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. Word processing files may be sent via email to: kbr@sam.sdu.dk. Manuscripts should be sent to Editor: Karsten Boye Rasmussen, Department of Organization and Management, University of Southern Denmark, SDU-OU, Campusvej 55, DK-5230 Odense M, Denmark.

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.

Editor
Karsten Boye Rasmussen, Department of Organization and Management, University of Southern Denmark, SDU-OU, 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 Halle Hall, 271 19th Avenue South, Minneapolis, MN 55455.
Phone: 612-624-7091
Email: wblock@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 © 2002 by IASSIST. All rights reserved.

CONTENTS
Volume 25 Number 1 Spring 2001

FEATURES

5 Understanding Barriers to the Use of Numeric Data in Learning and Teaching
Robin Rice

10 The Virtual Training Suite: Internet Skills for Teaching and Learning
Heather Dawson

15 MISSION (Multi-Agent Integration of Shared Statistical Information Over the [inter]Net) - The data archive perspective
Joanne Lamb

21 Research for Building a Better Data Community
Charles K. Humphrey

Editor's Notes

The articles in this issue are based on presentations from the IASSIST conference in Amsterdam in 2001. The articles show great involvement and enthusiasm in data archiving and dissemination. We begin with a paper from the session on "Learning and Teaching" by Robin Rice from the Data Library at Edinburgh. The title is "Understanding barriers to the Use of Numeric Data in Learning and Teaching" and starts by stating as a fact that "data resources are under-used in the learning and teaching environment". Robin Rice quotes that there is a lack of "statistical literacy". The Joint Information Systems Committee (JISC) funded this project on – among other issues – "the extent of use and the practicalities of using..."
data in teaching”. A survey was carried out and I have observed that one of the results was that only “one-quarter of those who teach with data” were familiar with the national data services. The late Per Nielsen from Denmark often mentioned that data archivists also should be “data pushers” – a marketing effort is needed! The article concludes by wondering how information technology will change into learning beyond the traditional classroom, the “virtual classroom” concept comes to mind.

This leads to the next article – from the same session – where Heather Dawson with assistance explains about “The Virtual Training Suite: Internet Skills for Teaching and Learning” aiming to “support lecturers, students, and researchers in finding and using resources on the Internet”. Again the JISC has funded the project, which has developed 40 free Web based Internet tutorials that cover a wide range of the academic subjects taught in UK universities and colleges aiming to provide a structured learning environment.

In the session “Tools for Data Services” Joanne Lamb from Edinburgh talked about MISSION (Multi-Agent Integration of Shared Statistical Information Over the [inter]Net”, in this article ” – the data archive perspective” has been added. This time the project is funded by the European Commission to develop software “to allow consumers of statistics to access these data in an informed manner with minimum effort”. The article also includes some technicalities about the architecture of the system as well as showing the use of metadata and connects to IASSIST well-known acronyms as XML, DDI and FASTER.

You could call it “navel-gazing” or “praxis related research” when we had a session on “IASSIST” at the IASSIST conference. From Canada Charles K. Humphrey writes on “Research for Building a Better Data Community”. And I simply have to quote him on: “I was exited by the research carried out by Karsten Boye Rasmussen and Repke de Vries about IASSIST as a virtual community”. In this community we are (still) facing the problem of “how to get researchers to conduct their projects so that their data products meet archival standards”. In a study on practitioners use of medical research findings one objective was to identify data products from the researchers. A survey with attitudinal items shows that (only) around 80 percent regard “data a valued by-product” and “secondary analysis as a valid research method”. Do we have a 20-80 problem here? Furthermore Charles Humphrey shows that about half (!) of the researchers regard it as a "waste of funds to save" data and do not agree in "archiving is integral". Indeed, a marketing effort with argumentation is required towards both producers and consumers.

Presentations from the conference are available for view at the IASSIST web-site http://www.iassistdata.org. From this top-page you can then click "Multimedia Presentations at the IASSIST 2001 Conference". Enjoy!

Karsten Boye Rasmussen
Understanding Barriers to the Use of Numeric Data in Learning and Teaching

Background
UK higher education is rich in numeric datasets. In the socioeconomic field, for example, there are large-scale, representative sample surveys (e.g., General Household Survey), current and historical population censuses, international comparative datasets, longitudinal surveys, economic time series, and data about markets, companies, and commerce. In the UK a centrally funded system of national data services for higher education provides for the dissemination of much of this research data, which is free at the point-of-use and accessible over the Internet (via JANET, the UK academic network).

However, these data resources are under-used in the learning and teaching environment. Despite the potential gain in numeracy, critical use of evidence and empirically-based knowledge by students conducting data analysis at both the postgraduate and undergraduate levels is infrequent, and obstacles exist that make integration of numeric data resources into coursework difficult. Employing numeric data effectively in teaching requires specialised skills and more time for preparation than the use of printed materials or bibliographic databases, and both students and teachers require a high level of support. As expectations about the use of information technology in learning and teaching rise, the barriers that inhibit the use of this wealth of data in the classroom and in student projects need to be lowered.

Understanding statistical evidence is important not just for postgraduates learning to be researchers and entering the professions, but for undergraduates as well. Milo Schield has written widely about teaching statistical literacy in higher education. He explains it as a different and more fundamental skill than producing or ‘doing’ statistics: “Statistical literacy focuses on making decisions using statistics as evidence just as reading literacy focuses on using words as evidence. Statistical literacy is a competency just like reading, writing, or speaking.” The need for application of such a competency in many fields is readily apparent.

The Numeric Data Project
This paper reports findings from a national collaborative project: “Using Numeric Datasets in Learning and Teaching,” funded by the JISC (Joint Information Systems Committee, which itself is funded by the Higher Education Funding Councils). The lifetime of the project is February 2000 to September 2001. Project partners are from three national data centres, EDINA, MIMAS, and the Data Archive, and two university data libraries, the University of Edinburgh and the London School of Economics. Additionally, a Task Force of experienced academics from across the UK was recruited as volunteers to guide the enquiry and its outcomes. This partnership reflects the novel perspective taken by the project to examine use of the nationally-funded data services with particular reference to local support needs of teachers and learners within their universities. The project is one of several funded under the JISC’s Learning and Teaching Development Programme (see http://www.jisc.ac.uk/dner/programmes/projects/ for a full list of projects).

A major objective of the project was to generate knowledge on issues such as the extent of use and the practicalities of using data in teaching, and the experiences teachers have of data support from both national data services and support staff in local institutions. Since user surveys tend to target those already registered for national services, there is no ready evidence about the larger population of UK university teaching staff on these issues. Therefore, a nationally representative sample survey was needed to discover the current “state of play” before recommendations about how to lower barriers could be made. The survey was designed to ask teaching staff about their use of numeric data in teaching and supervising students, their experience of national data services, barriers to using data in teaching, and the extent of support available within their institutions.

The teachers’ survey was enhanced by qualitative case studies of a diverse set of postgraduate and undergraduate classes using numerical data in teaching, which both inform the enquiry and also act as exemplars for other teachers. The full survey results and case studies are available on the project Web site at http://datalib.ed.ac.uk/projects/datateach.html. The final report with its recommendations, teaching resources, and other information is also available.
Survey Methodology
A sample postal survey was conducted of UK university teaching departments within the social sciences, plus other selected disciplines “outside” the social sciences, such as public health sciences. Two hundred sixty-seven department heads were randomly selected from a universe of 1590 (1 in 6 sampling fraction). The sampling frame was purchased from the marketing company Mardev, extracted from the Worldwide Academic & Library File. Department heads were asked to complete the four-page questionnaire themselves and to pass copies to relevant teaching colleagues to garner their participation. (A Web version was also made available for on-line input.) There were 206 responses collected from 110 departments. Fifteen records were removed as ineligible (e.g. non-teaching department). Following telephone, e-mail, and postal follow-up requests to sample members, the final response rate (110 / 252) was 44 percent of departments sampled.

Survey Results: Use of Data in Teaching and Learning
Due to the survey design and instructions to department heads, there was likely a skew toward data users among those in the sample who participated, as a result of self-selection. (Non-data users tended not to respond to the survey, as it was not felt to be relevant to them.) Seventy-nine percent of those survey respondents who taught or convened courses used data either “nearly always,” “often,” or “occasionally” (see Chart 1). The sample also seemed to over-represent senior staff (perhaps because the request was sent to department heads), teachers of methods courses, and those committed to quantitative analysis. Among those who used numeric data in teaching in some form, about two-thirds expected students to work with data on a computer, in “hands-on” fashion. As Table 1 shows, a higher proportion of methods courses were hands-on than subject courses. [The categories of “methods-based” and “subject-based” were coded during analysis, based on names of courses supplied by respondents.] Surprisingly, neither course level nor class size appeared to affect whether the course was hands-on.

<table>
<thead>
<tr>
<th></th>
<th>Methods</th>
<th>Subject</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on</td>
<td>85</td>
<td>54</td>
<td>64</td>
</tr>
<tr>
<td>Not hands-on</td>
<td>15</td>
<td>46</td>
<td>36</td>
</tr>
<tr>
<td>n=</td>
<td>46</td>
<td>100</td>
<td>146</td>
</tr>
</tbody>
</table>

Table 1: Whether course is “hands-on,” by course type

Although the survey was directed towards staff, not students, there was an attempt to understand the level of data use by students in their independent learning. Ninety-two percent of respondents who were either post- or undergraduate supervisors recommended the use of numeric data for students’ research at least occasionally (depending on the nature of the research project). Below are “typical” responses for each category.

- **Nearly always do (35 percent):** “Statements need to be backed up with evidence – often of an empirical nature.”

- **Often do (33 percent):** “Depends on topic, but statistical sources can contextualise a topic.”

- **Only occasionally (21 percent):** “Many students are more inclined to qualitative research.”

- **Never have and don’t plan to (6 percent):** “Not relevant to what I am teaching.”

- **Haven’t yet but would like to (2 percent):** “Not always appropriate and [I am] insufficiently briefed on numeric data available.”

Burden of Data Preparation
The survey instrument dealt directly with the issue of how burdened teachers felt regarding data preparation. As Chart 2 shows, a slight majority felt that data preparation was a burden, but warranted.
Respondents were also asked if they felt the need to update / refresh / revise the data used on a regular basis. Of those responding (78 percent of those eligible, n=142), 57 percent said yes, and only 14 percent said no. However, 29 percent said yes, but there was insufficient time to do so.

Data Sources and Use of National Data Services
The survey showed quite clearly that, although the use of numeric data among the survey respondents is high, the use of national data services that provide on- or off-line access to secondary datasets is not. Only one-quarter of the respondents who used data in teaching had “used or considered using” the national academic data services (namely the Data Archive, EDINA, and MIMAS) for teaching purposes.

So what are the sources of numeric data used in higher education classes? Most strikingly, half the teachers either required their students to collect their own data, or taught with data they collected themselves (see Chart 3). Nearly half, 44 percent, used print data sources, extracted from a monograph or serial. (Print publications obviously do not provide the material needed for a “hands-on” component, which gives students practice at manipulating data on a computer, unless the data are hand-entered.) The rest of the sources, including from a colleague, freely available on the Internet, or bundled with a textbook, were used by less than 20 percent of teachers who use data. Twice as many respondents received data from a government agency or “directly from the data producer” as were registered with a national data service.

These results indicate a need to further explore the nature of data sources needed by particular disciplines for teaching particular types of courses, and whether the national data services and local institutions are providing adequate collections. The findings also seem to undermine the notion that anything needed can be obtained freely on the Internet. Financial and company datasets, for example, are profitable information commodities, which require substantial academic discounts or subsidies to be affordable.

Would the national data services be more widely used if they were providing relevant collections to teaching departments? A closer look at the barriers to use of the national data services uncovers deeper issues than just ensuring that available sources exist.
Barriers to Using Datasets in Teaching

Those 46 respondents who were familiar with the national data services (one-quarter of those who teach with data) were asked to rank eight factors they thought might act as barriers in using national data services for learning and teaching purposes. Table 2 shows the median score for each barrier, in descending order, and also the mean score. The two top-rated barriers were “lack of awareness of relevant materials,” and “lack of sufficient time for preparation.” This issue was highlighted in a separate question, in which 57 percent agreed on the need to update /refresh /revise datasets used for teaching, but 29 percent had insufficient time to do so. The third greatest barrier was “registration procedures” [of the national data services]. However, the other barriers received high enough scores to also be considered seriously: namely, difficult data extraction interfaces, unsuitable file formats, inadequate dataset documentation, and lack of tailored teaching subsets.

<table>
<thead>
<tr>
<th>Barriers</th>
<th>Mean Score</th>
<th>Median Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of awareness of materials</td>
<td>6.5</td>
<td>7</td>
</tr>
<tr>
<td>Lack of time for preparation</td>
<td>6.4</td>
<td>7</td>
</tr>
<tr>
<td>Registration procedures</td>
<td>5.6</td>
<td>6</td>
</tr>
<tr>
<td>Interface</td>
<td>5.0</td>
<td>5</td>
</tr>
<tr>
<td>Format of datasets</td>
<td>4.8</td>
<td>5</td>
</tr>
<tr>
<td>Documentation</td>
<td>4.6</td>
<td>5</td>
</tr>
<tr>
<td>Lack of teaching subsets</td>
<td>4.4</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 2: Average ranking of barriers (8=highest score, 1=lowest).

In an open-ended question, users were asked for positive changes the national services could make to support teachers and learners in the use of datasets. Thirty-six out of 46 eligible respondents answered the question with a variety of useful suggestions. Answers were grouped into the following four categories (with examples of actual responses):

- **Easier access** - Able to get data without learning special software.

- **Simple registration for students** — “Make registration procedures simple and abolish restrictions on use (e.g. all students signing disclaimers).”

- **Create relevant and interesting teaching datasets** — “Rapid access to key summary economic data in form tailored for teaching.”

- **Effective publicity** — “The initiative needs to come from the National Services but better publicity would be a start.”

**Support Issues**

Prior to the survey, only anecdotal evidence was available to determine how teachers obtained support for classroom use of datasets. Members of the Task Force were familiar with the common reality of peer support for data use in both research and teaching via word-of-mouth. One member was aware that he was considered to be “the data guy” in the department, to whom others came for support. Although two data librarians were involved in the project, specialised data libraries and data librarians are not common in UK universities. Site representatives for the national data services can be based in the library, computing service, or elsewhere in an institution, but it was not known how much support they actually provide to users.

To provide a baseline measure on this issue, the survey asked each respondent, “From whom have you ever had support in obtaining or using data, whether for teaching or for research?” Of those who responded, more than a third (37 percent) had received no support at all. More than one source could be ticked; the average number of sources of support received was two. Peer support was the most common form, either from a project co-worker/assistant or another colleague (26 percent and 47 percent, respectively). The local computing service (26 percent) was roughly matched with the local library service (23 percent), which had helped about a quarter of respondents each. National service staff provided help to 10 percent of respondents, and their local site representatives only helped 7 percent of
them.

As an indicator of the satisfaction level with this status quo, users were asked to characterise the level of data support provided in their institution. The results are shown in Chart 4. Notably, only 14 percent agreed that local support was “very good across the board.” The majority, 62 percent, felt that support “tends to be ad-hoc.”

To follow this up, the survey instrument anticipated a number of local support activities and asked respondents to tick all “forms of locally provided support needed by academic data users.” Those who responded to this question (162 or 79 percent of total) reinforced the need for a number of forms of locally provided support, above all “Data discovery / locating sources” (66 percent). All of the answers shown in Chart 5 received “votes” from between one-third and two-thirds of those responding. The average number of needs ticked was three.

An open-ended follow-up question tended to reinforce the forms of support suggested in the questionnaire, although a significant minority felt that no additional support was needed, or expressed concern about where the resources would come from.

### Recommendations

The Task Force and the project team provided the following recommendations to the JISC (project funder) at the close of the project. Further elaboration may be found on the project Web site.

1. A broad initiative is recommended to promote subject-based statistical literacy for students, coupled with tangible support for academic teaching staff who wish to incorporate empirical data into substantive courses.

2. The development of high-quality teaching materials for major UK datasets needs to be funded adequately, in order to provide salience to subject matter and demonstrate relevant methods for coursework.

3. The national data services need to improve the usability of their datasets for learning and teaching.

4. A more concerted and co-ordinated promotion of the national data services should then follow, which is responsive to user demand.

5. Universities should develop IT strategies that include data services and support for staff and students, and integration of empirical datasets into learning technologies.

### Conclusion

UK higher education is undergoing many changes. The renewed attention to “learning and teaching” is an impetus for change in university teaching practices. Advances in information technology are creating new spaces for learning beyond the traditional classroom, and forms of teaching beyond the traditional lecture. Yet the pressures on academic staff who are still rewarded primarily for research rather than innovative teaching are great. To ensure that statistical literacy is taught effectively, new products and resources must be developed and adequate levels of support and technology provided.

1  With acknowledgments to the project team: Peter Burnhill (Project Director), Melanie Wright, Sean Townsend; Joan Fairgrieve for statistical analysis; and the Task Force on the Use of Numeric Data in Learning and Teaching. For membership see http://datalib.ed.ac.uk/projects/datateach/participants.html


The Virtual Training Suite: Internet Skills for Teaching and Learning

Introduction.
This paper describes the development of the RDN Virtual Training Suite which aims to support lecturers, students and researchers in finding and using resources on the Internet. It will provide an overview of the aims of the Virtual Training Suite and its content, focusing specifically on the social science related tutorials, which are available through SOSIG (The Social Science Information Gateway). It then gives some evidence of the way in which the tutorials are currently being used drawing upon the results of a recent evaluation study conducted by the University of Bristol and practical examples of its incorporation into the teaching and learning experience at the London School of Economics.

What is the RDN Virtual Training Suite?
The Virtual Training Suite is a series of 40 free Web based Internet tutorials which have been funded by the Joint Information Systems Committee (JISC), under their Distributed National Electronic Resource (DNER) programme, on behalf of Higher Education Funding Councils of England, Scotland and Wales. The tutorials were created by the Institute for Learning and Research Technology, University of Bristol with input from 30 universities, museums and research organisations across the UK. The first phase of the project was completed in July 2000 with the launch of the first 11 tutorials which included titles covering the physical sciences, social sciences and the humanities. A further 29 were launched in May 2001. Again these encompassed subjects from a range of disciplines, including engineering, statistics, government and social welfare.

The tutorials aim to cover a wide range of the academic subjects taught in UK universities and colleges. They also support the work of the Resource Discovery Network (RDN). This is a national Internet search service, which is being created, for academics and researchers based in British higher and further education institutions. It is composed of a co-operative network with a central organisation called the Resource Discovery Network Centre and a number of subject based independent service providers called ‘hubs’. There are currently 5 hubs. Each has responsibility for selecting, cataloguing and classifying Internet resources in a particular subject field. BIOME – health and life sciences; EEVL – engineering, mathematics and computing; Humbul – humanities; PSIgate – physical sciences and SOSIG – social sciences, business and law.

What are the aims of the Virtual Training Suite?
The Virtual Training Suite aims to offer a subject based introduction to locating and using Internet resources. Each tutorial addresses the needs of a particular subject community; taking them directly to the most important Internet sites and offering subject specific guidance on what they need to know.

It provides a flexible learning experience. All material is Web based and can be accessed from anywhere at any time of the day or night! Therefore users can directly control the pace at which they learn.

The tutorials are designed for use by both academic staff and students. They include specialist sections for both these categories of user in which guidance is geared to their particular needs. The student sections include information on finding materials for essays and citation style guides. The lecturer resource section contains tips on tracing online course materials and syllabi; case studies of how teachers have used the Internet and the facility to download the tutorial or print off posters for use in handouts.

The tutorials aim to provide a structured learning environment with clearly defined learning objectives. After completion of the appropriate tutorial, users should be:

- Aware of the range of types of material that can be found on the Internet and how these can be used to support their work
- Be able to identify the key Internet sites and resources for their subject area; know how to use effectively the main tools and techniques for Internet searching.
- Be able to critically evaluate the Internet resources that they find.
- The tutorials also seek to promote awareness and
effective use of the information gateways, which are being created by the RDN.

**How was the Virtual Training Suite created?**
The Virtual Training Suite was created by a consortium of authors from UK university libraries, academic departments, research bodies and organisations led by the Institute of Learning and Research Technology, University of Bristol. Contributors included: the National Institute for Social Work (Internet Social Worker); University of Manchester (Internet Anthropologist), Edinburgh University Data Library (Internet for Social Statistics) and the Data Archive, University of Essex (Internet for Social Research Methods). This collaboration was able to draw upon the existing expertise of subject specialists. They were able to identify key issues of concern for their subject community and gear examples towards them. For example, the Internet Medic identifies out of date material as a problem with many health related Internet sites and provides guidance to users on how they might check the currency. Close contact with the academic community also enabled testing and provision of qualitative feedback on the structure and content of individual tutorials.

**What do the tutorials contain?**
All the tutorials were constructed using CALnet software developed in house by the Institute for Learning and Research Technology at the University of Bristol. They share a common framework of 4 sections, which clearly structure the learning experience. Users can choose to follow the whole tutorial sequentially or select the section most suited to their individual needs.

- **Tour** – The first section entitled TOUR provides a guided tour of key Internet resources for the subject area. This highlights the range of materials available and directs users to the most important sites. For example, the Internet for Lawyers includes references to primary legal materials (legislation, treaties, law reports and judgements); secondary resources (journal articles, case commentaries and textbooks); Finding tools (indexes to legislation, library catalogues and directories); organisational homepages (professional bodies, law forms, government departments and law departments in universities); statistical data and teaching materials (lecture notes and syllabi). Provision has been made for updating these entries to take account of changes in URLs and other alterations in content.

- **Discover** – The second DISCOVER section introduces the user to techniques for effective Internet searching. It includes a comparison of the strengths and weaknesses of commercial search engines and information gateways with guidance on when it is most appropriate to use each. This section also has a particular aim in helping students to use the RDN hubs effectively. For instance the Internet Politician provides examples of how to use the advanced search form on SOSIG to truncate search terms and restrict searches to particular resource types.

- **Review** – the third section is entitled REVIEW. This teaches skills for critically evaluating the quality of Internet sites. This is an area of particular importance as the lack of quality control on the Internet means that users must be careful to assess the value and authenticity of sources before they use them in their work. The tutorials use case study scenarios to highlight common pitfalls and offer tips on how to begin to assess quality. The process is broken down into a series of simple questions which students can ask relating to Who? Where? And Why? the resource was placed on the Internet. You might want to take a look at the Internet Sociologist for an entertaining and effective example. This uses a site called ‘Kill the Television’ to discuss issues relating to bias on the Internet. It provides guidance to users on how they might check the authenticity of an author and look for more information on his/her motives in placing a resource on the Internet.

- **Reflect** – The final REFLECT section summarises the skills taught in the tutorial and provides case studies of how students, researchers and lecturers might incorporate usage of the Internet into their working practices. A good example is provided in the Internet Aviator where the undergraduate scenario tells the story of Ryan Ayre who is looking for material about the military usages of unmanned aircraft for his essay. It takes him through the stages of research, showing how he can use the Internet to find video footage, technical reports and relevant news items on the Internet. The tone is light hearted but it does communicate useful lessons about the range of material that can be found on the Internet and guidance on items which are currently available in paper only.

- **Links Basket** - Another notable feature of the tutorials is the links basket. As they move through the tutorials users can collect useful URLs in their ’shopping basket’. At the end they can pick these up and save them as a series of useable bookmarks.

- **Quizzes and practical exercises** also lighten the learning experience and offer students the chance to test their understanding of the material. Each section contains a selection of optional multiple choice, fill the gaps or more open ended questions.

- **Additional Features** – These are geared towards student needs and include a glossary of commonly used Internet terms and guides to citing Internet resources in essays.

- **Resources for Teachers** - The second phase of the tutorials introduced a supporting “Resources for
Teachers” section. This includes an introductory Powerpoint presentation, student workbook and handout and lesson plans. Teachers can use the “Print/Download” option to print out or download the whole tutorial or specific chapters within it. These can then be used as slides or handouts. Free posters may also be printed from the site.

**How is the Virtual Training Suite being used?**

The first independent academic evaluation of the usage and value of the Virtual Training Suite was completed by Lin Amber of the University of Bristol Information Services in March 2001. This provides both quantitative and qualitative data on initial usage.

In terms of qualitative statistics, initial usage has been encouragingly high. From March-December 2000 there were over 43,000 log ins to the site, with an average of 204 sessions per day. Interest in the future development of the project remained high as over 2,000 people signed up to receive notification of the launch of the second phase of tutorials. This means that usage is likely to rise further in the future. One of the most popular options was to download the tutorial (739 sessions recorded) and to print off posters (1473 requests). These trends were noted by the project organisers and as a result these features were incorporated into the second phase of the project and advertised more widely.

Analysis was also made from a total of 122 online feedback forms completed between December 2000 and March 2001. Although the number of respondents only represents a small percentage of the total number of users, it does provide some interesting information on the type of users of the tutorials and their opinions of the content. 25% of the users were librarians seeking material for inclusion in user education sessions, 14% lecturers, 11% undergraduates, 8% researchers, 10% post graduates. The rest fell into other categories such as school students. 56% of users classified themselves as independent learners. This shows that the tutorials are reaching a wide audience, representing all the academic departments. These are large in scale and often take place in lecture theatres seating over 100.

Feedback remains an ongoing process. Users are encouraged to complete on-line forms with their comments and these have now been highlighted more prominently to encourage greater response.

**Why is the Virtual Training Suite being used at the London School of Economics?**

The paper will now conclude by offering some qualitative examples of why the tutorials are currently being used by Library staff at the London School of Economics (LSE).

The LSE is a renowned teaching and research institution for the social sciences with a student body of over 6,000, of whom 93% are studying for post-graduate degrees. There are currently over 800 part time students. Usage of the tutorials was particularly attractive to us for a number of reasons.

- **The wide subject coverage offered.** The 40 packages match a large number of the areas, which are taught and researched by School staff and students. Subject based material was regarded by the users as more relevant than general courses as there was a feeling that it saved time by taking them directly to what they needed to know and avoided ‘irrelevant IT terminology’. The Virtual Training Suite also covers a number of subject areas, such as medicine, which have a marginal interest to LSE researchers. The paper based library collections do not cover these areas comprehensively so information about the most important Internet resources was welcome.

- **The flexibility of the packages.** Free Internet access made it possible for part time students and students on fieldwork placements who could not attend timetabled library training sessions to be offered some training. It also enabled large numbers of students to receive training, avoiding the problems of restrictions on numbers in training rooms.

- **The downloadable materials and posters.** These provided a readily-usable resource for preparing user education materials which was welcomed by library staff who were under pressure to provide more training for growing numbers of students.

**How is the Virtual Training Suite being used by the London School of Economics Library?**

The range and flexibility of the package means that it has been possible to actively incorporate it into the learning programme in a number of ways:

- **Large scale student inductions.** At the start of each academic year subject liaison librarians give presentations for new lecturers and students in particular academic departments. These are large in scale and often take place in lecture theatres seating over 100. They are intended as overviews to the services offered. Slides have been taken from the tutorials and incorporated into PowerPoint presentations to demonstrate the type of training courses the library offers.
• Hands on workshops. Liaison librarians regularly offer small workshops for research and masters students from their departments. These sessions provide hands on teaching for 10-15 individuals at a time. The tutorials have been incorporated into some of these as a useful resource for helping novice students to find out what is available on the Internet for their subject area and improving the efficiency of their Internet searching skills.

• Drop in sessions. In addition to targeted workshops. The library also offers regular drop in Internet training sessions of around 90 minutes throughout term time. These are open to all staff and students without booking. The trainer usually demonstrates search techniques and important sites and then the students are able to engage in hands on exercises. As part of these sessions students are encouraged to work through tutorials as well as notes and materials prepared by more general Internet courses such as the TONIC Netskills package. Trainers have found that while students are working independently at their own pace they have been able to offer a supportive atmosphere, which encourages learning.

• Training sessions for library staff. All new information desk staff at the LSE receive training sessions on Internet skills so that they can assist users in finding information. There is also a regular ongoing programme of refresher training sessions to remind established staff of new services. The tutorials have been used effectively in these. Staff have been encouraged to explore the tutorials and to note key sites in the TOUR sections as these often contain specialist subject directories or gateways which are of value in providing starting points for research. They are particularly useful in directing staff who do not have a wide experience of the subject area.

• Inclusion in handouts. The library produces a number of guides for staff and students on electronic services. An example of these are quick reference guides for particular subject which are intended to offer an overview of starting points for research in the subject area. They contain a list of the main classmarks for the subject area, lists of locations of important journals and information on key electronic resources such as CD-ROMs, databases and Internet sites. The Virtual Training Suite tutorials have been listed in the appropriate guides as a good starting point for beginning Internet research.

• Links on the library Web pages. As the Virtual Training Suite is intended as an independent learning package, links to it have been placed on the library training Web pages. This has encouraged us to begin a process of redeveloping the information that we offer there in order to provide more full-text, self-contained training materials. The library is involved in a collaborative project with the Teaching and Learning Technology Project based in the academic section of the LSE. This aims to develop online materials to cater more effectively for distance and part time learners who cannot attend formal training sessions. The first stage of the project has been the development of a package to train users how to search the library catalogue effectively. We also aim to develop more information on local fee based Internet services. This will supplement the information contained within the Virtual Training Suite.

Future Developments and Maintenance
The second phase of the tutorials was launched on May 8th by Michael Wills, Parliamentary Under-Secretary of State for Education and Technology via a live link up between simultaneous launch events held at the six universities of the RDN hubs. Already users are asking if the suite will be expanded. They have pointed out gaps in subject coverage, such as Art and Design, Music and Archaeology and asked for courses in these areas. The project team has petitioned the funders to see if more money can be made available to expand coverage. They are currently awaiting a response from them. In the meantime any suggestions for new subjects to be covered are welcome, although no promises can be made at this stage. Decisions are likely to be based on the response by users to the resource now that the full compliment of tutorials has gone live.

The other main query about the Virtual Training Suite is how it will be kept up to date. This is clearly an important issue where Internet resources are concerned. As a result the resource has been firmly embedded into the organisational structure of its parent service. The tutorials are all hosted by the relevant RDN hub, so for example all the social science tutorials are hosted by SOSIG (the Social Science Information Gateway). The hubs will be responsible for the upkeep of their set of tutorials by maintaining regular link checking procedures. Tutorial authors may also suggest edits to take account of new developments.

Ultimately we hope that word will spread about this new educational resource, and that it will enhance people’s experience of using the Internet as a source of information, raising the skills level of academics and students nationwide.

Footnotes
1 The Virtual Training Suite is located at: http://www.vts.rdn.ac.uk/
2 Further information on JISC is located at: http://www.jisc.ac.uk/
3 HEFCE homepage is located at http://www.hefce.ac.uk/
Further information can be found at: http://www.ilrt.bris.ac.uk/

The first eleven tutorials were: Internet Medic, Internet Lawyer, Internet Politician, Internet Psychologist, Internet Social Worker, Internet Sociologist, Internet Economist, Internet Manager, Internet for History, Internet for English, Internet Aviator.

Social Sciences
Internet Anthropologist
Internet for Development
Internet for Education
Internet Geographer
Internet for Government
Internet for Social Policy
Internet for Social Research Methods
Internet for Social Statistics
Internet for Women’s Studies
Reference
Internet Instructor

Humanities
Internet for History and Philosophy of Science
Internet for Modern Languages
Internet Philosopher
Internet for Religious Studies
Internet Theologian

Health and Life Sciences
Internet for Agriculture, Food and Forestry
Internet Bioresearcher
Internet for Nature
Internet Vet
Physical Sciences
Internet Chemist
Internet Earth Scientist
Internet Physicist

Engineering and Mathematics
Internet Civil Engineer
Internet Electrical Engineer
Internet for Health and Safety
Internet Materials Engineer
Internet Mathematician
Internet Mechanical Engineer
Internet Offshore Engineer

The RDN homepages are located at: http://www.rdn.ac.uk/

Further information on this can be obtained from: http://www.webecon.bris.ac.uk/calnet/

See http://www.sosig.ac.uk/vts/sociologist/index.htm

See http://www.eevl.ac.uk/vts/aviator/index.htm

The full text can be viewed online from the Virtual Training Suite Pages http://www.vts.rdn.ac.uk/

For further information on TONIC see http://www.netskills.ac.uk/TonicNG/cgi/sesame?ng

See http://www.library.lse.ac.uk/infoskills/catalogue/

* Paper presented at the IASSIST/IFDO Conference 2001, Amsterdam. Heather Dawson, British Library of Political and Economic Science, h.dawson@lse.ac.uk. With the assistance of Debra Hiom, Institute for Learning and Research Technology, University of Bristol, d.hiom@bristol.ac.uk and Emma Place, Institute for Learning and Research Technology, University of Bristol, Emma.place@bristol.ac.uk
MISSION (Multi-Agent Integration of Shared Statistical Information Over the [inter]Net) -The data archive perspective

by Joanne Lamb

Abstract
MISSION [1] is a multi-national project funded by the European Commission. It aims to provide a modular system of software that will enable providers of official statistics to publish their data in a unified, and unifying, framework, and to allow consumers of statistics to access these data in an informed manner with minimum effort. The objective of the project is to develop an integrated set of software modules, which will

- Allow suppliers of statistics to subscribe to an integrated network of datastores via an interface to their existing data while retaining control over all aspects of access to their data: their level of involvement; the data they supply; the users who can access it; and the level of resources to commit.

- Allow users to make declarative requests, with a minimum of understanding of statistics, or the domain area, and still retrieve meaningful results from our internal routines or through an interface with external statistical packages.

- Give the user a range of options for automatic harmonisation of statistical data, with clear indication on the interpretation of the results.

- Provide audit trails of data manipulation and analysis, so that methods can be retained, re-used and published.

- Maintain libraries of metadata that can be made available to other users.

- Provide a flexible architecture that allows third parties to act as Independent Metadata Providers, thus encouraging the free exchange of knowledge.

- Allow users to build up individual profiles, accessing data and methods most relevant to their needs.

- Offer a number of independent, interoperable systems that can run on different hardware platforms and access heterogeneous data storage systems.

MISSION is a development of a fourth framework project, ADDSIA [2], but it brings a number of new initiatives to the basic ideas of that project. These are:

- The use of agent technology to optimise queries;

- The use of the Unified Modelling language (UML) in designing the system;

- The development of the concept of Metadata Libraries that are independent of data sources, and which provide a middleman service to the user;

- Tools to enable expert users to develop and share their methodology.

The MISSION Project
MISSION is a European Union R&D project, number- IST-1999-10655. The project started in January 2000 and is due to finish in December 2002. We are therefore halfway through the project. MISSION grew out of the ADDSIA project and has the same partners, who are:

University of Edinburgh, Scotland, UK (co-ordinator)
Office for National Statistics, UK
Central Statistics Office, Ireland
Tilastokeskus (Statistics Finland), Finland
University of Athens, Greece
University of Ulster, Northern Ireland, UK
Desan Marktonderzoek BV, The Netherlands

The objectives of the project can be summarised as follows: We aim to build a software suite that will allow statistical data providers to integrate their publication of data on the Web. This software will have a number of features, based on the requirements of data suppliers and data users.

For suppliers, the system will allow them to subscribe to an integrated network of datastores via an interface to their existing data. They will be able to retain control over all aspects of access to their existing data.
Data Users will be able to make requests in a declarative manner, with a minimum of understanding of statistics, or the domain area, and still retrieve meaningful results. The users will be able to tailor their environment, from simple requests to detailed in-depth analysis. They will also be able to build up individual profiles, accessing the data and methods most relevant to their needs.

A key objective is to allow methods of data manipulation and analysis to be retained, re-used and published. This will be done using Libraries of metadata and of tables, both of which can be developed in the MISSION system and then published for re-use by other users. We have in fact separated the functions of data providers and metadata providers. This approach gives Mission a very flexible architecture, which will allow third parties to act as Independent Metadata Providers.

The MISSION system will be implemented on independent, interoperable systems running on different hardware platforms and accessing heterogeneous data storage systems.

The core of the MISSION approach is the notion of allowing a user to form a query over several datasets, using a centralised statistical engine, which would parse the request and send partial queries to different datasets held in the system. This was the key concept of ADDSIA, which has been further developed in MISSION.

To achieve this we identify three basic concepts: the Client, the Library and the Data Server. The Client is a piece of software that can be downloaded from a MISSION site and installed on the user’s machine. It connects to a host MISSION site, and offers both the user interface and the user’s workspace.

The Library is the core of the MISSION system. It is a repository for different types of statistical metadata: access, methodological and contextual. It also contains the statistical processing engine. Libraries communicate with each other via agents, and therefore once a user is connected to a host library, he or she potentially has access to all MISSION Libraries.

The Data server is the unit that gives access to the data. It provides the link from the data to the library. Data servers register with libraries, and then register their datasets. The access rights components of the data server enables the data suppliers to specify who can access which parts of their data.

Only aggregated data is sent to the Library. When a request is sent to the Data Server, the result of that query is computed, and this aggregated result is sent to the library. In this way, sensitive micro data is not sent outside the Data supplier’s site, and also the amount of data transferred over the Internet is reduced, thus making the retrieval efficient.

Figure 1 gives the overall picture of the MISSION system.

**Innovation in MISSION**

MISSION has extended this basic idea in a variety of ways. First, we are using agent technology to enhance the system in a number of places. In the formulation of a query, the user is presented with logical variable names and descriptions, and agents will search a number of libraries to discover datasets containing these logical variables. The agents will then process the metadata for these datasets and can use a number of techniques to optimise the user’s request. These agents first construct the correct query, and then plan the execution of that query. Further agents connect with the data.
servers. This use of agents is illustrated in Figure 2, where the first part of the diagram shows the query planning stage, and the second part shows the query execution phase.

Second, we have designed a graphical user interface, which will allow the user to specify and format his query using a table template to build up the query. This graphical specification is translated into the internal query language, and the results of the query populated the table frame that the user has built up. There will also be opportunities for the user to ‘post process’ this query, by graphically specifying the modifications, as shown in Figure 3.

A third aspect worth mentioning is the development of the MISSION model of metadata. Figure 4 shows the MIMAMED data model. MIMAMED stands for MicroMacroMetadata model, and is a model designed to
describe all three types of data, and to process the data and metadata simultaneously. In this model, developed by the University of Ulster [4], the ‘sunkey table’ is a reference table for a set of particular aggregated data – the result of a query. The Summary table is the actual data, all other tables contain metadata, which, when aggregates are combined, will also be processed. In addition, we have demonstrated that this model maps to other common models, such as data warehouse structures, cubes and the CRISTAL [5] model.

Finally, the MISSION project is considering how the system will be made available at the end of the project. We have made an in principle commitment to Open Source [6] publishing of the source code, subject to this being a legal option in the setting of a European Union R&D project.

Process to date
The aspects of the system described above will feature in the first prototype, due in September 2001. In February, we successfully tested the connectivity of the system, with a Client in Ulster linking to a Library in Edinburgh, which queried three Data Servers in Athens (running on different operating systems and relational databases). We have also completed the Data Server installation package. In May 2001, we repeated the experiment with a more complete query processing, and are about to complete the Library installation package and supply the demonstration sites with these test versions.

With the first prototype in September, the user will be able to:

will enable the expert user to share methods and publish results via the system. When looking at the requirements of the system for these features, we need to consider three different aspects:

- The model for handling and displaying ‘transformations’;
- Storing MISSION tables for publishing;
- More contextual metadata.

For the first point, we will build a metadata model developed round the transformations that we have identified in Figure 5. However, practically this will be stored as XML, for two reasons:

- First, it is a principle that all metadata held in the Client workspace will not require any software except a Java environment – so the data cannot be held in a database.
- Second, we are developing a metadata user interface, which will allow the user to search any XML file.

This metadata browsing tool will give us the ability to exploit XML files from different sources. We are particularly interested in the development of the DDI for aggregated data, and also in the standards table DTD.

For importing data files, PC-AXIS, and SPSS will be the first formats that we support. We are considering which other formats are likely to be in demand, and expect to see a demand for DDI in future.
The concentration for prototype 1 has been on the downloaded Client software. Users register to the Library, and have privileges to access certain datasets according to the data suppliers’ stipulation. Thus control of the use of data is left with the supplier. The degree to which data is confidential is also at the suppliers’ discretion, and we will advise caution in this area. While a single request may be easy to monitor, tracking a series of requests that may lead to disclosure is more difficult.

In contrast, the public user – i.e. a user who accesses the system via the web, without registration – has no direct access to data. Access to metadata is freely available, and also to pre-defined tables which are created through MISSION. Therefore the amount of information accessible to the public depends on the number of tables published in a library.

The second prototype is scheduled for March 2002, and the final version for December 2002.

Implications for Data Archives
We have depicted the MISSION Library as a separate entity from the Data Server, and we picture the two different modules being run by different organisations that have different expertise. The inspiration for the Independent Metadata Provider scenario came from the way in which social science currently uses quantitative data. Typically a researcher will get data from an archive, which will be well documented for secondary analysis. However, after a two or three year project, the findings are published in theoretical papers, and the modified data is destroyed for legal or economic reasons. The researcher may not legally be permitted to keep the modified data, since this would be for a purpose other than that of the original project. Alternatively, there may be a financial cost for using the data for another reason. If the researcher cannot keep the modified data for his own purposes, it is even more difficult for another researcher to pick up on this work and continue.

It has been observed that the manipulation of data prior to analysis encompasses the hypotheses of the research [7]. It is therefore important to capture not only the algorithm of the transformation, but also the reasoning behind it. If this can be presented to the analyst as a tool for aiding his own work, then the overhead of supplying this metadata is not seen as arduous. Once this reasoning has been captured, it is available to give a reasoned method for other to use.

The wider context
This section places MISSION in the wider context of statistical information systems research in CES and in relation to other EU initiatives in which we are involved.

The CES is participating in two R&D projects and three networking projects. While MISSION is concerned with data dissemination, the other project, IQML [8] is concerned with data collection.

The three networking projects will be described briefly. MetaNet - A network of excellence for harmonising and synthesising the development of statistical metadata - started in November 2000, and held its first conference in April 2001. The proceedings from the conference are due at the end of May, and will be available from the website [9].

AMRADS - Accompanying Measure to R&D in Statistics – is a project concerning issues of technology transfer from R&D projects to national statistical institutes, and between national statistical institutes. While the focus of this project is on official statistics, the issues it addresses – the transfer of new ideas and technology between research and services institutes – also has relevance for data archives. The project has six themes led by experts in the area, and CES is responsible for metadata. The first significant event in this project will be the ETK/NTTS2001 conference in Crete in June 2001, where AMRADS has had significant input to the programme.

COSMOS, a Cluster Of Systems of Metadata for Official Statistics, will start in September 2001. Clusters are specifically for EU fifth framework projects, to encourage the sharing of knowledge during the lifetime of the projects. In COSMOS,
lead by CES, we have five projects, whose acronyms are:

MISSION (see above)
IQML A Software Suite and Extended Mark-Up Language [XML] Standard for Intelligent Questionnaires
FASTER Flexible Access to Statistics, Tables and Electronic Resources
Metaware Statistical Metadata Support for Data Warehouses
IPIS Integration of Public Information Systems and Statistical

Details of the projects can be obtained from the Information Society Technology website [10]. Figure 6 shows the relationship between the projects. The two R&D projects are at the bottom of a hierarchy of generalisation. These feed into COSMOS, where the objective is to demonstrate interoperability between some components of the five projects. They, and COSMOS, feed into MetaNet, which is looking to build a conceptual framework of statistical metadata. Finally, MetaNet and metadata form one strand in the general investigation of issues concerning technology transfer from R&D in official statistics.

The expertise of librarians and archivists is complementary to that of producers of (official) statistics. We expect to input the metadata from providers automatically, and are keen to utilise as many standards of metadata that it is feasible to handle.

Finally, MISSION is also contributing to other activities aimed at getting a shared understanding of statistical metadata needed for the processing, documentation and preservation of statistical data using modern technology.

References
2. http://www.ed.ac.uk/~addsia


Conclusions
In conclusion, we would like to emphasise the following points. First, MISSION is an open system, which aims to help users exchange methodologies as well as utilise data from different sources. This gives the opportunity for data archives and data libraries to host metadata sites as well as access to data. We deliberately called these sites Libraries, since we feel that much of the metadata knowledge is held at the servicing level, rather than at the data provision level.
Research for Building a Better Data Community

I have come to believe that IASSIST members must seriously consider the value of conducting research about our own profession and field. If we do not initiate and value such research, the expectation that anyone else will undertake this task for us is unrealistic. I am not necessarily referring to “theory-driven” research, although this would be welcomed. Rather, I have in mind “issue-driven” research, that is, the kind of research that helps us understand the relationships, norms, and behaviours within our information and science cultures, including our own data subculture. For example, we need to conduct research on the issues behind data preservation and access. I am not talking about “how” to preserve data or provide access but instead investigating the norms and behaviours underlying the activities of data preservation and access.

A few recent events have led me to this new conviction. First, I was asked a couple of years ago by a major research council to review a grant application in which the principal investigators were proposing to study the economics of archiving data. In my enthusiastic endorsement of this application, I wrote that the principal investigators should expand their scope to explore the “data economy”, which I characterized as who gets access to which data, when and how. I thought that this would be a wonderful project for our profession. We would have research conducted about data archives and their role in the data economy. Here data archives would be the object of research rather than the sources of data for research. Unfortunately, the project was not funded. Nevertheless, it did stimulate my thinking about the possibilities of such research.

The next recent event took place at the 2000 IASSIST conference. I was excited by the research carried out by Karsten Boye Rasmussen and Repke de Vries about IASSIST as a virtual community. Their use of the IASSIST e-mail discussion list and the log files for the Web site and on-line issues of the IASSIST Quarterly clearly demonstrated ways of doing research about our organization and profession. A third and even more recent experience has been the research that I have been conducting in conjunction with the consultation underway in Canada about creating a national data archive.

Before talking about specific research findings, let me briefly describe through an example one way in which “issue-driven” research might be performed in our field. One challenge we face, which has not changed over the last forty years, is how to get researchers to think about archiving their data at the beginning of a project rather than at the conclusion of their research. And more than just thinking about archiving data earlier, how do we get researchers to conduct their projects so that their data products meet archival standards – rather than having to build an archival data product after the research has concluded. In other words, how do we mainstream data archiving in the research process? This is neither a new nor novel idea. However, is it an idea whose time has arrived? If our profession better understands the dynamics of current research practices that inhibit data archiving, can we be instrumental in bringing about the necessary changes to mainstream data archiving?

I am currently a co-investigator on a nationally funded project studying research utilization in the field of nursing. Specifically, we are studying how practitioners eventually use medical research findings about pain and pain control. How do research outcomes end up in practice? I introduced the principal investigator of this project to the value of data sharing and stewardship when she was a graduate student several years ago.1 I also helped her preserve the research data from her dissertation. Subsequently, she has become a successful researcher who is a key proponent of data preservation and data sharing in Nursing research. This is how I came to be invited as a member of her team.

Out of this project, I was the lead author of a paper presented at the 2000 Conference on Social Science Methodology in Cologne, Germany entitled, “Archivist on board: contributions to the research team”.2 This paper presented the role and value of an archivist on the research team. While I was not at the conference to present the paper, a colleague and co-author read it on my behalf. Ekkehard Mochmann was present at this session and came to the rescue of my colleague when one researcher, after hearing the presentation, protested that we were trying to turn researchers into archivists. My colleague reported that Ekkehard explained to the researcher that what we were proposing was to initiate partnerships between researchers

by Charles K. Humphrey*
and archivists. When first hearing of this exchange, I was struck by the immediate reaction of the researcher. What were the attitudes and values behind such a response? Why were the attitudes of this researcher so seemingly different from those of the principal investigator with whom I work?

Turning to some findings from the research we are conducting in conjunction with the national data archive consultation, some insights into attitudinal differences underlying the principles of archiving data can be found. One of the four surveys that were administered was a sample of researchers who received a grant from the Canadian Social Sciences and Humanities Research Council between 1998 and 2000. One objective of this survey was to identify the number of researchers who produce data products as part of their research and to determine how many researchers have ever archived or intend to archive their data. A second objective was to investigate researchers’ attitudes about data sharing and archiving.

Eleven items were used to gauge these attitudes. These questions touch upon the legitimacy of secondary analysis as a research method, on the value of data as a by-product of research, on the issues of data ownership and data sharing, on research council funding to prepare data for sharing, and on the impact that ethics review boards have on data sharing (see List 1). Because the wording of five items (5a, 5d, 5e, 5h, 5j) does not support the principles of data sharing or archiving, the response categories for these items were recoded to correspond with the direction that supports the archiving principle.

Figure 1 shows the combined percentage of respondents ‘agreeing’ or ‘strongly agreeing’ on each item. The items in this Figure have been arranged in decreasing order of support. Eighty-one percent agree that data should be a valued by-product of research, while only 21 percent agree that data do NOT belong to the researcher as her or his intellectual property. This decreasing order of agreement represents an increasing difficulty in support of sharing and preserving research data. Six steps of item-difficulty can be seen in this Figure. Eighty-one and 78 percent of the respondents accept the first two items, data as a valued by-product and secondary analysis as a valid research method, respectively. The second step consists of the items about research councils covering the costs to prepare data for sharing and

---

**LIST 1**

Attitudinal Items Underlying Support for Data Archiving

5a. Secondary data analysis is not a valid research method.
5b. Data should be considered a valued by-product of research.
5c. Data should be shared with other researchers, assuming it has been appropriately anonymized.
5d. Data belong to the principal investigator as her or his intellectual property.
5e. Data should only be shared if the principal investigator decides to share it.
5f. Archiving data should be an integral part of conducting research.
5g. Researchers who obtain information that cannot be easily reproduced from respondents are, to a degree, trustees of the data.
5h. Spending resources to prepare the data from my research so that other researchers can use it would be a waste.
5i. Research councils should include funds to cover the costs of preparing data for sharing.
5j. Ethics review boards make it impossible to share confidential data on human subjects.
5k. Ethics review boards need to be educated about the need to preserve data.

---

**FIGURE 1**

Attitudes Underlying Support for Data Archiving

<table>
<thead>
<tr>
<th>Item</th>
<th>Percent Agreeing or Strongly Agreeing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data valued by-product</td>
<td>81</td>
</tr>
<tr>
<td>Secondary analysis</td>
<td>78</td>
</tr>
<tr>
<td>Research councils fund</td>
<td>71</td>
</tr>
<tr>
<td>Researchers as trustees</td>
<td>68</td>
</tr>
<tr>
<td>Data should be shared</td>
<td>64</td>
</tr>
<tr>
<td>Educate ethics boards</td>
<td>52</td>
</tr>
<tr>
<td>Waste of funds to save</td>
<td>50</td>
</tr>
<tr>
<td>Archiving is integral</td>
<td>48</td>
</tr>
<tr>
<td>Ethics make a barrier</td>
<td>28</td>
</tr>
<tr>
<td>Share data regardless</td>
<td>27</td>
</tr>
<tr>
<td>Data don’t belong to PI</td>
<td>21</td>
</tr>
</tbody>
</table>

---

and archivists. When first hearing of this exchange, I was struck by the immediate reaction of the researcher. What were the attitudes and values behind such a response? Why were the attitudes of this researcher so seemingly different from those of the principal investigator with whom I work?

Turning to some findings from the research we are conducting in conjunction with the national data archive consultation, some insights into attitudinal differences underlying the principles of archiving data can be found. One of the four surveys that were administered was a sample of researchers who received a grant from the Canadian Social Sciences and Humanities Research Council between 1998 and 2000. One objective of this survey was to identify the number of researchers who produce data products as part of their research and to determine how many researchers have ever archived or intend to archive their data. A second objective was to investigate researchers’ attitudes about data sharing and archiving.

Eleven items were used to gauge these attitudes. These questions touch upon the legitimacy of secondary analysis as a research method, on the value of data as a by-product of research, on the issues of data ownership and data sharing, on research council funding to prepare data for sharing, and on the impact that ethics review boards have on data sharing (see List 1). Because the wording of five items (5a, 5d, 5e, 5h, 5j) does not support the principles of data sharing or archiving, the response categories for these items were recoded to correspond with the direction that supports the archiving principle.

Figure 1 shows the combined percentage of respondents ‘agreeing’ or ‘strongly agreeing’ on each item. The items in this Figure have been arranged in decreasing order of support. Eighty-one percent agree that data should be a valued by-product of research, while only 21 percent agree that data do NOT belong to the researcher as her or his intellectual property. This decreasing order of agreement represents an increasing difficulty in support of sharing and preserving research data. Six steps of item-difficulty can be seen in this Figure. Eighty-one and 78 percent of the respondents accept the first two items, data as a valued by-product and secondary analysis as a valid research method, respectively. The second step consists of the items about research councils covering the costs to prepare data for sharing and

---

**LIST 1**

Attitudinal Items Underlying Support for Data Archiving

5a. Secondary data analysis is not a valid research method.
5b. Data should be considered a valued by-product of research.
5c. Data should be shared with other researchers, assuming it has been appropriately anonymized.
5d. Data belong to the principal investigator as her or his intellectual property.
5e. Data should only be shared if the principal investigator decides to share it.
5f. Archiving data should be an integral part of conducting research.
5g. Researchers who obtain information that cannot be easily reproduced from respondents are, to a degree, trustees of the data.
5h. Spending resources to prepare the data from my research so that other researchers can use it would be a waste.
5i. Research councils should include funds to cover the costs of preparing data for sharing.
5j. Ethics review boards make it impossible to share confidential data on human subjects.
5k. Ethics review boards need to be educated about the need to preserve data.

---

**FIGURE 1**

Attitudes Underlying Support for Data Archiving

<table>
<thead>
<tr>
<th>Item</th>
<th>Percent Agreeing or Strongly Agreeing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data valued by-product</td>
<td>81</td>
</tr>
<tr>
<td>Secondary analysis</td>
<td>78</td>
</tr>
<tr>
<td>Research councils fund</td>
<td>71</td>
</tr>
<tr>
<td>Researchers as trustees</td>
<td>68</td>
</tr>
<tr>
<td>Data should be shared</td>
<td>64</td>
</tr>
<tr>
<td>Educate ethics boards</td>
<td>52</td>
</tr>
<tr>
<td>Waste of funds to save</td>
<td>50</td>
</tr>
<tr>
<td>Archiving is integral</td>
<td>48</td>
</tr>
<tr>
<td>Ethics make a barrier</td>
<td>28</td>
</tr>
<tr>
<td>Share data regardless</td>
<td>27</td>
</tr>
<tr>
<td>Data don’t belong to PI</td>
<td>21</td>
</tr>
</tbody>
</table>
about researchers serving as trustees of data that cannot be easily reproduced from respondents. Seventy-one and 68 percent endorsed these items, respectively. The third step is made up of the item stating that data should be shared if it has been properly anonymized (64 percent) and the item that ethics review boards need to be educated about the need to preserve data (62 percent).

A slightly larger step occurs with the next two items. Fifty percent agree that spending resources to prepare the data from their research would not be a waste and 48 percent agree with the statement that archiving data should be an integral part of conducting research. An even larger drop occurs with the fifth step. Twenty-eight percent disagree that ethics review boards make it impossible to share confidential data on human subjects, while 27 percent disagree that data should only be shared if the principal investigator decides to share it. As mentioned above, the smallest percentage of agreement (21 percent) is that data do NOT belong to the principal investigator as her or his intellectual property.

A scale was constructed based on the total number of items on which each respondent agreed with the principles of data archiving (see Figure 2). A score of zero indicates that the respondent did not support any of the items endorsing data archiving, whereas a score of 11 represents someone who supported all of the items. Seventeen percent were low supporters of data archiving (those with scores from zero to three), while 24 percent were high supporters (those with scores from eight to 11). Fifty-nine percent are in the middle. The correlation between this scale and a question asking how important it is for Canada to establish national services for the preservation of research data is 0.50, which is corroborative evidence that this scale measures some aspect of support for data archiving.

What do we make of these findings in light of the question asked earlier about how to mainstream data archiving in the research process? First, only around a quarter of Canadian researchers in this study appear to be strong advocates of archiving data. While only 17 percent seem to be protagonists, close to 60 percent are in the ambivalent middle. Apparently, this rather substantial group requires further education on the principles of archiving.

Secondly, one of the eleven items rather succinctly summarizes the notion of mainstreaming data archiving. This is the item that states, “Archiving data should be an integral part of conducting research.” Looking at the results of this item, 12 percent agreed strongly, 36 percent agreed, 30 percent were unsure, 18 percent disagreed, and 4 percent disagreed strongly. The 30 percent who are unsure is as alarming in this distribution as the 22 percent who disagree. One concern raised by this finding is that at least half of the researchers do not view data archiving as part of the normal practices of conducting research. The concept of archiving may be generally understood, but archiving as part of the research process has not become routine.

An explanation for these results may be directed at incomplete training of researchers in their graduate school years or at senior researchers who are not mentoring junior researchers about data archiving. Another possible explanation is the failure of our profession to promote archiving as part of the research process. If the importance of the practice is not taught as part of the research method, data archiving will not be discussed or perceived as a generally important activity.

Turning away from this specific example, I would like to conclude with a couple of observations about changing our thoughts in IASSIST toward research. First, our organiza-
tion is well positioned to conduct comparative, cross-national research. We are an international organization with opportunities to investigate the generalizability of national findings. Are the attitudes described above held only by Canadian researchers, or are these attitudes commonly found among researchers across societies? Will we discover underlying attitudes about archiving data that are held by researchers around the globe? These are challenges that we can undertake together. Secondly, we live in a world increasingly calling for evidence-based decision-making. We should approach the issues we face in our profession by building evidence through research. Not only will we be stewards of data, but we will be contributing to the knowledge about the research process.

Footnotes


3 “Strongly agree” and “agree” are the responses supportive of these principles.

IASSIST 2002
Accelerating Access: Enhancing Collaboration and Dissemination

Call for Papers
The International Association for Social Science Information Service and Technology (IASSIST) invites submissions for papers, panel discussions, poster/demonstration sessions and workshops for its 28th annual conference to be hosted by The Roper Center for Public Opinion Research and the Thomas J. Dodd Center at the University of Connecticut. Mark your calendar now and plan to join us June 11-15, 2002.

IASSIST conferences bring together data professionals, data producers and data analysts from around the world who are engaged in the creation, acquisition, processing, maintenance, distribution, preservation, and use of numeric social science data for research and instruction.

Conference Themes
Accelerating Access is a theme that characterizes the ever increasing demand for quantitative and qualitative data through the Internet. Our key goal as data providers is to ensure user communities can gain access to and effectively use quantitative and qualitative resources. The conference this year will focus on the capabilities both of Internet technology and of the wider data community in providing support to and partnering with the research and education infrastructure. The influence of evolving web environments for data producers, data archives/data libraries, and end-users will be a key area of exploration. The conference will include papers on collaboration between data professionals and researchers in metadata dissemination and in integration projects, particularly those that have evolved from emerging standards for describing data and documentation. We will also examine the legal and political barriers to web-based data access, and the role of virtual data access in new and emerging economies.

Focus of papers and presentations for review:
- The impact of “born digital” data
- Collaborative metadata dissemination and integration projects
- Data lifespan and integrity in web environments
- DDI standards and new developments
- Documenting qualitative resources to enhance access
- Future of data warehousing and data mining
- GIS and data access
- Impact of Internet technology on social science research methods
- New research/archive networks
- Political implications of data access in virtual communities
- Preparation of complex data for multiple user audiences
- Preserving and accessing research content
- Privacy and confidentiality
- Relationships among archives and the social science research community
- Universal access to public data
- Web based data access for research and teaching
- Library subscriptions, licensing and data acquisition policies
- Legal barriers to Internet-based data access

Proposals due by January 11, 2002
The deadline for paper, session, and poster/demonstration proposals is January 11, 2002. The Conference Program Committee will send notification of the acceptance of proposals by February 1, 2002. Session proposals should contain information on the focus of the session, the organizer/moderator and possible session participants. It will be the responsibility of the session organizer/moderator to secure session participants. Please send submission including proposed title and an abstract (recommended length 150 words) to: iassist@ropercenter.uconn.edu

Additional Information
For the most current information check the IASSIST 2002 conference web site at: http://ropercenter.uconn.edu/iassist2002

or contact the Program Committee at: iassist@ropercenter.uconn.edu

Important dates:
Deadline for proposals: January 11, 2002
Notification of proposal acceptance: February 1, 2002
Registration deadline: May 9, 2002
IASSIST 2002 Conference: June 11-15, 2002