

SUBJECT GATEWAYS : THE CLEVER WAY TO INFORMATION

By

Pankaj Kumar Das*

ABSTRACT

Technology advances over the past two decades have made data retrieval faster and easier, giving rise to a substantial industry providing access to professional, business and scientific information. Some progress has been made towards increasing the relevancy of the data with the induction of various search engines and subject directories. Despite these activities, information sources remain scattered, hard to find and difficult to access. Using technology, visionary institutions, associations and individuals build a kind of network resource discovery service, called "Subject Gateways" on the web, which is de-facto network use environment. This subject gateways evolved during the last five years among early digital library projects within the library communities of the various countries. Subject gateways allows libraries and related organizations to explore the usefulness of their subject expertise in the organisation of knowledge in the word of network-based, digital information.

This paper gives an overview of the subject gateways, its historical development, definitions, features and lastly the burning issues encompassing it. Also some of the international initiatives related to further research and trend are explored.

* Student, Documentation Research and Training Centre, Indian Statistical Institute, BANGALORE-560059, e-mail: pankajdas128@hotmail.com

0. Introduction

"The web is quickly becoming the world's fastest growing repository of data"

(Tim Berners-Lee, W3C director and creator of the WWW)

Information technology has made many advances over the years, with the most rapid ones coming recently. From Gothenburg's printing press to e-publishing. One might think that technology is making it easier for us to organize and store information, has made access and retrieval easier; however, it is not a straightforward case. Advancing technology has created a whole new problem – information boom.

Technological advances have made it easier and faster to analyze, collect, abstract, index, search, and use data and information. But the information may be scattered, not only in several different databases, but also on completely different systems, some, of which users may not be able to access. If information professionals find it difficult, then for users it is nearly impossible. So to arrest this titanic problem, the concept of " Subject Gateways" has emerged in the early 1990s.

1. Why Gateways?

“People are increasingly.....going to the Internet before they go to the library.”

“ Librarians are increasingly.....taking librarianship out of libraries and onto the Internet.”

The phenomenal growth of the Internet has posed an unmanaged and increasingly unmanageable environment. The current situation of Internet may be characterized as follows:

1. Internet as a whole is not well organized.
2. Users do not find what they are searching for, and still waste hours looking, are frequently arguments against the Internet.
 - 2.1 Huge amount of public information are handled by various search engines in a fairly basic manner.
 - 2.1 Search engines have no inbuilt mechanism for matching retrieval to the age, level or experience of the users or to the purpose of the study.
 - 2.2 Web publishing is often very careless in authenticating sources.
 - 2.3 Accessibility is sometimes illusory, and different file formats and protocols require additional software to read files and to achieve cross-platform performance.
3. Ease of technical facilities to copy and paste makes a copyright abuse inevitable.
4. In spite of attempts to create metadata structures, there is an almost complete absence of effective information control mechanisms.

Although Internet is a unique example of the successful linking of people and resources, but still it is somewhat lacking in the cost-effective manner.

2. What are Gateways?

Moffat describes the establishment of the gateways as **“a process of identification, filtering, description, classification and indexing before they are added to databases which is freely available via a WWW.”**

So we can say that gateways are the Internet search tools to help people find resources on the Internet, e.g.: electronic journals; software; data sets; e-books; mailing lists / discussion groups (and their archives); articles / papers / reports; bibliographies; bibliographical databases; organizational home pages; educational materials; news; resource guides.

Gateways offer linked collections of Internet resources via a database of resource description. This can be:

- ☞ Browsed – according to broad classification

- ✍ Searched –through index
- ✍ Quality controlled – due to selection

Thus it can be said that the gateways are “ a gathering place of discipline specific resources.”

3. What are subject gateways?

Subject gateways are also known as (1) subject-based information gateways (SBIGs); (2) subject based gateways; (3) subject index gateways, (4) virtual libraries; (5) clearinghouse; (6) subject trees; (7) pathfinders; (8) quality –controlled subject gateways, etc.

Subject gateways is nothing but the facility that allows easier access to networked-based resources in a definite subject area. The simplest type of subject gateways are sets of web pages containing lists of links to resources. “Subject gateways” as a term was popularized in the UK Electronic Libraries Programme (e-lib)

According to *Dempsey, L; Gardner, T and D. Michale, UKOLN, University of Bath, UK* **“Subject gateways are Internet services which supports systematic resource discovery. They provide links to resources(documents, objects, sites or services) predominantly accessible via the Internet. The service is based on resource description. Browsing access to the resource via a subject structure is an important feature”**

According to (Emma Place, ILRT, University of Bristol, UK) **“Subject gateways are Internet- based services designed to help users locate high quality information that is available on the Internet. They are typically, data bases of detailed metadata (or catalogue) records which describes Internet resources and offer a hyperlink to the resources.”**

Subject Gateways offer user an alternative to the generalized approach of the commercial global automatic “vacuum cleaner” type of search engines, as infoseek; altavista, etc. Subject gateways are characterized by two key factors:

- A. They are selective, pointing only to Internet resources that meet with quality selection criteria.
- B. They are built by subject and information specialists- often librarians.

Besides these there are few more, as-

1. Generally limited to specific subjects
2. Scope-policy declaring what subjects they are indexing
3. Defined target group- e.g. academics, researchers, etc.
4. Manually created records- rich resource description containing relevant information
5. Distributed cataloguing – a scattered group of subject specialists contribute to the databases

6. Quality criteria- there is an official set of quality criteria
7. Classification systems- used as underlying system for browsing possibilities
8. Use of open standards – to support co-operation with other services e.g. cross-searching.

4. Historical development of subject gateways

The subject gateways engaged in response to the challenges of “resource discovery” in a fast developing Internet environment in the early and mid 1990s. Due to the emergence of the network information retrieval systems (Gopher, WWW, Archie, netfirst, etc) and access protocols (ftp, gopher, telnet, http, etc) innovative information technologies and services emerge.

The Electronic Libraries Programme (eLib) of JISC of the UK Higher Education Funding Council set up in 1995 which includes besides other things, access to network resources (ANR) and subject gateways were funded as the part of ANR area and latter on it lead to the funding and establishment of eLib subject gateways- SOSIG, EEVL, OMNI, History, ADAM, and Biz/Ed.

5. Features of subject gateways

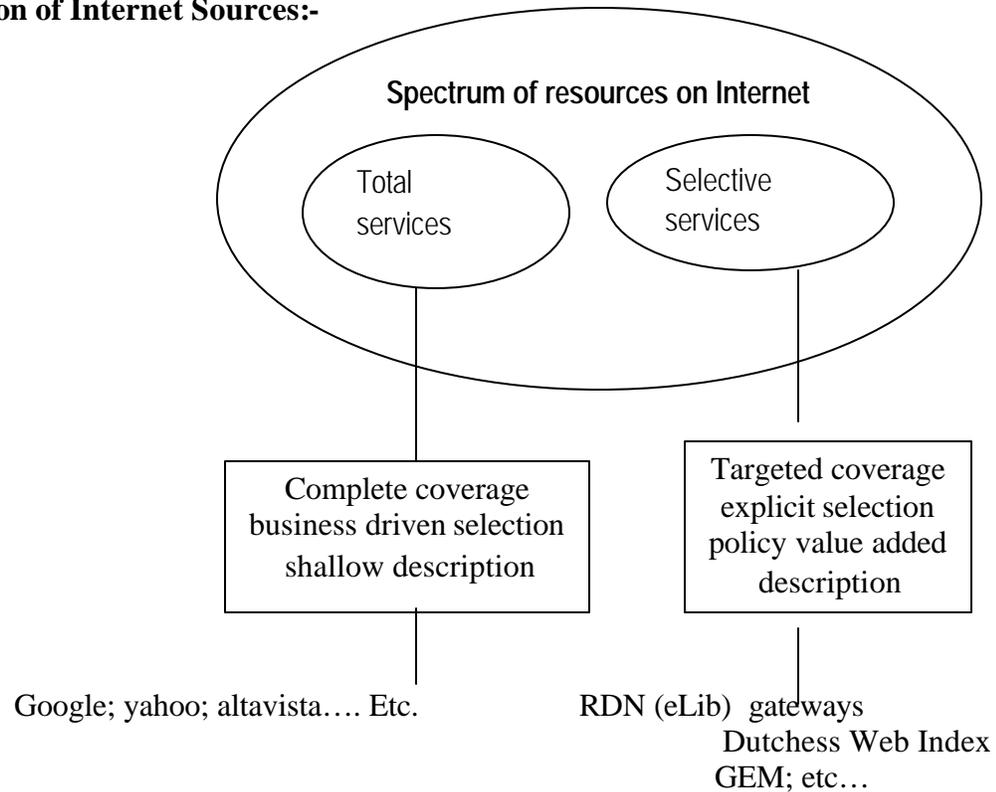
1. Each resource selected is evaluated against explicitly defined quality selection criteria.
2. Resources are classified using a range of schemes, e.g.-DDC, MESH
3. Metadata (standard resource description) are provided based on a particular standard e.g.- Dublin Core
4. Written resource descriptions are provided for each resource, often by library subject specialists.
5. Currency of resources is checked by link checking software (e.g. ROADS), databases such as OMNI.
6. Institutional commitments are the vital component for continued development of a subject gateway.

6. Difference between search engines & subject gateways

Search Engine	Subject Gateways
General resource is available.	It is a “gathering place of discipline specific resources”
It totally depends on the powerfulness of the search-engines algorithms.	High level of human input is there, as the selected resources must meet a number. Of criteria applied by a librarian or academic, who ensures that only high quality, relevant resources are included in the database.
The results can be overwhelming, unmanageable,	The results are specific, precise, and linked to

full of irrelevant references and are often too prolific to meet user needs.	relevant documents.
Records are created by an automatic process and typically consist of a mixture of metadata offered by the author of the page (if this is available) and text picked up from the page itself.	Records are created by a cataloguer, which is designed to highlight the main features of resource in an easily readable, concise fashion.
Entries are displayed more as “raw-data.”	Entries are described in a more “human-readable fashion”
It indexes pages.	It indexes resources.

Distribution of Internet Sources:-



7. Why libraries should go for subject gateways?

Libraries are the most suitable institution to undertake this work due to the following reasons:

The natural metaphor

- Browsing, reference desk

Expertise in relevant areas

- Classification, acquisition, keywords
- Information, seeking behaviour
- Guiding & helping users

8. Benefits of gateways for library

- Leading the way into the information age
- Communicating with "non-nerds"
- Access to huge-high quality collection
- Integrate into existing structures on the Internet.

9. Benefits of gateways for the users

- Diverse resources brought together

-Research, learning, leisure, enrichment-all brought together

Someone to ask – what's where?

-What's what?

-What's good?

10. Key initiatives for building tools and standards in subject gateways

ROADS (Resource Organisation and Discovery in Subject-based Services)

It is being funded by the JISC (Joint Information system Committee) through e-Lib programme (UK). It is an open source set of software toolkit, which enable the set up and maintenance of web-based subject gateways. A ROAD based information gateway is based on a database that contains information about Internet resources. The records in the database contain information such as the title of the resource, which maintains it how to access it, a classification number, and a description and keywords. The user is given access to this information while either browsing or searching the database. This is particularly important for geographically distant resources that might require some time and effort to access. The software includes the database technology, required to set up a gateways. For downloading the free online software visit its site URL: <http://www.ilrt.bris.ac.uk/roads>

DESIRE (Development of a European Service for Information on Research and Education)

This is one of the largest projects funded by the Telematics for Research Sector of the Fourth Framework Programme funded by the European Union. In particular, DESIRE intend to provide:

- ✍ tools for indexing and cataloguing information servers
- ✍ tools for management and maintenance of information servers
- ✍ demonstration and evaluation of tools and techniques for information catching and secure access to information servers
- ✍ background information for developers of networked information systems
- ✍ training materials

IN Oct'99 DESIRE published the "Information Gateway Handbook"- a guide for libraries interested in setting up large-scale subject gateways of their own. This handbook is freely available at the site: (<http://www.desire.org>) and describes all the methods and tools require to set up a large scale Internet subject gateways.

ISSAC

It is a research project of Internet Scout in the USA. Its aim is to create an architecture that enables repositories of metadata records to be cross-searched. (<http://www.scout.cs.uisc.edu/research/index-html/>)

IMesh

(A.)

It was formed as a result of a meeting at the second European Conference on Research and Advanced Technology for Digital Libraries, held at Crete, in Sept'98. One of the main aims of Imesh, "The International Gateway Community" is to explore the potential for collaborative development of gateway internationally. Many of the technologies required for cross-searching different gateways and from remote cataloguing into gateways already exist, but the strategic organization between gateways are lacking and IMesh aims to address it. In June'99 the first IMesh workshop was held in Warwick in the UK. It includes a mailing list (Imesh Mailbase), some informational web pages ((Imesh Website) and some collaborative activity.

(B)

IMESH Toolkit Project

This project was started in 1999 and funded by the National Science Foundation (USA) and JISC (UK) to develop the architecture toolkit for distributed subject-gateways. (building on ROADS and ASAAC) The Toolkit project is intended to produce a consistent framework for the development of subject gateway software. The focus is on the sharing of both software and metadata. Where possible, the toolkit will build on existing and ongoing work (ROADS DESIRE, Isaac Network and RDN). One of the main aims of the project is to reduce the entry costs for new subject gateways, including reducing the effort required to support specialized or local functionality.

Reynard project

It was decided in the Fifth Framework programme under the European Project, and its duration is from Jan-2000 to June-2001. Its aim is to provide a one-point access to, and at a consistent presentation of national subject services in Europe. It will exploit the existing services by way of creating a shared test environment within which national initiatives will experiment with co-operative efforts, devise models for sharing metadata, develop business models and foster standardization activities.

11. Some issues in subject gateways

11.1 Business issue

Subject gateways, and mostly dependent on project or research funding- "soft money", which is temporary and unpredictable in nature thus creating issues for long term planning, collaboration and service development. Some gateways had commercial partners, some were part of a wide service, and some stood alone, which offers various heterogeneous services directed at consumers, learners, and professional users, thus gateways differs widely in aims, and scope, which is the hindrance in its popularity.

11.2 Sustainability issue

In this age where survival of the fittest is the mantra, subject gateways still lacking a good rapport with the professional or scholarly publishing, or with public service broadcasting, and without their active participation the very purpose of subject gateway will die.

11.3 Collaboration issue

There is a various levels where collaboration may take place but which one is most suitable for our purpose is not clear as it is still not clear that whether collaboration will be fruitful between national initiatives or not where the extension of a particular services are same.? Whether it may be within a particular country, and then within subject? How much is the subject areas are susceptible to international treatment?

Another aspect here is that of “branding”. Gateways aspire to high quality “brands”. And the successful collaborative arrangements will be done only if they don’t compromise on the brand value of individual initiatives, which is unlikely to happen.

11.4 Interoperability issue

It is the heart of all the issues and the areas where interoperability and standard solution are required are: protocols, query languages and record syntaxes, metadata schemata, subject classification and cataloguing rules. Without solving these problems we cannot dream of a seamless flow of information in a subject gateways.

11.5 Integration and cross searching issue

The approach of searching remote databases often in addition to a local database is referred to as cross searching. If a system allowed several gateways in the same subject area to be cross-searched, then inevitably there would be some duplication of results, as some source would be catalogued by more than one gateway. There is another aspect, as different subject gateways are having different collection development policies, so the combined results of a cross-search will contain links to resources of a differing minimum “quality” resource description. So again it is somewhat similar situation to what today’s search engines are encountered with.

11.6 Quality issue

The sense of the word quality changes with the circumstances and in the context of architecture, standards and software solutions of subject gateways, the use of uniform resource names (URN) is also having two major issues. *First*, the infrastructure needed to support the full requirements for URNs is currently not available. *Second*, the support for URNs by the major web browser developers has been low (if not non-existent). This

means that end users are unlikely to start to create and distribute URNs if the majority of typical web browser users will be unable to resolve them. So there is no fun at all to build such an ambitious subject gateways if it will not be browseable by the users end effectively and smoothly.

11.7 Multilingual issue

As the Internet has been flourishing outside the English speaking countries, so we have to take care of metadata properly in a multilingual platform. Here, the character-set issue is most important. If all the cataloguers and users make use of multiple languages and character sets, it may run the danger of increasing “false positives”, where the same word means different things in different languages.

12. Conclusion

The successful future of subject-gateways on the web is clearly one of the collaborative dynamism ultimately delivering a service, which enables the web access to a comprehensive array of high quality data, and value-added service via a single, unified access point. A coherent and seamless service by a large number of subject gateways is a future ambition where additional value added features may be included to make subject gateways a “**mature gateways**”, which will be offering- local document storage and archives; mirrors of remote gateways; reference section or services; large specialized databases in a subject area; full text access to documents; community creating effort such as: discussion list, links to experts, event calendars, news, bookshops, or job indexes, etc...

The emergence of “**brokerage services**” which provides a unified use of heterogeneous other network services is clearly taking a stake in this arena.. The ultimate goal is of course to allow cross searching and cross browsing between gateways in the same, similar or different subject areas, across one or several languages and types of services. A seamless and coherent discovery architecture for a large number of subject gateways is a further goal, and to achieve it, an “**International z39.50 specification for Library and Resource Discovery**,” currently known as the “*Bath profile*” is working which will provide the means for effective searching across OPACs, gateways and other information services. The work is also being underway on PURL (Persistent Uniform Resource Names) to drive the quality standards. The subject gateways have pioneered a service model, which will continue to develop in years to come, although it has to overcome a titanic hurdles, but with the advent of newly IT (Internet Technology) it is very much in reality.

13. References

1. Dempsey, Lorcan.(2000) “The subject gateway: experiences and issues based on the emergence of the Resource Discovery Network”, *Online information review*, vol. 24 (1), pp. 8-23.

2. Koch, Traugott. (2000) “ Quality-controlled subject-gateways: definitions, typologies, empirical overview”, *Online information review*, vol.24 (1), pp.24-34.
3. Gardner, Tracy. and Iannella, Renato. (2000) “Architecture and software solutions”, *Online information review*, vol.24 (1), pp.35-39.
4. Heery, Rachel. (2000) “ Information gateways: collaboration on content,” *Online information review*, vol. 24 (1), pp. 40-45.
5. Hiom, Debra. (2000) “ SOCIG: an Internet hub for the social-sciences, business and law,” *Online information review*, vol. 24 (1), pp. 54-58.
6. Place, Emma.(2000) “ International collaboration on Internet Subject Gateways,” *IFLA Journal*, vol. 26 (1), pp.52-56.
7. Hawkins, Donald T. and Levy,Louise R. (1988) “ Knowledge Gateways: The Building Blocks,” *Information processing & Management*, vol. 24 (4), pp. 459-468.
8. Mackie, Morag and Burton, Paul F. (1999) “The use and effectiveness of the eLib subject gateways: a preliminary investigation,” *Program*, vol. 33 (4), pp. 327-337.
9. Janes, Joseph W. and Rosenfeld, Louis B. (1996) “Networked Information Retrieval and Organization : issues and Questions,” *JASIS*, vol. 47 (9), pp. 711-715.
10. URL: <http://www.dlib.org/dlib/december99/12dempsey.html>
11. URL: <http://www.dlib.org/dlib/january98/01kirriemuir.html>
12. URL: <http://www.ilt.bris.ac.uk/roads/papers/w4gpospa.html>
13. URL: <http://www.sosig.ac.uk>
14. URL: <http://www.desire.org>
15. URL: <http://www.w3.org>
16. URL: http://www.exploit-lib.org/issue3/multilingual_gateways/
17. Dempsey, L. and Herry, R. (1998) “Metadata : a current view of practice and issues,” *Journal of Documentation*, vol.54 (2), pp.145-172.
18. Kerr, L. and MacLeod, R. (1997) “ EEVL: an Internet gateway for engineers,” *Library Hi Tech*, vol. 15 (3-4), pp. 110-118.