SOSIG - a move towards subject-based services.

by Nicky Ferguson[1]
ESRC Visiting Fellow In Networked Information
Social Science Information Gateway Project,
University of Bristol. England.

HOLDING HANDS OR OPENING GATEWAYS?

The UK was, and is, very fortunate in that an early commitment was made to providing a nationwide integrated academic computer network - JANET. This made it easier for successful national initiatives to be devised and implemented. However JANET was based on X25 protocols, not the Internet protocols known as "IP". This hindered international integration It became clear that arguments over the relative merits of protocols were irrelevant because the Internet was going to be the de facto standard for international networking. So the UK started a rapid transition to driving on the same side of the road, in networking terms, as the rest of the world. Fortunately the existence of JANET means that, for academic institutions, this transition will be completed in a relatively short time.

In June 1992, some time before this transition was decided upon, but when it was already looking inevitable to many of us, I was appointed by the UK Economic and Social Research Council with a brief to support UK Social Scientists in the use of computer networked information. This was, and is, a very broad brief; and since there was not a precedent for a job of this kind I was to some extent improvising, making the rules up as I went along.

The infrastructure for networked communication had been built by the technicians. So, initially, it was the technicians who tended to use it. It wasn't until the infrastructure was there, and the user base expanded significantly away from the original designers and builders, that the way people were going to use it "in real life" began to emerge. It no longer seems surprising to us that that a medium designed for the rapid exchange of large data files for "serious" work is now most popularly used for exchanging short messages. It may well be that an infrastructure intended for file transfer of software and complex datasets will be used mainly for magazine publishing and the promotion and delivery of new service industries. As it has been with infrastructure so with information. Whilst the providers and users were the same small band it was not clear what the problems or the potential were in this area. Moreover, the provision of networked information has largely been the realm of the technical specialist, not the information specialist. There are clues here to understanding the subsequent rather uneven development in this area. It has become easier to provide information, and much development has gone into the *interface* for users, with tools such as Mosaic and Netscape, but the information processing side has lagged behind.

I entered this arena at a time when there was clearly sprouting enthusiasm amongst social scientists who had previously regarded computers with fear and thought of email as just another way of increasing their workload. This enthusiasm was fed by global consciousness-raising in many fora and by my own small efforts. But often this enthusiasm did not progress from my visit - it did not translate into "real work". It was fine to go step by step through the maze with hand-holding documentation and a supportive guide but not so easy to navigate to unknown territory after a few weeks had elapsed. Even for the brave there were more obstacles - the origins of the infrastructure mentioned above meant that relevant social science information was scattered and sparse, often seeming to occur incidentally. My workshops and demonstrations were offering a glimpse of the possibilities rather than handing out a tool which could immediately increase the efficiency and productivity of researchers. It was all very well to look at all this fancy stuff, and the feedback from my sessions was always very good, but people still regarded this "Internet stuff" as a plaything.

As information provision slowly became easier, some academics were of course delighted to rediscover the joys of the second hand bookshop. Spending an hour browsing the networks might, or might not, uncover the odd jewel amongst the dust and chaos. Once found, the jewel could be copied or printed out and squirrelled away with the other printouts and photocopies. But only the most organised users made a note of their path as they went, so after a few days directing a colleague to rediscover the jewel might be impossible. Even now with the facilities of browser programs one person's hot list is another person's cold shower. I'm not too interested in the schedule for evening classes in a college in the mid-west of the USA - but it might be very useful to the right user. Making some personal details about oneself available over the net might perform an important function - to personalise and humanise discourse in a collaborative project where the participants have never met, for example. But I do not want to keep tripping over these details when I'm looking for something else. So system administrators, responding to complaints, hit upon the extraordinary idea of organising information according to subject headings. It took a little while for everyone to realise that librarians have been doing something similar for years, and by that time a host of idiosyncratic infant subject classification schemes were sprouting. In setting up the Social Science Information Gateway, we resolved to attempt at least to share the underlying classification system with

other UK national service providers.

THE RETURN OF THE CATALOGUER

With the advent of client or browser software giving users a graphical interface to networked information, the possibilities for junk or vanity publishing seemed to expand dramatically, the idea of making pictures, text and sounds available across the world -do-it-yourself multi-media publishing - was irresistible. Combine this with the relative ease of creating *HTML* - hypertext mark-up language, the building block of the World Wide Web, and you have an explosive combination. While the development of publishing was bounding ahead, the users of information were not so well provided for. Browsing was more exciting - instead of showing users meteorological data in tables, I could now bring up on their screens satellite photographs in glowing colour. But even with subject categories and fancy graphics, all we have really done is to give the second hand bookshop a facelift. You may know which shelf to look on, if you're lucky some of the books may have glossy covers, but the essential problem of locating relevant and useful texts remains.

One way we have tried to deal with this at SOSIG is by providing a searchable catalogue of information about each of the over 500 resource centres at which we point. This is quite different from the various so-called robots or automated search mechanisms which rely on highly resource intensive scouring of the networks and fairly crude automated examination of the resources themselves. We rely on human intervention to describe, classify and organise social science resource centres. In this way we also introduce an element of quality control. For each resource centre which appears anywhere on the subject menus, a form or template has been filled out - this contains a description and keywords as well as appropriate technical information such as the URL (network address) and the UDC (classification) number assigned to that resource. The user can then search through this information using an on-screen form. A dynamic list of hits will then be returned listing appropriate resource centres, describing them and pointing directly to them. Various options are provided and others (including Boolean search options) will be added in the near future.

ROADS TO THE FUTURE

The ideal for such services is that they should be distributed - so that centres of expertise are responsible for relevant subject areas. To answer the obvious question that this raises about our own activities, it is probably neither feasible nor desirable in the long term for us to attempt to take responsibility for describing and organising all the social science resource centres in the world, it is surely better for centres of excellence within the different social sciences to take responsibility for their own areas and for us to co-ordinate these efforts, but as a medium term solution the current SOSIG is certainly preferable to a totally centralised model. Aiming for a distributed model, however, creates its own problems. It demands the ability to search across different servers which in turn implies that the resource descriptions will be in (preferably an internationally accepted) standard form. When the catalogue databases become large, as they undoubtedly will, manipulating the descriptions and templates will also become a problem if we rely on the relatively unsophisticated tools we use at present. In addition this system of describing and searching for networked resources should not be idiosyncratic - it should be adaptable and aim for future integration with other resources such as OPACs and citation indices.

For these reasons, in collaboration with UKOLN, the UK Office for Library and Information Networking at the University of Bath, and Loughborough University of Technology, we have recently been funded to develop a system for allowing linked and geographically distributed resource discovery services to be set up.

ROADS -*Resource Organisation and Discovery in Subject-based services*- will allow users to search across different subject-based servers and will develop searching mechanisms based on emerging Internet standards such as Whois++. It will also investigate integration with other standards such as Z39.50 and Marc (in its various incarnations). As well as expanding the knowledge base and the capabilities of services such as SOSIG, ROADS will provide a packaged solution for information providers who wish to set up a subject-based service. We also hope to encourage centralised national service providers to focus their effort on the (initially many) subject areas not covered by these distributed services, so that a good coverage can be achieved in a relatively short time. Thus ROADS will help to achieve the goal of a scaleable system for resource discovery, cataloguing, description, organisation and quality control.

We have no illusions that ROADS will be a so-called killer application for networked information - there will not be such an application, rather a number of different approaches will emerge and possibly merge. Moreover sometimes a user will not find the obscure object of desire; or perhaps wishes to *comprehensively* survey networked resources on a topic without necessarily having regard to quality or currency; or to search across different languages and character sets. For these reasons, the ROADS partners intend to collaborate with European partners, not only to develop ROADS further, but also to develop complementary systems, including a comprehensive automated indexing system for European World Wide Web servers. Thus, if the ergonomic nut crackers fail to break open the shell and reveal the kernel, we will provide the back-up of a well-designed hammer.

This European collaborative proposal, codenamed DESIRE, has recently been shortlisted for funding by the relevant European funding agency.

We hope that all three of these initiatives will promote the design and building of Subject-based Information Gateways (SBIG's), the implementation of which will result in a distributed resource discovery service based on rich descriptions and a quality controlled approach organised around subject centres of excellence. These efforts will be complemented by a comprehensive approach to European WWW index design, the implementation of which will result in a European discovery service based on automated indexing and an automated harvesting technology.

REFERENCES AND FURTHER INFORMATION

1 Paper presented at IASSIST95 May 1995 Quebec City, Quebec, Canada.

BUBL - The Bulletin Board for Libraries -

<URL: http://www.bubl.bath.ac.uk/BUBL/>

FIGIT - The Follett Implementation Group on Information Technology -

<URL: http://lamin.bath.ac.uk/FIGIT/figit-2-94.html>

Guardian (1995). *Online* March 30, 1995

NISS - National Information Services and Systems -

<URL: http://www.niss.ac.uk/>

ROADS - Resource Organisation and Discovery on Subject-based Services -

<URL: http://ukoln.bath.ac.uk/ukoln/roads/roads.html>

SOSIG - Social Science Information Gateway -

<URL: http://www.sosig.ac.uk/>

UKOLN - The UK Office for Library and Information Networking -

<URL: http://www.ukoln.bath.ac.uk/ukoln/>

Further reading may be found at:

<URL: http://ukoln.bath.ac.uk/ukoln/roads/related.htm>