

Networked Data Services

**Towards a Future Data Infrastructure
for the Social Sciences in the Netherlands**

An advisory report by the Social Sciences Council

Royal Netherlands Academy of Arts and Sciences
Amsterdam, September 2003

© 2003 Royal Netherlands Academy of Arts and Sciences

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photo-copying, recording or otherwise, without the prior written permission of the publisher.

P.O. Box 19121, 1000 GC Amsterdam, the Netherlands

Telephone: +31-20-5510700

Fax: +31-20-6204941

E-mail: knaw@bureau.knaw.nl

www-address: <http://www.knaw.nl>

ISBN 90-6984-399-4

The paper in this publication meets the requirements of  iso-norm 9706

Preface

This report analyses the present data infrastructure for the social sciences in the Netherlands and strongly recommends setting up a new and better one.

A good data infrastructure is of the utmost importance for the social sciences. It affects two important, successive steps in empirical social research: first, the collection of data, and second, the archiving and in some cases the enrichment of data. The latter step serves to make data available and accessible for secondary analysis. Although both aspects are of equal importance and closely connected, this report focuses primarily on the second aspect, archiving and data enrichment. There are two reasons for this focus. First, social scientists have become increasingly dissatisfied with the quality and, in particular, the fragmentation of this part of the infrastructure. Second, a major reorganisation of the Netherlands Institute for Scientific Information Services (NIWI), an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), could lead to the services of NIWI being discontinued. As the Steinmetz Archive – at present the most important data archive for the social sciences – is part of NIWI, this development would pose a major threat to the existing data infrastructure.

The Social Sciences Council (Sociaal-Wetenschappelijke Raad /SWR) has therefore decided to bring forward the implementation of an existing plan: to advise on the future data infrastructure for the social sciences in the Netherlands. Early in 2003, after consulting the president of the Academy and the chair of the General Board of the Netherlands Organisation for Scientific Research (NWO), the SWR decided to install an advisory committee. The committee was asked to evaluate the present data infrastructure for the social sciences in the Netherlands and to outline the contours of a new infrastructure, in case the conclusion was drawn that the present one is unsatisfactory.

The present report presents the findings of the committee. The conclusions of the report are crystal clear: the present infrastructure is not satisfactory and should be replaced by a more effective and efficient one.

The SWR is very pleased with this report and wholeheartedly endorses the conclusions and recommendations of the committee.

Implementing the recommendations of the committee will require the co-operation of a number of organisations. The SWR sincerely hopes that all these organisations will share the sense of urgency expressed in this report and will be willing to take all the necessary measures to come to a better infrastructure for the social sciences.

Dr. J.J.A. Thomassen,
Chair Social Sciences Council

Content

0 Summary	7
1 Introduction	9
2 Functions in the context of an optimal data infrastructure for the social sciences	13
3 Collection of data	15
4 Reuse of data	17
5 Towards a future data infrastructure for the social sciences	23
6 Investments in data infrastructure for the social sciences	27
7 Recommendations	29

List of references	31
--------------------	----

Annexes

1 Descriptions of concepts	35
2 Structured questionnaire	36
3 International comparison of data services centres	37

List of acronyms	39
------------------	----

The landscape of data-users and data-providers in the Netherlands is a fluid and rather complex one. The users and providers work at different locations, are part of different organisations and have differing objectives. The picture may be summarised as fragmented.

Data services for empirical social scientists have two main functions. The first is focused on data collection, most frequently by means of surveys or official public records. Related to this is the second function, which is focused on facilitating the reuse of data by archiving, documenting, preserving and distributing the collected data and servicing the social sciences research community.

The context of these functions has changed considerably over the last two decades and the related organisational structures have only been partially adjusted in response to the changed requirements. The changing context refers to the increased internationalisation of the social sciences and international collaboration, but also to the increased use of data, increased costs of data collection, improved methods of data collection and higher and increased demand for micro-data, which could not be met by data-providers.

On account of public discussions concerning privacy issues and lack and inaccessibility of data from the Netherlands in the 1980s and early 1990s, empirical social scientists engaged in economic research turned their backs on data from the Netherlands. This was in part stimulated by the availability, at low costs, of data from other countries. As research and data collection are financed out of taxpayers' money, data may be regarded as a public good. However, the debate about data as a public good has not yet reached a final conclusion. A response to the increased costs of data collection is the process of data enrichment, which in itself may provide an important stimulus for the development of the empirical social sciences.

In order to stimulate the future development of the empirical social sciences there is a clear need for the integration of national data service functions. This will have a significant added value for these sciences and will moreover unify the at present fragmented landscape of data-users and data-providers in the Netherlands as well. The proposed solution emerges from the presented analysis in a generic sense as it provides a different perspective, i.e. it represents a transformation from a supply and data production oriented perspective towards a demand and data services orientation. In a technical sense it makes use of network technology provided by information and communication technology to link centres in a national structure. Also in an organisational sense a perspective different from the traditional supply-oriented approach is needed. A network solution is being recommended in which all the actors become interdependent, and organisational (scope) and network economies are generated.

The following is recommended:

- Establishment of a data services centre. This centre would unite the present functions of the Steinmetz Archive, the Netherlands Historical Data Archive

and the Scientific Statistical Agency, with new functions, such as linking data at the request of topical groups, offering training and training facilities for new user groups such as young empirical social scientists and for various groups, e.g. employees of Statistics Netherlands, and housing the Data Protection Officer. The employees of the recommended data services centre would have technical, methodological, archival and communicational skills; the director would have a background in the social sciences. The data services centre would need to maintain close contacts with the topical expertise centres. The provision of data would be free of charge.

- Encouragement of topical expertise centres, serving different areas of scientific interest. Each topical expertise centre would act as a forum for empirical social scientists on the selected topic and therefore maintain close contacts with them. The topical expertise centres would be linked to the data services centre and be complementary through their emphasis on content. The employees of topical expertise centres would be researchers, skilled in methodology.

It is recommended that a scientific steering group be established in the very near future. Its main task would be actively to stimulate the simultaneous establishment of the data services centre and the topical expertise centres. The group should be small, in order to act goal oriented, capable of integrating the various activities. It would need to be composed of a selected number of representatives of the Netherlands Organisation for Scientific Research, the Netherlands Royal Academy of Arts and Sciences, the universities, and Statistics Netherlands. The time-horizon for the complete development and full integration would probably be a decade. It would develop its final form in three stages of roughly equal length. In the first stage of the development, the scientific steering group would give priority to establishing an operational data services centre and a limited number of topical expertise centres. The possibilities for setting up a topical expertise centre purposing on qualitative data should also be explored, which, if successful, would be incorporated in the second stage of the development. After the third stage, at least five topical expertise centres (with total estimated annual operating costs once fully operational of 2.5 million euros) and the data services centre (with total estimated annual operating costs once fully operational of 3.5 million euros) would be fully operational.

Background

The Social Sciences Council (Sociaal-Wetenschappelijke Raad/ SWR)¹, an advisory council of the Royal Netherlands Academy of Arts and Sciences (Koninklijke Nederlandse Akademie van Wetenschappen/ KNAW), intends to advise on the future data infrastructure for the social sciences. This intention has the approval of the President of the KNAW and the chair of the General Board of the Netherlands Organisation for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek/ NWO).

The main reason for advising on this matter is that prominent empirical social scientists have become aware of the fragmentation of the data infrastructure for the social sciences. They expect that with the reduction of this fragmentation, the improved data infrastructure would more adequately meet the expectations of the researchers.

There is a sense of urgency about this advise, because it is expected that the Netherlands Institute for Scientific Information Services (Nederlands Instituut voor Wetenschappelijke Informatiediensten/ NIWI), a KNAW institute, will be discontinued in its present form. A decision must therefore be taken on the tasks, organisation and positioning of the archiving function for social sciences data – at present mainly performed by the Steinmetz Archive of the NIWI – and on the type of institution that will host this archiving function in the future.

In addition, a network of labour market researchers in the Netherlands intends to set up a centre for labour market studies, which would also include facilities for the enrichment of labour market data. On the one hand, this initiative is regarded as an important investment for the social sciences, but on the other, it may lead to a further fragmentation of the data infrastructure for the social sciences.

In the European context, developments regarding the data infrastructure for the social sciences present major opportunities for international research cooperation in the future, but also call for a certain level of international coordination and allocation of tasks.

Mandate of the Committee ‘Towards a future data infrastructure for the social sciences’

The SWR has installed a Committee with Dr. A.J.A. Felling (member; Professor of Methodology at Nijmegen University), B. Henrichsen (member; Director of the Norwegian Social Science Data Services, Bergen, Norway), Dr. P. Hooimeijer (member; Professor of Demography at Utrecht University), Dr. G.A. van der Knaap (chair; Professor of Economic and Social Geography at Erasmus University Rotterdam), E. Mochmann (member; Managing Director of the Central Archive for Empirical Social Research, University of Cologne, Germany) and Dr. A. Vollering (secretary; Secretary SWR, Amsterdam).

¹ See the ‘List of acronyms’ for the acronyms used in this report.

The tasks of the Committee are:

- To analyse the functions to be fulfilled for an optimal data infrastructure for the social sciences.
- To determine the present quality level and institutional context of these functions.
- To evaluate, in the light of the functions that should be fulfilled, the long-run efficiency and effectiveness of the present data infrastructure for the social sciences. If the Committee expects the present data infrastructure to be inefficient and ineffective in the long run, it will outline the global contours of a more efficient and effective data infrastructure for the social sciences. It must be possible for this data infrastructure to be structurally embedded (i.e. intrinsically, organisationally, financially) in the Dutch context. The Committee also takes into account the possibilities of an increased division of labour between data services centres in Europe.

Under the Committee's terms of reference, at least the following disciplines are regarded as belonging to the domain of the social sciences in the broad sense: economics, educational sciences, political science, sociology and social geography. More specifically for its present task, the Committee is required to focus only on the empirical social sciences.

The orientation of the Committee should be towards:

- Quantitative rather than qualitative data.
- Social data (i.e. referring to people) and economic data (referring for example to enterprises), and not towards for example medical data.

Approach of the Committee

First of all, the Committee has studied relevant literature on data infrastructure for the social sciences (see the 'List of references'). Annex 1 provides descriptions of concepts used in this report.

During the period March-June 2003 the Committee interviewed a selected number of data-users and data-providers. For discussions with parties concerned, i.e. the KNAW, Statistics Netherlands (Centraal Bureau voor de Statistiek/ CBS), NWO, the Ministry of Education, Culture and Science (Onderwijs, Cultuur en Wetenschappen/ OCenW), and university research groups, the Committee used a structured questionnaire (see Annex 2).

The Committee met on 31 March 2003 and 26 May 2003 in Amsterdam. Due to time constraints and heavy schedules of the Committee members, the Committee frequently consulted by e-mail and telephone. These discussions were not only practical but also proved to be very effective.

The Committee presented its draft report to the SWR in August 2003.

Structure of the report

Chapter 2 describes the functions of data services for empirical social scientists in the context of an optimal data infrastructure for the social sciences.

Chapters 3 and 4 focus on these functions: the former on data collection and the latter on the reuse of data. In each chapter, an outline of selected relevant developments – of different content and with a different expected impact on the

functions – will first be given in the light of an optimal data infrastructure for the social sciences. Secondly, the present standard and institutional context of these functions are determined and evaluated in both chapters.

Chapter 5 broadly outlines the contours of a more efficient and effective data infrastructure for the social sciences. Chapter 6 focuses on investments that are involved with this data infrastructure.

The recommendations of the report are summarised in Chapter 7.

Functions in the context of an optimal data infrastructure for the social sciences

Strong segregation along disciplinary lines in data-use

The picture that emerged from the interviews with different data-users is a rather complex one. Economists and social geographers, for example, mainly draw on data from national and international data-providers: the CBS, Eurostat, Chambers of Commerce, government agencies and expertise centres, etc. For example, sociologists and political scientists mainly draw on data from national and international data-providers: data archives (national survey data from the Social and Cultural Planning Bureau [Sociaal en Cultureel Planbureau/ SCP] and the Dutch National Election Study [Nationaal Kiezersonderzoek/ NKO], and international survey data such as the Comparative Study of Electoral Systems/ CSES, the European Social Survey/ ESS, the Eurobarometers, the European Values Study/ EVS, and the International Social Survey Programme/ ISSP), the CBS, Eurostat and expertise centres, etc. Dutch sociologists and political scientists, however, are also closely involved in developing and providing content for the ESS, the EVS, CSES and the ISSP.

Data-users and data-providers

There is no clear dividing line between data-users on the one hand and data-providers on the other. A data-user is also a data-provider, e.g., when he first enriches his own data (meaning here: adding contextual data to his own survey data), then publishes on the basis of these data and afterwards offers the enriched data to other data-users. This data-user (=data-provider) can have several reasons for offering the data to other data-users.

On the other hand, a data-provider who does not perform empirical research in the social sciences as well is not a data-user. Such a data-provider is a professional data-provider. Examples of long-existing professional data-providers of Dutch data are the CBS and the Steinmetz Archive.

Scientific data processing functions

The concept of data infrastructure for the social sciences refers to the total production process from primary data collection to data analysis for scientific or policy purposes (including the related publication). It concerns the provision of services and generation of products at all stages of the scientific data production process, where each stage is designed by data-users who are also the actors demanding for specific types of services and products.

The scientific data processing functions follow the lines of empirical social sciences research:

1. Research design and operationalisation of a scientific hypothesis.
2. Data collection, including preparation for use and data documentation.
3. Data enrichment and servicing (archiving, making data available and accessible).

4. Elementary data processing (e.g. cross tables, graphs).
5. Exploratory data analysis (e.g., to what extent will the data-provide relevant information for empirical social scientists and policy-makers?).
6. Data analysis for scientific or policy purposes.

An actor can carry out the scientific data processing functions both sequentially and simultaneously (especially functions 5 and 6). Actors may also limit themselves to just one of the scientific data processing functions.

Each function can be carried out at different physical locations. In addition, similar processes can be carried out at more than one physical location.

There is a large variety of both data-demanding actors (i.e. data-users) and demand for services and products. This variety cannot be met in a singular organisational structure focused predominantly on the scientific data production process. Ideally, a data infrastructure should be designed to serve all these specific demands, i.e. this organisation of activities should be flexible and tailored to the particular stage in the process in order to meet demands and provide the right kinds of services and products.

Functions of data services for empirical social sciences

Data services for empirical social scientists have two main functions.

The first concerns the process of data collection itself. The data are most often obtained from surveys or official public records.

Related to this is the second function, which is focused on facilitating the reuse of data through archiving, documenting, preserving and distributing the collected data and servicing the social sciences research community. The main reasons for the reuse of data are:

- Scientific research must be replicable. The collected data must therefore be available and accessible for empirical social scientists wishing to replicate that research.
- Research and data collection are financed out of taxpayers' money. The data must therefore be available for other users.
- Reuse of data is often cheaper than new data collection.

Increased internationalisation of social sciences and international collaboration

Empirical social scientists have become more internationally orientated in recent decades. They often collaborate on data collection and analysis in a bilateral or multilateral setting at international level.

At national level, however, they hardly collaborate on data collection and analysis with the exception of political scientists involved with the NKO and sociologists involved with the Netherlands Kinship Panel Study/ NKPS.

Dutch empirical social scientists and their foreign colleagues have worked together to set up a number of international data-sets, e.g. ISSP, ESS, and EVS. A number of international social scientific networks are also developing. As a consequence, these empirical social scientists networks collect and share cross-national, longitudinal data.

Although NWO is not (yet) a professional data-provider itself, it has a number of social sciences data infrastructure funding programmes. NWO is an important actor in contributing to the national (e.g. the NKO and the NKPS) and international (e.g. the EVS, the CSES and the ESS) data infrastructure for the social sciences in the Netherlands.

With regard to the funding of international data collection, NWO also cooperates in the context of the European Sciences Foundation/ ESF.

And last but not least, NWO was a prominent player in an agreement with the CBS that resulted in an improved access to data for scientific analysis.

More enriched data

In addition to data collection by means of surveys, data enrichment has become increasingly important. Data enrichment means: adding contextual data to survey data. Data-users can perform data enrichment on data they have collected themselves or on data drawn from other sources, e.g. a data archive.

The reasons for the increasing importance of data enrichment are: (a) data collection is relatively labour-intensive; (b) technological developments (e.g. computer capacity, and satellite imaging for geographical data); (c) improved methodologies for data enrichment; (d) a greater variety of data have become available and accessible; (e) enriched data provide a better basis for scientific analysis.

Evaluative remarks

In the interviews, a number of data-users elaborated on the need to cluster research expertise with respect to the data enrichment of international topical data-sets. Although there already exist a number of topical expertise centres in the Netherlands, these users suggested that efforts should be made for a limited number of social sciences topics to catch up in the international context. In this

light, the interviewees pointed out that the availability of international data should also be taken into account.

Concomitant with the strong increase in the use of data during the last two decades, the Committee also observes an increased demand for specific data types that is not met by professional data-providers. In some areas of the social sciences, e.g. detailed regional labour market analysis, a reduction in the availability of data is even observable.

In addition to this, the Committee notes an increasing specialisation in different scientific areas. This is partly a response to government policy with respect to university research (i.e. the creation of research schools or the encouragement of centres of excellence) and partly a result of the growing number of empirical social scientists who are publishing actively.

These developments have among other things led to the creation of topical expertise centres at various stages of development. In this context it would be an oversight only to focus on the fragmentation that has developed over the last decades and not to take into account current and likely future developments as well. These developments are taking place not only within universities but also at research institutes associated with universities, such as the Centre for Labour Market Research and Education ROA (at Maastricht), or the Research Institute for Small and Medium-sized Business EIM (at Zoetermeer). Recent developments in the area of qualitative research may also be noted.

In order to develop a structure that is more than a number of isolated topical expertise centres lacking any complementary, it should be recognised that these centres potentially have a national or even an international added value. In order to take full advantage of this potential, the development of these centres should be encouraged while at the same time allowing for national integration to enable the wider science community to enjoy the positive effects of this development. This can be achieved by installing a national coordinating body to focus on the development of common standards and having the power to enforce this, while at the same time stimulating the development of expert knowledge at different locations. It goes without saying that part of such a structure would consist of a strong methodology group with expert knowledge of metadata handling. This new task could be performed by a national data services centre. In addition to this would be the complementary topical expertise centres, where small, specialised groups would share and develop substantive expertise, but would not perform scientific research themselves.

Topical expertise centres, constructed along the lines sketched above, could also be a highly effective stimulus for the further internationalisation of the social sciences.

Qualitative data and other types of primary data

More developed technologies – especially computer technology – have made it easier in recent decades for social scientists to handle a wider variety of data. This also applies to qualitative data. Qualitative data are (see: KWALON/ Platform voor Kwalitatief Onderzoek in Nederlands, 2002): (a) videotapes, audiotapes, transcripts of interviews, other discussions and observations; (b) notepapers that played a role in the process of data collection and analysis; (c) other texts and tapes that played a role in empirical research.

In contrast to the reuse of quantitative data, the reuse of qualitative data is still rare. This situation exists because the researcher who collects qualitative data usually claims the property rights to those data. There are many theoretical and practical reasons for this.

As a consequence, these data are not accessible for other researchers.

The first archive for qualitative data ('Qualidata') was set up in the United Kingdom in the mid-1990s. This has intensified the debate with respect to such matters as the quality of the data. Nowadays, a network of Dutch empirical social scientists who use qualitative data ('Kwalon') is striving to set up a warehousing function for their qualitative data. At present, however, they still lack a solid archive for this warehousing function.

More professional data-providers and increased quantity and quality of data

Most Dutch data used in economic and social geographic analysis are collected by public or semi-public institutions (e.g. the CBS, Chambers of Commerce and the statistical offices of large municipalities). Usually, these data are collected for administrative purposes and not (directly) for specific scientific purposes.

The number of professional data-providers in the Netherlands has grown in recent decades, not only because the number of data-providing units within public or semi-public institutions has grown, but also because the number of private institutions providing data has increased. The latter professional data-providers do not collect data for specific scientific purposes but when requested sell these data, e.g., to empirical social scientists. An example of this type of data is satellite images supplied by commercial companies that can be analysed by geographers for scientific or policy purposes.

The quantity and quality of data held in official public records have increased. E.g. government agencies like the Dutch Tax and Customs Administration (Belastingdienst), the Topographic Services The Netherlands (Topografische Dienst) and the unemployment offices hold more and better records.

The data provided by professional data-providers have become more available and accessible for empirical social scientists, notably during the last decade.

However, as has been already observed above, data have become less available in some areas of the social sciences. This is for example the case in the area of detailed regional labour market analysis.

Role of the CBS in providing data that are not primarily collected for scientific use

The main function of the CBS is data collection for policy purposes. These data are generally collected by means of surveys and from official public records.

The CBS provides some data to empirical social scientists that are not primarily collected for scientific purposes. The CBS is a major professional data-provider for economic data-use and – to a lesser extent – for social data-use.

The relationship between the CBS and the social sciences in the Netherlands has had its ups and downs. In the 1970s and 1980s, public discussions on data-use and disclosure risks had a major impact on the data protection measures taken by all Dutch data-providers. Furthermore, in the late 1980s the Ministry of Economic Affairs (Economische Zaken/ EZ), the soul funding agency of the CBS, confronted the CBS with severe budget cuts and suggested that the CBS should operate on a more commercial basis, charging data-users at cost. In the early 1990s NWO and the CBS agreed that NWO would pay the CBS NLG 1 million (about 450,000 euros) annually, and that CBS, in turn, would provide data at a reduced charge. At the same time, NWO set up an agency to act as an office for ordering CBS data under the CBS privacy restrictions. The latter has led to the establishment of the Scientific Statistical Agency (Wetenschappelijk Statistisch Agentschap/ WSA) in 1994. Ever since, the use of CBS data has increased. The WSA is therefore regarded as a success. After a few years, it started to perform other functions apart from its initial role: for data-users who have a subscription to the 'WSA data', the WSA also acts as an office for ordering data from other professional data-providers, including the Steinmetz Archive.

It is worth mentioning that the CBS is at present developing the Social Statistical Data (Sociaal Statistisch Bestand/ SSB) and the Economic Statistical Data (Economisch Statistisch Bestand/ ESB), in which data from official public records (e.g. tax payments and social security data) will be linked. As a result, a quasi-census for the Dutch population of persons (SSB) and of enterprises (ESB) will become available and accessible for empirical social scientists.

It is of interest to note that on 1 January 2004 the legal position of the CBS will change from that of a public institution into that of a semi-private institution, as part of a general process of withdrawal of the government from public services institutions leading to increased corporatisation of such institutions.

The data provided by public or semi-public institutions like the CBS are financed by taxpayers' money. These data serve not only governmental interests and needs but also the public at large. Therefore, these data may therefore be

regarded as a public good (see also Soete and Ter Weel, 2003) and the professional data-providers can provide these data free of charge to empirical social scientists.

Role of the Steinmetz Archive and the NHDA in providing data that are primarily collected for scientific use

The archiving function for social sciences data in the Netherlands is presently mainly fulfilled by the Steinmetz Archive. The Steinmetz Archive, established in the early 1960s by a group of empirical social scientists, places the data to the disposal of data-users and therefore enables data reuse. It thus provides data that are primarily collected for scientific purposes. The main function of the Steinmetz Archive is that of providing data for the reuse through archiving, documenting, preserving and distributing the collected data and servicing the social sciences research community.

The Steinmetz Archive acts as the national archive for social sciences data. This means a researcher who has finished data collection submits the data to the Steinmetz Archive for finding. An important data-provider that has filed its data at the Steinmetz Archive is the SCP.

The Steinmetz Archive – once a jewel in the crown of the empirical social sciences – gradually lost its splendour. It is not exactly clear what caused this. One reason may be that in the mid-1990s, the Steinmetz Archive was absorbed into a larger organisational unit within the NIWI, an institute of the KNAW. Being a small part of the NIWI, the Steinmetz Archive became less visible than before. Furthermore, the job cuts at the Steinmetz Archive over the past decade made it impossible for the organisation to function fully: the quality of service gradually decreased, making the Steinmetz Archive less visible than before. Another factor could be that the WSA nowadays operates as a one-stop shop for Steinmetz Archive data, which has hardly contributed towards the visibility of the Steinmetz Archive.

The table below gives an overview of selected social scientific data archives in Europe. It illustrates the current backward position of the Steinmetz Archive compared to those in other countries.

Country	Permanent positions (FTEs)	Temporary positions (FTEs)	Population size (million)
Denmark	* 14.5	?	5
Finland	10	4	5
Germany	**33	17	82
Netherlands	*** 4.2	-	16
United Kingdom	* 45	?	58

* Including 5 FTEs in the historical department in both Denmark and the United Kingdom.

** Germany excluding researchers but including 8 FTEs in the historical department.

*** Netherlands excluding a small but unknown number of FTEs in the NHDA.

? = Unknown.

The first two columns (concerning permanent and temporary positions) refer to the social sciences archives in the selected European countries. The third column shows the population of these countries.

Although the situation concerning preservation and the providing of access to research data in each country is different (see Annex 3 for a short report on an international comparison by the Social Sciences and Humanities Research Council of Canada and the National Archives of Canada, 2001b), and there is therefore no precise relationship between the number of permanent positions in a data archive and the population of the country, the above overview clearly indicates that the present Steinmetz Archive is, to say the least, far from robust. The small number of positions at the Steinmetz Archive is a matter of serious concern as it implies long-term instability concerning the reuse of data collected for scientific purposes.

Although the archiving function for social sciences data in the Netherlands is at present mainly performed by the Steinmetz Archive, a small part of this function is carried out by the Netherlands Historical Data Archive (Nederlands Historisch Data Archief/ NHDA), a professional data-provider. Like the Steinmetz Archive, the NHDA is a small unit with a very limited number of permanent positions within the NIWI, a KNAW institute.

The present objective of the NHDA is to archive and disseminate data-sets containing historical data; the NHDA has restricted itself to collecting data-sets concerning Dutch history, and/or data-sets created by Dutch researchers. These historical data-sets can be relevant for empirical social scientists in the Netherlands.

Data protection precautions

Because of privacy problems in the population census of 1971, the Dutch population has become very concerned with privacy aspects of individual data collected on people and enterprises and stored both in private and in a large number of public databases. Therefore, Dutch data-providers have become very careful about providing their data for public use – including scientific use. The principle of data protection has in itself become an important social asset.

There are now legal and ethical provisions under which the data are collected out of taxpayers' money. Data can remain publicly available, subject to conditions of fair use. Implementation of this principle does, however, involve complex issues. A major issue for empirical social scientists is to determine at which stage(s) in the scientific data process data need to be protected. To ensure confidentiality, care should be taken with the protection of the privacy of respondents, both persons and enterprises, at all stages.

The following precautions can be taken to ensure confidentiality (see also Sociaal-Wetenschappelijke Raad, 2003):

– Appoint a Data Protection Officer (*functionaris voor de gegevensverwerking*).

The notion of the Data Protection Officer was introduced in the present Dutch

data protection law (see: Minister van Justitie en Minister van Grote Steden- en Integratiebeleid, 2000). Data-users can use direct identifying data provided that they report their research to the Data Protection Officer.

- The data-user signs a contract with the data-provider containing the following clause: ‘As soon as the data-user receives the data-set from the data-provider, he distinguishes within the data-set a set of communication data and a set of research data, keeping the communication data for no more than six months after he received the data’.

These precautions are suggested in the proposed code of conduct for the use of identifiable data in scientific research. At present, the Association of Universities in the Netherlands (Vereniging van Samenwerkende Nederlandse Universiteiten/ VSNU) is implementing this code of conduct for all the universities in the Netherlands.

Evaluative remarks

As a consequence of the difficult data supply in the early 1990s, some empirical social science researchers engaged in economic research turned their backs on the hardly accessible data from the Netherlands. These researchers denounced the lack of data accessibility and availability. Ever since, they have performed social sciences research with more accessible data from other countries. Although their motivations were sincere and their research did still contribute to the development of the social sciences, one may wonder whether or not this was a desirable situation. It is clear that it is not a desirable situation from the viewpoint of the Dutch government. The number of studies related to Dutch society has been reduced and less insight about its structure and developments has become available for economic and social policy decision-making than hitherto. National data should be better preserved and made accessible as sources of evidence for policy-makers. It would be best for all if the situation of the early 1990s were never to happen again.

As already mentioned in Chapter 1, the NIWI is expected to be discontinued in its present form. The Committee considers there to be an urgent need to recapitulate the tasks, organisation and positioning of the archiving function for social scientific data at present fulfilled by the Steinmetz Archive. The same holds for the archiving function for historical data at present fulfilled by the NHDA.

The Committee considers it highly worthwhile for the WSA to act as an office for ordering CBS data under the CBS data protection restrictions. However, the CBS data and the WSA subscription are still too expensive for empirical social scientific research, because data provided by the CBS should be regarded as a public good. The CBS should therefore provide these data free of charge to empirical social scientists.

Under the same line of reasoning, the WSA may be regarded as a public good as well. Empirical social scientists (in fact: the organisational entity they belong

to, like the university department or research school) should not have to pay for the WSA subscription (25,000 euros a year).

Furthermore, the Committee notes an overlap between the functions performed by the WSA acting as an office for ordering data from the Steinmetz Archive and those performed by the Steinmetz Archive: a data-user can order data from the Steinmetz Archive directly or via the WSA. In the interviews, all data-users (representatives of the economic sciences, educational sciences, political sciences, social geography and sociology) and data-providers stated that this situation was unclear.

In the light of the above evaluative remarks, the Committee concludes that it is logical to unite the functions of the WSA, the Steinmetz Archive and the NHDA into a single data services centre. This centre should be an integral part of the scientific data production process, as these services will then prove themselves useful for empirical social scientists, and will add value to their work. In order to be a serious discussion partner for data-users as well as professional data-providers, the director of the data services centre should have a background in the social sciences.

The archiving function should be set up more seriously: also qualitative data, enriched data from the topical expertise centres, and e.g. older CBS data-sets should be archived by the data services centre.

Other tasks for the data services centre could also be considered, e.g. linking data at the request of topical groups, providing training and training facilities for different empirical social sciences research groups, acting as a partner in communications with professional data-providers (such as the Topographic Services The Netherlands) and data-users, supervising the contracts between data-providers and data-users to uphold the data protection legislation, and accommodating the – independent – Data Protection Officer.

And last but not least, the data services centre should act as the representative of the Netherlands in international settings (e.g. Council of European Social Science Data Archives/CESSDA) and monitor the development of the collection of foreign data archives.

The Committee regards its objective – in line with the objective formulated in the 2001b report of the Social Sciences and Humanities Research Council of Canada and the National Archives of Canada – as follows:

‘to create a “trusted system” that provides the research community with an accessible and comprehensive service empowering end users to locate, request, retrieve and use data resources in a simple, seamless and cost effective way’.

The ‘trusted system’ should have long-term stability: it must be intrinsically, institutionally and financially embedded in a solid manner. Therefore, the Committee recommends the following:

- The Steinmetz Archive, the NHDA and the WSA should be combined and turned into a data services centre.
- Topical expertise centres should be set up or encouraged. They should be linked with the data services centre in order to create a comprehensive national structure.
- A scientific steering group should be established to promote national integration.

The tasks of the **data services centre** would include to:

- Archive survey data-sets (an existing activity of the Steinmetz Archive and the NHDA), older CBS data-sets (a new activity for the Steinmetz Archive or the NHDA), qualitative data-sets (a new activity for the Steinmetz Archive or the NHDA) and enriched data-sets (a new activity for the Steinmetz Archive or the NHDA).
 - Represent the Netherlands in international forums (e.g. CESSDA) and monitor the development of the collection of foreign data archives.
 - Link data at the request of topical groups. Topical groups would be chaired by a steering group member or the director of the data services centre.
 - Provide training and training facilities for different empirical social sciences research groups, such as researchers from universities and research institutes, CBS employees and young researchers from research schools.
 - Communicate with professional data-providers and data-users, e.g. on acquisition of data; set standards; transfer (technical) knowledge with data-providers; document data according to metadata standards; update data guides (data guides on Internet and on paper); translate data guides in English; handle the orders from data-users; and organise meetings and summer schools for topical groups.
 - Supervise the contracts between data-providers and data-users to uphold the data protection law.
 - House the Data Protection Officer. The latter performs an independent role.
- These represent considerably more tasks than those at present of the WSA, the Steinmetz Archive and the NHDA.

In order to guarantee continuity of this data services centre, it must have a critical mass of employees. These employees have technical, methodological,

archival and communicational skills. They would not be researchers. The director would have a background in the social sciences.

The data services centre, mandated by NWO and the KNAW, must be located within the existing structure of Dutch scientific organisations, preferably close to an existing research group that already has demonstrated considerable expertise in data methodology. However, if a university were physically to host the data services centre, the latter would need to be financially and otherwise independent from its host so as to ensure that the hosting university did not prevent other universities from using the data services products.

The report by the Social Sciences and Humanities Research Council of Canada and the National Archives of Canada (2001b) recommends the creation of a new national research data archival service in Canada. The tasks of the recommended service in Canada are roughly comparable with those of the recommended data services centre in the Netherlands. The annual operating costs for a comprehensive Canadian facility and network are benchmarked in the area of 3 million Canadian dollars. With this in mind, the annual operating costs for the recommended data services centre may be estimated at 3 million euros a year. Furthermore, the additional annual operating costs of the data services centre for an extra topical expertise centre (see below) is estimated at about 0.1 million euros.

The provision of data were, however, be free of charge.

The data services centre should develop common standards jointly with the topical expertise centres, stimulate and harmonise the development of course ware on data methodology issues and organise topical seminars. In this way an effective and efficient national data infrastructure might emerge.

Topics for **topical expertise centres** would include value studies, election studies, and research on small and medium-sized enterprises. The tasks of a topical expertise centre are to:

- Be a forum for empirical researchers on the selected topic.
- Organise meetings and summer schools for topical groups.
- Train CBS employees and young researchers.

The topical expertise centres would complement the data services centre and be linked to it in such a way that for the outside world they operated as a single networked organisation.

The employees of topical expertise centres would be researchers, skilled in methodology, but at the same time specialised in the subject area of the centre itself, for example in educational, geographical or political science.

In order to stimulate data reuse, the topical expertise centres would of course be obliged to archive their linked and enriched data-sets at the data services centre in accordance with jointly developed methodologies and rules.

On basis of the Canadian report (2001b) and information obtained in the interviews, the annual operating costs for a topical expertise centre would be 0.5 million euros.

Each topical research centre would need to maintain close contacts with the empirical researchers on the selected topic.

In addition, a **scientific steering group** should be set up with a selected number of representatives from NWO, the KNAW, the universities and the CBS. The main task of this group would be actively to stimulate the establishment of the data services centre and to stimulate and coordinate the development of the topical expertise centres. The group should be of a limited size in order to realise the stated objectives. The chair should have a background in the social sciences.

It is of course difficult to estimate the amount of money needed for the future data infrastructure for the social sciences. Furthermore, what is needed is not always available.

In order to have a somewhat broader perspective on investments in data infrastructure for the social sciences, it is useful to compare the needs for large and very large investments for the different types of sciences. RAND Europe (2001) has made an inventory of the present and future needs for investment in three large scientific clusters: an alpha cluster (humanities), a beta cluster (medical sciences, technical sciences, science, chemistry, biology, geosciences, environmental sciences, agricultural sciences), and a gamma cluster (social sciences). The report states that the present investment rate for the gamma cluster is about 7 percent. Since the present budget for the social sciences is about 300 million euros a year, the present investment therefore amounts to between 20 and 22 million euros a year.

The report also states that the investment rates during the next decade should be 7 percent for the alpha cluster, 15 percent for the beta cluster and 10 percent for the gamma cluster. It should be noted that these were also the investment rates of ten or fifteen years ago. Obviously, although the *de facto* investment rates have fallen far below the level of investment rates of ten or fifteen years ago, the desired investment rates have not changed.

The Committee has considered the following estimated investments for the data infrastructure for the social sciences in three long-run options:

	Estimated investments	Funding by *
Option A: Continuation of the present situation	2 million euros a year	KNAW: 1 million euros a year NWO: 1 million euros a year
Option B: Basic data services centre, no topical expertise centres	3 million euros a year	KNAW: 1.5 million euros a year NWO: 1.5 million euros a year
Option C: Elaborate data services centre and five topical expertise centres	6 million euros a year	KNAW: 1.75 million euros a year NWO: 1.75 million euros a year Ministry of ocenw: 1.25 million euros a year Ministry of EZ: 1.25 million euros a year

* Suggested funding of data services centre: 50 percent KNAW and 50 percent NWO. Suggested funding of topical expertise centres: 100 percent NWO (of which 50 percent Ministry of ocenw and 50 percent Ministry of EZ).

The Committee recommends following the long-run option: i.e. establishing a data services centre and five topical expertise centres.

The time-horizon for the complete development and full integration would probably be a decade. At the end of that decade, at least five topical expertise centres would need to have been established (once fully operational, total estimated annual operating costs are 2.5 million euros), and the data services centre would need to be running (once fully operational, estimated annual operating costs are 3.5 million euros).

However, it would be important for the data services centre to be running properly in the short run – say within three years – since the centre can be considered to be at the heart of the data infrastructure for the social sciences. It is therefore necessary for the scientific steering group to give priority to setting up the data services centre. The minimum estimated annual operating costs of the data services centre are 3 million euros a year, while each topical expertise centre would increase these costs by 0.1 million euros a year. Furthermore, in the short run the scientific steering group should encourage the setting up of a limited number of topical expertise centres. If for example two topical expertise centres were set up after two years, the estimated annual operating costs of the data services centre would be 3 million euros a year in the first two years and 3.2 million euros a year (being 3 million euros a year plus twice 0.1 million euros a year) in the subsequent years. Of course, in the case of this example, the estimated annual operating costs of the (two) expertise centres are 1 million euros a year (being twice 0.5 million euros a year).

Therefore, in this first stage of the development, the scientific steering group must give priority to establishing a running data services centre and a limited number of topical expertise centres. The possibilities for setting up a topical expertise centre purposing on qualitative data should also be explored, which, if successful, would be incorporated in the second stage of the development. After the third stage, at least five topical expertise centres and the data services centre would be fully operational.

Recommendations

The Committee recommends the following actions be taken by the KNAW, NWO, the CBS, the Ministry of OCenW, and the Ministry of EZ.

The KNAW should:

- Transfer the present tasks and budget for the Steinmetz Archive and the NHDA to the data services centre.
- Arrange an extra budget in order to make a long-term financial contribution to the data services centre.

NWO should:

- Transfer the present tasks and budget for the WSA and for CBS data to the data services centre.
- Arrange an extra budget in order to make a long-term financial contribution to the data services centre.

The KNAW and NWO should:

- Establish a scientific steering group.

The CBS should:

- Transfer its present task of archiving the older CBS data-sets to the data services centre in accordance with (a) the scientific steering group; (b) the data services centre; (c) the topical expertise centres.

The Ministry of OCenW should:

- Arrange an extra budget in order to make a long-term financial contribution to the topical expertise centres.
- Coordinate the instalment of the data services centre.
- Coordinate the instalment of the topical expertise centres.

The Ministry of EZ should:

- Arrange an extra budget in order to make a long-term financial contribution to the topical expertise centres with respect to small and medium-sized enterprises.

List of references

Books, articles

- Arts, J.J.L. 2003. *Toegang tot publiek gefinancierde data*. The Hague: Sdu Uitgevers.
- Boddy, M. 2001. *Data policy and data archiving; Report on consultation for the ESRC research resources board*.
- Borghans, L. 2003. Nederlandse data zijn te duur. In *Economisch Statistische Berichten*. 88:132-133.
- Centraal Bureau voor de Statistiek. 2001. *Statistieken die tellen; Strategie voor de middellange termijn, 2002-2005*. Voorburg: Centraal Bureau voor de Statistiek.
- Centraal Bureau voor de Statistiek. 2002. *Jaarverslag 2002*. Voorburg: Centraal Bureau voor de Statistiek.
- Centraal Bureau voor de Statistiek. 2002. ssb Special. In *Sociaal-economische maandstatistiek 2002/12*.
- Economisch Instituut Tilburg. 1995. *Op weg naar een data-archief voor economen*. Tilburg: Stichting Economisch Instituut Tilburg.
- Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS). 2002. *Service; Für Wissenschaft und Gesellschaft*. Cologne: GESIS.
- Kaiser, J. 2003. NIH Sets Data Sharing Rules. In *Science*. 299:1643.
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. 2001. *Wege zu einer besseren informationellen Infrastruktur*. Baden-Baden: Nomos Verlagsgesellschaft.
- KWALON/ Platform voor Kwalitatief Onderzoek in Nederland (dr. T. Hak, dr. F. Wester). 2002. *Kwalidata: Een archief van kwalitatief onderzoek*.
- Minister van Justitie en Minister van Grote Steden- en Integratiebeleid. Wet van 6 juli 2000, houdende regels inzake de bescherming van persoonsgegevens (Wet Bescherming Persoonsgegevens). In *Staatsblad van het Koninkrijk der Nederlanden*. Jaargang 2000 302. The Hague: Sdu Uitgevers.
- Mochmann, E. 2002. *International social science data service; Scope and accessibility*. Cologne: Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln.
- Norwegian Social Science Data Services (NSD). 2001. *A node in the information society*. Bergen (Norway): NSD.
- OECD Follow-up Group on Issues of Access to Publicly Funded Research Data. 2003. *Promoting access to public research data for scientific, economic and social development*. <<http://dataaccess.ucsd.edu/FinalReport.htm>>.
- RAND Europe. 2001. *Investeringsbehoefte uitrustings wetenschappelijk onderzoek*.
- Reidpath, D.D., and P.A. Allotey. 2001. Data sharing in medical research: an empirical investigation. In *Bioethics*. 15:2.
- Sociaal-Wetenschappelijke Raad. 2003. *Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek*. Amsterdam: KNAW.
- Social Sciences and Humanities Research Council of Canada, *National Archives of Canada*. 2001a. *National research data archive consultation; Phase one: needs assessment report*. <http://www.sshrc.ca/web/whatsnew/initiatives/data_archives_e.asp>.

- Social Sciences and Humanities Research Council of Canada, National Archives of Canada. 2001b. *National data archive consultation; Building infrastructure for access to and preservation of research data in Canada*. <http://www.sshrc.ca/web/whatsnew/initiatives/data_archives_e.asp>.
- Soete, L., and B. ter Weel. 2003. ICT and access to research data: an economic review. *Background paper for NWO expert meeting "Costs of open access to research data" on 8 May 2003, The Hague, The Netherlands*.
- The Royal Statistical Society & the UK Data Archive. 2002. *Preserving & sharing statistical material*. Essex, United Kingdom: UK Data Archive, University of Essex.
- Tweede Kamer der Staten-Generaal. 2002. *Wet op het Centraal Bureau voor de Statistiek*. Vergaderjaar 2001-2002, 28 277, nrs. 1-2. The Hague: Sdu Uitgevers.
- Wouters, P. 2002. *Policies on digital research data: an international survey*. Amsterdam: NIWI.
- Wouters, P., and P. Schröder (eds). 2003. *Promise and practice in data sharing*. Amsterdam: NIWI.
- Wetenschappelijk Statistisch Agentschap (WSA). 2001. *WSA geëvalueerd; Evaluatierapport 1998-2001*.
- Wetenschappelijk Statistisch Agentschap (WSA). 2002. *Lage drempels voor data van hoge kwaliteit; Meerjarenplan WSA 2003-2006*.
- Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln. 2002. *Eine Kurzinformation zum Serviceangebot des Zentralarchivs*. Cologne: Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln.

Articles without author

- 20 June 2003. We belemmeren niet, maar stimuleren; Interview met Joris Nobel (CBS). In *Onderzoek Nederland*.

Web sites (consultation of home pages most recently on 30 July 2003)

- CBS. <<http://www.cbs.nl/en/>>.
- CESSDA. <<http://www.nsd.uib.no/CESSDA/>>.
- CSES. <<http://www.umich.edu/~cses/>>.
- ESS. <<http://www.europeansocialsurvey.org/>>.
- Eurobarometer. <http://europa.eu.int/comm/public_opinion/description_en.htm>.
- Eurostat. <<http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=Eurostat>>.
- EVS. <<http://www.europeanvalues.nl/>>.
- ISSP. <<http://www.issp.org/>>.
- NHDA. <http://www.niwi.knaw.nl/us/dd_nhda/dd_nhda.htm/>.
- NKO. <<http://www.bsk.utwente.nl/skon/>>.
- Social Science Data Archives. <<http://www2.fmg.uva.nl/sociosite/databases.html>>.
- Steinmetz Archive. <<http://www2.niwi.knaw.nl/en/maatschappijwetenschappen/steinmetzarchieff/toon>>.
- SWR. <<http://www.knaw.nl/swr/>>.
- WSA. <http://wsa.magw.nl/index_uk.htm>.

Annexes

Descriptions of concepts

Empirical social sciences research is the activity using scientifically accepted methods for social and economic data analysis that leads to opinions regarding populations on a non-identifiable, aggregated level, answering one or more intrinsically linked scientific questions.

Empirical researchers in the social sciences in the Netherlands work in a university setting or in other non-commercial research institutes.

Data-use can be described as all activities by empirical social scientists with respect to social and economic data. These activities include (see: Minister van Justitie en Minister van Grote Steden- en Integratiebeleid, 2000):

- Collecting, recording, ordering data.
- Saving, updating, amending data.
- Claiming, consulting, using data.
- Providing data.
- Distributing data or making data available.
- Bringing together or collating data.
- Protecting, erasing, or destroying data.

Structured questionnaire

(in Dutch)

De onderstaande vragen hebben als leidraad voor de gesprekken gediend.

Vraag 1a

Wat is het belang van empirisch onderzoek voor de ontwikkeling van uw discipline/ vakgebied in de loop van de tijd geweest?

- Kunt u voor het (huidige) empirische onderzoek globaal schetsen:
 - in hoeverre primaire data resp. secundaire data worden gebruikt?
 - uit welke bron de onderzoekskosten van het datagebruik worden gefinancierd; met andere woorden: welke organisatie(s) betaalt(/betalen) het verzamelen en verwerken van primaire data, het aanschaffen van secundaire data, etc.? Te denken valt aan bijvoorbeeld: universiteit, NWO, EU.

Vraag 1b

In hoeverre werken de Nederlandse onderzoekers in uw discipline/ vakgebied samen in het datagebruik? Zijn er bijvoorbeeld onderzoekers van Nederlandse universiteiten die elkaar hebben gevonden in bepaalde wetenschappelijke vraagstellingen en op grond daarvan gezamenlijk data (laten) verzamelen? Als er sprake is van een dergelijke samenwerking, op welke wijze is deze dan tot stand gekomen?

Vraag 1c

In hoeverre speelt internationalisering een rol in het empirische onderzoek in uw discipline/ vakgebied? In hoeverre werken Nederlandse en buitenlandse onderzoekers in uw discipline/ vakgebied samen in het datagebruik? Zijn er bijvoorbeeld Nederlandse en buitenlandse onderzoekers die elkaar hebben gevonden in bepaalde wetenschappelijke vraagstellingen en op grond daarvan gezamenlijk internationale data (laten) verzamelen? Als er sprake is van een dergelijke samenwerking, op welke wijze is deze dan tot stand gekomen?

Vraag 2

In hoeverre kent u (dat wil zeggen: heeft u te maken gehad met) aanbieders van secundaire data, en wat is het algemene beeld dat u heeft van deze aanbieders van secundaire data?

Te denken valt aan aanbieders van secundaire data als: Centraal Bureau voor de Statistiek, Sociaal en Cultureel Planbureau, Wetenschappelijk Statistisch Agentschap, Steinmetz-archief, expertisecentra op uw vakgebied of (zeer) verwante expertisecentra binnen het universitaire bestel, Eurostat, andere aanbieders van internationale data, etc.

Vraag 3

Welke (ontwikkeling van de) data-infrastructuur in Nederland zou geschikter zijn voor (de toekomstige ontwikkeling van) uw discipline/ vakgebied?

Als u van mening bent dat er een meer geschikte data-infrastructuur moet komen, zou u dan enkele namen van (jonge) onderzoekers binnen uw discipline/ vakgebied kunnen geven die hierbij betrokken moeten worden?

Annexes

The Social Sciences and Humanities Research Council of Canada and the National Archives of Canada have examined all existing national research data archives focusing on the social science or humanities (2001b). Its Chapter 6 reads as follows:

‘The Working Group examined all existing national research data archives focusing on the social science or humanities ... This investigation included face-to-face interviews with data agency directors, comparative analysis of policies and regulations, examination of services, mandates, budgets and governing structures. Chief among the lessons learned are the following:

- Many countries have long recognized the need for a research data archive to assist and support the work of the research community. Several of the data archives examined have been in existence for 30 years or more;
- Although many services of a research data archive, particularly those related to access and training, are best distributed among a number of locations, for reasons of economy, practicality and effectiveness, preservation, network management and standards development functions are best performed within one facility;
- No two research data archives are the same. Each was established within a specific national or disciplinary context that reflected the particular needs of the research community it serves. They range in size from small, disciplinary specific, limited service organizations to large, multidisciplinary, full service, internationally networked, R&D focused, national institutions;
- Successful research data archives are directly attached to a country’s research infrastructure, rather than to its archival community. They are characterised by a service orientation that emphasises access to, and preservation of, the most useful data for research, rather than capturing records of the past;
- Research data archiving is a complex and highly technical business. Successful data archives employ dedicated, professional data experts and place considerable emphasis on training the next generation of research data managers. Developing highly qualified personnel serves the needs of both the research community and many other areas of the public and private sectors that have to deal with large volumes of data;
- There is a direct correlation between the funding stability of a research data archive and its success in supporting the research community. By its very nature, archiving is a long-term enterprise. The most useful data archives are those that are assured of their continuing existence;
- Although research data archiving requires long-term funding commitments, the institutional costs are always only a very small fraction of the costs of data collection;
- Building trust with both users and producers of research data is vital. If users cannot rely on the timely and efficient delivery of high quality data, and if depositors are not convinced that their intellectual rights and the protection of their participants will be upheld, no one will trust or use the services provided;
- The most successful data archives have both institutional independence and flexibility. They work in close co-operation with numerous government departments and universities but are not dependent upon any particular one for

financial stability or decision-making. Independence is necessary to ensure that the data access needs of the research community remain the first priority, rather than the record keeping needs of government departments or traditional cultural archives. Flexibility is important for the adoption of new technologies and the ability to respond to the changing needs of researchers.

The Working Group's detailed survey of 36 institutions produced three generalised approaches to preserving and providing access to research data. Each represents the organizational characteristics of today's national data archiving services:

- **A small scale, specialised topical data archive**, usually hosted by a university department, with limited data handling capability, employing off-the-shelf technology. Clientele are often restricted to one, or a small group, of research disciplines, and annual operating budgets range from between \$200K to \$400K.
- **A medium-sized, agency-based data archive**, whose parent organization is usually a national research institute or government department. Often located on a university campus to better serve its core research clientele, these archives base their mandate, and subsequent collection activities, on that of their parent agency. Services are moderately extensive, and staff members sometimes take leadership roles in relevant national and international organizations. Annual budgets range from \$500K to \$1.5M.
- **A comprehensive research data archive**, servicing a wide variety of communities, including academic researchers, NGO and government policy analysts, public archival agencies, and individual citizens. Often established through legislation, such data archives are recognized as national institutions responsible for the general principles and specific duties outlined in their founding Acts. Through one or more physical locations, and extensive use of the Internet, a comprehensive range of services are provided, often including specialised training, educational outreach, technical support and R&D. Data management capabilities are extensive and often developed in-house. Such agencies have established working relationships with other national institutions and government departments, and staff members are often leaders of international associations and actively engage in international data exchanges. Annual budgets range from \$3M to \$6M.'

List of acronyms

CBS	Centraal Bureau voor de Statistiek (Statistics Netherlands)
CESSDA	Council of European Social Science Data Archives
CSES	Comparative Study of Electoral Systems
ESB	Economisch Statistisch Bestand (Economic Statistical Data)
ESF	European Science Foundation
ESS	European Social Survey
EVS	European Values Study
ISSP	International Social Survey Programme
KNAW	Koninklijke Nederlandse Akademie van Wetenschappen (Royal Netherlands Academy of Arts and Sciences)
Ministry of EZ	Ministerie van Economische Zaken (Ministry of Economic Affairs)
Ministry of ocnw	Ministerie van Onderwijs, Cultuur en Wetenschappen (Ministry of Education, Culture and Science)
NHDA	Nederlands Historisch Data Archief (Netherlands Historical Data Archive)
NIWI	Nederlands Instituut voor Wetenschappelijke Informatiediensten (Netherlands Institute for Scientific Information Services)
NKO	Nationaal Kiezersonderzoek (Dutch National Election Study)
NKPS	Netherlands Kinship Panel Study
NWO	Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research)
SCP	Sociaal en Cultureel Planbureau (Social and Cultural Planning Bureau)
SSB	Sociaal Statistisch Bestand (Social Statistical Data)
SWR	Sociaal-Wetenschappelijke Raad (Social Sciences Council)
VSNU	Vereniging van Samenwerkende Nederlandse Universiteiten (Association of Universities in the Netherlands)
WSA	Wetenschappelijk Statistisch Agentschap (Scientific Statistical Agency)

