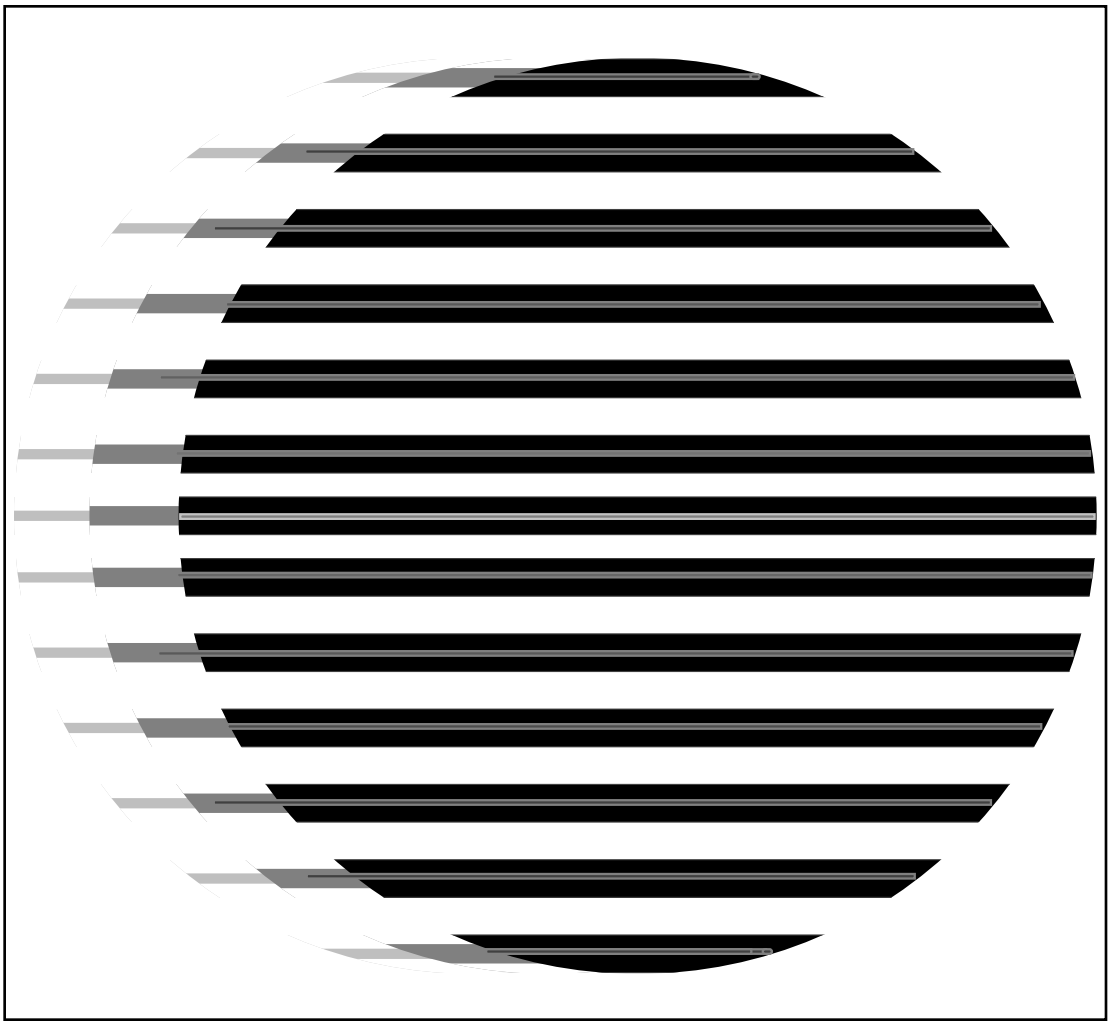
IASSIST

QUARTERLY

VOLUME 27 Spring 2003 NUMBER 1



Printed in the USA

IASSIST QUARTERLY

The **IASSIST QUARTERLY** represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The **QUARTERLY** reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of **IASSIST**.

Information for Authors:

The QUARTERLY is published four times per year. Authors are encouraged to submit papers as word processing files. Hard copy submissions may be required in some instances. Manuscripts should be sent to Editor: Karsten Boye Rasmussen .

The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press *Manual of Style* or Kate L. Turabian's *Manual for Writers*. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". *Journal of the American Society for Information Science 30(2): 77-82, March 1979*. Announcements of conferences, training sessions, or the like, are welcomed and should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event.

Editors

Karsten Boye Rasmussen,
Department of Organization
and Management,
University of Southern
Denmark, SDU,
Campusvej 55, DK-5230
Odense M, Denmark
Phone: +45 6550 2115
Email:kbr@sam.sdu.dk

Production

William Block,
Minnesota Population Center,
University of Minnesota,
537 Heller Hall
271 19th Avenue South.
Minneapolis, MN 55455.
Phone: 612-624-7091
Email:block@socsci.umn.edu

Walter Piovesan
Maps/Data/GIS Library,
Simon Fraser University,
Burnaby, B.C
Canada V5A 1S6.
Phone: (604) 291-5869.
Email:walter@sfu.ca

Title: Newsletter - International Association for Social Science Information Service and Technology

ISSN - United States: 0739-1137 **© 1997 by IASSIST. All rights reserved.**

CONTENTS

Volume 27 Number 1 Spring 2003

FEATURES

5 **Integrating Public Domain Data to Construct Community Profiles**
K.Reed, B. Blunsdon, N. McNeil & S. McEachern

09 **Data Processing in FSD: Challenges in a New Archive**
H. Keckman-Koivuniemi & M. Kleemola

12 **A Reflection on the Past Decade**
Chuck Humphrey

Editor's Notes

Welcome to the IASSIST Quarterly - IQ vol. 27 issue 1. Three articles are presented in this issue.

The article from Ken Reed, Betsy Blunsdon, Nicola McNeil (Deakin University, Victoria, Australia) and Steven McEachern (University of Ballarat, Victoria, Australia) was presented at the IASSIST conference (May 2003, Ottawa) in the session with the title "Strengthening Numbers: measurement, aggregation and policy". The Reed, Blunsdon, McNeil & McEachern article has the title "Integrating public domain data to construct community profiles". The paper describes the construction of a database of community characteristics at the postcode level. A community is described by variables in the areas of economic, social, political and cultural characteristics. These publicly available data from a wide variety of sources are in the project collated and integrated. The data are supporting multi-level research enabling the contextualization of individual behavior by data on the context in which people live their daily lives, such as the neighborhood, school, community or region. There is a great deal of data available in the public domain, but these are collected for a wide range of purposes, and with great variation in the units of analysis. The project on local areas in Victoria, Australia relates to other neighborhood projects in the USA (Los Angeles and Chicago). The database has information on the years 1996 and 2001, which are years in which there was both a national census and a national election. From a variety of sources data is included on businesses, crime and suicide, licensed premises, information about religious institutions, schools, recreational facilities, cultural organizations and services.

The second paper is also from the May 2003 IASSIST conference in Ottawa. The paper gives an insight into a new archive by the authors Hannele Keckman-Koivuniemi and Mari Kleemola, both research officers at Finnish Social Science Data Archive. The title of the paper is "Data Processing in FSD: Challenges in a New Archive". The Finnish data archive started in 1999 situated at the University of Tampere and funded by the Ministry of Education. The archive has about thirteen employees and has archived about 400 studies. The archive is a social science data archive and receives data files in SPSS, Excel, SAS or ASCII formats. Intensive data processing is carried out at the FSD in order to make the data corresponding the original questionnaire. The data is documented in SPSS files, the integrated documentation is obviously considered important, but a more extensive documentation is also made through the use of the DDI.

In the last article from one of the IASSIST key figures Chuck Humphrey - past president of IASSIST and an active person at almost every IASSIST conference – you are invited to take a look into the major accomplishments of IASSIST in the past decade. This article is for both newcomers and old members of IASSIST. While looking back Chuck Humphrey is at the same time into the process of setting the goals for the future of IASSIST. In the past IASSIST has been active in helping to set up data archives and in obtaining members outside the traditional geographical areas of IASSIST (North America, Europe (Western) and Australia). The Outreach program that started in 1996 has brought in many new members to IASSIST conferences and spread the information about IASSIST. In the communication room the website on the Internet was the biggest move in the past. Apart from delivering services directly to the members, IASSIST has also been active as participants and initiators of metadata standards like the Data Documentation Initiative (DDI). Many more points are made in the article. Read the article and contribute by articles to the IQ or by stating your opinion on the IASSIST mailing list.

Remember to visit the IASSIST website on www.IASSISTdata.org.

Papers for the IASSIST Quarterly are most welcome. Please contact the editor (kbr@sam.sdu.dk).

Karsten Boye Rasmussen, November 2003.

Integrating Public Domain Data to Construct Community Profiles

*by Ken Reed, Betsy Blunsdon, Nicola McNeil & Steven McEachern **

Introduction

In recent times there has been an increase in the availability of and requirement for public domain data that can be used to construct community level profiles in order to study contextual level phenomena. In this paper, we will review the developments that have given rise to this requirement, and explain how we have constructed a database of community characteristics at a postcode level. We shall then review applications of this type of data, and the opportunities for both aggregation and extension.

The demand for community level data has grown following developments in academic research and also in policy related work. Academics in many fields, including the government, are now examining aspects of the social world at multiple levels and therefore need accurate data at these levels (Sampson 1988; Griffen 1997; Raudenbush and Sampson, 1999; Sampson and Raudenbush 1999).

There has also been an increased demand for community level data within local government areas in order to better understand and address local issues, and to be used for evidence based planning. The approach we will outline provides an opportunity to do this, particularly when integrating a variety of data sources and utilising multiple methods of data collection.

The need to measure 'context'

Researchers in many fields are now studying social phenomena at multiple levels of analysis. It has been previously recognised that to understand the social world in detail requires a multi-level approach, yet until recently this has not been possible. New theoretical and empirical developments in the area of multi-level modelling make it possible to measure 'contexts' (such as 'environment', 'organisation', 'community' or 'neighbourhood') in addition to 'individuals'. Hierarchical or 'multi-level' modelling techniques estimate the effects of variables at different hierarchical levels. These techniques were developed primarily in educational research, to estimate separate parameters for the effects of the school, teacher and pupil on a student's academic achievement (Goldstein 1995; Singer 1997). The technique of multi-level modelling involves estimating higher-level effects from the data, then using these higher-level estimates as constants in the analysis of units at lower levels of analysis. This enables effects at lower levels to be determined while taking into account the tendency of (lower-level) individuals in similar (higher-level) contexts to exhibit similar responses (Goldstein 1995; Bryk and Raudenbush 1992).

These recent developments in multi-level research have seen major advances in many fields, including Sociology, Psychology, Social Work, Organisational Theory and Criminology. For example, understanding aspects of social capital, such as the tendency to volunteer, involves understanding both aspects of individuals and the characteristics of their communities. Communities can differ in terms of density of social relationships, infrastructure characteristics (such as number of business premises versus residential, parkland, number of schools), homogeneity, heritage, community safety and structural aspects (such as age and gender distribution and household composition). Similarly, individuals may differ in terms of demographics, personality, attitudes, values, behaviours or motivation.

It is not surprising given these developments that there has been growth in the number of multi-level and ecological studies in a broad range of fields. Two examples of such research designs are the Project on Human Development in Chicago Neighborhoods and the Los Angeles Family and Neighborhood Survey (LA FANS). Both projects attempt to disentangle the intricate effects of individual, family and community factors on the human development of children and juveniles, by undertaking a longitudinal investigation of individuals, communities and individuals within their community.

The aim of the Los Angeles Family and Neighborhood Survey[1] is twofold – firstly to study the impact of neighbourhood, family and peers on children's development, and secondly to study how and why families choose to live in certain neighbourhoods. The focus of this study is to provide in-depth understanding of these relationships to inform policy development, particularly welfare policy.

The Chicago Neighborhoods Project[2] aims to reveal the causes and implications of deviancy amongst residents of several Chicago neighbourhoods. Over an 8-year period, researchers will gather official data about neighbourhood characteristics, including the economic, social, political, business and cultural structures within the neighbourhood. Information will also be collected through a series of co-ordinated longitudinal studies that trace 7,000 randomly selected children through to young adulthood, looking at personal characteristics, changing circumstances and the exhibition of delinquent behaviours.

In Australia, there has been no large scale, systematic study that attempts to describe the characteristics of geographic regions in this manner. The main difficulty for Australian researchers attempting this type of undertaking is that there was no consensus amongst government agencies about how the geographic element of this data would be collected. For example, the Business Register identifies the geographic location of businesses by the postcode they are located in, while Federal and State Government election results are recorded by electoral district and crime statistics are recorded by local government area. In an attempt to address this inconsistency, the Australian Bureau of Statistics has recently revised a classification scheme that will enable the production of statistics that can be integrated. Statistical units, such as households and businesses, are assigned to a geographical area, with the data collected from these statistical units then compiled into pre-defined geographic aggregations that, subject to confidentiality restrictions, are then available for publication.

With multi-level research comes the need to measure characteristics of communities and neighbourhoods with much more precision than has been required in the past. This provides the opportunity to integrate a higher number of community and neighbourhood indicators into comprehensive data sets of characteristics.

Sources of data at different levels
The data we have sourced and integrated has been collected at the postcode level. That is, the data can be treated as roughly representative of geographical communities. The postcode has both advantages and disadvantages as a unit of analysis. On the one hand it is an administrative boundary that may not exactly reflect the social ecology of the residence; on the other hand, it offers an advantage in that there are many different types of data available at the postcode level.

The type of data available includes: demographic data (in Australia this is available for 1996 and 2001, both years in which there was a census), data on the incidence and perceptions of crime, the number and locations of licensed premises, the number and location of churches and religious institutions (as well as other cultural institutions), the number of schools, the number and type of recreational facilities, social services (such as law courts, legal centres, community centres) and the number and location of businesses, government offices and agencies. A detailed list of the type and source of data is available in Table 1 (http: //ReedTable1.html)

Aggregation and extensions
Community data collected at the postcode level can be aggregated into local council areas or regions and into state level areas. This type of dataset also provides the opportunity for extension through integration with other survey data and qualitative techniques. By constructing a sampling frame for household or individual surveys from electronic phone directories, there is also the opportunity to identify neighbourhoods within postcode levels (by street location, for example). Telephone survey data can also be supplemented at the level of the respondent through more detailed personal interviews, following the analysis of the survey data.

Contextual and situational data can be further extended through systematic social observation (SSO) and ethnographic fieldwork. Mapping the variation in community characteristics provides a rich information source for case selection in ethnographic fieldwork. Information about the characteristics of communities and neighbourhoods can also be integrated with the detailed descriptions of ethnographies and social observations in order to further understand complex social situations and interactions. An example of this type of extension is found in Sampson and Raudenbush (1999), a study of social disorder in Chicago. They combine social observation, census data, police records and a survey of residents to test a theory of structural constraints with local collective efficacy in order to understand the sources and consequences of social and physical disorder in urban areas.

Sourcing, collating and integrating community level data over time also provides enormous potential for identifying macro-level structural change. Our database enables us to aggregate the data of local government areas for 1996 and 2001. This enables the analyses of variation between these areas at each point in time, and reveals changes within council areas over the two periods. This has important policy applications for community capacity building, community regeneration and social capital.

Applications
It is widely accepted that an understanding of the social world requires theories at multiple levels to achieve complete explanation. While a 'complete' explanation is a *direction* for progress rather than an attainable goal, this type of data set will provide much information about macro-level contexts such as environment, community and neighbourhood. It also makes possible analyses of the landscapes in which many social issues arise. The

following examples illustrate how this data might be applied for multi-level research.

Our first example is a study of the propensity to volunteer in community-based fire fighting. The study will take into account the effects of the individual, life-situation and community on tendency to volunteer. People may choose to become volunteer fire fighters for a variety of personal reasons, including feeling part of a community and a sense of obligation, or for instrumental reasons, such as maintaining property security and developing skills that enhance employment prospects. Analyses at the individual level may capture such differences in perception, but not the effects associated with influences derived from living within different types of community. Community level data such as that described will give us insight into the relationship between types of community characteristics and the tendency to volunteer. This contributes to theoretical development and policy and planning issues involving volunteer activity and community development.

A second example is the study of the effects of social relationships and networks on personal outcomes, in this case health. Like the problems described thus far, a central issue in this investigation is distinguishing between the effects at the individual level from effects at the community level. For example, if we compared two communities, one characterised by dense social networks and the other by a high amount of individualism, we cannot determine whether it is living in that type of community or having many social relationships that produces individual outcomes, including better health. To this end, community level data provides the opportunity to study the interaction effects of individual and community characteristics on personal health.

In these cases we treated postcodes as roughly representing geographical communities. We employed a sample design capable of allowing statistical inferences that estimate the effects at both the postcode level and again at the individual level. As referred to previously this offers both advantages and disadvantages. The community level data that we have described is available at the postcode level while individual data can be aggregated from sources, such as the Census, to provide measures of affluence, demographic structure and residential stability.

Conclusion

Recent developments in multi-level theory and policy and planning requirements make it timely to collect and integrate data from a number of sources to better understand the context in which social action occurs. The data set that we describe is constructed by collecting data from a number of different sources, almost all of which is available at postcode level. This data can then be aggregated and extended to a higher level of analysis, such as the region or the state, or disaggregated into neighbourhoods or streets, if combined with further data collection methods. This extended data set, or community profile, can then be used to identify and pinpoint the effects of social change across a range of social levels, from a state wide macro-level down to the micro-level of a particular street if need be.

These applications stem directly from the recent abundance of publicly available data and the development of effective ways in which the community profile may be used. Academic researchers, policy workers and local or state governments may immediately benefit from the adoption of this method of social inquiry. At present, the only limiting factor for this method is the potential incompatibility of the various data sets available. However, as seen by the recent actions of the Australian Bureau of Statistics, this is a problem that has been recognised and is being addressed.

In closing, it can only be expected that the kind of multi-level data collection and analysis that we have presented today, and that can be seen in the Los Angeles Family and Neighborhood Survey and the project on Human Development in Chicago Neighborhoods projects, will become an established and invaluable part of social research and planning.

References

Bryk, A. and Raudenbush, S. (1992) *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, California: Sage.

Goldstein, H. (1995) *Multilevel statistical models*. London: E Arnold, Halsted Press.

Griffen, M (1997) "Interaction between individuals and situations: Using HLM Procedures to Estimate Reciprocal Relationships." *Journal of Management* 23 (6):759-773.

Raudenbush, S. and Sampson, R. J. (1999) "'Ecometrics': Toward a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods." *Sociological Methodology* 29:1-41.

Raudenbush, S. and Sampson, R. J. (1999) "Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods (1)" The American Journal of Sociology, 105: 603-636.

Sampson, R. J. (1988) "Local Friendship Ties and Community Attachment in Mass Society: A Multi-Level Systemic Model." *American Sociological Review* 53:766-769.

Singer, J. D. (1997) "Using SAS PROC MIXED to fit multilevel models, hierarchical models and individual growth models." *Journal of Educational and Behavioral Statistics* 23: 323-355.

Footnotes

[1] Refer to http://www.lasurvey.rand.org

[2] Refer to http://phdcn.harvard.edu/about/about.html

* Paper presented at the IASSIST Conference, May 2003, in Ottawa, Canada. Ken Reed, Betsy Blunsdon, Nicola McNeil (Deakin University, Victoria, Australia), Steven McEachern (University of Ballarat, Victoria, Australia). (Ken Reed: kreed@deakin.edu.au).

Data Processing in FSD: Challenges in a New Archive

*by Hannele Keckman-Koivuniemi & Mari Kleemola**

The Finnish Social Science Data Archive FSD began operation in 1999 as a separate unit of the University of Tampere. It is funded by the Ministry of Education. At present the archive has thirteen full-time employees.

FSD has archived about four hundred studies, of which one third are international in scope. FSD obtains about one hundred new studies annually from a variety of sources: ICPSR, public and private research organisations and individual researchers. Studies from other archives (for example ICPSR) are documented in Finnish but the data are not processed.

The data that FSD archives must be from an area or discipline in the field of social sciences. It should also fulfil certain technical and legal requirements: aspects of copyright and ownership should be clear, there should be no legislative impediments to archiving (e.g. data protection, privacy protection), the original purpose of data collection should not prevent archiving and, last but not least, the information content and technical properties should make the data suitable for archiving.

Data are processed intensively

The archiving process begins when the depositor delivers machine-readable data to FSD. The data are most often received in SPSS or Excel format and occasionally in SAS or ASCII format. Depositors are asked to give detailed information about the collection procedure, resulting publications and the research project in general. FSD prefers to obtain supplementary documentation both in electronic and paper formats. The completed and signed material description and material deposit agreement forms are necessary as well.

All national studies acquired by FSD are processed intensively – a very time-consuming undertaking. There are two reasons for doing this. First, the number of archived studies so far is manageable. Second, researchers do not usually offer their data. Rather, FSD staff identify studies from, for example, scientific journals and then request the data from the researchers. This way all archived studies are by presumption "important" and deserve intensive processing.

A goal of intensive processing is to make the content of the archived data correspond as closely as possible to that of the original questionnaire. Documents received from the depositor, such as the questionnaire, original observation matrix, variable lists, printed books and articles play a key role in this work. All alterations are carefully documented in the syntax.

FSD uses SPSS in data processing and preserves data in portable format. We have chosen SPSS because it provides - at least at present - the best option for preserving data and identification information (that is labels) in the same file. Since SPSS is widely used, converting material to future formats should be manageable. Most of our customers are familiar with SPSS and about 90 % of researchers wish to have data in this format.

Checking the Data

The original data file is preserved without modification as long as necessary for the archiving process. So far we have not deleted any original datasets. A copy of the original file is used to produce a version suitable for secondary use. We focus on the content, not the "look and feel" of the studies.

During the checking process mistakes are corrected and verifications, amendments and additions are made. We confirm that the number of variables and cases match the documentation supplied. We check frequencies to verify variables, valid and invalid values (e.g. missing data, not applicable responses etc.). Variables are renamed corresponding to question numbers. Background variables without their own question numbers are renamed bv1, bv2 etc. FSD has standardised labels of some background variables (e.g. original question: How old are you? --> variable label: Respondent's age). Variable and value labels are constructed based on the questionnaire. Variable labels often become quite long as they may contain the whole question text. We try to keep variable names and labels consistent within studies of the same series.

Questionnaires often include questions that are directed only at respondents who meet certain requirements. The archive checks these filter conditions. If the data include answers from people who do not belong to the specified

target group, the responses are classified as missing data.

In order to make sure that the content of the archived data corresponds as closely as possible to that of the questionnaire some variables may have to be dropped or added. A variable is dropped if it is undefined or data security aspects so require. Constructed variables, such as combined variables and sum variables, are usually dropped. However, those constructed variables that are integral to the usability of the data, especially weight variables, are kept - providing that the documentation provided is explicit enough. New variables are added only if usability so requires.

Data Protection
Confidentiality aspects require that personal data are deleted. It is recommended that depositors remove these types of data (names, addresses, birthdays etc.) before delivering the material to the archive. Under certain circumstances, FSD stores materials which contain personal data. In these cases, the archive anonymizes the data according to its own guidelines and depositors' instructions.

Variables indicating place of residence and business are also problematic. On one hand there is always a risk that a single respondent might be identified, on the other hand the deletion of these types of variables prevents secondary users from conducting regional comparisons, especially if no other regional variables are used. Usually these types of variables are dropped. If necessary, they can be restored. Variables of larger regional units (provinces, districts) are kept.

Version Control
We aim to process the datasets only once. Still, some of them have to be reprocessed. Sometimes additional information about variables is given by depositors, errors are detected or processing procedures updated. For example, datasets processed in the early days of FSD already need repairing.

The first final version of a dataset is called version 1.0. If subsequent changes are more or less cosmetic (typing errors etc.) the new version will be called 1.1. In the case of significant changes (e.g. a variable added) the new version will be named 2.0.

We add these alterations to the end of the same SPSS syntax file used in processing the dataset in the first place. This is not a long-term solution but has worked so far. We track changes and versions as well as all files in our operational database.

Documentation
FSD uses DDI standard for creating data documentation. At present FSD produces study descriptions in Finnish and English and PDF codebooks in Finnish for all our national studies. All documentation is available on the Internet. Datasets are translated into English on request.

The data and documentation are fully compliant with the search engine NESSTAR. FSD also takes part in the MADIERA project (Multilingual Access to Data Infrastructures of the European Research Area) which started in December 2002.

Challenges for the Future
We need to - and will - review and update our data processing instructions and procedures in the near future. It is obvious that FSD will not have the resources to continue processing all datasets intensively, but will have to introduce several different levels for processing and documenting datasets. First we need to define the minimum level of processing required to preserve dataset quality for the long-term.

One problem common to all data archives is how to get principal investigators to provide enough details about the data and documentation. The intensity of our processing requires a large amount of information about each study. Researchers are often unwilling to take the time necessary to dig up and assemble the details and the lack of information slows the archiving process significantly. Using different processing levels would mean a faster archiving process and quicker publication of data.

Another challenge is version control. We cannot continue controlling versions by amending the syntax file originally used to moderate the dataset because the original data might - and probably will - be unreadable in the future. The SPSS syntax should merely be a tool, not documentation needing to be preserved. Also, as our purpose is to preserve the content, not the "look and feel" of the data, we are not planning to preserve the original datasets forever.

The data for a growing number of studies are collected by computer-aided interview systems like CATI and CAPI. At the moment we process these datasets manually. This needs to be automated as the number of archived CATI and CAPI studies grows. Computer aided interviews also mean that there are no traditional-style survey questionnaires - but our current data processing procedures assume that a printed questionnaire exists.

NESSTAR and MADIERA will enable online data downloading in the future. This may create an entirely new set of problems. Another challenge will be data protection. The measures we take today may not be sufficient in the future. The scientific community is increasingly aware of these challenges. Last but not least is the question of long-term preservation of data and other electronic documents.

In this article we have only touched the surface, the list of challenges surely does not end here. FSD has received a lot

of information of good or best practices in data processing from other data archives in Europe and North-America. Co-operation on the international level has been, and will continue to be, crucial for all data archives, especially new archives like FSD.

* Paper presented at the IASSIST Conference, May 2003, in Ottawa, Canada.

Address: FSD, University of Tampere, 33014 Tampere, Finland, Fax: +358 3 215 8520, Internet: www.fsd.uta.fi. Email: hannele.keckman-koivuniemi@uta.fi - Tel: +358 3 215 8530. Email: mari.kleemola@uta.fi - Tel:. +358 3 215 8528.

A Reflection on the Past Decade

*By Chuck Humphrey**

The following discussion offers a perspective about the major accomplishments of IASSIST over the past decade. Particular attention is paid to areas that are recognized as strengths of the organization and in which there have been notable successes. Many of the actions and outcomes in these areas flow from the goals embraced by IASSIST at the beginning of the 1990's. It is hoped that this discussion will assist in reviewing these goals and in setting directions for the future.

Most of the evidence in this summary is found in the business and program of the annual conferences of IASSIST. For many of our members, the IASSIST conference is a time to learn about new products, standards, services, and technology. These meetings keep IASSIST members at the forefront of our field. Concomitantly, the direction of the profession both is reflected in the issues of these conferences and is shaped by the discourse and outcomes of these meetings. Therefore, IASSIST conferences are an important source of evidence when considering the progress of this organization.

Six topics are discussed below that touch on the values expressed in the goals of IASSIST. While these topics are not strictly a re-expression of these goals, one can nevertheless find elements of the original goals in each of these six areas.

The Pulse of the Membership

The size of the membership has remained fairly stable over the past decade. While some have allowed their membership to lapse, overall there have been slightly more gains than losses. New members came from two primary sources. First, there has been an increase in the number of Europeans joining IASSIST, many of whom work in national data archives. Secondly, professionals new to data services, many working in academic libraries, account for the other membership gains.

A consorted effort was made in the early 1990's to expand membership among the staff of data archives. In 1993 under the direction of the UK Data Archive Director, many Archive staff attended IASSIST in Edinburgh. This experience demonstrated the value of IASSIST to data archive staff and equally important, these new members brought a fresh enthusiasm to the organization. In subsequent years, several national data archives have supported their staff to attend the IASSIST conference. One encouraging indicator of this new commitment is that European attendance at IASSIST conferences held in North America has increased in recent years.

In conjunction with this growth of European members, many participants joined IASSIST after attending the ICPSR Summer Program workshop on social science data services. For this group, IASSIST serves as an organization where they can continue their professional development and can network with other data professionals. Many of these new members are the only individuals on their campuses providing data services. IASSIST is the one organization to which they belong where they find colleagues doing the same kind of work that they do.

During the past decade, structural changes occurred in several universities that altered the institutional location of data services. Computing centres that had been actively supportive of social science data services, for example, Princeton University, Northwestern University, and the University of Alberta, began divesting themselves of client services. The paradigm of central computing centres shifted from a services-based to a utility-driven organization. Under this model, computing became another outlet in the offices and labs on campus. Along with the light switch and the power and phone outlets, institutions added a computing network outlet. While computing centres became uninhabitable for data services, academic libraries welcomed the IT expertise of data services staff. In many instances, libraries absorbed data services that had become orphaned by the movement toward utility-driven computing centres.

IASSIST members played a supportive role in the relocation of data services in several universities during these times of structural change. Our members work in a rich variety of organizational settings and consequently, there is a wide range of models for providing data services. One model that was imitated by a number of institutions was an amalgamation of government information, maps,

and data into a single administrative unit. This saw the convergence of support for statistical and spatial data.

Another significant environmental change having an impact on the institutional location of data services was the digital library initiative. This movement started to take root in academic libraries at the same time that several social science data services were moved into libraries. The digital library seemed to legitimize the incorporation of data services in the eyes of some academic library directors. Other directors recognized data as a valuable research resource that belongs in the library. The confusion between data services and the digital library unfortunately has never been resolved. For one thing, the digital library has been largely dominated by digitization projects and has failed to establish a strong connection with data services. Consequently, attention to the creation of digital collections has often overlooked data services as part of this movement.

As interest in social science data increased in areas of the globe outside of Australia, Europe and North America, IASSIST introduced two new initiatives to support these activities. A spin-off from these programs has been the addition of new members to IASSIST. First, an Outreach program, which is discussed in more detail below, was formally established during the 1996 conference in Minneapolis. This initiative seeks to support staff in countries where social science data services are just taking root. The second initiative was to create a new region for Africa within IASSIST. With a leadership base in the SADA and a growth in social science data activities on this continent, the prospects for new members from this region are encouraging

During this period, some turnover occurred in the membership due to career changes for some people and retirements for others. A testimony to the quality of the people in IASSIST is that contacts and friendships have been maintained with many of those who have changed careers. We also have the exemplar role model of a few retired members who have remained active in IASSIST!

As an organization, IASSIST needs to support professional data staff through training and upgrading, promoting social networking, contributing to standards development, disseminating information and knowledge, generating collaborative work, initiating and coordinating research, and offering a forum for professional issues. Many of these activities are conducted on a peer-to-peer basis, which makes IASSIST membership all the more important to professionals in data services.

Communications

Communication occurs on many fronts. There is member-to-member contact, official communiqués from the leadership of IASSIST, knowledge dissemination through the IASSIST Quarterly, and the public face of IASSIST. These communications require tools and organizational structure to occur, and the work and leadership of the Publications Committee has been essential is this regard.

IASSIST has provided an electronic means for members to communicate with one another through an email discussion list, which started in 1991 initially hosted by Princeton University before moving to Yale and then to Columbia, where it currently resides. This is a closed list to members and is used primarily by most members as an extended reference service where advice or information is sought from the wide realm of expertise among the membership.

In recent years, a membership directory has been produced annually from the records of the Treasurer and distributed to members in good standing. The Administrative Committee of IASSIST has decided not to give or sell mailing labels from this directory to advertisers, protecting the membership from unwanted junk mail. The primary purpose of the directory is for member-to-member contact.

In 1995, the IASSIST website was introduced to provide the organization with a public presence on the Internet. IASSIST members are good citizens of the Internet and the website was intended to reflect this spirit through the organization's support of open access to information. The website has undergone two phases of development. The initial phase focused on providing a description of the organization to the public. It tried to communicate what IASSIST is and why it is important. The second phase remodeled the site and incorporated material that members will also find useful. This design tried to balance the promotion of the organization with tools created by members that are important in the work that we do. As development continues with this phase, creative ways of using this medium are being sought to strengthen the organization's outreach to the wider data community and to deliver training and educational programs.

The IASSIST Quarterly (IQ) continues to be a print publication that is mailed to members in good standing. The production of the IQ, however, has long involved electronic stages in its preparation. In recent years, this electronic copy has been converted to PDF format and made available on the IASSIST website. The digital version of the IQ on the website is made available at approximately the same time as the print edition to the membership. Open access to the IQ contributes to the wider value of the Internet (as a public service of IASSIST), provides the organization with greater visibility, and shares the expertise of the membership more globally.

Other structural changes have happened in the organization that also improved communications within IASSIST. In 1998, a Treasury Group was organized creating three assistant treasurers located in three regions: the U.S.,

Europe and Canada. This new structure within IASSIST allowed for a flexible way of dealing with the bridge financing of conferences in these regions and for collecting membership fees. Beginning in 2000, the Membership committee was restructured around the Regional Secretaries. This change was introduced to provide a more direct contact with members in a region. Both of these initiatives were begun to establish a closer relationship between activities within IASSIST and its members.

The communication initiatives discussed above have been important for the life of this organization. Another assessment is found in the research conducted by Karsten Boye Rasmussen and Repke de Vries where they investigated IASSIST as a virtual community. Their findings suggest that IASSIST can make better use of network communication technology. Physical distances on this planet are an obstacle to our members meeting together, which is further discussed below under the topic of regionalism. Discovering a balanced use of network communications with in-person conferences is a priority of the organization.

Metadata, Standards, and Technology
Technology is a means to an end for those of us working in data services. There is no doubt, however, that technology at times seems to be in the driver's seat. This was particularly true in the early 1990's when many data services struggled with the migration of their data and documentation from mainframe environments to network technology.

IASSIST conferences throughout this past decade have held sessions dedicated to the issues of technological change in data services. In 1994 at the San Francisco IASSIST conference, prototypes of data extractors using a Web interface were demonstrated. These early examples confirmed the need for data documentation standards to facilitate Web extraction services.

Discussions around codebook standards evolved from roundtable discussions into interest groups in 1993 when a working group on Codebook Documentation of Social Science Data and a group to create documentation guidelines for data producers were established. In 1995, the Data Documentation Initiative (DDI) merged earlier working group interests and became the focal point for discussions about documentation standards within IASSIST. Many IASSIST members have contributed to the DDI standard and continue to lead in its development. The maturation of DDI is one of the major success stories of this organization.

The work of our European members through CESSDA and national data archive initiatives made further outstanding contributions in metadata and extraction tools. Among these projects was the integration of a major social science thesaurus with data catalogues, multi-catalogue searching over the Web using multiple languages, and Nesstar. These major contributions have been the focus of discussion at many IASSIST conferences over the past decade. Like DDI, they are successes to be celebrated.

The Ottawa IASSIST in 2003 held a session in which major data archives described how recent technology changes have been integrated into their data processing operations. This is another example of the convergence of changes in metadata, documentation, and technology. These new tools are changing the daily business of the major data archives.

The intellectual contributions of IASSIST members to these very significant developments over the past decade are achievements in which the organization can take great pride.

Confidentiality, Intellectual Property and Privacy
The issue of confidentiality and access is a long-standing topic in IASSIST. A delicate balance exists between the protection of the privacy of the individuals from whom data have been collected and claims for access to such data for legitimate research purposes. The millennium bug seemed to draw new public attention to the vulnerability of the massive amounts of data existing on individuals. One consequence has been a negative public reaction to the potential uses of digital information on individuals. Protective legislation has appeared recently in many jurisdictions that pose serious threats to legitimate research use of such data.

These concerns have been the topic of many plenary and concurrent sessions at IASSIST conferences. Institutional practices and procedures have been presented, such as the Norwegian model of a research data ombudsman. Technological fixes have been debated, including synthetic files and research data centres. At the Ottawa conference the issue of privacy was expanded to incorporate spatial as well as statistical data. These topics will not go away any time soon and IASSIST needs to continue addressing them to better inform the public, policy-makers, researchers, and the rest of the data community.

Data commodification surfaced over the past decade with the commercial success of the Internet. While the recent collapse of the dot-com industry has temporarily tempered interests in commercializing data, the issue itself will not go away. The next resurgence of e-business will drive a new wave to commodify data. Open scientific research is threatened by ownership claims to data. Consequently, this organization will need to steer the discussion about data ownership to one of data stewardship, and to be prepared to address ownership issues in terms of the barriers that it presents both to data access and to preserving data.

New Frontiers in Research Data

The frontiers of social science research data continue to expand. One important new area has emerged from advances in computational methods in qualitative research. In fact, developments in this area have led to the formation of data services dedicated to qualitative data. Of particular note is Qualidata in the UK Data Archive, which is responsible for the acquisition, preservation, and dissemination of qualitative research data. Recent IASSIST conferences have held workshops and sessions on both the archiving of and providing data services for qualitative data. For example, a workshop was presented at the Ottawa conference in 2003 entitled, "Everything you ever wanted to know about preparing qualitative data, but were afraid to ask." In a sense, we have witnessed qualitative data come of age over this past decade with an expectation that secondary uses of this type of data will increase the overall value of qualitative data.

With the popularization of PC-based Geographic Information Systems in the 1990's, spatial data have become another rapid growth area in social science data. The affiliated use of geo-referenced statistical data with corresponding spatial data has added new demands for aggregate statistics from data services. Starting slowly in the early 1990's, GIS is approaching tidal wave strength as applications for spatial analysis sweep across social, health, economic, business, and educational research. IASSIST conferences have incorporated spatial data as a focal topic in recent years. The theme of the 2000 conference in Evanston was "Data in the Digital Library: Charting the future for social, spatial and government data." In 2003, a plenary was dedicated to spatial data in additional to concurrent sessions that dealt with GIS data issues.

Historical data have always been part of social science data interests in IASSIST. More recently, the development of public use microdata files from historical censuses has taken on new prominence. Following the success of the IPUMS project at the University of Minnesota, a new international historical census microdata program has emerged, IPUMS-International. These projects are contributing a wealth of new data for researchers. An additional spin-off from these projects is a network of collaboration between researchers and professionals in data services.

All three of these new data frontiers share metadata, preservation and access issues with existing social science data. The developments in DDI have been applied in the IPUMS project. Furthermore, enhancements to the DDI standard have resulted from the challenges of documenting these historical data sources. Similarly, the lessons, methods, and tools used in documenting quantitative data are being adopted with qualitative data. Again, unique aspects of qualitative data are contributing to metadata practices established for quantitative data. The archiving of spatial data lags behind other social science data types. However, many GIS researchers are now looking upon the preservation of spatial data as an extension of other social science data. IASSIST can play an important role in this development.

These three data areas warrant special mention because of their growth and contribution to new research in the social sciences. This does not, however, detract from the continued developments in quantitative data in the social sciences. Projects in these areas continue to see major advancements and include issues such as synthetic data, probabilistic linkage of large administrative databases, and the management of large consumer expenditure databases. Activities in these areas will continue to challenge IASSIST members to find appropriates ways of preserving and providing access to these data.

Outreach and Regionalization

Beginning in 1996, the IASSIST Administrative Committee initiated intentional outreach to support participants financially from countries just developing data archives and services. Post Cold War changes in Europe have resulted in the emergence of a number of new national data archives. Similarly, the post apartheid period in South Africa saw the development of the SADA. IASSIST has been supportive of the staff from these new archives. Some financial assistance has been made available to encourage their participation in our conferences. In addition, conference programs have included opportunities for speakers from new data archives. This has helped build contacts for those new to data services and archiving as they work to establish a network of professional connections. The work of the International Outreach Committee, in collaboration with IFDO, UNESCO, and the conference host-institutions, has been another IASSIST success story.

With social science data services and archives expanding around the globe, the reliance on an annual conference to keep the community connected is not as effective as it once was. While the size of the community remains relatively small, the distances of worldwide participation create serious difficulties. One approach worth examining is the formation of strong regional IASSIST communities structured around the Membership Committee and Regional Secretaries. In a sense, Europe has been functioning like this since the formation of CESSDA and its introduction of expert seminars. Similarly, the biennial meeting of the ICPSR Official Representatives has served a similar function in convening mostly North Americans outside of IASSIST conferences. Ways of engaging members to meet within regions without eroding the attendance at annual IASSIST conferences need to be explored.

Conclusion

IASSIST remains a membership-based organization committed to institutional solutions to preserving and

providing access to data. The mission of IASSIST has not diminished over the past decade. If anything, its mandate has expanded as a result of the changes discussed above. Professional development will remain as important tomorrow as it has been over the past ten years. IASSIST clearly has a significant role to play in this area. Furthermore, a "voice for data" will be needed on the international scene tomorrow as much as it is today. The representation by IFDO and IASSIST this past year to get research data incorporated within the UNESCO Charter on the Preservation of Digital Heritage is an example of the type of leadership required at the international level. Currently, advocacy for research data at the international level has very few voices, and the voices that do exist are without coordination. The Committee on Data for Science and Technology (CODATA) and IASSIST hold many of the same interests but rarely if ever share the same platform. IASSIST needs to look for strategic alliances with other organizations concerned about the preservation and access to research data.

The talents and capacity of the membership of IASSIST have made this a successful organization. We may be small in numbers; but everyone in our number, counts. The recruitment of gifted new members, whose vitality will carry organization's mission forward, is crucial to the future of IASSIST.

IASSIST

INTERNATIONAL ASSOCIATION FOR SOCIAL SCIENCE INFORMATION SERVICE AND TECHNOLOGY

• • •

ASSOCIATION INTERNATIONALE POUR LES SERVICES ET TECHNIQUES D'INFORMATION EN SCIENCES SOCIALES

The **International Association for Social Science Information Services and Technology (IASSIST)** is an international association of individuals who are engaged in the acquistion, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompases hard copy as well as machine readable data

Paid-up members enjoy voting rights and receive the **IASSIST QUARTERLY**. They also benefit from reduced fees for attendance at regional

and international conferences sponsored by **IASSIST**.

Membership fees are:

Regular Membership: $50.00 per calendar year.

Student Membership: $25.00 per calendar year.

Institutional subcriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subcription: $75.00 per calendar year

Membership form

I would like to become a member of IASSIST. Please see my choice below:
Options for payment in Canadian Dollars and by Major Credit Card are available. See the following web site for details: http://datalib.library.ualberta.ca/membership/membership.html

- ☐ $50 (US) Regular Member
- ☐ $25 Student Member
- ☐ $75 Subscription (payment must be made in US$)
- ☐ List me in the membership directory
- ☐ Add me to the IASSIST listserv

Please make checks payable, in US funds, to IASSIST and Mail to:

IASSIST,
Assistant Treasurer
JoAnn Dionne
50360 Warren Road
Canton, MI 48187
USA

Name: _____

Job Title: _____

Organization: _____

Address: _____

_____

_____

City: _____ **State/Province:** _____

Postal Code: _____ **Country:** _____

Phone: _____ **FAX:** _____

E-mail: _____ **URL:** _____

Return Undelivered Mail To:

IASSIST QUARTERLY
c/o Wendy Treadwell
1758 Pascal St. North
Falcon Heights, MN 55113
USA