Internet subject gateways

Lyn Robinson[1] and David Bawden[2]

[1] School of Library, Archive and Information Studies, University College London
[2] Department of Information Science, City University London

Subject gateways, or quality gateways, are a tool for Internet resource discovery which have come to the fore in the past two years, as a means of rapid identification of 'quality' resources. This review discusses and exemplifies these tools, where possible identifying general points or trends, with particular reference to some which are being further developed under the remit of the UK Resource Discovery Network. Relevant URLs are given whenever relevant; these were all checked in the first two weeks of August 1999.

By subject gateway, we mean something more than the lists of resources which are now very common on the Internet. We include within the scope of the term those lists of resources which have:
• a clearly expressed subject scope, defining what resources may be considered for inclusion
• explicitly defined criteria of quality, used to select resources for inclusion
• some form of annotation or description of resources
• some categorisation, classification or indexing of the collection
• clearly defined responsibilities for their creation and maintenance

These requirements mean that subject gateways are very much 'value added' products, requiring a good deal of intellectual effort in their construction. They are therefore particularly useful tools for the searcher, but correspondingly rarer than simpler listings.

History and current status

In the UK, the Joint Information Systems Committee (JISC) of the UK Higher Education Funding Councils funded the establishment of a series of subject gateways, the Access to Network Resources (ANR) projects, as part of the eLib (electronic libraries) programme. These gateways were:
• ADAM   art and design
• CAIN   conflict studies
• biz/ed   business and economics
• EEVL   engineering
• IHR-Info   history
• OMNI   biomedicine
• RUDI   urban design
• SOSIG   social sciences

Because of the nature of the funding authorities, these gateways were to concentrate on sites of relevance to users in UK higher education, and to favour, though by no means to restrict themselves to, UK sites. However, their significance extends beyond the UK context, because of the innovative nature of these gateways, with procedures for ensuring quality and consistency in the choice of resources for inclusion, and in organisation and maintenance of the gateways themselves. These gateways are discussed generally by Kerrimiur et. al. (1999), while Macleod, Kerr and Guyon (1998) discuss in detail the development of the EEVL engineering gateway.

Some, though not all, of these gateways use the ROADS software, the product of another ANR project, for creation and maintenance of a database of resource descriptions, and hence for creating selective subject gateways, and allowing cataloguing, searching and browsing of resources. Thus, although other subject gateways were set up in the UK and elsewhere, the eLib gateways very much set the standard for this kind of search aid.

It was announced during early 1999 that, after the cessation of funding of the eLib gateway programme, that JISC would fund a new Resource Discovery Network in place of the existing eLib subject gateways, with an expanded remit, to provide services to users outside the academic community. This is discussed later.

Access to subject gateways
Subject gateways may be identified in the usual ways in which any Internet resource may be found: by use of a search engine or general directory, such as Yahoo, from printed listings, or by personal recommendation. With the increase in the number and value of these tools, specific listings of subject gateways are now being created: some libraries and information centres, which initially attempted to list useful Internet resources, have now moved to simply listing subject gateways.

Two particularly useful listings of subject gateways are PINAKES, from Heriot Watt University, and OLIG (Oxford Libraries Internet Gateway) from the Bodleian Library at Oxford University:
• http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html
• http://www.bodley.ox.ac.uk/olig/

There are several updating services, which include information on the launch of new subject gateways, and on modifications and changes to existing ones: two useful resources of this kind are the Internet Resources Newsletter, from Heriot Watt University, and the Internet Scout project, from the University of Wisconsin:
• http://www.hw.ac.uk/libWWW/irn/
• http://scout.cs.wisc.edu/index.html

Nature and characteristics of subject gateways
There is no precise definition of a subject gateway. However, as noted above, they share several characteristics. They are, necessarily, subject specific, including resources pertaining to some restricted topic, and/or type of user. Unlike search engines, whose indexes are automatically constructed by software agents following automated identification of resources, gateway resources are selected intellectually by human experts. Explicit, and often strict, criteria are applied in the selection of resources, and the resources thus chosen are described and classified and/or indexed by the same human experts. Resources are selected on two general counts: appropriateness of subject content, and quality of the information resource. Appropriateness of subject coverage, or 'scope' of resource, is usually regarded as easier to assess than quality.

By definition, one criterion for inclusion is that the resource must be accessible from the Internet, but this covers a wide variety of format. The SOSIG selection criteria, for example, specifically include the following types of resource:
• electronic journals
• digitised books
• reports and papers
• scholarly mailing lists and archives
• educational software
• bibliographic databases
• electronic newsletters
• datasets
• bibliographies
• home pages of key social science organisations

Some gateways go beyond a simple identification and collection of existing resources. RUDI, for example, 'gathers and re-publishes multimedia material [and] researches and creates new resources.

In addition to this wide coverage of resources, some gateways offer other services; for example, RUDI, with a declared aim of maintaining 'a community of participation in which users are also authors', provides web site hosting for organisations in the urban design area, and offers special facilities for sharing 'transient and topical data', in an 'environment of participation', including e-mail lists, a discussion forum, and a comprehensive diary of events.

There are, of course, very many subject listings of Internet resources which do not exhibit the characteristics noted above; the resources may, for example, be selected according to the *ad hoc* preference of the compiler, or they may be listed without any further sub-division by subject, or without any description of the resource. Though such listings may be useful, particularly if no other listing of resources in their subject area exists, they do not qualify as subject gateways, as the term is used here.

Each of these five points - subject scope, quality criteria, resource description, categorisation, and responsibility for maintenance- will now be considered in turn.

1.   subject scope

Subject gateways are, by definition, subject specific. The degree of specificity varies, but most correspond to academic disciplines or areas of professional practice - biomedicine, engineering, philosophy, library/information science, and so on. Examples are given in the list below. In some cases, the subject scope is apparently considered so obvious that no scoping definition is given; this is more common for 'traditional' academic subjects. Examples are:
- 'the field of Librarianship and Information Science'   (PICK)
- 'online resources for historians'   (IHR)
- 'chemistry on the Internet'   (Chemdex)

For other fields of study, definitions are given, e.g. for RUDI, the urban design gateway:
> Urban design in this context includes the physical design, management, planning and use of buildings and landscape in terms of their relationship to public and open space.

and for Business information Sources on the Internet:
> ..useful resources which contain company, business news, or market information, and also sites which provide significant guides to these types of resource. The list is selective, not exhaustive.

An interesting exception is BUBL (originally the Bulletin Board for Libraries) which includes a catalogue of Internet resources over a very wide subject range, but selected so as to be of value to library/information specialists, by virtue of their being academic or professional resources, containing substantive information likely to be of value in a library context.

Some gateways apply criteria, other than subject coverage, for selection. SOSIG, for example, in a particularly detailed scope policy (www.sosig.ac.uk/desire/escope.html), excludes such material as:
- resources which contain information on primarily illegal areas such as bomb-making or paedophilia
- resources consisting solely or mainly of advertising
- information intended for use only by an individual or local group

'considers' (with an implication of disfavour) such things as:
- pages maintained by students
- commercial and fee-based resources
- resources requiring the user to register
- resources using advanced WWW technology

and 'prefers'
- information which is scholarly rather than popular
- resources originating in Europe.

2.   quality criteria

Assessing quality on the Internet is a rather more difficult matter than the equivalent task in 'traditional' resources, a is made clear, for example, by the notes on this point for the SOSIG social science gateway:
> 'Apply content criteria by looking at and evaluating the information contained in the resource. Remember that anyone can publish on the Internet, so information has often not been through traditional quality 'filters' such as publishers, editors or peer-reviewers. Resources may not be what they appear to be or what they say they are, as on the Internet there is no guarantee that the resource is accurate or honest.'

Quality issues have been a source of discussion and concern since the Internet first became a serious tool for information provision, and a good deal of expertise and good practice in this area has been created; see, for instance, the books by Cooke (1999) and by Alexander and Tate (1999). The general principles, which are accepted by all subject gateways, are well summarised by another quote from SOSIG:

> 'The information content is a primary consideration when evaluating Internet resources for [the gateway, which] aims to point to primary information and not just lists of links. Information should be valid, accurate and current, and should come from a valid and authoritative source. The information should be comprehensive, for example giving full text of documents as opposed to just abstracts, and organisational information as opposed to merely contact details. Resources should not duplicate information held elsewhere in [the gateway]. '

An overview of quality criteria for subject gateways is available at
- http://www.sosig.ac.uk/desire/qindex.html

Specific quality criteria for particular subjects are set out by the gateways dealing with those subject areas. Examples are:
- http://www.sosig.ac.uk/desire/ecrit.html
  social sciences
- http://www.omni.ac.uk/agec
  biomedicine
- http://www.ub2.lu.se/eel/qualcrit.html
  engineering
- http://www.adam.ac.uk/adam/reports/select
  art and design

The Sosig social science gateway, a particularly detailed example, uses the following 15 main criteria for quality assessment:

- Validity; do the resources fulfil the stated purpose?
- Authority and Reputation of the Source; who provided the information?
- Substantiveness; is the information substantive?
- Accuracy; is the information accurate?
- Comprehensiveness; to what level of detail does the resource go?
- Uniqueness; is the information on the site unique?
- Composition and Organisation; is the information well composed?
- Ease of Navigation; is it easy to navigate the resource?
- Provision of User Support; are there instructions?
- Use of Recognised Standards

- Appropriate use of Technology; how appropriate is the format?
- Aesthetics; has consideration been given to the appearance of the site?
- Information Integrity (work of the Information Provider); is the information current and up to date?
- Site Integrity (work of the Web-Site Manager); is the site current and up to date?
- System Integrity (work of the Systems Administrator); is the technical performance of the resource acceptable?

Each of these, in turn, has detailed sub-criteria, for example:

Composition and Organisation; is the information well composed?
> Does the text follow basic rules of grammar, spelling and literary composition?
> Does it include jargon?
> Is the information within a resource phrased unambiguously?
> Is the information clearly organised?
> Is there a good structure?
> Is the information within a resource arranged logically and consistently?
> Is the information broken down into logical parts?
> Is the resource well laid out?
> Is the resource organised by the needs of the user?
> Is the information broken down into digestible parts?
> Is the content clearly described?
> Are the headings clear and descriptive?
> Is there evidence of internal standardisation (e.g. use of a 'style sheet'?)

By way of contrast, Business Information Resources on the Internet (a site generally recognised for its quality and usefulness) has a much simpler criteria statement, appropriate for a resource complied by one individual:
> Only those resources which seem to me to have useful content are included. I take into account criteria such as: coverage, currency, reliability, etc. However, inclusion in this list is not a guarantee of its quality ! I look at each site before including it, and briefly again each time I update, but cannot claim to have explored each one in depth.

3.   resource description
The descriptions of resources provided by subject gateways vary from a simple annotation, e.g. from Business Information Sources on the Internet:
> **Europe: YellowWeb Europe**
> Classified directory of about 45,000 European web sites. In Spanish, german, French, Italian, Dutch.

or a short critical evaluation, e.g. from Cranfield's Aerospace Resources:
> **Virtual Museum of the Invention of the Airplane**
> This is an amateur site, produced by an academic at the Beckman Institute for Advanced Science and Technology, University of Illinois. It is still under construction, so there are several gaps. The construction is a little poor in places, with over-large graphics and a lack of textual links to ease navigation. However, the site is still useful with information and pictures related to the history of aviation and inventors in that field.

to an informative summary of content, with additional descriptive keywords, e.g. from OMNI:
> **Biochemical Journal**
> **Description**: This online version of the Biochemical Journal provides full text and graphics for subscribers to the print version, as well as abstracts, for all articles from issue 316-3 (15 June 1996) onwards. Articles published from issues 316-1 (15 May 1996) onwards are available in pdf format. There is also a service which emails the table of

contents to users every two weeks. The Biochemical Journal covers biochemistry, and cell and molecular biology. Published by Portland Press.

or from SOSIG:

Conflict in Kosovo: failure of prevention ? An analytical documentation, 1992-1998
Description: This working paper, by Stefan Troebst (1998) is the first in a series published by the European Centre for Minority Issues (ECMI) based in Flensberg, Germany. It "… documents attempts by international organisations and NGOs to mediate in the Albanian-Serbian conflict over Kosovo from the beginning of the war in Bosnia and Hercegovina in the spring of 1992 to the Drenica massacre of early March 1998. The author pays particular attention to recommendations by diplomatic actors and other third parties seeking to improve the tense status quo, to find interim solutions or to achieve a resolution of the conflict". There is an online abstract, with the full text version made available as a .pdf file.
Keywords: Social sciences, Kosovo, Balkans, ethnic cleansing, Moldova, Transdniestria, conflict, peace-keeping, international relations, politics, economics, sociology, Serbia, Yugoslav Federation

4.   categorisation

Categorisation of resources again varies in complexity. At its simplest, it may involve a series of broad categories for browsing, with full-text search of resource descriptions. Some gateways go further than this, however, in providing additional indexing and classification for each resource, using either standard publicly-available classifications and thesauri, or vocabularies derived specifically for that gateway.

Most common is a set of broad categories devised for the gateway, often mixing subject categories with format of material, for example, that of the Internet Directory for Botany:
- arboreta and botanical gardens
- biologists' addresses
- botanical museums, herbaria, natural history museums
- botanical societies and organisations
- university departments, other institutes
- lower plants
- vascular plant families
- checklists and floras, taxonomic databases
- conservation, threatened plants
- economic botany, ethnobotany
- paleobotany, palynology, pollen
- gardening
- journals, books, literature databases, publishers
- link collections, resource guides
- listservers and newsgroups
- images
- software
- other resources

In this case, as in several others, the potential confusion caused by separate categories for subject and format is overcome by assigning resources to all appropriate categories.

Other gateways distinguish subject categories from format. For example, the Chemdex chemistry gateway has nine main categories. One titled 'Chemistry' is a subject listing, with 20 sub-divisions: analytical, biological, catalysis, etc. The others describe the originators of a resource - universities/institutes, institutions, government, companies - and the format of the resource - communication, databases, software, other links.

Standard classification schemes are used by some gateways, e.g. UDC by SOSIG, and Dewey by PICK and ADAM. The OMNI biomedicine gateway uses the standard controlled vocabularies of its area - the National Library of Medicine classification and the MeSH thesaurus - for browsing by general and specific concepts respectively. The EELS engineering subject gateway uses the Ei Classification Scheme, from Engineering Information Inc., though this does not cover all subjects within the gateway's database; the EEVL gateway uses its own classification scheme, with a 'loose adherence' to the Ei scheme (Macleod, Kerr, and Guyon 1998).

Most developed in this respect is the BUBL gateway, which offers three means of subject access: 10 general subject headings, e.g. general reference, creative arts, engineering and technology, which are a synthesis of a number of subject classifications used in UK higher education, such as subject groups used for assessment of research and teaching quality; Dewey class numbers; and a series of subject terms initially based on Library of Congress Subject Headings, but customised and expanded.

Most gateways offer searching and/or browsing access to a single resource catalogue, but some provide more specific access to particular types of resource. EEVL, for example, allows searching of:
- a catalogue of 'quality' engineering web sites
- a collection of UK-based engineering web sites, with less stringent quality criteria
- electronic journals in the field of engineering
- engineering newsgroups
- several bibliographic databases

## 5.	responsibility for maintenance

Most subject gateways are created and maintained by small teams of subject specialist information professionals, based in institutions with subject relevance, often academic libraries. Usually, they are the product of a single institution, although collaboration is an increasing trend: EEVL, the engineering gateway, for example, has been hosted by the Library of Heriot-Watt University, in partnership with the Edinburgh, Napier, Cambridge, Nottingham Trent, and Cranfield Universities, Imperial College, and the Institution of Electrical Engineers. Some, such as OMNI and BUBL, make use of volunteers for selection and description of resources, giving careful attention to their selection and training. A few are still maintained by enthusiastic individuals, e.g. the Human Languages Page, and Business Information on the Internet. Resources of this kind, evidently labours of love, may express a 'personality' absent for more anonymously maintained gateways, but are clearly vulnerable to a change in circumstances for their creator.

One consequence of the nature of the maintainers is that the gateway may focus on the needs of a particular group of users, or on information provided by the maintaining institution. For example, the PICK library/information science gateway is produced by the Thomas Parry Library, which serves the Department of Librarianship and Information Science, at the University of Wales, Aberystwyth, and hence its 'core users are the students and staff of this department'. Similarly, the IHR historical resources gateway is maintained by the Institute of Historical Research, and the PORT maritime studies gateway by the National Maritime Museum; both gateways include a strong emphasis on resources produced by, and of importance to, the respective institutions. As Macleod, Kerr and Guyon (1998) note, the strengths of these gateways lie very much in their 'subject specialism, individual characters, and dedicated specialist staff'.

The Resource Discovery Network
The significance of the eLib gateways in setting standards for subject gateways will be clear from what has been said above. In early 1999, a new initiative was announced, funded by JISC and to be run jointly by the UK Office for Library Networking (University of Bath) and the Arts and Humanities Data Service (King's College London).

The Resource Discovery Network (RDNet) will be a network comprising a central network centre, and a distributed collection of hubs. The hubs approximate to the existing subject specific gateways, though with widened remit; it will also be possible to search for resources across several hubs simultaneously.

Four hubs announced as of the time of writing were:
BIOME   life sciences and health
               (successor to OMNI)
EMC   engineering, mathematics and computing
               (successor to EEVL and Aerospace Resources)
SOSIG/BizEd   social sciences, business and law
HUMBUL   humanities
Hubs covering other subject areas will follow. (Information on this, and other aspects of RDNet, can be found at www.rdn.ac.uk)

These services, even more than the eLib gateways, will be created and maintained by consortia of relevant institutions. For example, the partners for the Biome hub are:
• Universities of Nottingham, Oxford and Reading
• Natural History Museum
• Royal College of Veterinary Surgeons
• Royal Free Hospital
• BBSRC Research Institutes Librarians group
• CTI (Computers in Teaching Initiative) Centre for Biology

The widened remit of the hubs, as compared with the eLib gateways, seems likely to include both a more general subject coverage, and also an extension of purpose beyond the needs of the academic community. Developments in RDNet should provide the impetus, and set standards, for development of the next generation of subject gateways.

List of gateways

The following is a representative, not comprehensive, list of subject gateways, as of August 1999; it illustrates the range of subjects covered by gateways of the sort described here.

BUBL: resources in many subject areas, selected for library use
http://www.bubl.ac.uk/link/

SOSIG: social sciences
http://www.sosig.ac.uk

OMNI: medicine and healthcare
http://www.omni.ac.uk

EEL: engineering
http://www.ub2.lu.se/eel/

EEVL: engineering
http://www.eevl.ac.uk

ADAM: art and design
http://www.adam.ac.uk/adam/

PICK: library/information science
http://www.aber.ac.uk/~tplwww/e/

Aerospace Resources on the Internet: aerospace information
http://www.cranfield.ac.uk/cils/library/subjects/airmenu.htm

AHDS: arts and humanities
http://ahds.ac.uk/

Biz/ed: business and economics
http://www.bized.ac.uk/

Business Information on the Internet
http://www.dis.ac.uk/business

CAIN: conflict studies
http://cain.ulst.ac.uk/

ChemDex: chemistry
http://www.shef.ac.uk/~chem/chemdex/

EdWeb: educational reform and information technology
http://edweb.cnidr.org/

ELDIS: development and the environment
http://nt1.ids.ac.uk/eldis/

GEM (Gateway to Educational Materials): educational resources
http://thegateway.org/

Geo-Information Gateway: geography, geology, the environment
http://www.geog.le.ac.uk/cti/info.html

History: historical studies

http://www.ihrinfo.ac.uk/
Human Languages Page: language-learning and linguistics
http://www.june29.com/HLP/

HUMBUL: humanities
http://users.ox.ac.uk/~humbul/

InfoLaw - legal resources
http://www.infolaw.co.uk/

Internet Directory for Botany
http://www.helsinki.fi/kmus/botmenu.html

The Math Forum: mathematics
http://forum.swarthmore.edu/library/

MCS: media and communication studies
http://www.aber.ac.uk/~dgc/gate.html

NetEc: economics
http://netec.mcc.ac.uk/NetEc.html

NOVAGate: forestry, food, veterinary, and agricultural sciences
http://novagate.nova-university.org/

Philosophy Around The Web: philosophy studies
http://users.ox.ac.uk/~worc0337/phil_index.html

Philosophy in Cyberspace: philosophy
http://www-personal.monash.edu.au/~dey/phil/

Port: maritime studies
http://www.port.nmm.ac.uk/

Psci-com: science
http://omni.ac.uk/psci-com/

SciCentral: science
http://www.scicentral.com/index.html

Psych Web: psychology
http://www.psywww.com/

RUDI: urban   design
http://rudi.herts.ac.uk/

TIPTOP: physics
http://physicsweb.org/TIPTOP/

World Wide Arts Resources : the arts
http://wwar.com/

References

JE Alexander and MA Tate (1999), Web Wisdom; how to evaluate and create information quality on the web, Lawrence Earlbaum, New York

A Cooke (1999), Guide to finding quality information on the Internet: selection and evaluation strategies, Library Association, London

J Kerrimiur et. al. (1999), UK subject gateways to information on the Net, Electronic Library, 17(2), 91-94

R Macleod, L Kerr and A Guyon (1998), The EEVL approach to providing a subject based information gateway for engineers, Program, 32(3), 205-223