

The Social Science Information Gateway: putting theory into practice

[Debra Hiom](#)

Institute for Learning and Research Technology
University of Bristol
Bristol BS8 1TN

Abstract

The Social Science Information Gateway (SOSIG) was established in 1994 to provide access to networked resources for social science researchers, academics and librarians. This paper discusses issues of quality, access, resource identification and description, that SOSIG has faced in the four years since its establishment. It also reviews the involvement of SOSIG in the European Union's DESIRE project.

Introduction

The information needs of social science researchers can increasingly be met via the Internet. Electronic information resources relevant to the social sciences, including data and datasets, are being made publicly available world-wide and the Internet offers the potential for researchers to access these from their desktops. However, the proliferation of resources coupled with a lack of Internet skills leads some researchers into difficulties when attempting to search for - and retrieve - useful information.

The Social Science Information Gateway (SOSIG) was established in 1994 as a pilot project to provide fast and easy access to relevant, high quality networked resources for social science researchers, academics and librarians. The gateway provides access to Internet resources via an online catalogue where each resource had been classified and described by an information professional. The SOSIG team locate, assess and describe high quality networked resources from around the world, adding value and saving time and effort for researchers and users of their research by providing the facility to browse and search resource descriptions and connect directly to resources of interest. The catalogue currently contains over 3500 descriptions of resources in over a hundred and sixty subject headings ranging from Anthropology to Statistics.

Four years is a long time on the Internet and so, for an Internet service SOSIG is very well established. But compared to a traditional library catalogue, for example, SOSIG is still very young and we have many plans for expansion and enhancement. Any service and especially any Internet based service requires research and development to keep it running effectively. SOSIG has a policy of continually reviewing and improving the service based on user feedback and in response to fast-changing technologies and standards developments. SOSIG's participation in the EU-funded [DESIRE \(1998?\)](#) (Development of a European Service for Information on Research and Education) project has enabled the service to concentrate some of its research and development activities in the area of resource discovery, as well as provide additional functionality to the service itself. This paper looks at some of the theoretical work that underpins the service and how that impacts on the day to day running of the gateway.

The Theory

The EU-funded DESIRE project has provided the opportunity for some crucial research and development work to be

undertaken in the area of resource discovery. DESIRE is a major international project aiming to build large scale information networks for the research community, examining many related issues such as discovery of information, security, caching and training. The project is supported by the European Union's Telematics Application Programme involving 22 partner organisations from across Europe. The strands of work that SOSIG is involved in within DESIRE concentrates on issues of cataloguing and indexing information on the Internet and training users to make the greatest use of the tools and services produced in the project. The partners within the cataloguing and indexing information strand are:

- Koninklijke Bibliotheek (National Library of the Netherlands)
- Loughborough University
- Netlab at the University of Lund in Sweden
- SOSIG at the University of Bristol
- UKOLN (UK Office for Library and Information Networking) at the University of Bath

The deliverables from this work are currently being consolidated in a single web site to act as a dissemination tool for this component of the DESIRE project ([Cataloguing and indexing. 1998](#)).

Issues of quality

Almost anyone can publish on the Internet, and the ease with which even the most inexperienced can put material on the Web has led to a proliferation of resources. The quality of the information however, is often indeterminate. Without publishers, editors, peer-review and other 'filters', much of the information is popular in nature as opposed to academic, and much of it may be invalid, inaccurate or irrelevant to academics.

DESIRE conducted a 'state of the art' review into the development of quality systems and selection criteria for subject information gateways ([Hofman & Worsfold, 1998](#)). The review covered:

- quality definitions and practices used in the commercial sector
- selection criteria currently being used by subject gateways
- selection criteria used by other Internet services
- selection criteria used in libraries

The review fed into the development of a generic list of selection criteria that could be used by services and institutions to inform the development of their own set of selection criteria. This work also led to the development of a graphical and textual representation of the quality processes involved within a subject information gateway. The model can be used in the specification, implementation, development and evaluation of a gateway. The list of selection criteria and the quality model are currently being revised and the updated versions will be made available from the DESIRE Cataloguing and Indexing Web site.

Resource description

There are many lists of Internet sites available, subject gateways are distinguished by the added value of detailed descriptions of the content, origin and nature of each site. This creation of descriptions is also known as resource description or metadata. DESIRE carried out a survey of the current resource descriptions formats being used, as well as looking at some of the issues surrounding the description of Internet resources ([Dempsey & Heery, 1998](#)).

Classification

A number of Internet catalogues and services are using traditional library classification schemes to organise their information and aid retrieval. A study of the classification schemes currently in use in Internet search and discovery services is available in the DESIRE report The role of classification schemes in Internet resource description and discovery ([Koch & Day, 1998](#)). This review of classification schemes included universal schemes such as LC (Library of Congress), and UDC (Universal Decimal Classification), national schemes such as Nederlandse Basisclassificatie (BC), subject specific schemes such as National Library of Medicine (NLM) and home grown schemes. It also reviewed attempts to apply automatic classification of information.

Distributed cataloguing

Because of the open nature of the Web there is considerable potential for distributed collaborative cataloguing of networked resources. Information gateways can be built by teams of staff who are geographically dispersed but who can add resources to a database from their desktops via the WWW. DESIRE looked at and described a number of different strategies for distributed cataloguing. One model with which DESIRE experimented was setting up distributed databases in different countries. As part of this work the Koninklijke Bibliotheek have set up an experimental ROADS based subject gateway for Dutch Arts and Social Sciences Resources for the purpose of cross searching geographically dispersed databases. A demonstrator has been set up to show the potential of cross searching distributed gateways. The demonstrator allows combined searches of three information gateways: SOSIG, Biz/ed (Business Education on the Internet) and the Koninklijke Bibliotheek's experimental Arts and Social Sciences gateway.

Automated indexing

DESIRE is also working on an automated approach to resource discovery which is complementary to the quality controlled catalogue approach. The effort is centred on the creation of a European Web Index (EWI) which will provide a search interface to Internet documents published in Europe (harvesting not only on-line documents from the Web, but also those available via other Internet protocols. The harvested information will be made available through the Z39.50 protocol. There has also been some work on the integration of robot gathered databases and subject information gateways.

The practice

SOSIG practice has informed much of the more generic theoretical work which we have undertaken in DESIRE. Similarly SOSIG has used the results of the DESIRE research to inform the development of the gateway. Issues discussed above have underpinned SOSIG change and developments in the following areas:

Issues of quality

A fundamental distinction between SOSIG and the search engines lies in the way in which they build their collections. SOSIG is selective - it only points to resources that have been vetted for quality by a team of information specialists. The search engines aim to be comprehensive in their coverage and use robots to try to automatically index every resource on the Internet, regardless of its content and source. SOSIG used the generic quality criteria generated from the DESIRE report to develop a scope policy and a set of formal selection criteria for the gateway. The DESIRE report has also been used to develop selection criteria for other information gateways.

Resource description

SOSIG creates metadata for resources that it selects using the pre-defined set of quality criteria. The service uses a generic metadata format for describing resources using a simple record structure based on attribute-value pairs. This format is designed to be full enough to support effective user retrieval and selection of resources and is capable of being mapped to a variety of other formats, e.g. the Dublin Core.

Classification

The DESIRE classification report caused SOSIG to review the scheme that was currently in use on the service (an abridged form of the UDC (Universal Decimal Classification) scheme). The selection of UDC numbers used by SOSIG has now been extended to cover a wider selection of social science subjects and been used to provide browsable sub sections within the main subject categories. This will allow for more specific browsing and help to manage the increase in the number of resources.

Distributed cataloguing

SOSIG has piloted two approaches to distributed cataloguing (based on investigative work carried out through DESIRE).

1. **Section editors** In May 1997 ten universities across the UK began a one year pilot project as Section Editors for SOSIG. One or more librarians at each institution were appointed to spend half a day per week building up the SOSIG collection. An initial training workshop was held to familiarise the Section Editors with the SOSIG selection and cataloguing skills. An administration centre was also built to support remote cataloguing into the SOSIG database using the Web. The administration centre provides access to the template creation and editing tools as well as supporting tools and documentation.

The Section Editors are responsible for the identification, evaluation and cataloguing of resources under a particular subject area. The SOSIG core staff continue to be responsible for the overall development of the collection as well as for routine maintenance tasks such as link checking. The input from these subject specialists has done much to increase the depth and breadth of the SOSIG collection as well as foster valuable links with the library community.

2. **Correspondents** A second model for distributed cataloguing was the recruitment of Correspondents to help build up the number of catalogued resources from other European countries. A call for Correspondents was requested through articles in key journals and through messages to a number of European mailing lists. Correspondents are volunteers who are willing to submit suggestions on a regular basis of resources to be added to SOSIG. Unlike the Section Editors the correspondents do not have direct access to the SOSIG database, their suggestions are handled by the SOSIG core staff who create the catalogue entries in the database. A set of tools and guidelines were produced to support the Correspondents. This approach has had varied success as it relies on the good will of individuals who are receiving no support from their institution for this work. However it does seem to have improved the profile of SOSIG in Europe and we receive many more 'Add a New Resource' suggestions from European organisations albeit on a one off basis.

Automated Indexing

SOSIG is experimenting with the Combine technology developed by DESIRE to create a companion database to the main quality controlled section on SOSIG. The 'All Social Science' database is generated by feeding the harvester with selected URL's from the main SOSIG database, the harvester will then index pages found from those URL's generating a large database of (mostly) social science material.

Another complementary approach to increasing the depth of the SOSIG collection is to use automatic methods of producing catalogue records, using metadata provided by information providers. This would allow gateways to catalogue to a deeper level within a site or organisation. There is an increasing awareness within the Internet community of the importance of providing good metadata to aid the discovery and use of their Web documents and information resources. SOSIG is hoping to work with the social science community by harvesting metadata produced by "trusted information providers". This will help to increase awareness of the output of the UK academic and research community.

Other areas of development

There are several other areas of development not directly related to the work on DESIRE that SOSIG is currently working on.

Thesaurus

SOSIG recently introduced a thesaural tool on the service to help users refine their searches. The thesaurus is based on HASSET (Humanities and Social Science Electronic Thesaurus) which was developed by the UK Data Archive at the University of Essex. We are currently developing a second implementation of the thesaurus based on feedback received from our user community. SOSIG will also be collaborating with the Data Archive to expand HASSET into a general tool for the social science community.

Mirroring

Recent talks with the US-based InterNIC (who are the creators of the popular Internet review guide - The Scout Report) have resulted in an agreement to mirror each others' services, to help to save international bandwidth and

provide faster access for our users.

Subject cache

Caching is intended to reduce the amount of traffic on the Internet whilst also speeding up response times for users. This is achieved by storing copies of popular Internet pages at locations closer to the end user. SOSIG is experimenting with the creation of a subject based cache that would be populated by periodically harvesting the non-UK resources in the database. Some initial work on this has already begun.

Conclusion

In an ever changing environment such as the Internet, research and development is critical for the continued success of subject gateways in satisfying their users expectations. SOSIG is working to bring together work on the gateway with other initiatives such as caching, mirroring and training to build an Internet 'community centre' for social scientists.

References

[*Cataloguing and Indexing*](#) (1998) Utrecht: Surfnet bv. Available from: <http://www.desire.org/results/discovery/> (Accessed: 10 June 1998)

Dempsey, L. & Heery, R.A. (1998) [*Review of metadata: a survey of current resource description formats*](#). Bath: UKOLN. Available at: <http://www.ukoln.ac.uk/metadata/desire/overview/> (Accessed 10 June 1998)

[*The DESIRE Project*](#).(1998?) Utrecht: Surfnet bv. Available from: <http://www.desire.org/> (Accessed: 10 June 1998)

Hofman, P. & Worsfold, E. (1998) [*Selection Criteria for Quality Controlled Information Gateways*](#) Bath: UKOLN. Available from: <http://www.ukoln.ac.uk/metadata/desire/quality/> (Accessed 10 June 1998)

Koch, T. & Day, M. (1998) [*The role of classification schemes in Internet resource description and discovery*](#). Bath: UKOLN. Available at: <http://www.ukoln.ac.uk/metadata/desire/classification/> (Accessed 10 June 1998)

How to cite this paper:

Hiom, Debra (1998) "The Social Science Information Gateway: putting theory into practice" *Information Research*, 4(1) Available at: <http://informationr.net/ir/4-1/paper48.html>

© the author, 1998. Last updated: 10th June 1998

Check for citations, [using Google Scholar](#)

[Contents](#)

7 6 4 1

[Home](#)

[Web Counter](#)
