

Gerry McKiernan

Open Archives Initiative Data Providers. Part I: General

As observed by Ginsparg, the technological advances of the Internet and the lack of initiative on the part of conventional journals in response to the electronic revolution rendered the development of his e-print archive “an accident waiting to happen” (McKiernan, 2000, p. 128).

arXiv.org

More than a decade ago, Paul Ginsparg, then a particle physicist with the Los Alamos National Laboratory, New Mexico, developed arXiv.org, an Internet-based service that enabled authors to store and access preprints – pre-publication versions of their work – from a central location. While the system was intended originally only to organize a haphazard and unequal distribution of electronic preprints in the field of high-energy physics, it quickly became a primary means of communicating ongoing research within several select scientific communities. Perhaps more significantly, arXiv.org also directly and indirectly inspired others to establish their own e-print servers or to develop other alternative forms of scholarly publishing.

In this first of a series, we profile more recently established Open Archives Initiative (OAI) Data Providers whose content is not only “harvestable” by OAI Service Providers (www.openarchives.org/service/listproviders.html) (see also McKiernan, 2003a, b, 2004b), but perhaps more importantly, offer open access to institutional and discipline information resources in a wide variety of publication and media formats.

Digital Library of the Commons

The Digital Library of the Commons (DLC) (<http://dlc.dlib.indiana.edu>)

“provides free access to an archive of international literature on the commons, common-pool resources and common property.” The “Commons” is “a general term for shared resources in which each stakeholder has an equal interest.” “Studies on the commons include the information commons with issues about public knowledge, the public domain, open science, and the free exchange of ideas . . .” “Common-pool” resources (CPRs) are defined as “natural or human-made resources where one person’s use subtracts from another’s use, while “common property” is defined as “a formal or informal property regime that allocates a bundle of rights . . . Such rights may include ownership, management, use, exclusion, access of a shared resource” (<http://dlc.dlib.indiana.edu/cpredef.html>).

The DLC site includes such features as: “advanced searching; browsing by region, sector, and author name; an author submission portal for uploading a variety of document formats; and a service that uses email to alert subscribers to new documents in their area of interest.” To “search the digital library and archives”, the user can use any free-text term or phrase, or perform a subject search using appropriate words and phrases from a “Common Pool Resources Keyword Thesaurus” (<http://dlc.dlib.indiana.edu/thesaurus/thesaurus.html>). The thesaurus not only provides access to a controlled vocabulary, but also includes entries for related words (e.g. “MACKEREL” sa [see also] “fisheries”) (http://dlc.dlib.indiana.edu/thesaurus/thes_m.html).

Terms and phases may also include standard subdivisions (e.g. “case study”, “history”, “study and teaching”). In the basic search, only the title, assigned keywords, and abstract text are searched. The user can combine two or more terms in a Boolean and relationship (“match all, in any order”), in a Boolean or relationship

(“match any”), or search the terms as a phrase (“match as a phrase”) by selecting the option from an adjacent drop-down menu.

The DLC also provides an “advanced search” (<http://dlc.dlib.indiana.edu/perl/advsearch>) to search (or limit) by one or more fields, notably:

- author(s);
- year (or range of years);
- title;
- series;
- keywords;
- language;
- region or sector;
- discipline;
- country;
- agency;
- Conference; and/or
- submission date.

For a keyword field search, users can consult the “Keyword Thesaurus” from which terms and phrases can be copied and pasted into the search field. If desired, users can search or limit a query to documents written in one (or more) common Western languages (i.e. English, French, German, Italian, Portuguese, or Spanish). And likewise, users can conduct or limit a search to one (or more) regions (e.g. Africa, Central America and Caribbean, North America) or resource sectors (e.g. Agriculture, Information and Knowledge, Urban Commons). The DLC site includes an information page, which, through examples, defines the nature and scope of each sector (<http://dlc.dlib.indiana.edu/contentguidelines.html>).

Users can also search or limit a query by discipline by inputting a known discipline (e.g. African Studies, Chemistry, Economics) or consult, copy, and paste from a hotlinked “Discipline Thesaurus”. The user can also search the collection or limit it to a “country” (or countries) discussed in

the source document by entering the name of the particular country or by copying and pasting those of interest from a hotlinked "Country Thesaurus". As with the basic search, users can combine two or more terms in a Boolean and relationship ("match all, in any order"), in a Boolean or relationship ("match any"), or search the terms as a phrase ("match as a phrase") by selecting an option from a drop-down menu found adjacent to the input window for each field.

Using a drop-down menu beneath the last field option (i.e. "submission date"), users can specify that "Retrieved records must fulfill 'all' of these conditions" (default), or that all of the fields be searched in a Boolean or relationship ("any"). Search results can be displayed in order "by author's name" (default), "by title", "by year (most recent first)", or "by year (oldest first)".

The DLC can also be browsed by "region" or "sector" (see Figure 1), or by "author" or recent submission dates ("recent submissions").

Among the most notable of the features, content, or services of the DLC is its author self-archiving service (<http://dlc.dlib.indiana.edu/submit.html>) or (<http://dlc.dlib.indiana.edu/userservices>.

html to providing access to this "Working paper archive of author-submitted papers, as well as full-text conference papers, dissertations, working papers and pre-prints, and reports," the DLC also offers access to:

- a variety of general and specialized bibliographies on the Commons;
- Commons-related full-text articles and books not incorporated within the DLC;
- electronic journals and newsletters that focus on Common Property Resources;
- the home page of the International Association for the Study of Common Property (IASCP); and
- links to Commons-related Web sites.

The range of subject bibliographies covers such diverse topics as "co-management", "indigenous knowledge and institutions", and "property rights" (www.indiana.edu/~iascp/subbibs.html), while a searchable database titled *The Comprehensive Bibliography of the Commons* (www.indiana.edu/~iascp/lforms/searchcpr.html) includes more than 40,000 citations, 5,000 of which include full abstracts (March 2004). The specialized bibliographies, as well

as the database content, were compiled by Charlotte Hess (hess@indiana.edu), the DLC Director and Librarian for the Workshop in Political Theory and Policy Analysis at Indiana University (www.indiana.edu/~workshop/). The DCL site also provides access to commons-related full-text articles and books (not in the DLC) on a number of topics that include "agriculture", "fisheries", "forestry", as well "information and knowledge", "land tenure and use", and "wildlife" (www.iascp.org/articles.html).

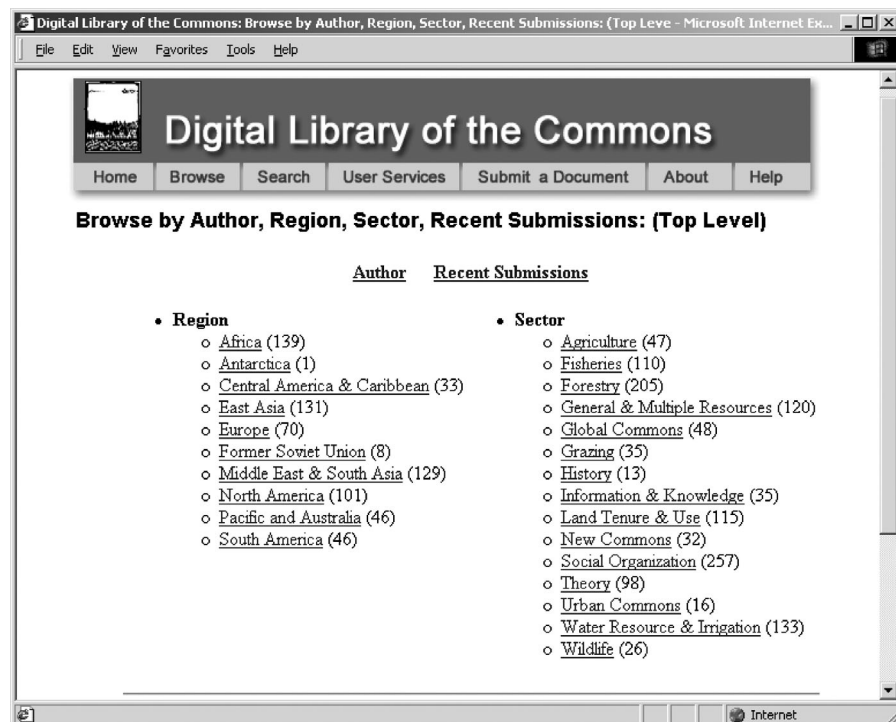
In addition to its director, staff of the DLC include John Walsh, Manager, Electronic Text Technologies, Indiana University Digital Library Program; Ying Jin, Software Analyst/Programmer, Indiana University Digital Library Program; and Laura Wisen, DLC Coordinator and Library Coordinator, Workshop in Political Theory and Policy Analysis (<http://dlc.dlib.indiana.edu/about.html>). The DLC is a collaborative project of the Workshop in Political Theory and Policy Analysis and the Indiana University Digital Library Program with funding from a variety of sources.

In March 2004, DLC is scheduled to upgrade to GNU EPrints software (v.2) (<http://software.eprints.org>). GNU EPrints (v.2) was developed by Christopher Gutteridge, system administrator with the Systems Group, School of Electronics and Computer Science, University of Southampton (UK), and doctoral student in the Intelligence, Agents, and Multimedia Group, with the assistance of Mike Jewell, a doctoral student in the Image, Speech, and Intelligent Systems Research Group at the School (<http://software.eprints.org/credits.php>). GNU EPrints was developed as part of the Open Citation Project (<http://opcit.eprints.org>), a Digital Library Initiative Phase 2 (DLI2), International Digital Libraries Project funded by the Joint Information Systems Committee (JISC) of the Higher Education Funding Councils, in collaboration with the National Science Foundation (NSF) (McKiernan, 2004b).

E-LIS: e-prints in library and information science

E-LIS: e-prints in library and information science (<http://eprints>).

Figure 1
Screen print of "top level" browse options in the Digital Library of the Commons



rclis.org) is an open access archive that offers the full text of published or unpublished scientific or technical documents in librarianship, information science and technology, and related disciplines. E-LIS uses the GNU EPrints software (see above) and thus allows its repository content to be “visible, accessible, harvestable, searchable and useable by any potential user with access to the Internet” (McKiernan, 2004a). As of mid-February 2004, there were more than 450 items deposited in the E-LIS collection (see Figure 2).

Within E-LIS, users are provided with three search options: basic, “simple”, and “advanced”. In a basic search – the default search option found on the site’s homepage – users can “search titles, abstracts or keywords”. Using the “search” option linked from the homepage (“simple search”), users can search these fields separately (“title/abstract/keywords”), or in combination with the name of an author or editor (“authors/editors”) and/or a “year” (or range of years). For the former two fields, users can specify that the terms (or phrase) be searched as a Boolean or relationship (“match any”), or as a Boolean and (“match all, in any

order”) by selecting from the drop-down menu located to the right of the query window. In addition, users can specify that the field data be combined in a Boolean and relationship (“retrieved records must fulfill ‘all’ of these conditions”), or in a Boolean or relationship (“retrieved records must fulfill ‘any’ of these conditions”) by selecting the option from a drop-down menu found beneath the search fields. Likewise, users can “order the results” “by title” (default), “by year (most recent first)”, “by author’s name”, or “by year (oldest first)” by selecting the option of interest from a secondary menu.

E-LIS also provides an “advanced search” (<http://eprints.rclis.org/perl/advsearch>) that allows the user to search (or limit) by one or several fields, notably:

- title;
- authors;
- abstract;
- keywords;
- subjects (JITA classification (see below));
- EPrint type (e.g. “conference paper”);
- conference;

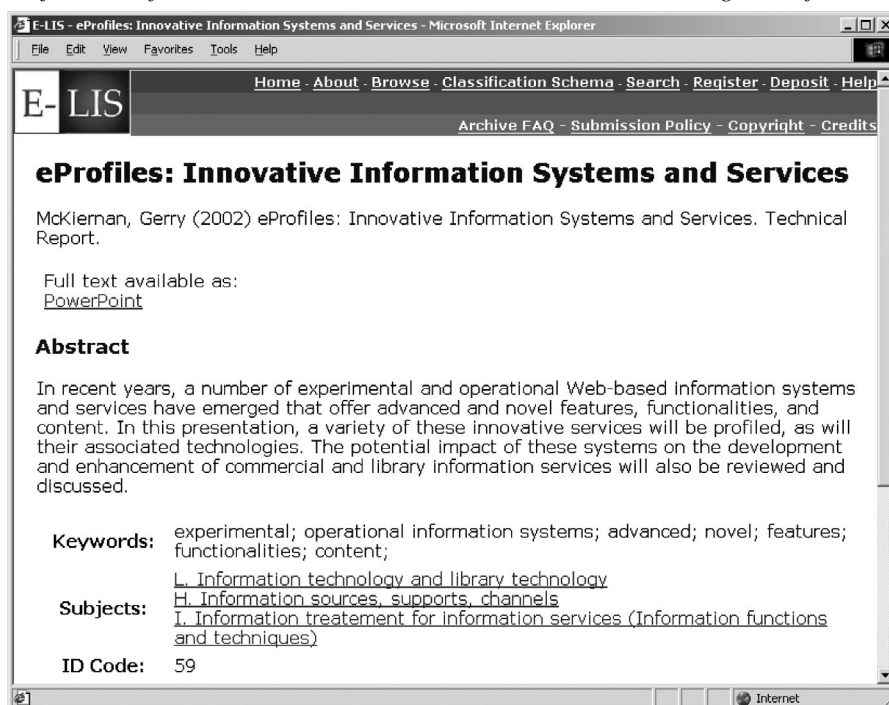
- department;
- editors;
- status;
- refereed (“no preference”, “yes” or “no”); and
- year (or range of years).

In addition to offering basic and advanced search features, E-LIS allows the user to browse its collection using a specialized classification scheme known as the JITA Classification System of Library and Information Science (<http://eprints.rclis.org/JITASchema.html>). JITA includes the following broad categories and associated letter codes:

- *A. Theoretical and general aspects of libraries and information.* Library and information science as a field. Information theory and library theory. Relationship of library and information science with other fields such as computing and communication science and cognitive science.
- *B. Information use and sociology of information.* Use and impact of information. Bibliometric methods. Information in society. Information society. Information economics. Information policy. Information dissemination and diffusion. Information needs and information requirements analysis. User interfaces, usability.
- *C. Users, literacy and reading.* Use studies. User studies. User categories: children, young people, social groups. User training, promotion, activities, education. Literacy. Reading and story telling.
- *D. Libraries as physical collections.* World libraries. National libraries. Public libraries. Academic libraries. School libraries. Government libraries. Private libraries. Special libraries. Science libraries. Technical libraries. Health libraries, Medical libraries. Archives. Museums.
- *E. Publishing and legal issues.* Mass media. Printing, electronic publishing, broadcasting. Bookselling. Intellectual property: author’s rights, ownership, copyright and copyleft. Intellectual freedom. Censorship.
- *F. Management.* Co-operation. Marketing. Finance. Public relations. Personnel management.

Figure 2

An E-LIS record includes the title, author, date, abstract, assigned keyword and subject classification codes, and alternative Web locations, among other fields



Funding. Local government. Reorganisation. Unitary authorities.

- *G. Industry, Profession, and Education.* Information industry. Software industry. Computer and telecommunication industry. Organisations. Staff. Biographies. Curricula aspects. Education. Training.
- *H. Information sources, supports, channels.* Periodicals, newspapers. Gray literature. Archival materials. Rare books and manuscripts. Print materials. Microforms. Non-print materials. Audio-visual, multimedia. Electronic media. CD-ROM. On-line hosts. Databases and database networking. OPAC's. e-journal. E-books. E-resources. Web pages. Portals. Repositories (OAI-compliant and not).
- *I. Information treatment for information services (information functions and techniques).* Cataloguing, bibliographic control. Content analysis: abstracting, indexing, classification. Index languages, processes and schemes. Data and metadata structures. Knowledge representation. Information transfer: protocols, formats (XML), techniques. Information presentation: hypertext, hypermedia. Image systems. Filtering. Reference linking. Design, development, implementation and maintenance of information systems and services.
- *J. Technical services in libraries, archives and museums.* Acquisitions. Serials management. Withdrawals. Stocktaking. Record keeping. Paper preservation. Digitization. Digital preservation. Circulation. Document delivery. Interlibrary loans.
- *K. Housing technologies.* Resource centres. Library, archive and museum buildings. Furniture. Vehicles. Architecture. Planning, design, removal. Safety. Disaster planning.
- *L. Information technology and library technology.* Telecommunications. Computer networking. Internet, including WWW. Computers. Scanners. Digital cameras. Photocopiers. Computer and network security. Authentication, and access control. Software. Software methodologies and engineering. Automated language processing. Automatic text retrieval. Data base

management systems. Object-oriented DBMS. Intelligent agents. Library automation systems. OPAC systems. Search engines.

The JITA Classification Schema was developed from the NewsAgentTopic Classification Scheme (maintained by Mike Keen at Aberystwyth, UK) and the RIS classification scheme of the *Review of Information Science* (www.inf-wiss.uni-konstanz.de/RIS/) originally conceived by Donald Soergel (University of Maryland). "JITA" is an acronym derived from the first initials of the first names of the schema developers: José Manuel Barrueco Cruz (Universitat de València, Spain), Imma Subirats Coll (Government of Catalonia, Spain), Thomas Krichel (Palmer School of Library and Information Science, Long Island University, USA), and Antonella De Robbio (Università degli Studi di Padova, Italy).

E-LIS has been established as a community service by Research in Computing, Library and Information Science (RCLIS) (rclis.org) to promote open access to papers in these fields. RCLIS itself seeks to build "a database about all current and past research in computing and library and information science" and to make it "freely available for public and private, commercial and non-commercial purposes." Librarians and other scholars can deposit a variety of documents (e.g. preprints, postprints, conference paper, presentations, technical reports, and working papers) in E-LIS in a range of formats (PDF, PostScript, TeX, LaTeX, HTML, XML, ASCII (text), MS Word, PowerPoint, or RTF). Registration is required prior to the first deposit, and all submissions are reviewed by an editorial board before formal incorporation with the E-LIS collection. A RCLIS initiative related to E-LIS is DoIS: Documents in Information Science (dois.mimas.ac.uk), a "database of articles and conference proceedings published in electronic format in the area of Library and Information Science" (www.uksg.org/serials/coll.pdf).

The E-LIS service is hosted by CILEA, the Consorzio Interuniversitario Lombardo per l'Elaborazione Automatica, a non-profit consortium of nine universities in the Lombardia Region of Italy, and has been funded by Ministero de Educación, Cultura y

Deporte's Dirección General de Universidades, Programa de Estudios y Análisis (Spain).

The E-LIS repository can also be searched and browsed using an alternative interface provided by the Collection of Computer Science Bibliographies site (<http://iinwww.ira.uka.de/bibliography/Misc/E-LIS.html>).

INFOMINE: scholarly Internet resource collections

INFOMINE (<http://infomine.ucr.edu/>) "is a virtual library of Internet resources relevant to faculty, students, and research staff at the university level. It contains useful Internet resources such as databases, electronic journals, electronic books, bulletin boards, mailing lists, online library card catalogs, articles, directories of researchers, and many other types of information" (<http://infomine.ucr.edu/?view=about>). Inaugurated in January 1994 by the Library of the University of California, Riverside, INFOMINE was "one of the first Web resources of any type offered by a library", "one of the first Web-based, academic virtual libraries", as well as "one of the first to develop a system [that combined] the advantages of the hypertext and multi-media capabilities of the Web with those of the organizational and retrieval functions of a database manager" (see also Mitchell and Mooney, 1996; Mason *et al.*, 2000).

INFOMINE has incorporated a wide variety of Internet resources into one (or more) broad subject categories or formats that include:

- biological, agricultural and medical sciences;
- business and economics;
- cultural diversity and ethnic resources;
- electronic journals;
- government information;
- maps and GIS;
- physical sciences, engineering, computing and math;
- social sciences and humanities;
- visual and performing arts.

As of mid-February 2004, there were more than 115,700 resources incorporated within INFOMINE (<http://infomine.ucr.edu/>), with more than 31,000 prepared by information specialists ("expert created content") (<http://infomine.ucr.edu/?view=about/>

content.html) and the remainder “robot/crawler created”. Indeed, INFOMINE includes “focused, automatic Internet crawling as well as automatic text extraction and metadata creation functions to assist ... [its] experts in content creation and users in searching” (<http://infomine.ucr.edu/?view=about/about.html>).

Within INFOMINE, users are offered a variety of basic and advanced search and browse options. From its homepage, users can perform a keyword search using terms or phrases to search all but one of the full record fields (“description”). “Different searches can be combined and/or complex searches run by using various Boolean and proximity operators in search statements. [NEAR] ... [range = near 1 through near 20], [NOT] ... , [AND] ... as well as ... [OR] are the operators supported by INFOMINE” (<http://infomine.ucr.edu/?view=help/index.html>). Search results are displayed in order by relevance “score” (default) and include a hotlinked entry for the resource, a summary description, an associated relevancy score, a link to the full INFOMINE record for the resource (“more info ...”), and a link to a Web form that allows the user to “comment on this resource”. Among its fields, the full INFOMINE record provides:

- Resource title (e.g. “All that JAS: Journal Abbreviation Sources”).
- a summary description or excerpt (e.g. “All That JAS: Journal Abbreviation Sources” is a registry of Web resources that list or provide access to the full title of journal abbreviations or other types of abbreviated publication titles (e.g. conference proceedings titles). Selected OPACs that offer abbreviated title searching have also been included. In addition, access to full journal title directories or lists is also provided.”)
- Assigned Library of Congress subject heading(s) (e.g. “periodicals – abbreviations of titles”)
- Assigned Library of Congress Classification range (e.g. “Z 5051 – 7999”).
- Assigned “keywords” (e.g. “abbreviations; ejournal abbreviations; JAS; journal abbreviations;

journals; OPACS; periodicals; reference resources; serials”).

- Authors (e.g. “Iowa State University Library, Ames; McKiernan, Gerry”).
- Other useful information about this resource (e.g. “URL: <http://www.public.iastate.edu/~CYBERSTACKS/JAS.htm> | Access: free | Resource Types: Acronyms & Abbreviations; Ejournal; Reference resources.
- INFOMINE catalog information (e.g. Record Number: 24310 created by: wendie of UC Riverside at 2000-09-08 00:00:00. Modification: Last modified by linear of UC Riverside at 2003-12-12 10:23:27 Record Origin: This is a local, expert-created record. Subject Categories: Electronic Journals Social Sciences & Humanities”).

At the top of the original search results page, the user is offered the option of adding “robot-selected resources” (“add robot-selected resources?”) to the original search results of “expert-selected resources” by clicking on an associated button. A black mortarboard icon indicates expert-selected resources (“selected by librarians”), while those that are robot-selected are indicated with a robot-like icon (“selected automatically”). Users can also display the search results as an alphabetized title listing by clicking on a “titles display” button, or modify a search or initiate a “new search” by clicking on the respective buttons located in the right hand corner of the results page adjacent to the titles display option. Alternatively, users can access an “advanced search and browse” page by clicking on the “all subject categories” link found above these latter options.

In the bottom half of the original search results page, users are provided with the option of modifying the original search statement by adding (or removing) candidate search fields (“search in these fields”) and/or broad subject categories (see above) (“search in these subject categories”) by checking or unchecking appropriate options. In addition, users can “limit [the] search to” “record origin” (“expert created” (default) or “expert + robot”), and/or “resource access” (“all” (default), or “free” or “fee”), and/or “resource

types” (i.e. “instructional”, “databases”, “bibliographic databases”, “datasets”, “ebooks”, “ejournals”, “statistics”, “subject directories”, “digital libraries”, “virtual libraries”, and/or “search engines”) by selecting the aspect of interest. Users are also provided with a variety of “display options for search results” that enable them to “display” results in a “regular” brief record format (default) or alphabetically by title (“titles”), and rank them by “relevance” (default) or alphabetically by title (“titles”).

In addition to basic search options, INFOMINE users are offered advanced search and browse options within defined subject categories and formats (see above). In the advanced search mode, users are provided with field, format, and display options identical to those provided at the bottom of a search results page (see above and Figure 3). Users are also provided with advanced browse options for the resources organized within a subject category. For any category, users can “browse within the highlighted indexes by clicking on them” (see bottom Figure 3). These include:

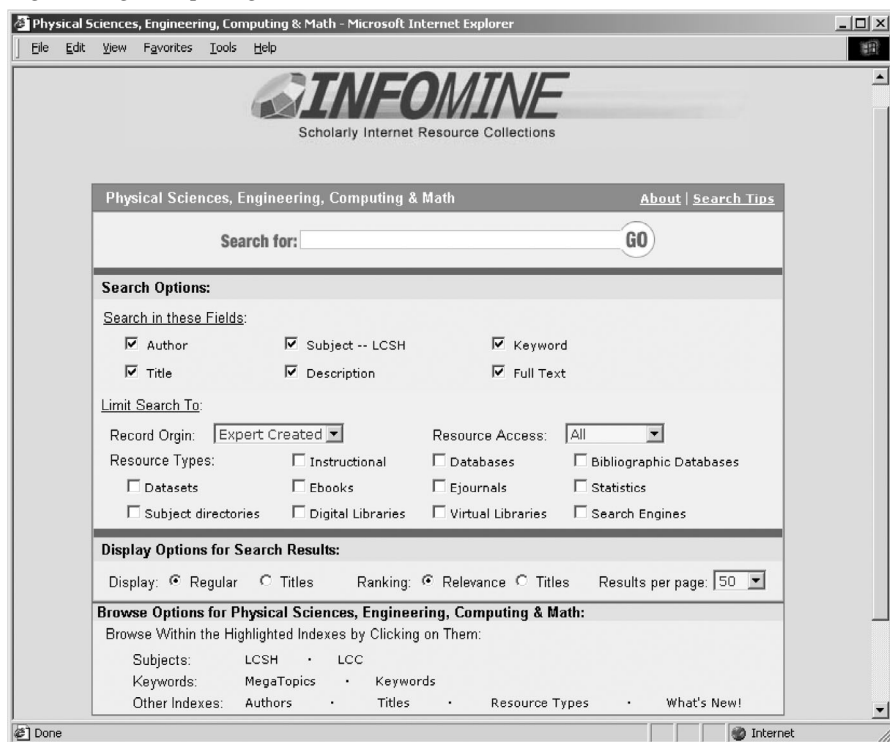
- “Subjects:” “LCSH” and “LCC”.
- “Keywords:” “megatopics” and “keywords”.
- “Other indexes:” “authors”, “titles”, “resource type” and “what’s new”.

The metatopics feature is among the most novel in INFOMINE and is intended to allow the user to “explore the phrases in INFOMINE”. Currently, the title, author, keyword, subject (LCSH), and description metadata are indexed in the megatopics function, which requires a Java-capable browser”. The metatopics function is based on the Phind phrase browser originally developed at the Department of Computer Science at the University of Waikato (infomine.ucr.edu/?view=projects/phind/).

As to be expected, “subjects”, “keywords” and “other indexes” browsing options are specific to the particular subject category. For example, browsing the Library of Congress Subject Headings (LCSH) in the “Physical sciences, engineering, computing and math” subject group will only display the physical sciences,

Figure 3

Screen print showing advanced search page for the “physical sciences, engineering, computing and maths” collection in INFOMINE



engineering, computing, and mathematics LCSH terms and phrases assigned to the resources incorporated within this INFOMINE discipline group, and likewise for the other indexes. Users can browse the subject headings, Library of Congress classification ranges, authors, titles, and other indexes for the entire INFOMINE collection by clicking the “all subject categories” advanced search and browse hotlink located above a search results listing.

The INFOMINE development team is led by Steve Mitchell (smitch@citrus.ucr.edu), Science Reference Librarian, INFOMINE Co-Coordinator and Managing Editor, IMLS Grant Project Director, Library, University of California, Riverside and colleagues at UCR. In addition, more than 24 individuals affiliated with the library systems of the University of California system, Wake Forest, and the University of Detroit serve as contributors to the INFOMINE collection (<http://infomine.ucr.edu/?view=about/participants.html>).

iVia (<http://infomine.ucr.edu/iVia/>) is an open source Internet subject portal or virtual library system, was developed and serves as the platform for

INFOMINE. “As a hybrid expert and machine built collection creation and management system, it supports a primary, expert-created, first-tier collection that is augmented by a large, second-tier collection of significant Internet resources that are automatically gathered and described.” “The software enables institutions to work cooperatively or individually to provide well-organized, virtual library collections of metadata descriptions of Internet and other resources, as well as rich full-text harvested from these resources.” “iVia as funded by the National Science Digital Library program (National Science Foundation) (<http://nsdl.org>), the National Leadership grant program of the Institute of Museum and Library Services (US) (IMLS) (www.imls.gov), the Fund for the Improvement of Post-Secondary Education (US Department of Education), and the Library of the University of California, Riverside” (Mitchell *et al.*, 2003). The iVia software is available free-of-charge under the iVia Software Affero General Public License (AGPL) (<http://infomine.ucr.edu/iVia/ivia.php?section=2>) (<http://infomine.ucr.edu/download/>).

A new project funded by IMLS known as Data Fountains (http://infomine.ucr.edu/Data_Fountains/) will create an array of iVia installations that can be used by cooperating projects to generate metadata for Internet resources new to their collections or to augment older metadata. Data Fountains is “A national, cooperative information utility for shared Internet resource discovery, metadata application and rich, full-text harvest of value to Internet portals, virtual libraries and library catalogs with portal-like capabilities” that uses the full and semi-automated approaches to crawling and classification incorporated within iVia. As with iVia, the Data Fountains software is open source.

Open video project

“The purpose of the Open Video Project (www.open-video.org) is to collect and make available a repository of digitized video content for the digital video, multimedia retrieval, digital library, and other research communities.” The repository can be used by researchers “to study a wide range of problems, such as tests of algorithms for automatic segmentation, summarization, and creation of surrogates that describe video content; the development of face recognition algorithms; or [the creation and evaluation of] . . . interfaces that display result sets from multimedia queries” (see also Marchionini *et al.*, 2002; Slaughter *et al.*, 2000). In addition, the Open Video Project will of value to various educational communities.

Inaugurated in 1998 with an initial content of nearly 200 video segments, the repository has grown to include a select collection of videos presented at various Conference on Human Factors in Computing Systems sessions; select videos from the Informedia Project (www.informedia.cs.cmu.edu) at Carnegie Mellon University; the Digital Himalaya Project; and the Internet Moving Images Archives based on the Prelinger Archives (www.prelinger.com); as well as the University of Maryland HCIL Open House Video Reports (www.cs.umd.edu/hcil/pubs/video-reports.shtml).

With the Open Video Project, users can perform a basic keyword search using the query box located on the project homepage, or a “detailed search”

(www.open-video.org/detailed_search.php), which allows users to specify a search field (“title”, “description”, “keywords”, or “transcript”) by selecting the option of interest from a drop-down menu. In addition, users can limit a search by various “attributes” that include “genre”, “duration”, “format”, “color”, “sound”, “language”, and “creation date” (see Figure 4).

In addition, the Open Video repository can also be browsed by “genre”, “duration”, “color”, “sound”, or collection (e.g. “Internet moving images archive”) (www.open-video.org/index.php) (see Table I).

In addition to search and browse options, the Open Video Project home page includes a “Special collection spotlight”, linked “Project news” entries, and a “Featured video” collection that includes entries for a featured “new” video, “other new videos”, and a select “popular” video and entries for “other popular videos”.

On the Open Video Project homepage, the number of records in the video repository that match a particular attribute is indicated in parentheses to the right of a collection grouping (e.g. “Genre>Documentary [494]”). On browsing, the user is presented with a

page that provides a variety of other browse and search options. In a browse display, the results are displayed in order by relevancy (default); the user can resort the results by selecting from available options offered in a drop-down menu (i.e. “title”, “year”, “duration”, “popularity”). For each entry, the genre, assigned keywords, duration, and popularity (number of downloads) are displayed to the right of a large thumbnail image of a sample image from the video set.

Users can redisplay the default sorted set in groups of ten records (default), or display results in groups of 20, 30, 50, or “all” records by selecting from an adjacent drop-down menu. The default can be accepted layout or changed by clicking the icon associated with one of three other layout options. For any layout, the total number of retrieved records is noted above the results display (“Search results: (494 videos found)”). The entire collection can be browsed in sequence by clicking the appropriate navigation arrow found next to the search results total.

Within the browse (or search) results page, users can also “modify [a] search” by inputting additional keywords of interest, accepting or changing the genre, duration, and/or format defaults, and/or accepting the options for color and/or sound.

On clicking the hotlinked thumbnail from the original results display, a full record is displayed in a separate window (see Figure 5), and provides the following “Video information” and “Digitization information”:

- year (if known);
- genre (e.g. “documentary”);
- keywords (e.g. “NASA, space”);
- duration (“00:00:30”);
- color (e.g. “yes”);
- sound (e.g. “yes”);
- amount of motion (e.g. “medium”);
- language (“English”);
- sponsor (e.g. “National Archives”);
- contributing organization; University of Maryland, College of Information Studies);
- transcript available (e.g. “no”);
- digitization date (e.g. “1998”); and
- digitization organization (e.g. “University of Maryland”).

Figure 4

Matrix of “attributes” by which an Open Video repository search can be limited

Genre	Duration	Format
<input checked="" type="radio"/> Any Genre	<input checked="" type="radio"/> Any Duration	<input checked="" type="radio"/> Any Format
<input type="radio"/> Documentary	<input type="radio"/> Less than 1 minute	<input type="radio"/> MPEG-1
<input type="radio"/> Educational	<input type="radio"/> 1 to 2 minutes	<input type="radio"/> MPEG-2
<input type="radio"/> Ephemeral	<input type="radio"/> 2 to 5 minutes	<input type="radio"/> MPEG-4
<input type="radio"/> Historical	<input type="radio"/> 5 to 10 minutes	<input type="radio"/> Quicktime
<input type="radio"/> Lecture	<input type="radio"/> More than 10 minutes	
<input type="radio"/> Other		
Color	Sound	Language
<input checked="" type="radio"/> Either	<input checked="" type="radio"/> Either	<input checked="" type="radio"/> Any Language
<input type="radio"/> Color	<input type="radio"/> Sound	<input type="radio"/> English
<input type="radio"/> B&W	<input type="radio"/> Silent	<input type="radio"/> French
		<input type="radio"/> Spanish
Creation Date		
<input type="radio"/> From <input type="text"/> example: 1965		
<input type="radio"/> From <input type="text"/> to <input type="text"/> example: 1965 to 1990		

Table I

Users can browse the Open Video Project by genre, duration, and availability of color or sound

Genre	Duration (minutes)	Color	Sound
Documentary	Less than 1	In color	With sound
Educational	1-2	In black and white	Silent
Ephemeral	2-5		
Historical	5-10		
Lecture	More than 10		
Other			

“INFOMINE: promising directions in virtual library development”, *First Monday*, Vol. 5 No. 6, June, available at: www.firstmonday.dk/issues/issue5_6/mason/ (accessed 18 January 2004).

Mitchell, S. and Mooney, M. (1996), “INFOMINE: a model Web-based academic virtual library”, *Information Technology and Libraries*, Vol. 15 No. 1, March, pp. 20-5, available at: <http://infomine.ucr.edu/pubs/italmine.html> (accessed 17 January 2004).

Mitchell, S., Mooney, M., Mason, J., Paynter, G.W., Ruschinski, J., Kedzierski,

A. and Humphreys, K. (2003), “iVia open source virtual library system”, *D-Lib Magazine*, Vol. 9 No. 1, January, available at: www.dlib.org/dlib/january03/mitchell/01mitchell.html (accessed 18 January 2004).

Slaughter, L., Marchionini, G. and Geisler, G. (2000), “Open video: a framework for a test collection”, *Journal of Network and Computer Applications*, Vol. 23 No. 3, July, pp. 219-45, available at: <http://ils.unc.edu/%7Egeisg/info/Jnca0112.pdf> (accessed 17 January 2004).

FURTHER READING

Ginsparg, P. (1994), “First steps toward electronic research communication”, *Computers in Physics*, Vol. 8 No. 4, July/August, pp. 390-6, available at: <http://xxx.lanl.gov/ftp/hep-th/papers/macros/blurb.tex> (accessed 19 January 2004).

Gerry McKiernan (gerrymck@iastate.edu) is a Science and Technology Librarian and Bibliographer, Iowa State University Library, Ames, Iowa, USA.