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, should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event.

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

Production
Fadi Dagher
423 West 120th St., #107
New York, NY, 10027
USA
Email: fd2102@columbia.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.
Welcome to the third issue of the IASSIST Quarterly vol. 29.

IQ volume 29 is still called 2005, while we are now enjoying being in 2006. From the IQ-team we do not expect or even hope to be totally current, but we are happy to be gaining on some of our rather persistent backlog. In the past year our gaining has been explicitly due to the work of the co-editors Louise Corti and Wendy Watkins, but I also want to thank all the authors of articles for the IQ as well as the reviewers / proof readers working behind the scenes plus all the people working on the technical side with layout, printing, distributing, etc. of the IASSIST Quarterly. This work is all undertaken with good will and voluntarily!

The first article in this issue is from the IASSIST conference in Edinburgh in May 2005. Bobray Bordelon is the Pliny Fisk Librarian of Economics and Finance/Data Services Librarian at Princeton University and presented at the conference his paper: “Cross-national & intergovernmental data: paying for one-stop shopping”. The point of departure for the article is that since most organizations lack the resources to build its own interface and provide the ongoing maintenance it is relevant to explore the choices for a commercial aggregator. Bobray Bordelon examines and characterizes four “aggregators”: Datastream International, EIU World Data, Global Financial Database, and Global Insight.

The second article is from Louise Corti - Associate Director & Head ESDS Qualidata, Outreach & Training UK Data Archive – on “Qualitative Archiving and Data Sharing: Extending the reach and impact of qualitative data”. Louise Corti argues that while there is a well-established tradition in social science of reanalyzing quantitative data, there is still not yet a well developed paradigm, nor a pervasive research culture of sharing or secondary analysis of qualitative data. So the article is a promotion effort in that respect and calls for further internationalization of the debate that has begun in some countries. To get started on this debate you can have head start by – among other things in the article – reading about the six main perceived barriers that have been identified through contact with researchers, together with some pragmatic solutions to confronting these barriers.

Lastly we have an article from our very own world. IASSIST maintains a very active list-server. The data – e.g. mails – from that list-server was collected some years ago and some categorization was performed on that collection. The article “Self Reflection of Virtuality in a Professional Association: a Compact Description of Mailing List Data” is a description of the IASSIST organization moving in the direction of a virtual community. We try to offer an answer to the jesting question: “Is there IASSIST life between IASSIST conferences?”. The ongoing IASSIST list-server, weblog and periodic IASSIST Quarterly should be a positive answer to that matter. The investigation was carried out by Repke de Vries, Department of Public Services at The Royal Library of the Netherlands, and myself.

A crew of people are taking care of the IASSIST website so it is constantly evolving. You can take a tour at http://iassistdata.org and visit the IASSIST weblog (blog) - IASSIST Communiqué – at http://iassistblog.org. The IASSIST website contains information on previous and coming conferences, including past presentations, as well as easy access to the published articles of the IASSIST Quarterly.

Articles for the IASSIST Quarterly are most welcome. Papers can be generated from IASSIST conferences, from other conferences, from local presentation, discussion input, etc. Contact the editor via e-mail: kbr@sam.sdu.dk.

Karsten Boye Rasmussen, June 2006
Cross-National & Intergovernmental Data: Paying for One-Stop Shopping

Introduction
Many intergovernmental and governmental organizations produce their own interfaces for statistical data. With the exception of the Common Database of the United Nations’ independent cooperation for integrating data between chosen to produce their own interface to example is the Economic and Social Data may also rely on commercial sources. Is a researcher likely to want to combine it logical to do so? Is the native interface good enough to use on its own and does it is easily allow the combining of data from various sources? Are the data from the original producer easily extractable? Does the original producer allow alteration of the data? Does your organization have the staff and expertise to design and maintain its own interface? Does your organization have the expertise to match changing data elements over time? Is there a commercial service, already in existence, that already does what you want? If so, how do the short-term and long-term costs compare? How would you treat documentation? How does the commercial vendor treat documentation?

Commercial Aggregators
Since most organizations lack the resources to build its own interface and provide the ongoing maintenance, the choices for a commercial aggregator will be explored. There are many elements to investigate when considering a commercial solution:

- Variety of data (due to cost, we want one source to cover as many organizations as possible; ideally it should be international and cover all countries and territories)
- Source of data should always be specified
- Fixed beginning date or Rolling Period
- Start date (should be easy to determine when each series starts)
- Frequency of data (daily, weekly, monthly, quarterly, annual)
- Update frequency
- Downloading capabilities and format (continuous series; text, tab, or comma delimited; EXCEL, SAS, SPSS, STATA)
- Documentation (online, direct links, paper)
- Interface/ease in finding data
- Medium of storage and delivery (INTERNET, WEB, CD, tape, terminal based)
- Cost
- Customer help (toll free phone number, responsive e-mail, online chat)
- Lease or purchase

Analyzing the Aggregators
Four major aggregators will be examined: Datastream International, EIU World Data, Global Financial Database, and Global Insight.
Another issue to consider is what metadata are provided. Datastream provides the start date, frequency, unit and scale of measurement, adjustment factor, conversion method, and original source. Global Insight and the EIU provide similar information with the EIU also providing the name of the analyst. Global Financial Database provides the most detail and its detailed source lists are of great value to historians and economists trying to conduct further research. One often is given the history of an indicator.

What is missing?
The sources tend to omit actual definitions of the terms being used. The EIU is the only one of the four sources to provide definitions. For the novice trying to choose between several choices, a hyperlink or mouse-over could provide much assistance. It would be useful to provide the prefaces to the works being used or the methodological documentation. For example, including International Financial Statistics Country Notes, Balance of Payments Manual, and Guide to Direction of Trade Statistics would assist the researcher in understanding inconsistencies and subtle distinctions in series.

What do we call an item?
Even when the same source is used, the vendor may not use the original terminology. The International Monetary Fund tries to standardize terminology between nations by having a consistent vocabulary. Since items are sometimes measured quite differently, the IMF tries to bring together similar concepts by grouping items by line number. For example, line 14a of International Financial Statistics represents the amount of money in circulation outside of banks. The United Kingdom,
Japan, Brazil, and Canada refer to this as “Reserve money of which currency outside DMBs (Deposit Money Banks)”. The United States calls it “Reserve money of which currency outside banks”. The Euro nations, Switzerland, and Sweden refer to it as “currency issued” while Denmark defines it as “Reserve Money of which outside BIs”. Datastream refers to it as “currency in circulation” or “currency outside banks”. Global Insight uses four terms: “currency in circulation outside DMB”, “currency outside banks”, “currency outside banking institutions”, and “currency in circulation”. If this was not confusing enough, one would have to turn to the Country Notes and not the International Financial Statistics Yearbook to determine what constitutes a “DMB” in each nation. A few examples are Qatar considers DMBs to be locally owned commercial banks (including Islamic banks) and branches of foreign banks; Canada includes only chartered banks; and Georgia only commercial banks. In Comoros, it is one specific bank -- Banque pour l’industrie et le commerce-Comores.

The confusion continues when a source is listed but the notes provide contrasting information. For example, the EIU provides a definition of M1 for the United States and cites its source as International Financial Statistics. However a note states the data uses national concepts rather than IMF concepts. Throw in similar series having different base dates and one begins to understand why researchers are confused and crave simplicity.

What documentation should be provided?
At minimum the following should always be listed:

- Start date
- Frequency
- Update Schedule
- Formats available for downloading
- Unit of measure
- Scale
- Adjustment factors (if any)
- Status (active/discontinued)
- Conversion Method (if applicable)
- Source
- Name of analyst for forecast
- Definition
- Links to old data if discontinued
- Comparability charts

How can we help researchers?
Talk to vendors about what is needed. Research institutions help pay the bills. Use the collective power of groups such as IASSIST and the American Library Association to persuade vendors to improve their products. Provide guides that point out common mistakes and fallacies. Long ago, Princeton University provided a guide for Datastream when the company itself was providing no documentation. This guide not only provided search examples but also pointed the user to where to look for items and the many misnomers of the program. Provide training to staff and researchers. Have the actual documentation available at a common reference point. While many of the sources are standard monographs that you may not have room for in a reference collection, have the basic guides to documentation and try to keep major statistical works on site. While the user accessing this information off site will not have immediate access, the serious researcher would know where to turn. Perhaps most importantly, we can know the sources’ advantages and limitations and be able to guide the researcher to their appropriate use.

* Bobray Bordelon is the Pliny Fisk Librarian of Economics and Finance/Data Services Librarian at Princeton University. bordelon@princeton.edu. The article is based on a presentation delivered at the 31st IASSIST Conference in Edinburgh Scotland on May 25, 2005.
Qualitative Archiving and Data Sharing: Extending the reach and impact of qualitative data

Introduction
Archived qualitative data are a rich and unique, yet too often unexploited, source of research material. They offer information that can be reanalysed, reworked, and compared with contemporary data. In time, too, archived research materials can prove to be a significant part of our cultural heritage and become resources for historical as well as contemporary research.

But while there is a well-established tradition in social science of reanalysing quantitative data, there is not yet a well developed paradigm, nor a pervasive research culture of sharing or secondary analysis of qualitative data. The lack of discussion in the current literature on the benefits and limitations of such approaches is evident. In the UK, Finland and France some interesting debate has begun, but it is still early days. Further internationalisation of some of the arguments would certainly be welcomed, and I would encourage IASSIST members who are considering archiving qualitative data to consider kick starting some written discussion in their own countries.

This contribution provides an overview of some of the perceived barriers to re-use and highlights some of the positive pragmatic measures that are being taken to enable both sharing and re-use of qualitative data. I draw on the recent experiences of ESDS Qualidata and a new research council funded programme intended to investigate innovative ways of extending the reach and impact of qualitative data.

A brief history
Readers will be well aware that research data archiving has been around for some years. The data archiving movement began in the 1960s within a number of key social science departments in the United States who stored original data of survey interviews. The movement spread across Europe and in 1967 a UK data archive (UKDA) was established by the UK Social Science Research Council (SSRC). But the emphasis was strictly quantitative. The SSRC’s successor, the ESRC introduced a formalised Datasets Policy in 1996 that contracted all award holders to submit data for possible accession to the UK Data Archive. The word ‘data’; was typically, and perhaps conveniently, taken by researchers to refer purely to numeric data.

Pressure from a small insistent minority fought for qualitative data to be explicitly embedded into the ESRC portfolio of data resources. The Qualidata Centre set up in 1994 in the Sociology department at Essex complemented the UK Data Archive with a joint mission to actively acquire, curate, disseminate and promote the raw data from social science research.

From 2003, Qualidata became an integral part of a larger joined up one-stop-shop for data sharing, archiving and dissemination, under the ESRC/JISC supported Economic and Social Data Service (ESDS).

Over the past ten years ESDS Qualidata has contributed to the elucidation of some of the key perceived barriers to re-using data - through extensive contact with 2000 or more qualitative researchers and through the experiences of handing many disparate data collections. These arguments have been rehearsed in a number of publications by Qualidata staff (e.g. Corti and Thompson, 2004; Corti, L., Witzel, A. and Bishop, L. 2005; Bishop 2005; Corti 2000).

While ESDS Qualidata has conquered some of the ‘mainstream’ methods for archiving and sharing qualitative data, it has made significant efforts to spark more general academic debate. However, there is still a significant under-use of archived qualitative data when compared with survey data. Moreover, there is still a noticeable imbalance in attitudes towards sharing and re-using data across disciplines and types of methodological approaches.

The key research issues facing re-use of qualitative data
There are some insistent voices who suggest there is a widespread reluctance to deposit qualitative data with a research archive. While this was partially true some ten years ago, today we see a new generation of qualitative researchers who are more inclined to either embrace or gracefully accept the ESRC’s Datasets Policy and its efforts to promote the value of sharing data (ESRC 2005). At the UK Data Archive, where some 150 qualitative datasets are catalogued, user figures have soared, particularly for use in research methods teaching.

Nevertheless, there are still barriers. The six key main perceived barriers that have been identified through contact
with researchers in the UK over past ten years can be summarised as follows:

- The practice of secondary analysis of qualitative data is not yet a common place research activity. The literature is not forthcoming on methodological guidance on how to approach the revisiting of data. Corti and Thompson (2004) provide the first inclusive and state-of-the-art chapter on the topic invited for a high profile methods reader by Seale et al. Progress is also hindered by preconceptions and sometimes less than innovative approaches to qualitative research. A cultural shift is required and we believe that this has been progressively happening since 1994.

- Problems of the implicit nature of qualitative data collection and analysis, of context and reflexivity, which are sometimes proclaimed to be indefinable. What are needed here then are practical strategies. Indeed, for research conducted in teams, data and fieldwork experiences are commonly shared, and for Principal Investigators who remain one step away from the field, it is imperative that they rely on their research staff on the ground to capture, document and communicate the nuances of the research process. It is vital to capture better and more systematically the context and the interrelationships among data and between data and other academic products, like analyses and write ups.

- Lack of time to get fully acquainted with research materials created by someone else. Social historians have been more forthcoming in revisiting data sources because of their willingness to embrace the slow and rigorous but commonly accepted practice of document analysis and the need to evaluate methodically the very sources they are revisiting. However it can be terribly time-consuming to locate suitable data sources, and to locate, for example, paper materials that may reside in traditional archival locations with limited access. New ways and tools that more efficiently expose the content and context of digital data sources need to be developed, in order to reduce such researcher burden.

- Constraints of informed consent. Informed consent is an ethical and legal requirement of the research process. It must be thought through at the time of research proposal planning and writing and be tailored towards the specific research questions and the sample. Often consent is not addressed until late in the research process by many researchers, and verbal consent alone is typically not sufficient for longer-term sharing and for effective use of research findings by the original researcher. Failure to realise the need to gain informed consent means that research efforts and the opportunities for archiving and secondary analysis are jeopardised from the start. But researchers require more guidance on this area to better understand the nature and implications of consent and confidentiality. Additionally, pragmatic strategies are also required to aid the commonly accepted practice of anonymisation or pseudonymisation. Bishop provides a succinct reply to some of the recent scepticism of the possibilities of re-use (Bishop 2005)

- Insecurity about exposure of one’s research practice, IPR or threat of misinterpretation. This may be relevant in some specific cases (e.g. an anthropologist's life work), but for the sake of data quality or auditing as Hammersley describes it (Hammersley 1997), exposure of data and methods is no bad thing. Capturing evidence, or specific reasons, as to why data cannot be shared is valuable.

- Finally, lack of a wide range of publicly available catalogued research data. While in the UK, the Economic and Social Data Service (ESDS) has done much to facilitate common resource discovery points of access through the use of standards at the study description level, ways to delve deeper into the qualitative data resource have not been as forthcoming as they have for survey data. The NESSTAR system is a good example of how data can be browsed online through the use of detailed data description down to the survey question (variable) level (Nesstar 2005). As the pool of rich and diverse shareable data expands, the greater the need also for interoperable and standardised description that will allow searching and location of key data across distributed sources. The means of enabling this information stock and flow to reach fruition needs to be investigated and common community methods agreed.

Prerequisites for making data shareable

There are two major issues that appear to be at the heart of making data fully shareable. The first is producing rich and full documentation about the data and the research processes used to conceptualise, collect, manage, process and analyse data. Full documentation enables effective resource discovery (i.e., catalogues) of distributed data sources and enables more informed re-use. The second challenge for sharing data is that of exposing data in the most flexible way possible so as to enable multiple methods of accessibility and innovative uses, for example, combine and link: activities that are the very core of some of the initial considerations of e-social scientists.

Both challenges require that:

- data are collected to a high standard using appropriate sampling strategies, rigorous data gathering methods and, where appropriate, systematic interview transcription
• research methods and practices (including the consent process) are fully documented
• the context of the data collection and analysis is captured
• the richness of the structure and features of data and are made available (use of mark-up)
• the interrelationships between data and analyses (intra-project) are made available (issues of representation)
• data are disseminated in sensitive ways that satisfy the ethical and legal requirements to which they are bound.
• data are represented in appealing and digestible ways, such presenting academic findings alongside evidence from the raw data (that is more than anecdotal quotes)

Enabling these requirements entails practical as well as conceptual challenges. And fundamentally, the underlying need is for the formulation, adoption and community subscription to commonly agreed methods, standards, and ontologies for data description and exposure. Previous work in this area has been spearheaded by ESDS Qualidata and the UK Data Archive in social science archival documentation and data processing for qualitative data (Corti, 2002).

Creatively exploring the barriers and looking forward
ESDS has found that data creation workshops have been very useful in helping unpack some of the specific issues and problems arising in the course of projects that are considering data sharing. The Medical Research Council (MRC) has also taken up an interest in data sharing of primary data from population and clinical trials data, with qualitative data firmly on the agenda. (Corti and Wright 2003). More recently the Natural Environment Research Council (NERC) has also implemented a formalised data management policy for a joint Council programme on Rural Economy and Land Use (RELU) that specifically includes all the qualitative data from the Programme. Context and consent are the two words that crop up most frequently in the debates.

For researchers, like myself, who have been seeking academic funding opportunities to confront such consent, context and technical issues, 2003 – 2004 saw a bumper harvest of such prospects. In the past it has been unusual for ESRC to fund research and development or consider dedicated methodological initiatives. But, over the past three years we have seen a much welcomed move towards dedicated funding for methods. These strands of money have enabled some innovative investigations to be undertaken, particularly for qualitative data. Five main pots of ESRC funding appeared on the scene, thanks to a number of champions to the cause of methods and data analysis: E-social science; the Research Methods Programme; the National Centre for Social Research; the Qualitative Longitudinal Study and the QUADS scheme.

Innovation: The QUADS scheme

QUADS is the ESRC Qualitative Archiving and Data Sharing Scheme, running from April 2005 until October 2006. The aim of the scheme is to develop and promote innovative methodological approaches to the archiving, sharing, re-use and secondary analysis of qualitative research and data. A range of new models for increasing access to qualitative data resources, and for extending the reach and impact of qualitative studies will be explored. The scheme also aims to disseminate good practice in qualitative data sharing and research archiving. This is part of the ESRC’s initiative to increase the UK resource of highly skilled researchers, and to fully exploit the distinctive potential offered by qualitative research and data.

The QUADS is a small initiative (some £500,000 over 18 months) but is dedicated to the mission of learning more about sharing, representation and re-use of qualitative data, in all of its disparate shape and forms. Five small exploratory projects have been funded together with a Co-ordination Role. The Co-ordination team based at ESDS Qualidata have been charged with the task of providing a pivotal role in fostering communication and understanding between the five demonstrator projects. Communication of the Scheme’s innovative efforts to the broader spectrum of qualitative researchers is much needed. But equally it is must be appreciated that there exist various communities of practice with different data needs and methodological approaches to sharing and secondary analysis of qualitative research and data. Fruitful collaboration is required which can be achieved through guided discourse to inform and help guide the progress of QUADS demonstrators, and to encourage the broader acceptance and take up of data sharing and re-use.
Key areas for QUADS projects

Four key areas of needs and commonality identified across all the QUADS projects point to: defining and capturing data context, audio-visual archiving; consent, confidentiality and IPR; and web and metadata standards.

The debate on capturing context has been around for some time now on the qualitative data archiving scene. QUADS aims to devise and recommend a minimum set of contextual constructs that would be necessary to document a collection of qualitative data to enable informed secondary use. Regarding audio-visual data, they are being handled by many of the projects and the scheme is providing an opportunity to share expertise on presenting and re-using such sources. On the hot topic of consent, confidentiality and copyright, while ESDS Qualidata maintain up-to-date detailed information many the QUADS projects do have specific consent and copyright issues, and it will be invaluable to see how these are confronted by the different projects during the demonstrator period. They will afford unique case studies that can be used in the future.

QUADS Coordination will hold an end of scheme hands-on demonstrator workshop, where projects will be able to talk about their investigations and developments and demonstrate any working QUADS products to an open invitation audience. Furthermore, QUADS Coordination through its web site will mount papers, tools and training materials arising during the course of the projects in an easy-to-navigate manner. A session at the NCRM Summer School is currently being planned. QUADS is an exciting pilot initiative that is exploring experimental methods. The team is therefore very happy to hear from anyone who is already working in this area or who would lie to contribute to these exciting projects.

But what about these standards?

In order to approach primary data now and in the future in years, we need that data to be accurately, richly and contextually described. And in turn, re-presentation of original data, methods and analytic interpretation and their interweaving requires agreed and exemplary standards and procedures. Fielding’s scoping study that examined issues for the role of qualitative data in e-social science (Fielding 2003) aptly confirmed that ‘it is timely to anticipate emerging innovations in qualitative methods, including new data forms, sources, possibilities for research archiving and data mining and the potential for increased participation and access’.

Representation can be viewed across a spectrum starting from the simple publishing of anonymised digital qualitative data sources or banks (which are typically not present) through to the ability to link qualitative data to other distributed data sources (e.g. audio-visual or geo-
coded data sources) and to creative and exciting ways of visualizing data. However, it is important to take a step back and see what is currently exposed. The researcher will find very little qualitative data even exposed to the web in any meaningful way. While there are archives of qualitative data to be found across the world, the majority is not even in digital format and “digitizing” these collections is often seen as merely providing an online catalogue of digitised metadata. The issue of how to make these data resources accessible to users has hitherto been a central concern for ESDS Qualidata who has continually been seeking ways to meet users’ requirements.

Standards of relevance are those for: building sustainable web sites; harmonious data descriptions to enable rich resource discovery (metadata); and marking-up data content. ESDS Qualidata recognised the need for standards and tools back in 2000 – tools that allow data to be published to the Web and support online interrogation of data via standard Web browsers. In 2000, Qualidata undertook pioneering work in this area through developing the Qualidata Online system and a methodology for sharing data (Corti and Barker, 2002, 2003).

The need to keep pace with the development numeric data browsing systems, that are now quite far advanced, is important not only for the UK Data Archive but also for other groups who wish to publish and share qualitative data.

Community efforts
QUADS Co-ordination is very aware that there are an increasing number of projects in the world that are looking at sharing qualitative data, typically via the web, and particularly in the wake of the e-science rush. But they are not linked up in any formal way. QUADS Coordination will be building an interactive map will be built to show the location of such initiatives and the key contacts. Additionally training and advice is being given on best practice in metadata creation and web standards for qualitative data. It is hoped that many of the e-science and methods research groups will be amenable to agreeing on some basic sets of standards.

*For more details about ESDS Qualidata see www.esds.ac.uk/qualidata and QUADS see: http://quads.esds.ac.uk/ or contact Louise Corti, UK Data Archive, University of Essex, Colchester CO4 3SQ corti@essex.ac.uk.

An earlier version of this article appeared in Issue 1 of Qualiti, the Newsletter of the National Research Methods Centre, University of Cardiff. www.cardiff.ac.uk/socsci/qualiti/newsletter.html

References


Corti, Louise (2002). Qualitative Data Processing Guidelines. Qualidata, UK Data Archive, University of Essex, Colchester


ESDS www.esds.ac.uk

ESDS Qualidata www.esds.ac.uk/qualidata

ESDS Qualidata Online http://www.esds.ac.uk/qualidata/online/

ESRC Datasets Policy www.esrcsocietytoday.ac.uk/ESRInfoCentre/Images/Annex%20C_tcm6-9738.pdf

Fielding, N.(2003), Qualitative Research and e-social science: appraising the potential, University of Surrey, http://www.ncess.ac.uk/docs/qualitative_research_and_e_soc_sci.pdf

NCeSS Hub and Nodes www.ncess.ac.uk

NCRM Hub and Nodes www.ncrm.ac.uk

Nesstar (2005) http://nesstar.esds.ac.uk/webview

QUADS http://quads.esds.ac.uk

Footnotes
^ See www.esds.ac.uk/qualidata/access/internationaldata.asp for an overview of progress on national data archives acquiring qualitative data
Self Reflection of Virtuality in a Professional Association: a Compact Description of Mailing List Data

Abstract
This article presents an investigative description of the utilization of a mailing list in our own professional membership organization: IASSIST. As an electronic form of communication, the mailing list has supported and supports the IASSIST in moving the organization in the direction of a virtual community. The mailing list offers an answer to the jesting question: “Is there IASSIST life between IASSIST conferences?” This work contributes to methodology by offering a refined typology for the description and analysis of mailing lists in general, as well as a specific subject categorization for the IASSIST mailing list based on findings in the mailing list communication. The article gives a compact quantitative description of the key figures from the use of the IASSIST mailing list based upon the typology and categorizations. The analyzed data consist of emails from 32 months before the year 2000 millennium turn. A follow up on the analysis with more present email data is considered.

Introduction
Virtual organizations have been identified as real (Davidow and Malone, 1993) or real organizations sometimes viewed as imagined (Hedberg et al., 1997). The concept of the virtual community has now existed for a good 10 years (Rheingold, 1993). The virtuality emerges due to intense use of information technology corresponding to organizational arrangements that potentially and practically break the boundaries of time and space. In our time of virtuality, people no longer have to share the same space or be in the same time, as direct electronic communication can span the space, and relayed (asynchronous) electronic communication can span the time. A voluntary association – such as the IASSIST – is considered as both an organization and a community, and by applying electronic communications, such associations have the potential of growing into a virtual community.

The aim: community and virtuality
The object for investigation – IASSIST – is a small voluntary professional organization. The International Association for Social Science Information Service and Technology is an organization of professionals – typically from data archives and libraries – supporting research and education. The concept of “social science” is viewed in its most wide-ranging sense. IASSIST is a network and shows network externalities. “The more the merrier,” but naturally this is balanced with the group of members actually being a group with common issues. Size has to be balanced with the necessary homogeneity within the group. The IASSIST organization was founded in 1974, and presently has about 300 members. The growth in members is mostly a result of the fact that “international” 25 years ago primarily meant “US-Western-European,” but now more fully encompasses the globe. However, the IASSIST organization has not had a strong intention of membership growth into areas outside its professional base of data archivists, data librarians, and some social scientists. All of these affiliated with mostly university and/or research institutions. The existing IASSIST network can be viewed as comprising most of the relevant potential membership, but there is small but constant growth in regarding data materials as a resource available through libraries and archives.

A community is a togetherness that shares. Historically, the sharing within communities can take different forms, from the communion (with focus on sharing idealistic/religious beliefs) to the commune (that also practices sharing of material effects). All communities share meaning through communication. The knowledge shared in the community can be exemplified or statistically described. In this context the starting point is a look at how the sharing takes place.

Historically, social groupings have shared the same geographical locality and togetherness in time. Information technology breaks that barrier by making some sharing look and feel as reality, but it is “virtual reality.” Virtuality is not reality, but it could be a dream or an imaginative creation. It can also be said that it surpasses reality – “it is surreal.” The surreal found in arts and literature surpasses the objective reality as it includes individual subconscious elements to provide a more full understanding of our reality – even an antagonistic image of reality. Half a century ago the mass media turned the world into a “global village” (McLuhan, 1964). Virtuality contains this implosion in a synchronous form; virtuality provides an even more elaborated ability to react and utilize the new media of communication in obtaining closeness across barriers of
time and space

**Selecting media for observation**

The communication between members of IASSIST has, from the start of the organization in 1974, taken place through conferences and also a newsletter/periodical. The membership has been among the “first movers” in the use of information technology (IT) for communication because their professional work included intensive use of IT and new media have been added to the list. From early on IASSIST has adopted the use of a mailing list and, later on, also a supporting web site. These media are briefly examined here for their capability to support virtuality and thus for founding the basis of an investigation into the virtuality of IASSIST:

**Conference:** The conferences are where IASSIST members meet face-to-face. Conferences are where people are situated in the same time and place and under a common heading and, furthermore, mostly detached from their obligations of everyday tasks. We may ask: “How and where do IASSIST members exchange views and share knowledge between the yearly conferences?” Or this could be formulated: “Is there life in IASSIST between conferences?”

**Newsletter:** Some life is added to the organization four times a year. Since 1977 there has existed a communicative channel for the membership associations in the form of a traditional printed and (surface) mail delivered membership newsletter (The IASSIST Quarterly or IQ). The content of the publication is primarily papers from presentations at the IASSIST conferences and the publication is not directly a medium for deliverance of data for research, although many articles have that subject as their focus. The articles can thus act as endorsements of data and directions for gaining access to data, as well as describe systems for data deliverance. Secondly, the IQ does not contain much communicative bi-directional interaction. Some references between articles in the IQ are found, but the periodical contains no actual debate. Although the newsletter is now also available in electronic form at the IASSIST web site (http://www.iassistdata.org/), the newsletter is considered without substantial independent importance for support of the virtual community. The argument is that the communication is one-way and that it started – and still exists – in the low-tech form of printed paper. However, the ease of access facilitated by the availability of the IQ on the web and the fact that IQ documents the objectives that members of IASSIST are pursuing in their work life count as assisting factors for IASSIST being a community.

**Web site:** The IASSIST web site has been conventional by creating access to some formal documents and more permanent announcements, especially the conferences. A bigger move forward was made when the IQ newsletter was published on the Internet (issues are available from 1993 onward). Furthermore, the web site features electronic copies of presentations made at the conferences, as the computer files (PowerPoint) are collected and presented in the original conference structure of days and sessions. The authors of this article have carried out some empirical research considering preferences for services at the web site. (This work is intended to be published later.) After the collection of the empirical data utilized in this article and after the survey of preferences mentioned above, the IASSIST has also recently opened a web log as a publishing and discussion board. This is obviously relevant for the concept of virtual community, but is not considered further in this article that is concentrating on the facilities available at the turn of the century.

**Mailing list:** In the early 1980s, communication by email was already an established fact among the IASSIST membership. Because of the international cooperation in the organization and the connection of data archives and libraries to mainframes for universities and research institutions, most members were connected to the forerunner of the Internet (ARPANET). The distribution of emails among members involved many copies to other members and was then consequently structured by establishing a mailing list for the membership. This mailing list has the central capability of permitting two-way communication.

Is the mailing list sufficient for constituting the infrastructure of IASSIST as a virtual community? The phrase “virtual community” was first used by Howard Rheingold (1993) in a book taking the point of departure in a mailing list (the WELL, “Whole Earth ‘Lectronik Link”), so there is precedent for virtuality obtained through a mailing list. Consequently, we regard the IASSIST membership as a virtual community – in relation to the fact that the non-virtual communication is sparse – and we regard the mailing list as a valid medium for investigation.

**Mailing lists**

A mailing list is basically a communication duplication facility for email. Emails sent to the mailing list are being distributed or sent on to all members (“subscribers”) of the mailing list. Hardie and Neon (1994) distinguish mailing lists into three types based on the applied filtering of information. The first one is the unmoderated list where everything sent to the list immediately is replicated to the subscribers, with no waiting time, but some messages might be annoying. The second one is the moderated list, where a moderator has to accept the input to the list; this requires work and introduces a varying time buffering of the accepted messages. However, the editing is necessary for the list not to be overwhelmed by unwanted commercials (spam). The third type of mailing list is the
digest list which sometimes resembles a newsletter by having several subjects included and commented on by an editor. There will be only a few emails and they will appear with some regularity (e.g., monthly). The digest list is a one-way distribution list because communication travels only from the editor to the members. The categorization above is based upon whether all subscribers can make postings to the list directly, indirectly, or not at all. This categorization is paralleled by database users having levels and combinations of “read” and “write” permissions. The next aspect to consider is whether subscription to the mailing list is open to everybody or to a defined group of people.

The investigated mailing list of IASSIST is a moderated list, with only the membership as possible subscribers, and a subscriber can always both submit and receive emails from the list.

**Literature**

Mailing lists have been the subject of some earlier investigations. Fox and Roberts (1999) investigated the use of a mailing list amongst medical doctors in practice (GPs). The article demonstrates some anecdotal content analysis of the emails as citations from emails are presented, however the article contains neither statistical analysis nor description. Another approach was used in the study by Hannah (1999), where questionnaires were emailed to the subscribers of a mailing list to investigate the observed benefits of the mailing list. This was carried out without investigating the activity on the mailing list itself. Xu (1998) studied several mailing lists used by system librarians. Emails from one of the lists were examined – for a two-month period of time – and questionnaires were sent to several categories of users (or user roles). An equal short period of mailing list traffic was examined by Burton (1994). Empirical later investigations of several mailing list and their members and in particular their non-participants (“lurkers”) are found in Stegbauer & Rausch (2002).

**Available data**

The current investigation of a mailing list is also an endeavor into the investigation and demonstration of the obtainable level of information from the Internet without actually asking for individual approval and consent from the subjects being investigated. Often the data of the mailing list is placed on the Internet and easily and directly available. In this case, clearance to access the IASSIST archives of the mailing list was given to the researchers by the organization. But many mailing lists are publicly open for retrieval and they often have searching facilities for the identification of threads of interest, and contributions to mailing lists are sometimes stored for easier retrieval and presentation through the use of a web-application. In the use of data in this article, precautions have been taken not to directly reveal the identity of people and their expressed opinions. However, the information for this kind of monitoring of activities on the Internet and especially on mailing lists are available. King (1996) discusses the ethical aspects of the availability of communication data and Burton (1994) also addresses the ethical aspects by sending out information to the mailing list under investigation.

**Themes of analysis**

The analysis will present descriptive answers to the following dimensions and questions that became apparent when describing the mailing list from some obvious viewpoints of interest:

**Active-Passive:** Who is sending to the mailing list? The senders are all identifiable as email addresses, thus permitting determination and comparison of the active addresses. Secondly, the members of the mailing list are also identified as email addresses (as the addresses being sent to) and these are the members of the IASSIST. The senders can be investigated and compared to the passive non-senders. (General communication and mailing list as media).

**Nationality:** With some accepted uncertainty, email addresses may indicate nationality. The main validity problem is that USA becomes the default used when a nationality is not directly given. But this is considered a minor problem in this context because most members outside the USA can be said to belong to educational institutions that in their email are directly attached to a nation. Had it been commercial institutions, this solution might not have had sufficient validity. (Internationality of an organization).

**Officer-Member:** Are certain membership groups more active in posting information to the mailing list? A list of persons performing official functions (officers) is available as another mailing list is used for administrative purposes. Some differences in communication patterns between officers and regular members are expected. (Hierarchy in the organization).

**Modes:** What modes of communication take place on the mailing list? Some emails stand isolated (“single”) while other emails are connected and can be combined into threads with regard to the same subject and within a defined period of time. Within a thread, the single email is classified according to its role in the thread as “initiation” or “reply.” (Communication initiator or follower).

**Data and method**

The underlying unit of analysis is a single email and all emails are stored and retrievable from the list server from December 1991 and onward. A technical shift occurred in May 1997, so to secure the comparison issue the investigation period includes 32 months (from May 1997 until December 1999). Because the same person could...
have several email addresses and a person could have changed his or her email address during the 32-month period, a considerable process of match-merging by user-written matching software took place for performing a valid aggregation and shift of analysis unit from emails to persons (participants on the IASSIST mailing list).

Findings and figures
The total material consists of 691 emails that have been sent from 162 persons. The membership on the mailing list consists of 265 persons, i.e., 103 persons did not post email to the list during the investigated period of time. The 691 emails sent from 265 potential posters of mail results in an average of 2.6 emails per person (or approximately one email to the mailing list per year per person).

When comparing the figures for within and outside North America, it appears that the ratio for mail per person is 3.1 versus 1.4. The higher figure among members from North America supports the findings of US-dominance in earlier mentioned studies (Xu, 1998; Burton, 1994).

The membership of 265 persons can be subdivided into 31 persons belonging to the group of IASSIST officials and 234 regular members. Seven of the officers have not participated in the mailing list. The ratio of emails from 24 active officers (280 mails) is 11.7, which is significantly higher than the 138 active non-officers sending 411 mails (ratio 3.0). A regression model shows that the binary office variable is highly significant and explains 20.1 percent of the variation, and that the addition of the geographic variable and the interaction term only accounts for an extra 3.6 percent of the variation.

The figures above account for the fact that 96 non-officer members are inactive in submitting to the list. Inequality in electronic communication has in several contexts been the subject for studies (Sproul and Kiesler, 1991, p. 60), and terms like “quiet observers” (Ha, 1997; Xu, 1998) or “lurkers” (Fox and Roberts, 1999; Stegbauer & Rausch, 2002) have been introduced. However, it is only reliable to conclude that persons responding to the list are reading the emails – or more precisely just that email. However, the rational behaviour of a mailing list member who never reads the mails would be to unsubscribe to the list. Inactive subscribers can be considered content and regard the list as providing useful information as in direct performance improvement via computer-mediated communication (Rice, 1994).

The emails were combined into threads or single emails, where a “single email” has no response from the list. The separation into threads was done by user written software that stripped the title field down to essential information, and when titles matched and emails were close in time they were considered to belong to the same thread. A thread with only one email is a “single.” So if an email is not a “reply” it is a starting email that can be categorized as either a “single” (without any later responding emails and no part of a thread) or an “initiation” (the start of a regular thread). Three hundred of the emails were single and had no follow up; the remaining 391 emails were combined into 125 threads (as shown in table 1 below). The content of the emails are not analyzed in this context. However, some single emails never expected any response as they are often email announcements (e.g., emails announcing the availability of a new dataset).

The officers were initiators of 213 (149+64) occurrences while the regular membership started 212 (151+61).

The emails in table 1 are produced by 24 officers and 138 regular members. The officers are thus characterized as more frequent starters of emails (an average of 8.9 emails per officer) compared to the regular membership (with an average of 1.5 emails). With respect to posting replies to the list, the difference between the two groups is not that extreme (averages 2.8 and 1.4). Furthermore, there is a clear relationship between being an active initiator and being an active replier. The number of replies correlated to the number of starts from the same person was much higher among the officers than among the regular membership (Pearson 0.66 versus 0.20). This means that a small group of the officers are very active in their use of the mailing list.

Conclusion
The article has demonstrated a utilization of data being available on the Internet – the data driven approach – and we have not created or collected other data for this

<table>
<thead>
<tr>
<th></th>
<th>Starting</th>
<th>Following</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single</td>
<td>Thread</td>
<td></td>
</tr>
<tr>
<td>Officer</td>
<td>149</td>
<td>64</td>
<td>67</td>
</tr>
<tr>
<td>Regular member</td>
<td>151</td>
<td>61</td>
<td>199</td>
</tr>
<tr>
<td>All</td>
<td>300</td>
<td>125</td>
<td>266</td>
</tr>
</tbody>
</table>

Table 1. Distribution of emails to mode and membership type
particular research task. We regard this as defendable ethical research as individuals are not being exposed, but we welcome a debate on surveillance of individuals without their given consent through the materials available or traces left on the Internet. The findings of the descriptive analysis include explanation of email participation, and it was found that the crucial variable for high activity was whether a person had official duties within the organization.

The mailing list forms a virtual community – and like for a virtual organization – a virtual community is also characterized by blurred boundaries. In further investigations we have used (but not yet published) questionnaire data to look for evidence of further virtuality in terms of boundary crossing where non-formal (non-paying) members enjoy close to the same benefits as the regular membership. Furthermore, a follow-up study on the mailing list (5 years after) is being considered

* This article appeared in another form in the proceedings from the 11th International Conference on Human-Computer Interaction (2005). Some of the figures were earlier presented at the IASSIST 2001 Conference in Amsterdam as “Professional associations in transition to virtual communities for collaboration: the case of IASSIST.”

Repke de Vries, Department of Public Services at The Royal Library of the Netherlands.

Karsten Boye Rasmussen, Department of Marketing and Management at University of Southern Denmark. Contact e-mail: kbr@sam.sdu.dk.

References


The International Association for Social Science Information Service and Technology (IASSIST) is an international association of individuals who are engaged in the acquisition, 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 encompasses 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 Subcription: $75.00 per calendar year (includes one volume of the Quarterly)

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 US
Shelagh Mackay
University of California
Davis, CA, USA
95616-8617

Name:
Job Title:
Organization:
Address:
City: State/Province:
Postal Code: Country:
Phone: FAX:
E-mail: URL:
Return Undelivered Mail To:

IASSIST QUARTERLY

c/o Fadi Dagher

423 West 120th St., #107

New York, NY 10027

USA