

Providing Local Support for Academic Data Users

by Robin Rice, Data Librarian, University of Edinburgh Data Library.

Introduction

United Kingdom higher education institutions provide user support for data services in a variety of ways. These are reflected in the diversity of professionals who serve as local organisational representatives (ORs) for the Data Archive. We might work in a data library, a university library, a university computing centre, a central research office, or an academic department.

What is involved in providing quality local support for data services?

Data users are involved in a research process, either as experienced researchers or as learners (or something in-between, in the case of many post-graduates). Supporters can facilitate that work by understanding the research process and where the use of data fits in. In order to provide the most effective help to users, supporters may need to do one or more of the following:

- conduct a reference interview to define the information need
- be knowledgeable about data sources available via the Data Archive and other data providers
- help the user overcome various barriers to access
- and be proactive in raising awareness within their institutions for the data services they can provide.

The Data reference interview

Those ORs who are practising librarians will know how to conduct a reference interview. It is well-known in library lore that users tend not to ask for the information they really need when approaching a librarian for assistance. Through a process of active listening and prompting for more information using both open and closed questions, the librarian is able to translate the user's initial query into a question that can be answered by available sources. By avoiding shortcuts, jumping to conclusions or making assumptions, the librarian spends a little extra time getting the query right, and saves the time of the user by finding the answer to the right question instead of an irrelevant one. This is especially true for data-related queries, which tend to be difficult, time-consuming, and sometimes even bewildering to the

non-subject expert trying to provide assistance to the experienced researcher.

There are a few angles that data supporters can take in leading a reference interview. First, be sure to determine the level of enquiry the user is committed to undertake (tactfully, of course). Is the work for a complex analysis or a class paper? Will it be part of a planned research project lasting several months, or background statistics for a quick and dirty proposal with an encroaching deadline? Is the user experienced in using secondary data or a novice? Is he or she well-grounded in the subject or methodology, or traversing new terrain? Which statistical packages is he or she prepared to use? You are likely to encounter some users again (and again) as they progress through the stages of their research; others may have just a one-off request. Finding out this information at the outset will help you anticipate future needs, and will prevent wasted efforts such as referring users to published sources when they want the raw data, or ordering data files when all they want is a printed table from a book.

Other aspects of the data request you may need to know apart from the general subject, are:

- What is the unit of analysis (country, area, household, individual)?
- What are the key variables needed?
- How recent does the data need to have been collected, and over what time period?
- Will a cross-sectional survey do, or is a longitudinal panel study required for the analysis?
- What geographic measures are needed?

All of these facets are easily found in the BIRON catalogue, and can help narrow down the choice of usable studies. Even when the user knows what study they want to use – “I need the Labour Force Survey” – it can help determine which in a series of datasets to order. (A title search for the LFS in BIRON brings up 92 studies with various timeframes and versions.) Also, it never hurts to satisfy yourself that the named study will indeed have the information required before going through the ordering process.

Asking questions of the user before trying to answer a query can also buy you time as you try to think of a starting point in your search. If they are referring to a published article that used a mysterious dataset, ask for the citation. Do not be afraid to ask

even 'dumb' questions about the user's request. Ask them to unpack acronyms and explain jargon to you, so you are not guessing at their meanings. Take notes if it is a complex query, or if you are unable to do a search right away, writing down key phrases as the user speaks. Remember that most people using data are doing original research, so it is not surprising that they are very knowledgeable about something you may know nothing about. A good reference interview brings out the subject expertise of the user just enough to define the query so you can find an appropriate source. Not only will you learn something as you struggle to understand the research question, you may be able to impart valuable information back to the user about various interfaces, registering to use data, available formats, and expected time scales for receiving the files. If you try to ensure that neither party is making unwarranted assumptions, your reference interview is bound to be successful.

Sources and referrals

As the OR for the Data Archive, sooner or later you will be asked about data that is outwith the Data Archive's holdings. This is why it pays for local data supporters to be knowledgeable about other data sources available nationally and locally. Familiarise yourself with SOSIG¹ (Social Science Internet Gateway), now a hub of the Resource Discovery Network, and Statbase², a web-based statistical delivery system of the Government Statistical Service, and other data resources on the Internet. Make a point of knowing and being known by other data service representatives on campus, such as MIMAS³ (Manchester InforMation and Associated Services) for Census and other online data and EDINA⁴ (Edinburgh Data and INformation Access) for some geo-spatial datasets. Find out which datasets your university subscribes to through these national services, and which additional datasets have been purchased for research projects through data vendors. Are they being made available to all users through a network or a computer lab, or are they only used by staff in a single department? Be aware of how data tend to be used in your university for research, and which departments are the biggest data users. How important is quantitative analysis in their research? How about qualitative, and spatial analysis? Are these emphases reflected in teaching, and at what levels? Who in your university can help users with

¹ <http://www.sosig.ac.uk>

² Use 'Statstore' option to drill down to actual data tables.
<http://www.statistics.gov.uk/statbase/mainmenu.asp>

³ <http://www.mimas.ac.uk/>

⁴ <http://www.edina.ed.ac.uk>

statistical software, research methodology, survey design, government documents in the library, making thematic maps? Perhaps there are even some gaps in the support network of which you need to be aware.

Knowing 'what's out there' and knowing what kinds of support you can and cannot give in the research process are both important for providing users with seamless referrals. Whether you can help them personally or not, you are getting them closer to a resolution of their data-related problem. If their problem cannot be solved through existing support mechanisms, you can call on outside help. Certainly the Data Archive staff can help with problems relating to datasets held there. Other national services have helplines as well. Government agencies, including the Office for National Statistics (ONS) and local authorities are often helpful and have even been known to send extracts of data needed by users as e-mail attachments. If you have done a thorough reference interview, you will even be in a position to post the difficult query to an appropriate email list – hopefully one you are already on or can easily find on Mailbase. The Data Archive OR list may be used for peer support, as can the member list for IASSIST⁵ (International Association for Social Science Information Service and Technology), made up of experienced data librarians and archivists from all over Europe and North America.

However, do not assume too quickly that you cannot provide the help yourself, even if you feel it is 'out of your league.' Sometimes the user gets stuck on a simple problem, which only appears difficult. **Read the documentation** – this is especially true for codebooks and user guides, which make for rather dry reading and the user may have skimmed or skipped the relevant sections. Perhaps there is something about an interface that is slightly tricky to use that you can discover if you try stepping through it yourself. Use common sense, and try not to be intimidated. Sometimes taking the user's contact details and sending them away is all that is needed to clear your mind for discovering a solution to the data-related problem.

Barriers to access

If a user approaches you for help, you can be sure of one thing. They have already tried to solve the problem on their own before bothering you with it. They may even be already frustrated by their efforts. If they do not already know you, even tracking you down as the OR may have involved significant effort. While all information can be associated with barriers

⁵ <http://datalib.library.ualberta.ca/iassist>

to access, numeric data in particular is difficult to obtain, confusing to understand – especially if the documentation is scant and requires specialised software to analyse. Even worse, the people whose role it is to support academic data use in universities may be under-supported themselves, lacking in training and time to devote to the special requirements of data users, because this role may be simply tacked on to an existing full-time job with conflicting priorities.

Yet it is the local supporter who can relieve the frustration of new and experienced data users and make the use of secondary data sources seem worth the trouble. The Data Archive OR can help data users identify a useful source, ease them through a tedious registration process, streamline ordering procedures, perhaps by setting up an FTP server to receive data files over the Internet to speed delivery and avoid media charges to users. The OR can even anticipate demand for datasets by ordering popular titles when they become available and having the user obtain permission to use data files from a locally maintained 'data library.' With practice, the OR can also help troubleshoot problems after the data arrives, including formatting issues, file storage and compression, and interpreting documentation. Certainly the progression of ever more 'user-friendly' interfaces has a role to play in lowering barriers to data access, but the increasing 'information overload' experienced by academics and students ensures that the local human supporter has a key part to play for some time into the future.

The commitment given to a funded research project often ensures that the data user is able to overcome obstacles to data use, such as locating a suitable source, waiting for a data access request to be processed, and learning new skills needed to make use of the data, with or without substantial local support. In the realm of learning and teaching, however, this is far from given. The barriers are simply multiplied for these users. The presence of helpful local supporters can make or break the use of secondary data in the classroom and in independent learning because an alternative can usually be found to using a particular secondary data source, and often must be within the time limits of the course.

In the case of classroom teaching, developing a newer or more interesting teaching dataset may be too much of a burden without any assistance, so the tired old datasets are trotted out again and again because they are already customised for the learning purpose at hand. In a wide range of undergraduate and postgraduate classes the use of secondary data promises enormous gains for numeracy, computer skills, and critical thinking, but until more barriers are

lifted at the national or local level, use of Data Archive holdings to enhance learning and teaching will remain under-used. (N.B. There is a new Mailbase list⁶ directed towards enabling the use of data in learning and teaching called *datateach*. Please join if you are interested in exploring this issue with others.)

Raising local barriers to accessing data may be easier than you think. There is a lot you can do without training in or knowledge of data analysis or statistical software. A user requirements survey conducted at the London School of Economics in 1997 prior to establishing a data support service, had some interesting findings. Out of nine named "Factors hindering data user in research," lack of skill was the least of the problems, and lack of availability was the biggest, followed by difficulty in obtaining data⁷. So with an ability to search BIRON and to provide or fill out forms for ordering the studies, the data supporter can overcome the largest of the difficulties as perceived by the academic researcher who then knows what to do with the data once it is received. The same study found that even among data users, a lack of awareness or lack of use of national data providers such as the Data Archive and MIMAS was widespread.

Outreach

There are a variety of ways the OR can raise awareness of data services and providers such as the Data Archive in their local institution. Certainly your department's or library's Web site can link to data resources and services wherever they are. These 'local portals' can complement national efforts such as subject-based portals for resource discovery or learning and teaching. Providing annotations along with the links, or classifying them into groups, can add value and save the user's browsing time. Of course, once created, the list will need to be updated periodically and have the correct links maintained. Paper handouts about local and national services and key datasets can be placed in visible locations such as the library or computing labs. You can also write articles about new datasets or developments such as NESSTAR in a university newsletter that reaches academic staff.

Along with these general forms of outreach, the OR can provide more customised outreach for existing users. The Data Archive can provide ORs with a list of users in their own institutions. Data

⁶ <http://www.mailbase.ac.uk/lists/datateach>

⁷ Brackenbury, Simon (1997) *Priorities for establishing a data support service: results from a user requirements survey* p. 14. Unpublished paper, British Library of Political and Economic Science, December.

users can be invited to join a current awareness e-mail list in which new datasets, upcoming training events, or other occasional items of interest to data users are announced. (If you feel your head of department needs to be informed about the value of providing data services, you may want to ensure that they are on the list as well!) As you make contacts with staff in different departments, you can offer to give departmental or classroom presentations about the Data Archive and other data services. These can be formal, with PowerPoint slides or Web demonstrations (most of the national data services can offer you slides for local presentations) or informal, such as being introduced at a staff coffee meeting. Finally, remember that the most important personal outreach you can effect is the word of mouth referral passed on to potential users from your satisfied users.

Information seeking: a theory

Brenda Dervin, a communication theorist, has developed a model to explain information-seeking behaviour that has attracted much attention in library and information science literature because of its relevance to libraries, information systems, and the reference encounter. The theory breaks with more traditional models of communication about sending and receiving information, by focusing on what Dervin calls Sense-Making. There are three crucial dimensions in the sense-making approach, making up a model of SITUATIONS-GAPS-USES. She defines these dimensions as follows:

SITUATIONS: The time-space contexts at which sense is constructed.

GAPS: The gaps seen as needing bridging, translated in most studies as 'information needs' or the questions people have as they construct sense and move through time-space.

USES: The uses to which the individual puts newly created sense, translated in most studies as information helps and hurts.⁸

According to this model, the reference question cannot be separated from the situational context in which information seekers find themselves, which has many implications for library or data services. This includes the importance of the reference interview (the same question asked by a first-year undergraduate and an experienced researcher may have different 'answers'); and the importance of

providing information at the point of need, e.g. not just in a start of term packet.

The concept of gaps is helpful in understanding that information-seeking is not just about curiosity or wanting answers to questions, but that every information-seeker is someone who has become cognitively blocked in a situation and wants to move again⁹. We already know from library studies that users first seek information from their own resources--what is available to hand, then from a friend or colleague, and only failing that, from an information service or library. You may be able to avoid being only the last resort of these users by

becoming someone they know personally or someone their colleague may know. Your own local Web pages can also become a part of the information they consult because it is convenient. The notion of gaps also underlines the importance of providing referrals (bridges) for problems you cannot solve for the user.

Finally, the reference interview should help you understand the nature of the intended uses the data or information is put towards, reinforcing the situational context of every user's query. There is a wide range of uses for Data Archive resources, such as:

- seeing how a questionnaire has been worded to design one's own survey
- getting quick frequencies of responses to survey questions of interest
- finding citations of publications based on particular studies
- exploiting the richness of a complex survey to analyse relationships of cases or variables that no one has ever studied before (secondary analysis itself)
- learning how various methodologies or sampling techniques are conducted.

It is hoped that this article will help the Data Archive OR reflect on what is needed at their local institution to attain a more user-centred local data support service.

This article is based upon a presentation given at the Organisational Representatives' Meeting held at the University of Essex in November 1999. The author can be contacted by e-mail – R.Rice@ed.ac.uk or via the Edinburgh University Data Library Web site <http://datalib.ed.ac.uk>

⁸ Dervin, B. (1983) *An overview of sense-making research: concepts, methods and results* Paper presented at the annual meeting of the International Communication Association, Dallas, TX, May. Available: <http://communication.sbs.ohio-state.edu.sense-making/art/artdervic83.html>

⁹ Lipow, Anne G. (1999) *Serving the remote user: reference service in the digital environment* Keynote address to the Ninth Australasian Information Online & On Disc Conference and Exhibition, Sidney, Australia, January. Available: <http://www.csu.edu.au/special/online99/proceedings99/200.htm>