intention of further fostering international co-operation.

Annex: Principles and Guidelines for Access to Research Data from Public Funding

I - Objectives

These *Principles and Guidelines for Access to Research Data from Public Funding* (hereafter the “*Principles and Guidelines*”) provide broad policy recommendations to the governmental science policy and funding bodies of Member countries on access to research data from public funding. They are intended to promote data access and sharing among researchers, research institutions, and national research agencies, while at the same time, recognising and taking into account, the various national laws, research policies and organisational structures of Member countries.

The ultimate goal of these Principles and Guidelines is to improve the efficiency and effectiveness of the global science system. They are not intended to hinder its development with onerous obligations and regulations or impose new costs on national science systems.

The specific aims and objectives of these Principles and Guidelines are to:

- Promote a culture of openness and sharing of research data among the public research communities within Member countries and beyond;
- Stimulate the exchange of good practices in data access and sharing;
- Raise awareness about the potential costs and benefits of restrictions and limitations on access to and the sharing of research data from public funding;
- Highlight the need to consider data access and sharing regulations and practices in the formation of Member science policies and programs;
- Provide a commonly agreed upon framework of operational principles for the establishment of research data access arrangements in Member countries; and,
- Offer recommendations to Member countries on how to improve the international research data sharing and distribution environment.

The *Principles and Guidelines* contained in this document should assist governments, research support and funding organisations, research institutions and researchers themselves in dealing with the barriers and challenges in improving the international sharing of, and access to, research data. These Principles and Guidelines should be considered in light of, and applied to, the following major issues inherent in providing data access:

- Technological issues: access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, broad international agreement on interoperability, and effective data quality controls;
- Institutional and managerial issues: while increased accessibility is important to all science communities, the diversity of the scientific enterprise suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers;
- Financial and budgetary issues: scientific data infrastructure requires continued and dedicated budgetary planning and appropriate financial support. The use of research data will not be maximised if access, management, and preservation costs are an add-on or after-thought in research projects. It is important to note, however, that the cost of storing and managing data has decreased dramatically in recent years, and lack of knowledge about such changes can, in itself, be a barrier to advancement;
- Legal and policy issues: national laws and international agreements, particularly in areas such as intellectual property rights and the protection of privacy, directly affect data access and sharing practices, and must be fully taken into account in the design of data access arrangements; and,
- Cultural and behavioural issues: appropriate educational and reward structures are

a necessary component for promoting data access and sharing practices. These considerations apply to those who fund, produce, manage, and use research data.

In working towards better access to research data in the context of these *Principles and Guidelines*, Member countries will need to determine the appropriate balance between the costs of improved access to this data and the benefits that result from such access. The efforts to improve access, of course, need to be carried out within existing financial limitations.

II - Scope and definitions

These *Principles and Guidelines* are meant to apply to research data, whether already in existence or yet to be produced, that are supported by public funds for the purposes of developing publicly-accessible scientific research and knowledge. The *Principles and Guidelines* are not intended to apply to research data gathered for the purpose of commercialisation of research outcomes, or to research data that are the property of a private sector entity. Access to such data is subject to a range of considerations that are beyond the scope of this document. Moreover, in some instances, access to or use of data may be restricted to safeguard the privacy of individuals, protect confidentiality or proprietary results, or national security.

Research Data

In the context of these *Principles and Guidelines*, “research data” are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.

This term does not cover the following: laboratory notebooks, preliminary analyses, and drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (e.g., laboratory

samples, strains of bacteria and test animals such as mice). All of these products or outcomes of research require different considerations than those dealt with here.

These *Principles and Guidelines* are principally aimed at research data in digital, computer-readable format. It is in this format that the greatest potential lies for improvements in the efficient distribution of data and their application to research because the marginal costs of transmitting data through the Internet are close to zero. While the economics of distribution of data in analogue formats are significantly different from data in digital formats, most of the principles articulated here will be applicable to the traditional analogue environment as well.

Research Data from Public Funding

Research data from public funding is defined as the research data obtained from research conducted by government agencies or departments, or conducted using public funds provided by any level of government. Given that the nature of “public funding” of research varies significantly from one country to the next, as do existing data access policies and practices at the national, disciplinary and institutional levels, these *Principles and Guidelines* recognise that such differences call for a flexible approach to improved access to digital research data.

Access Arrangements

Access arrangements are defined as the regulatory, policy and procedural framework established by research institutions, research funding agencies and other partners involved, to determine the conditions of access to and use of research data.

III - Principles

A – Openness

Openness means access on equal terms for the international research community at the lowest possible cost, preferably at no more than the marginal cost of dissemination. Open access to digital research data from public funding should be easy, timely, user-friendly and preferably Web-based.

B – Flexibility

Flexibility requires taking into account the rapid and often unpredictable changes in information technologies, the characteristics of each research field and the diversity of research systems, legal systems and cultures of each Member country. Specific national, social, economic and regulatory implications should be considered when organisations develop research data access arrangements, and when governments develop policies to promote data access and review the implementation of these Principles and Guidelines.

C – Transparency

Information on data and data-producing organisations, documentation on the data and specifications of conditions attached to the use of these data should be internationally available in a transparent way, ideally through the Internet. Lack of visibility of existing digital data resources and future data collection poses serious obstacles to access. Factors to consider in ensuring transparency include:

- Information on data-producing organisations and their holdings, documentation on available data sets and conditions of use should be easy to find on the Web.
- Research organisations and government research agencies should actively disseminate information on research data policies to individual researchers, academic associations, universities and other stakeholders in the publicly-funded research process;
- Whenever relevant, all members of the various research communities should assist in establishing agreements on standards for cataloguing data. The application of existing standards should be considered, whenever appropriate, in order to avoid placing additional burdens on research resources and work loads of researchers and their institutions.
- Information on data management and access conditions should be communicated among data archives and data producing institutions, so that best practices can be shared.

D – Legal Conformity

Data access arrangements should take into

account the legal rights and legitimate interests of all stakeholders in the public research enterprise.

Access to, and use of, certain research data will necessarily be limited by various types of legal requirements, which may include restrictions for reasons of:

- National security: data pertaining to intelligence, military activities, or political decision making may be classified and therefore subject to restricted access.
- Privacy and confidentiality: Data on human subjects and other personal data are subject to protection under national laws and policies. Such data may be vulnerable to breaches of confidentiality and privacy, and therefore should only be obtained by fair and lawful means, and when appropriate, with the full knowledge or consent of the persons involved. Data custodians should apply anonymization or confidentiality procedures that ensure a satisfactory level of confidentiality while preserving as much data utility as possible for researchers.
- Trade secrets and intellectual property rights: data on, or from, businesses or other parties that contain confidential information may not be accessible for research.
- Protection of rare, threatened or endangered species: in certain instances there may be legitimate reasons to restrict access to data on the location of biological resources for the sake of conservation.
- Legal process: data under consideration in legal actions (sub judice) may not be accessible.

Subscribing to professional codes of conduct may facilitate meeting legal requirements.

E – Protection of Intellectual Property

Data access arrangements should consider the applicability of copyright or of other intellectual property laws that may be relevant to publicly-funded scientific databases. Factors to consider include:

- As public/private partnerships in the funding of research and related data production are increasing, balanced public/private arrangements should facilitate broad access to

digital research data where appropriate. The fact that there is private sector involvement in the data collection should not, in itself, be used as a reason to restrict access to the data. Consideration should be given to measures that promote non-commercial access and use while protecting commercial interests, such as delayed or partial release of such data, or the voluntary adoption of non-commercial use licensing mechanisms. Such measures can allow the primary participants to fully exploit the research data without unnecessarily shutting off access.

- Because both public research and governmental data collection are serving the same broader public interests and are both publicly funded, in those jurisdictions in which government data and information are protected by intellectual property rights, such legal protection should not be used to impede access to them or their re-use, particularly for public research and other public-interest application.

F – Formal Responsibility

Access arrangements should promote explicit, formal institutional practices, such as the development of rules and regulations, regarding the responsibilities of the various parties involved in data-related activities. These practices should pertain to authorship, producer credits, ownership, dissemination, usage restrictions, financial arrangements, ethical rules, licensing terms, liability, and sustainable archiving. Further, consideration should be given to the following:

- Many of the problems related to access, dissemination and sharing of, data result from the lack of explicit institutional agreements on the terms of access and use. With data management becoming ever more complex in certain areas of research, traditional informal arrangements between researchers may no longer be adequate and may need to be complemented by formally agreed practices and procedures.
- Responsibility for the various aspects of data access and management should be established in relevant documents, such as descriptions

of the formal tasks of institutions, grant applications, research contracts, publication agreements, and licenses.

- Long-term sustainability of the infrastructure required for data access is particularly important. Research institutions and government organisations should take formal responsibility for ensuring that research data are effectively preserved, managed and made accessible in order that they can be put to efficient and appropriate use over the long term.

G – Professionalism

Institutional arrangements for the management of research data should be based on the relevant professional standards and values embodied in the codes of conduct of the scientific communities involved.

Factors to consider include:

- The use of codes of conduct for professional scientists and their communities could help simplify and reduce the regulatory burden placed on access.
- Mutual trust between researchers, and trust between researchers, their institutions and other organisations, plays an important role in the establishment and maintenance of such codes of conduct.
- In current research practice, the initial data-producing researcher or institution is sometimes rewarded with temporary exclusive use of the data. The rules for such incentive arrangements should be developed and explicitly stated by the funding sources in co-operation with the affected research communities.

In certain areas of science, a lack of planning for and execution of the proper documentation and archiving of data sets is one of the key impediments to realising maximum value from the investment in research data. Project and program planning activities, at all levels, should expressly acknowledge data issues at the earliest stages to take into consideration funding and technical assistance for the essential organisation and curation of those data sets. Attention should be paid to incentives and the development of

professional expertise in all areas of research data management.

H – Interoperability

Technological and semantic interoperability is a key consideration in enabling and promoting international and interdisciplinary access to and use of scientific data. Access arrangements, should pay due attention to the relevant international data documentation standards. Member countries and research institutions should co-operate with international organisations charged with developing new standards.

Although science is becoming a highly globalised endeavour, incompatibility of technical and procedural standards can be the most serious barrier to multiple uses of data sets. Factors that should be considered include:

- A first requirement for interoperability should be explicit mention of the standards employed.
- Adoption of the practices of disciplines most advanced in this respect should be promoted, in particular by the international professional organisations dealing with science and the collection and preservation of data for scientific and technological purposes.
- The work of organisations engaged in setting more general information and communication technology standards should also be considered.

I – Quality

The value and utility of research data depends, to a large extent, on the quality of the data itself. Data managers, and data collection organisation, should pay particular attention to ensuring compliance with explicit quality standards. Where such standards do not yet exist, institutions and research associations should engage with their research community on their development. Although all areas of research can benefit from improved data quality, some require much more stringent standards than others. For this reason alone, universal data quality standards are not practical. Standards should be developed in consultation with researchers to ensure that the

level of quality and precision meets the needs of the various disciplines. More specifically,

- Data access arrangements should describe good practices for methods, techniques and instruments employed in the collection, dissemination and accessible archiving of data to enable quality control by peer review and other means of safeguarding quality and authenticity.
- The origin of sources should be documented and specified in a verifiable way. Such documentation should be readily available to all who intend to use the data and incorporated into the metadata accompanying the data sets.
- One good practice currently being adopted in the life sciences is to link access to data sets with access to the original research materials, and to link copied data sets with originals. This facilitates validation of the data and identification of errors within data sets. Although such linking is not always possible, because the original materials or context of the data collection may no longer exist, information and access linking of this type should be given consideration.
- Citations to data, and the recording of citations in indexes, are important indicators of data quality. At the moment, however, there is a lack of consensus on how to effectively cite data. Scientific organizations and research associations should engage this issue and develop appropriate practices.

J – Security

Specific attention should be devoted to supporting the use of techniques and instruments to guarantee the integrity and security of research data. With regard to guaranteeing the integrity of a data set, every effort should be made to ensure the completeness of data and absence of errors. With regard to security, the data, along with relevant meta-data and descriptions, should be protected against intentional or un-intentional loss, destruction, modification and unauthorised access in conformity with explicit security protocols. Data sets and the equipment on which they are stored should be protected as well

from environmental hazards such as heat, dust, electrical surges, magnetism, and electrostatic discharges.

K – Efficiency

One of the central goals of promoting data access and sharing is to improve the overall efficiency of publicly-funded scientific research to avoid the expensive and unnecessary duplication of data collection efforts.

Consideration should be given to the following:

- Data access arrangements should promote further cost effectiveness within the global science system by describing good practices in data management and specialised support services.
- [Add text from p. 2 here] Use of accepted retention protocols and thorough documentation of data will help to reduce unnecessary duplication of effort as well as to establish the necessary selectivity in preservation.
- Cost-effective production, use, management and archiving of data may require specialised support services, for example through the employment of non-academic specialists on specific research projects or the engagement of data management specialist organisations. Current practice suggests that the pairing of researchers with data specialists as a research project moves forward can be the most effective way to support individual researchers, the broader scientific community and improve the return on public investment in research.
- Insufficient incentives for researchers or database producers may lessen their efforts on data-related activities. The development of new reward structures and the adaptation of existing ones, including recognition of data management activities in tenure and promotion review, should be considered as a way to address this problem.

L – Accountability

The performance of data access arrangements should be subject to periodic evaluation by user groups, responsible institutions and research funding agencies. Although each party is likely

to use somewhat different evaluation criteria, the sum total of the results should provide a comprehensive picture of the value of data and of data access regimes. Such evaluations should help to increase the support for open access among the scientific community and society at large.

The following should be considered in establishing evaluation criteria:

- Overall public investments in the production and management of scientific data.
- Management performance of data collection and archival agencies.
- Extent of re-use of existing data sets.
- Knowledge generated from the re-use of existing data.
- The use of targeted foresight exercises to determine the nature and scope of data preservation activities and the types of data most likely to be needed in the future.

Even if gaining clear insight into the cost, benefit and performance of data access arrangements will not be an easy task, those in charge of data access arrangements should put effort into showing the benefits of open data access to justify and help ensure sustained support from all levels of government,

M – Sustainability

Due consideration should be given to the sustainability of access to publicly funded research data as a key element of the research infrastructure. This means taking administrative responsibility for the measures to guarantee permanent access to data that have been determined to require longterm retention. This can be a difficult task, given that most research projects, and the public funding provided, have a limited duration, whereas ensuring access to the data produced is a long-term undertaking. Research funding agencies and research institutions, therefore, should consider the longterm preservation of data at the outset of each new project, and in particular, determine the most appropriate archival facilities for the data.

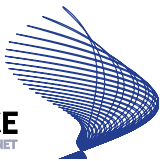
A Final Note – Timely Planning of Access Arrangements

Access arrangements, whether at the governmental or institutional levels, should be developed in

consultation with representatives of all directly affected parties. In collaborative research programmes or projects, and especially in international scientific co-operation or in research projects based on public private partnerships where there are differences in regulatory frameworks, the parties involved should negotiate digital research data sharing arrangements as early as possible in the life of the research project, ideally at the initial proposal stage. This will help ensure that adequate and timely consideration will be given to issues such as the allocation of resources for sharing and sustainable preservation of research data, differences in national intellectual property laws, limitations due to national security, and the protection of privacy and confidentiality.

Access arrangements also should be responsive to factors such as the characteristics of the data, their potential value for research purposes, the level of data processing (raw versus partially processed versus final), whether they are homogeneous data from a facility instrument or sensor versus heterogeneous field data collected by single researchers, data on human subjects or physical parameters, and whether the data are generated directly by a government entity or as a result of government funding. These variations in the origin or type of data should be taken into consideration when establishing data access arrangements.

NORFACE
ERA-NET



NORFACE - New Opportunities
for Research Funding
Co-operation in Europe

A Strategy for Social Sciences

www.norface.org



SIXTH FRAMEWORK
PROGRAMME

Women's Studies

Ecology

Geography

Pedagogy

Cognitive Sciences

Social Anthropology

Social Statistics and Informatics

Law

International Relations

Business and Administrative Sciences

Demography

Education Research