2012-11-05

# The History and Current State of the Information Portal in Libraries

Richard Hacken
hacken@byu.edu

Mathew Miles
*Brigham Young University - Idaho*, mmiles@byui.edu

## The History and Current State of the Information Portal in Libraries

The impetus for linking computers together to serve scholarly workstations has been to forge an international network to which (and from which) information might flow freely. But the dream of such a data nexus – along with all the tools of connectivity, discovery and documentation a scholar is likely to need – has been developing in steps, piecemeal, and is still not yet fully realized.

### 1. DEFINING THE "PORTAL" AND ITS VARIANTS

Architecturally and therefore literally, a "portal" is some sort of gateway or even grand entrance – especially into an important edifice such as a fortress, mosque or cathedral – whose features make it possible to control entry and exit and are often highly ornamented. The figurative meaning of "portal" in the virtual realm may benefit more from the borrowed concepts of entrance into an important space and entry control than from ornamentation (though the latter could be tied to attractiveness or user friendliness). Thus in the broadest sense, while keeping in mind the architectural parallels, we could say that a "portal" represents a unique electronic port of entry to a wide range of resources, services or both that are needed within a particular information community.

The closest general term to "portal" is perhaps "information portal," though a given portal may contain ancillary tools, often as part of a complementary "application portal" (see section 1.1), that are not directly associated with information retrieval and dissemination. A "library portal" would be one type of "information portal" and would necessarily consist, at least in the current understanding, of at least metadata and data that can be locally or remotely hosted.

The "scholarly portal" or "scholarly information portal" distinguishes the academic from the commercial portal. The "enterprise portal" or "corporate portal" is congruent with the design of a scholarly portal but with a very different bottom line and user community. Both scholarly and enterprise portals increasingly rely on de-centralized provision and management of content, which ideally keeps the content current and minimizes local labor costs at any one institution.

We are tempted nowadays to use the phrase "web portal" as a synonym for "information portal," especially since web access is ubiquitous and forms the central engine of today's portals. As we shall see in the following historical section, however, early types of portals were local and mostly self-contained – though some of them could be accessed remotely, if only in awkward ways. Early library catalogs are one example of such early non-web portals. At the present time, though, a web portal can grant the information community (library users, company employees, etc.) access to the panoply of services involving electronic mail, discussion forum, digital publication space, proprietary databases, search engines, and more.

One of the strengths of a web portal is that it typically allows the user to authenticate a single time to access all the resources offered. This saves users the time and frustration of having to authenticate each time they access a different resource. Authenticated users are allowed to register and connect to the portal for using a set of proposed services, notable among them the customization of their own workspace, which is organized with the assistance of portlets, forming a basic man-machine interface (IHM : "*interface homme machine*"). For staff and, if desired, partners and affiliates, the web portal is thus the entry point for data retrieval within an organization. The aim of the portal is to situate the user centrally within the information system. In fact, the tools bundled in the portal allow the user:

- to have a single access point for multiple sources of information,
- to acquire information that is organized and structured,
- to shape and customize information to the user's interests and proposed tasks.

A "digital library" is one specific type of portal that can involve a combination of searchable metadata and retrievable digital objects, though it may or may not limit itself to locally held objects and may include some of the peripheral services listed under "web portals" above. The term "digital archive" is sometimes used for an analogous online collection. As one answer to the scholarly communications crisis (which is aggravated in a time of financial hardship), an "institutional repository" is a type of local digital archive for a home institution's professional publications – to the degree copyright clearance is possible.

A "federated search," involving the retrieval and combination of results from multiple information providers, may be a variety of search portal in its own right, but it is often used as one cornerstone in a more variegated online library services approach. That is to say, one type of "portal" may be contained within another "portal." Perhaps the term "P2P" that now refers to a "peer-to-peer" network might come to include the nested connotation of "portal-to-portal" access as well.

A "metadata repository" is a broad term that could include catalogs, digital libraries or portals that allow for searching of metadata. The item thus identified might be either digital or analog, either local or remote.

"Information systems" is perhaps the broadest term of all, which would subsume all types of "portals," "catalogs," "digital libraries" and "metadata repositories" along with other cyber-tools for information query, storage and retrieval.

Smith (2004) describes the term "portal" as problematic, since it has been applied to widely different systems – from static and simple web pages to dynamic and sophisticated virtual conglomerates. He argues for adoption of "a definition that distinguishes portals from all other types of information systems and a General Portal Model (GPM)." (Smith, 2004, p. 93). His suggested definition is that of "an infrastructure providing secure, customizable, personalizable, integrated access to dynamic content from a variety of sources, in a variety of source formats, wherever it is needed." (Smith, 2004, p. 94).

Kalyanaraman and Sundar (2008) seek to add some conceptual clarity to the portal concept by making functional differentiations. They uncover five related metaphorical notions for the portal – gateway, billboard, network, niche and brand – that in turn suggest five leading aspects of portal utility – respectively, control, content, community, customization and commerce.

For purposes of clarity, we will distinguish in the following article between the types of portals discussed in this section. Unless otherwise stated, we will generally be referring to an information portal tied to web services.

### 1.1. IDENTIFYING VARIANT FUNCTIONS OF PORTALS

We generally distinguish functionally between three broad portal families:

a. the information portal, granting access to multiple information sources (documents, reports, articles, internal or external databases, etc.) aggregated in a single point;
b. the application portal, granting access to the various applications of the organization with corresponding data relevant to the user profile; and
c. the appraisal portal, sometimes called the standby portal, enabling analysis of portal usage as a means of enhancing the organization's knowledge base.

### 1.2 NOTIONS OF PORTAL USER PROFILES AND CUSTOMIZATION

The web portal concept is closely linked to the notion of a user profile. Ideally, each user has access to resources of the information system that fit the user's unique profile, all in accordance with security policies defined by the organization. At the same time, the user profile can guide the way to customization (Chevalier et al., 2007), and so we speak of an "online work environment" or "virtual office." Following this terminology, the environment is constructed of modular bricks (generally called "portlets" or "webparts") from which the user can choose to organize relevant scholarly structures within his or her workspace.

Popular customizations being built into portals include personalized search history lists, RSS feeds and the ability to store bibliographic entries for later printing, e-mailing or saving.  The latter illustrates how a popular commercial idea (the virtual "shopping basket") has been adapted to scholarly uses.

## 2. HISTORY OF PORTALS

The emergence and spread of new Information and Communication Technologies (ICT) – notably of the Internet, of research networks and enterprise intranets – have abetted the growth of a phenomenon now justly referred to as the "information explosion." Scholars would be casualties of the information explosion were it not for support systems that can help to pinpoint and extract relevant info particles from the vast oceans of data (Mothe et al., 2003). The sorcerer's apprentice of technology that has overloaded our neural networks, has ironically, also provided cyber-networks to help control the flood it enabled. Over the past few years, the use of web technologies within organizations (companies, universities, libraries, and so forth) has become an essential element in managing different sectors of information.  The evolution of portals of all kinds shows an increasingly strong trend toward more loosely coupled modular systems.

### 2.1 LIBRARY CATALOG MODELS

#### 2.1.1 EARLY LIBRARY CATALOG MODELS

In the commercial online realm, some of the first portals, such as Yahoo!, were web directories; others were search engines like AltaVista.  Just as these were essentially searchable catalogs of "holdings," i.e., of documents inside the World Wide Web, likewise early library catalogs (many predating Yahoo! and AltaVista) contained metadata for individual library holdings.  As such, the holdings were limited to locally owned content. Search strategies were often command driven, requiring the user to memorize a command structure [e.g., "((A=Tolstoy or A=Tolstoi) and (A=Leo or A=Lev)) and  (T="voyna i mir" or T="war and peace")"].  In contrast to researchers in house, remote users – if authorized – required a complex combination of protocols to connect to the catalog.  Relating to esthetics as well as utility, early catalogs were not yet amenable to a graphic user interface (GUI). Yet these catalogs could be considered proto-portals, since they combined in one place the query potential for finding millions of analog items.

The overwhelming emphasis in early library catalogs was on metadata rather than data.  The metadata and the materials to which they pointed (mostly analog) had to be maintained locally.

### 2.1.2 CURRENT LIBRARY CATALOG MODELS

The website for a university or research library today invariably offers much more than a catalog. Advancements from the early catalogs to the current models are in the breadth of ancillary tools and guides (at times leading to an inadvertent de-emphasis on the catalog itself) as well as increasing use of remote metadata retrieval and burgeoning linkable presentation of digital objects. In this section we limit our observations to the catalog itself; more broadly conceived websites with additional portal elements will be discussed under "hybrid portals".

The current state-of-the-art catalog can offer attractive and intuitive interfaces that include images and external links while minimizing user aggravation with search jargon. Just as importantly, the scope of MARC records has expanded to locate holdings beyond the walls of the library and beyond tangible analog materials: e-books, web pages, electronic journals, streaming audio and video, and other local or remote information objects. The user is not only guided to material locations onsite but can also receive (view, see, hear, read) virtual copies of many items at the computer. The search interface yields to, and becomes, the study surface of that which was sought.

From the maintenance point of view, at least some of the metadata and holdings (analog or digital) are hosted elsewhere.

## 2.2 METADATA REPOSITORY MODELS

In the realm of design, metadata repositories have evolved in parallel fashion as library catalogs. They differ from library catalogs in not indexing a single local library or collection but an aggregate of holdings at numerous other sites. The main progress has been from metadata-only searches to the current search-and-link connections to remote digital holdings.

### 2.2.1 EARLY METADATA REPOSITORY MODELS

OCLC and RLIN are examples of early metadata-only repositories. These bibliographic utilities host metadata from a variety of sources, but do not own any content of their own. Typically the metadata does not link you out to content. It simply provides information about where the content is located.

### 2.2.2 CURRENT METADATA REPOSITORY MODELS

The National Science Digital Library is a repository that houses and hosts metadata and indexes but no content. It gathers the metadata in a variety of ways, including OAI harvesting, ftp, email, direct entry and web crawling. This models works well for those that have the resources to host a very large database. It begins to break down as the amount of metadata increases past a critical point. The advantage of not having to host data may be offset by a lack of control over the data and the necessity of relying on other institutions for accuracy. (For more information on the NSDL see "hybrid portals" below.)

Within Algeria, the "Réseau Inter Bibliothèques Universitaires" (RIBU) is a metadata repository under development for a consortium of ten Central Algerian academic institutions. Sponsored by two European institutions, the Free University of Brussels and the Université d'Aix Marseille I, it was launched in 2005 with a standardized system for library management (SYNGEB). The objective is to establish an electronic union catalog for this part of the country, and the holdings of the Tébessa university library will become a part of this collaborative effort.

### 2.3 FEDERATED SEARCH MODELS

Returning to the architectural metaphor, federated searching developed from the desire to enter and profit from a number of important spaces through one entryway. ("Information mall" doesn't have the right feel.) Figuratively, the user can query a range of relevant databases simultaneously. The model has evolved and improved with developments in technology and the Internet. Current federated searching holds out promise as a model for the future, though it leaves something to be desired at present.

#### 2.3.1 EARLY FEDERATED SEARCH MODELS

Dialog Corporation's DialIndex was a very early example of federated searching. It allowed trained information seekers to sift through years of multiple databases via Telnet with a single search interface. The drawbacks were its very high cost and its lack of scalability in such a way that it could be distributed to the public. The user was required to come to the library and allow a librarian trained in Dialog intricacies – as an intermediary for the actual information seeker – perform the DialIndex search for a fee.

#### 2.3.2 CURRENT FEDERATED SEARCH MODELS

The current principle of federating searching involves software "translators," usually provided by a vendor like WebFeat Knowledge Prism or ExLibris Metalib, that search the indexes of other content providers who host the indexes, metadata and content. This is the most modular, scalable and loosely coupled system to date; it is made possible by the use of standard web protocols that have been widely adopted. The federated search does not require the housing of content, indexes or metadata and thus shares similar advantages and disadvantages as current metadata repositories – with the added caveat that at present the result sets reflect the "lowest common denominator" among individual database features from which the metadata derives.

At the turn of the millennium, the Grainger Engineering Library at the University of Illinois at Urbana-Champaign wished to provide “web-based asynchronous simultaneous searching of multiple secondary information sources and integrated reference linking between bibliographic resources.” (Mischo et al., 2002, p. 119). Today we might call that “federated searching.” The designers decided to create a custom federated search, for which they combined two different search technologies. One employed lightweight scripting against off-the-shelf web-based services (see section 4 below for more about web services), while the second was a more sophisticated, distributed approach requiring advanced programming techniques. The advantages of the first were simpler implementation and greater scalability. The disadvantage was the lack of flexibility in allowing the search code to interact effectively with each individual database. The second, more advanced programming method was easier to adapt to the idiosyncrasies of each individual database. The Grainger Library also implemented OpenURL resolving with these federated databases, which allowed greater access to full-text articles without having to perform multiple searches. Resource limitations in smaller, less well-funded libraries preclude such a model.

### 2.3.3 Beyond Federated Searching (Centralized Indexes)

Federated searching has made a good effort at giving the user the ability to search multiple databases from disparate sources simultaneously. But there are a number of weaknesses inherent in this model. As mentioned previously, the federated model requires the creation of a “translator” for each database to be searched. This translator receives the search request from the user and passes it on to the specific database in the required format. A separate result set from each database is then returned. This set of results usually consists of the first 10-25 records of the total even if the number of hits retrieved is much greater. This limitation is imposed due to the performance constraints of retrieving the entire results set from each database. If a search retrieved hundreds or even thousands of hits, the time required to retrieve the results from multiple resources would be prohibitively slow. Since the initial retrieval is of a small subset of records, any type of de-duping is only performed on this first subset.

The second problem is the loss of any type of advanced search functionality. Each database may have unique features such as proximity, adjacency, and wildcard or stem searching. Though there are similarities in the way these features are implemented across databases, there is no common standard; this makes it difficult for the user to include advanced search syntax in a federated search.

Two new products, Serials Solutions Summon and Ex Libris Primo Central, attempt to improve upon federated searching to overcome its limitations by creating centralized indexes that include records from a wide variety of databases along with local catalog data and other locally created and hosted collections. In this model, an automation provider (Serials Solutions, Ex-Libris) enters into agreements with content providers (Proquest, Ebsco, Gale) to index their data. The automation provider then indexes the contents of the catalog data and local collections of a particular library and hosts a single index that includes the metadata from all the database providers along with the library's local content. This overcomes the weaknesses of the federated search. The user is now able to use more advanced search syntax. This service also de-dupes and relevancy ranks the data. The results return more quickly in a single results set, instead of generating a separate set of results from each resource.

There are three downsides to using a centralized indexing service. First of all, initial price estimates suggest that it will be much more expensive than federated searching. The second problem is that the library will need to maintain some sort of gateway between its indexes and the centralized indexes to keep the centralized indexes up to date. The third problem is that both Serials Solutions and Ex Libris must negotiate a contract with every content provider whose content they wish to index. It's very possible that your library subscribes to content for which the automation provider does not have an agreement.

## 2.4 DIGITAL LIBRARY MODELS

Strictly speaking, a "digital library" would be the fully automated flip-side version of an analog library, i.e., it would house both a full range of targeted digital documents and online metadata, with links between the two.

"Institutional repositories" of locally produced digital publications, found at increasing numbers of universities or corporations, could be regarded as a special subset of digital libraries.

BAMBI (Better Access to Manuscripts and Browsing of Images), a European project for transcribing and annotating scanned ancient Greek and Latin texts, is a digital library model that includes collaborative functionality. It is aimed at both general users of a library who want to examine manuscript sources and, more importantly, at professional students of texts: philologists or critical editors of classical or medieval handwritten works: papyrologists, epigraphists, palaeographers, and codicologists (Calabretto et al., 1998). BAMBI has true Web 2.0 functionality, meaning the ability to tag or annotate using a web interface; it represents asynchronous work.

DEBORA Project (Digital Access to Books of the RenAissance) is another European project involving collaborative functionality. Its goal is to allow different users and scholars to participate in the digitization, metadata creation and posting on the Internet of rare sixteenth century books. The collection being created within DEBOPRA consists of digitized images of books from libraries in Lyon, Rome and Coimbra. (Le Bourgeois et al., 2001). With remote access, collaborators can use system tools for image processing, for metadata creation (tagging, annotations, rating of materials) and error correction. Since a special Java-based client – which is not widely available – is used for viewing and annotating, the use is mostly limited to scholars from the participating libraries.

The University of North Texas Libraries have assembled what they call a Texas History Portal, yet its architecture and functionality seem to brand it a digital library. Digital assets and metadata expressing the history of the state of Texas are self-contained on the UNT server. They flow into three different interfaces for distinct classes of potential users: Young Scholars, Researchers and Educators. Each of these areas provides auxiliary materials such as lesson plans and PowerPoint presentations. The most difficult part of creating this project was to design and build connections between the archival system of holdings and the tripartite front-end presentation. Metadata creation presented similar challenges: labor-intensive human input is very expensive (Nordstrom et al., 2004).

Gallica, meanwhile, the online digital library of the Bibliothèque de France, remains the gold standard. It has been in operation since 1997, has easy access to magnificent and rare texts within its own collection, and enjoys the budget support of the French state. It has expanded to include tens of thousands of books, journals, newspapers, maps, scores, engravings, photographs and sound recordings. Gallica with its high-end costs does not represent a model that many institutions can follow.

## 2.5 THE MOVE TOWARDS HYBRID PORTALS

As web portals evolved, companies began adding more services like email, chat and news to attract and retain users. The academic web portal concept received more attention with the growing success of other types of portals on the web. Design concepts led to a type of one-stop research environment characterized by distributed access over multiple platforms with ease of use and desirable levels of authentication and data security. The hybrid academic portal has become a tool that integrates on a single site all the work applications that the user needs; it can be targeted to on-campus users as well as to external members of the academic community. It offers consolidated access to relevant information sources and saves research time with its bias toward finding rather than searching.

As we mentioned above, some so-called "digital libraries" tilt heavily towards a model that combines metadata-to-data links with ancillary services. The National Science Digital Library (NSDL), for example, defines itself as a "metadata repository, search and discovery services, rights management services and user interface portal facilities." (Lagoze et al., 2002, p. 201). Though all of its data is offsite, NSDL defines the entirety of its project as a digital library with the various user interfaces being called portals. The success of the $24 million project was based on the idea that the ability to partner with, and aggregate, a variety of information sources would lower entry costs. Another aspect of lowering costs was the use, within the search phase, of simpler web-based protocols such as http and html rather than "more complex and rich mechanisms such as SGML and Z39.50." (Lagoze, p. 202).

There are at the present time innumerable examples of a library catalog expanded to hybrid portal proportions, or, as it is often informally referred to, the library's "home page." Most academic or research libraries are now taking advantage of modular features. One such library site, which we offer as typical of many, serves the library of the London School of Economics. Besides linking to the online catalog, this portal links to other electronic resources (many of them password protected for LSE users); to the archival collections; to news and calendar events; to service and local facilities information; to interactive Web 2.0 "ask a librarian" features, subject guides, online exhibitions and more. In true hypertext format, each secondary level, such as the archival collection, then offers a new range of choices within the new context: archives catalog, library catalog, pamphlet collection, rare books, and focused ancillary information for the archival user. An overview of the various trunks and branches of the web tree comes in the form of an A-Z site index, and a search box allows for pinpoint accuracy. An alternate means of reaching the user community is through "Services for you," a frame that supplies targeted and contextualized support to staff, students, alumni, visitors, or users with disabilities.

From a design and implementation perspective, the library portal of the Université Laval in Québec, Canada, is centered on a modular series of "portails thématiques" (subject portals). Each of these allows the addition of content with a simple and intuitive content management system (CMS). Similar to the editing interface of LibGuides by Springshare, this allows for a distributed model of labor among librarians trained in a subject but untrained in computer programming. One disadvantage is an irregularity of contents subject by subject, since they are not the product of an integrated system but are created by many hands in traditional ways that may be idiosyncratic or uneven (Chicoine, 2008). From the user's perspective, there is an opportunity to enter a customizable environment and seek information and guidance within a niche topic of interest (Teasdale, 2008).

For glimpses into the future of the hybrid scholarly portal, the Société Jouve (Jouve Company) in France has developed a number of academic information portals, each often referred to as a "Service Commun de Documentation" (SCD, roughly translated as "Collaborative Information Service"), at universities such as Nantes and Valenciennes. These sites were designed around the use of "open source" software.

## 3.0 CURRENT STATE OF INFORMATION PORTAL MODULES

With the growth of the "social web" or Web 2.0, further tools have made it possible to bring users into the library or to send library applications out to the users in virtual locations where they are likely to be found. In essence, these modular novelties are expanding the girth and reach of a library's information portal.

### 3. 1 EMBEDDING LIBRARY CONTENT OUTSIDE THE LIBRARY SITE

With the emergence of social networking sites such as Facebook, Twitter and Myspace as well as free web portals such as iGoogle, users have the ability to create their own customized web space. Within these personalized spaces the user can add widgets that pull content from many different websites without having to navigate to each one individually. Libraries can take advantage of the opportunity to offer content to the user within these personal web spaces.

#### 3.1.1 FACEBOOK

Many articles have been written recently about libraries' presence in Facebook. Research has offered information on the do's and dont's of communicating with users{{498 Connell,Ruth Sara 2009; }}. There are also accounts of the attitudes of users toward the library's presence in Facebook{{498 Connell,Ruth Sara 2009; }}. It appears that users will take advantage of library services offered through Facebook when those services are properly implemented. For example, the Penn State library reported that 29% of their reference questions came through Facebook. {{498 Connell,Ruth Sara 2009; }} (pg 28). Libraries have reported other positive benefits: Rutgers University Library reported that Facebook was instrumental in rebuilding a positive working relationship between library student employees and the student newspaper.{{525 Glazer,Harry 2009; }}

There are a variety of ways to communicate with users via Facebook. The library can create a Facebook group that students or faculty subscribe to. Subscribers to the group can receive updates as they are posted. Users can also send email messages to the contact number published on the library group site; this is one avenue for reference queries. Another option is for the library to create a distributable application that users can freely add to their personal profiles. For example, Mississippi State University created a Facebook application that contains catalog searching, links to services and news feeds all in a single widget.{{498 Connell,Ruth Sara 2009; }}

### 3.1.2 GOOGLE GADGET

Many people have created personal profiles on iGoogle (http://www.igoogle.com), which is a free portal to which users can add any of the freely available gadgets. It also integrates with gmail, analytics or any of the other Google services. Some examples of gadgets are news feeds, access to content from other websites, games, and video and audio feeds. Google makes available a development environment that allows anyone the ability to create and publish a gadget.

When creating an iGoogle account, a user receives the Google search and several other gadgets by default, such as a weather gadget, a news gadget and a calendar gadget. The user also has the ability to create tabs and select a theme for the site. At the McKay Library at BYU Idaho, we have created a gadget that allows the user to search the local catalog or conduct a federated search without having to navigate to the library website. Many libraries have created similar gadgets that search their local collections.

### 3.13 EMBEDDING LIBRARY CONTENT IN CAMPUS PORTAL

In 2003 Joe Zhou wrote an article entitled “A History of Web Portals and their Development in Libraries.” In this article he summarizes several books recommending what academic libraries should do to be successful in the coming portal world. The first recommendation was that “…the library portal should be integrated with the campus Web portals or have the capacity to be fully integrated in the future. “{{496 Zhou,Joe 2003; }}(p.6) The second was that the library portal should “…include courseware tools for faculty and students and incorporate the library’s major public services into course design.”{{496 Zhou,Joe 2003; }}(p.6)

When the campus portal MyBYUI was being developed at BYU Idaho, the library was consulted about interest in providing content. We worked with developers to provide the ability for students to view lists of books they had checked out as well as any fines they might have. We also provided the ability for them to add links to any of the library's databases.

As many academic institutions, BYU Idaho uses Blackboard as their course management software. The university also consulted the library about providing a presence in Blackboard. After discussing a number of approaches, the library decided that a link to the McKay library website would be provided by default at the top level whenever a new course was created.

## 4. HISTORY OF WEB PORTAL LANGUAGE DEVELOPMENT

The following table shows the main techniques used in traditional Web development:

| | Technical Programming | Technical Descriptions |
|---|---|---|
| On server | CGI | SQL |
| On cliant | Java, Javascript | HTML, xml, xhtml, css, smil |

Main techniques used in traditional Web development

**CGI**

The CGI (Common Gateway Interface) is a program that processes the information received from the client and generates a response, in HTML, corresponding to the customer's request. This system can be used to query a database using SQL.

**Java**

This is a compiled programming language (developed by Sun), which unlike CGI, will be executed on the client machine. It may, once the program and the data transferred, not to contact the server.

**HTML**

The HyperText Markup Language is a language of description used by CERN in 1990 and still used today for most Web pages. Specifications, such as the XML and XHTML are currently defined by the W3C (World Wide Web Consortium).

**XML**

The eXtended Markup Langage is a generalization of HTML that allows everyone to create their own markup language corresponding to their needs. Associated with Java, it allows you to create database systems as powerful as the combination of CGI, SQL and HTML limiting the use of bandwidth. Data is transmitted only once and can then be reprocessed by the client machine instead of the server. The XML can create more than one document that can be used both on the desktop as an "organizer" through document processing by different style sheets in CSS format.

The HTML has been reformulated to meet the standard XML as the XHTML (eXtensible

HyperText Markup Language).

Thus, the SMIL language is another XML standard, allowing you to create multimedia presentations. It has been developed by the W3C.

**NEW TRENDS FOR SEMANTIC WEB PORTAL DOCUMENTARY**

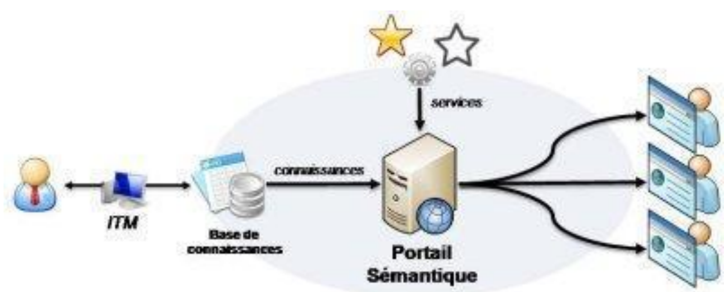
The latest trend in the evolution of documentary languages is the Semantic Web.

**THE SEMANTIC WEB** has emerged as the next major evolution of the Web [BL99] [BLHL01]. It is based on the idea of information processed by computer applications "agents". The objective of the semantic web is to make web content understandable to machines and carry the current Web to its full potential. The process consists in exploiting the following: [Tim Berners Lee].

- Ontologies "specification of a conceptualization" [Gru93] vocabulary of concepts, relations and axioms related to a certain area.
- A common language for describing ontologies and annotations using terms from these ontologies,
- Reasoning engine to infer the annotations from the axioms expressed in ontologies.

**SEMANTIC PORTAL CONCEPT**

The semantic portal is a website that provides a single gateway to a wide range of resources and services centered on a knowledge base. So users can move, search, and navigate their information space by exploiting the semantics of knowledge bases and using a set of high-level services.




The documentary semantic portal is based on applications types of semantic web. We describe two possible applications of Semantic Web:

1. Information research driven by ontology
2. A Semantic Web portal capabilities with HTL

### 1- Information research driven by ontology

The information search guided by ontologies: One of the contributions of the Semantic Web lies in the search for information. Currently looking for information on the Web is mainly based on full text technology, as in the major search engines, such as Google and Bing.

Research on the Semantic Web may, in addition to operating full text as in the current Web, exploit annotations of documents and ontologies. This allows access to resources according to their content (if annotations represent the content) rather than keywords. Such information retrieval, called 'guided by the ontologies' or 'semantics', has been studied in many projects and different languages, such as Shoe [HHS98] [LSRH97] Ontobroker [FDES98] (frame Logic) or WebKB [Mar97]. These projects have demonstrated the value and contribution of the research approach of semantic information.
Document annotations and the query are expressed using the vocabulary of ontology reasoning, engine search results and annotations that match the query by exploiting the inferences contained in the ontology. Figure 1 describes the principle of seeking information guided by the ontologies.

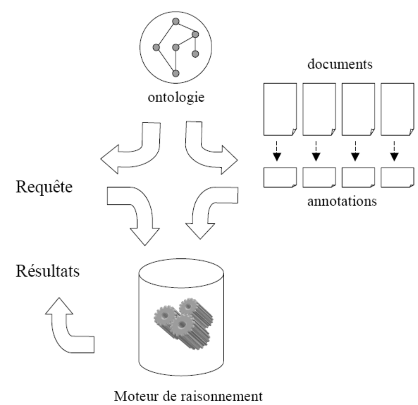



Figure. 1: Finding Information driven semantic ontologies

## 2- A semantic Web portal with HTL capabilities

We present a new form of Semantic Web portal using Natural Language Processing (NLP). Our system provides the means to annotate documents with metadata to enrich a knowledge base depending on the domain ontology matching, and especially to update linguistic resources used with the newly extracted information in order to improve performance of the system as a whole. Therefore, this system helps communities on the Web, including those working in the field of economic and scientific monitoring, creating web portals centered on the semantic scope. The end user will be able through the application interfaces to visualize data from the knowledge base to make complex and intelligent queries and to publish the results. It is important to note that the platform described later in this article complies with the standards and the Semantic Web languages (eXtensible Markup Language, XML Topic Map, Resource Description Framework (S) and OWL - Web Ontology Language).

**HISTORICAL DEVELOPMENT OF LANGUAGE FOR SEMANTIC WEB**

Different knowledge representation languages tailored to the Semantic Web were created. First, as SHOE [LSRH97] and Ontobroker [FDES98] proposed an extension of the HTML syntax for annotating web pages with "metadata" semantic and RDF. In order to ensure uniform treatment of all documents written in these languages, standardization work is now well underway: SOAP and RDF are W3C Recommendations TopicMaps an ISO standard ...

We describe here some kind of semantic web languages such as:

• Assertion languages (RDF and maps topical)

• a definition language for Web ontology (OWL);

• different description languages and composition of services (UDDI, etc.).

**4.1. RDF and RDFS**

- **RDF (Resource Description Framework)**

This is the emerging standard proposed by the W3C for the representation and exchange of metadata on the Web [RDF99] and is based on a model of triples (resource, property, value) and has an XML syntax. Any object in the Web identified by a URI is a resource and can be described in RDF, it may as well be a text document as an image or a sound recording.

An RDF description is a set of statements each specifying the value of a property of a resource. An RDF statement is a triple (resource, property, value), where value is either a resource or a literal. The RDF data model is as set of close semantic networks. A set of RDF statements can be seen as a multi-labeled directed graph whose vertices are resources or literals and whose arcs are labeled by the resource properties.

The RDF data model is defined independent of its syntax. " The exchange of RDF data on the Web has led to the model to associate an XML syntax. Figure 3 shows an example of the syntax description in RDF graph and its XML serialization. This is an annotation describing the resource that Inria has a partnership relationship with Lirmm whose activity is research.

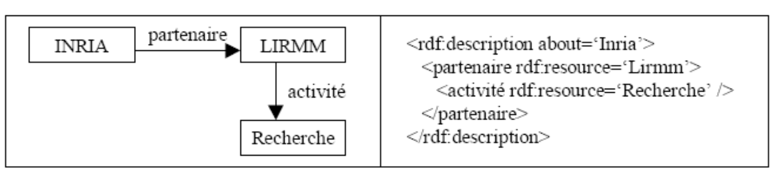



FIG. 3 – RDF annotation Example

- **RDFS (Resource Description Framework Schema)**

This is the accompanying standard RDF for representing ontological knowledge on annotations [RDF00]. It is dedicated to the representation of ontological knowledge used in RDF statements. RDFS is a set of declarations of classes and properties. It aims to extend the language by describing more precisely the resources used to label graphs. For this, it provides a mechanism to specify the classes whose resources are instances, such as properties. RDFS is always written using RDF triples, defining the semantics of new keywords such as:

- <ex:Livre rdf:type rdfs:Class> the resource ex: Book has type rdfs: Class, and is a class;
- <bibliothèque:Document153 rdf:type ex:Livre> resource library: Document153 is an instance of class ex: Book as we have defined;
- <bibliothèque:Document rdfs:subClassOf ex:Livre> class library: Document is a subclass of ex: Book, all instances of the library: Book are instances of ex: Book;
- <ex:localisation rdf:type rdfs:Property> says <ex:localisation> is a property;
- <ex:localisation rdfs:range ex: Library> says that any resource used as end of an arc labeled e.g. location will be an instance of class ex: library.

The need to further specify the classes is the origin of language dedicated to the definitions of classes: OWL.

For example, the RDF annotation, Lirmm resource is an instance of the class Institute, in RDFS schema definition, the activity property is an instance of the class Property.

Propriate and the subClassOf subPropertyOf represent subsumption relations that organize classes and RDF properties in hierarchies: class C subsumes D if C is more general than D, ie all resources of type C contain all resources of type D.

For example, class subsumes the class Inanimate Entity Institute.

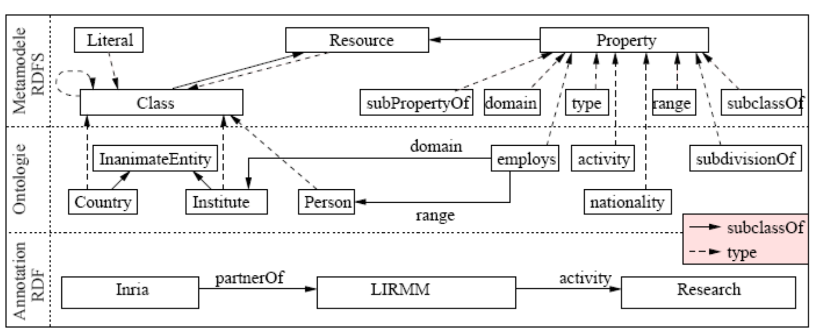



FIG. 1.12 – Le métamodèle de RDFS, une ontologie sous forme de schéma RDFS et une annotation RDF basée sur ce schéma.

### 3.1 Languages, assertions and annotations

Assertions affirm the existence of relations between objects. They are therefore suitable for the expression of annotations to be associated with web resources. We will discuss them here mainly because RDF appears to have advantages for computer manipulation.

#### 3.1.1 RDF

RDF is a formal language that allows us to state relationship "resources". It will be used to annotate documents written in unstructured languages, or as an interface for documents written in languages with semantic equivalent (databases, for example). RDF has a syntax and semantics.

An RDF document is a set of triplets of the form Subject, Predicate, Object. This document will be encoded in a paper machine RDF / XML or N3, but is often represented as a graph. RDF also provides some reserved keywords, which can give particular semantic resources. Thus, we can represent sets of objects (rdf: Bag), lists (rdf: sequence), relations of any variety (rdf: value) ...

RDFS (RDF Schema for) aims to extend the language by describing more precisely the resources used to label graphs. For this, it provides a mechanism for specifying the classes C whose resources will be labeled by C bodies like properties. RDFS is always written using RDF triples, defining the semantics of new keywords such as:

• <ex:Livre rdf:type rdfs:Class> the resource ex: Book has type rdfs: Class, and is a class;

• <bibliothèque:Document153 rdf:type ex:Livre> resource library: Document153 is an instance of class ex: Book as we have defined;

• <Library: Document rdfs: subClassOf ex: Book> Class Library: Document is a subclass of ex: Book, all instances of the library: Book are instances of ex: Book;

• <ex:localisation rdf:type rdfs:Property> says <ex:localisation> is a property;

• <ex:localisation rdfs:range ex: Library> says that any resource used as end of an arc labeled eg location will be an instance of class ex: library.

#### 3.1.2 Topic maps

Topic maps are an ISO standard HyTime from whose aim was to annotate multimedia documents. Derived from SGML, it is assigned an XML syntax (XTM). In addition, a group of ISO is responsible for defining a query language for maps topical (TMQL). Topical cards are built around four primitive notions:

1) The "topics" that can be understood as individuals with knowledge representation languages.

2) • The names given to the topics: one of the original card is the separation of topical concepts and their names. This allows to have multiple names for the same concept and names shared by several concepts.
3) • Instances are "proxies" external entities so they can be indexed by topics (where literal entities if they are representable).
4) • Bearings, which are sometimes seen as a fourth dimension, can specify the context in which a relationship is valid.
For example, the topic of the document is instantiated by myBook, it is called "book for history" whose scope is one of my discussions with lunch and colleagues "livre834" in discussions about U.S. immigration.
In the new syntax topical maps, they are represented by graphs including 3 types of nodes (topic, association, scope) and a number of types of bows (instance, occurrence, scope, name). Relations are represented by nodes, whose outgoing arcs label their role.
Furthermore, different interpretations are given according to the labels on arcs and nodes. Suffice it to say that topical cards do not have clear semantics, and designers tend to assume that the richness of language lies in the possibility of multiple interpretations.
This does not make it a very desirable candidate for the semantic web. However, there are tools to build a useful topical board that are used in a number of applications.

### 3.2 Ontologies, languages and definitions

OWL is dedicated to class definitions and types of properties, it allows a large number of manufacturers to express very fine properties-defined classes. OWL has been split into three distinct languages:
1) OWL Lite contains only a small subset of manufacturers available, but its use ensures that the comparative types can be calculated;
2) OWL DL includes all manufacturers, but with specific constraints on their use to ensure the decidability of the comparison types. For cons, the complexity of the language seems to necessitate a heuristic approach;
3) FULL OWL, for which the problem of comparing types is likely undecidable. The syntax of an OWL document is supplied by the different manufacturers used in this document. It is most often given in the form of RDF triples. Each manufacturer has associated semantics, in model theory. It comes directly from description logics. The semantics associated with the keywords of OWL is more accurate than that associated with the RDF document representing an OWL ontology.

### 3.3 Languages and services composition

We present the objectives and main features of specific languages to web services.

#### 3.3.1 UDDI

UDDI Protocol (Universal Description, Discovery and Integration) is a platform for storing descriptions of Web services available, like a directory style "Yellow Pages". Research services can be performed using a system of keywords provided by the organizations providing the services. UDDI provides a system of “White Pages” (addresses, phone numbers, identifiers ...) to obtain the coordinates of these organizations. A third service, "Green Pages" provides detailed technical information about services and allows us to describe how to interact with services thereafter by pointing to a RosettaNet PIP or "service interface" WSDL. The vocabulary used for the description follows a precise taxonomy to enable better categorization of services and organizations.

#### 3.3.3 WSDL

WSDL is an XML-based language used to describe the interfaces of web services, that is to say, by representing the abstract operations that the service can perform, independent of the implementation made. It has no way to describe more abstract services via conversation and transaction messages, but is generally used as a bridge between these representations of high level and low level.

#### 3.3.4 DAML-S

DAML-S is a service description language based on XML, using the model of description logics. Its advantage is as a high-level language for the description and invocation of web services in which semantics is included, unlike e.g. UDDI. DAML-S is composed of three main parts:

• Service Profile, which allows the description, promotion and service discovery. Research on services can be done by taking any element Service Profile as a criterion.

• Service Model, which shows the operation of the service, describing in detail and relatively abstract what to do to get there. It also allows a tight control of the running of the service.

• Service Grounding will present clearly and in detail how to access a service. It is in this part of the protocol that message formats are specified among others.

For now, DAML-S is a language that is still being specified, but the main lines are already drawn. deDAML-S can then be restricted to an abstract description and semantic services, also allowing us to express constraints on the parameters and use constructors.

### 3.3.5 XL

XL is a platform for Web services, XML-based, using a high-level language specific (XL), and taking into account W3C technologies (WSDL, SOAP) to enable interoperability with other XL applications written in a language other than XL. Any web service is regarded as an entity receiving and transmitting XML messages: in return XML messages with (purchasing a book) or without (consultation on the weather) changing the context. The data types used are those of XQuery, also developed by the W3C, which inspired syntax XL.

### 3.3.6 ebXML

Standard industrial strongly and specifically oriented e-commerce, ebXML (Electronic Business XML Initiative) has been developed since 1999 and is a system of standardization exchange for international trade business (UDDI would be closer to a directory).

It is primarily a set of protocols, with standards since 2001, relying on tools such as PPC (equivalent WSDL improved), R & R (Registery & Repository, functioning as UDDI), or SOAP (for the transfer of messages).

ebXML has been seen by many as a successor of EDI (Electronic Data Interchange) for several years that allow companies to exchange data in a secure and reliable way, but are often seen as expensive and difficult solutions to implement.

### 3.3.7 RosettaNet

RosettaNet has many similarities with ebXML. However, although ebXML corresponds to a more horizontal design web services, RosettaNet has a vertical view, that is to say, the information exchanged are generally about a product or a set of particular products.

RosettaNet currently includes more than 400 major players in the electronics industry and semiconductor and IT services companies and Information Technology.

RosettaNet is based on "Partner Interface Processes" (PIPs), which are XML messages exchanged between the systems that define the way forward for trade between the partners. Each PIP includes a document containing the vocabulary used in order to avoid any misunderstanding (the "business document"), and another defining the choreography of the message exchange (the "business process").

## 4 FUTURE RESEARCH FOR THE SEMANTIC WEB

The Semantic Web projects a number of issues and perspectives that deserve further research to facilitate the understanding of what may be the semantic web.

### 4.1 Modularization languages

We have seen that RDF focuses on assertions about the relationship between objects, while OWL is interested in describing classes of objects. It is a fairly natural division between factual knowledge and ontological knowledge. This structuring of the knowledge has been made by both description logics (A-Box and T-Box) and conceptual graphs (graph and support) from their common ancestor, the semantic networks. In terms of use, this separation is just as important. The design of ontologies is a specialty in the field, while factual knowledge, using a given ontology, is the responsibility of an informed user. It would have been natural to partition RDF and OWL following these specifications, but the need to increase the expressiveness of each language seems to have been strongest. The extension of RDF to RDFS in the same graph produces two very different levels of abstraction, and this lack of clarity is structured as one of the main criticisms that had been made to the semantic nets (including lack of formal semantics, default remedies which RDF). Similarly, we can encode OWL DL factual knowledge that is the responsibility of RDF. So there is a lack of clarity on the objectives of these languages, exacerbated by their multiplicity (RDF, RDFS, OWL Lite, OWL DL, OWL FULL).

A clearer division between RDF and OWL has led to the development of ontologies on the one hand, and on the other RDF documents whose resources would be classes or properties described in a document OWL. The joint use of two languages, yet natural, has so far not been studied. Even if the semantics of the RDF + OWL is defined immediately by semantic languages that compose it, major theoretical problems arise: If algorithms are known for reasoning about RDF (graph homomorphism), and reasoning in some subsets of OWL (work done for description logics), the juxtaposition RDF + OWL has not been studied (equivalent problems can be found in BD for the inclusion of queries, but remains a work in progress). Even OWL (Lite, DL, FULL) could be challenged by the corresponding complexity of the sub-languages RDF + OWL. Thus, one might prefer OWL DL language whose disjunction is excluded and has semantics of intuition-negation where classes can be considered as instances. Such language is not currently defined. To qualify, it would have had to develop a more modular approach to semantic web languages for which work would be welcome.

### 4.2 Inference Engines

The development of effective tools for reasoning in the semantic web will be a decisive factor for the adoption of a particular language. These are inference engines that will encapsulate applications in more advanced systems to query the web and act on the answers.

However, for the simplest of these languages (RDF), subsumption is NP-complete. Efficient algorithms have yet to be developed to calculate the homomorphisms of graphs that meet this problem. These algorithms will allow us to give an order of magnitude, to calculate a graph of homomorphisms to 500 vertices by 3000 vertices in a reasonable time. The problem at present is quite different. Even if we assume that the order of magnitude of a graph question is 50 nodes, the evidence base is the set of RDF documents available on the web. Today there are more than 3 billion HTML pages referenced by Google, and without indication of the success of RDF, we can ask how RDF documents will be available tomorrow. Although we believe the achievement of efficient algorithms possible, only an experiment on a large amount of real data can help validate this intuition.

However, in the RDF + OWL we deem desirable, other problems arise. Even adding only the negation of the atomic type, the subsumption problem is P2-complete. Local processing of information during the execution of the algorithm is no longer feasible.

### 4.3 Processing of languages

It's a safe bet that the knowledge available on the web in different forms (languages) will have different models (ontologies). Moreover, some applications will need to merge these sources of knowledge or adapt them to their needs. This activity is currently carried out on an ad hoc basis. It will be necessary to take full advantage of the knowledge available in the semantic web to transform and import in different contexts (languages, ontologies). There are a variety of such transformations (fusion catalogs, database extraction, normalization theories) that require various properties. A first research effort is expected to characterize these changes and their relationships. It should also be possible to define a standard way of transforming and running “Semantics". Currently, there is no infrastructure for processing RDF and XML still used for this purpose.

Finally, since these operations are to be performed by machines (indiscriminately), it is essential for the credibility of the semantic web that we can prove the correctness of transformations from their specifications.

### 4.4 Robust Inferences

A typical property of the Web is the amount of information found there. Unfortunately, there is a lot of misinformation, outdated, redundant or incomplete. A human surfer is generally able to discern these problems and overcome them without much thought, but not semantic web applications. It is therefore necessary to develop ways of thinking that take advantage of the semantic web, that is to say, that are as faithful as possible to the language specifications used without being affected by these problems. In a word, we need robust inference engines.

They could use a variety of techniques adapted to the context of the Semantic Web. Or anytime reasoning under resource constraints that could manage the huge size of the web.

### 4.5 Languages rules

Another need for those working on the Semantic Web is to develop a rule language. If an organization says on its X RDF document that a train is going from city A to city B, while Y states that an agency will process the travel from city B to city C, then it must be inferred that there is a path from A to C. However, this information can be found neither on the site of X, nor that of Y. An immediate solution can be declared in a document OWL property that path is transitive, but the problem becomes insoluble when one wants to take into account the existence or duration of the match. It is necessary to use a rule ": IF ... THEN! ..." That could represent the following: Train Journey SI-train

This type of rule has been studied as an extension of the simple conceptual graphs, and the results are immediately transferred to an extension of RDF. These rules are provided with a semantics which corresponds to logical formulas of the form!:

"X (P (X) Æ ($ Y Q (X, Y)))

where P (X) is a conjunction of atomic formulas whose variables are those appearing in X and Q (X, Y) is a conjunction of atomic formulas whose variables are those appearing in X and Y. Note that these formulas correspond to TDGS (Tuple Generating Dependencies) in databases. The use of such rules generates a very expressive language, and is unfortunately undecidable despite decidable subsets.

It should be noted that the expression of such a rule language would make a good candidate for a meta-language, for example, to provide a language for defining ontologies for new manufacturers by defining operational semantics.

This work, like all work on the language, can’t be done in isolation. It is normal that it be prosecuted in connection with international working groups helping to advance the state of the art. By consistent effort it could be produced by a small community with regard to inference engines and transformation. But it requires a significant investment in the medium term.

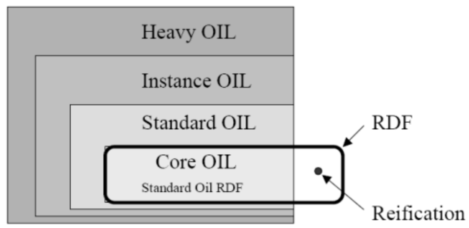



FIG. 1.13 – OIL et RDF(S)

Here is an example of an extract ontology defined in OIL:

Defined classdef system

subClassOf Document

Slot-constraint eats

Book value-type

Defined classdef software

Book subClassOf

SubClassOf NOT computer science

Slot-constraint reads

Value-type information system

Classdef tasty-information system

SubClassOf information system

## 5. SUMMARY

Many use the term “portal” indiscriminately to refer to very different data systems, information-gathering conglomerates and knowledge mills. Yet, a closer look at the distinct features of each type of portal – with its unique advantages and disadvantages – is a step towards meeting the demands of library users while still keeping within budget constraints. A convergence of technologies – along with a willingness on the part of libraries to explore hybrid systems – may satisfy scholars to a degree undreamed of in earlier years. Some of the most promising portal innovations for the future involve inexpensive web service modules that are easy to implement and to manipulate. Who knows what the advent of artificial intelligence and smart systems of Web 3.0 (Semantic Web) will contribute to the landscape of portals? The future is bright for the information user, and the library can be an integral part of that promise by keeping a critical but willing eye on portal developments.

## 5. REFERENCES


Alonso, G., Casati, F., Kuno, H. and Machiraju, V. (2004). Web Services: Concepts, Architectures and Applications. Springer, Berlin.

Arsanjani, A., Hailpern, B., Martin, J. and Tarr, P. (2003). “Web services: Promises and compromises.” ACM Queue, vol. 1, no. 1, pp. 48-58.

Calabretto, S., Pinon, J. M. and Bozzi, A. (1998). “BAMBI: système de manuscrits anciens pour historiens,” Les bibliothèques numériques, vol. 2, no. 3-4, pp. 31-50.

Chaoying M., Bacon, L., Petridis, M. and Windall, G. (2006). “Towards the design of a portal framework for web services integration,” Proceedings of Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services, February 19-25, Guadeloupe, p. 163.

Chevalier, M., Julien, C., Soulé-Dupuy, C. and Vallès-Parlangeau, N. (2007). Personalized Information Access Through Flexible and Interoperable Profiles. Berlin, Springer.

Chicoine, P. (2008). “La bibliothèque 2.0 èmerge à Québec,” Argus, vol. 37, no. 2, pp. 11-12.

Hailong S., Chunming, H., Jinpeng H. and Qin, L. (2007) “Design and implementation of an enhanced grid service container in the CROWN grid,” International Journal of Web and Grid Services, vol. 3, no. 4, pp. 403-423.

Hunt, L., Lundberg, M. and Zuckerman, B. (2005). “InsciptiFact: A virtual archive of ancient inscriptions from the Near East,” International Journal on Digital Libraries, vol. 5, no. 3, pp. 153-166.

Ignat, C., Steinberger, R., Pouliquen, B. and Erjavec, T. (2006). “A tool set for the quick and efficient exploration of large documents collections,” Proceedings of the Symposium on Safeguards and Nuclear Material Management. 27th Annual Meeting of the European Safeguards Research and Development Association (ESARDA-2005), London, 10-12 May 2005, available at http://langtech.jrc.it/Documents/0505_ESARDA-JRC-LT_Explore-doc-collections.pdf (accessed 30 March 2009).

Jimenez, G. and Espadas, J. (2006). “Implementation of an e-services hub for small and medium enterprises,” Proceedings of Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services, February 19-25, Guadeloupe, p. 187.

Kalyanaraman, S. and Sundar, S. S. (2008). “Portrait of the portal as a metaphor: Explicating web portals for communication research,” Journalism and Mass Communication Quarterly, vol. 85, no. 2, pp. 239-256.

Kolbitsch, J. and Maurer, H. (2006). ”The transformation of the web: How emerging communities shape the information we consume,” Journal of Universal Computer Science, vol. 12, no. 2, pp. 187-213.

Kreger H. (2001). “Web Services Conceptual Architecture (WSCA 1.0),” IBM Software Group, available at http://www.cs.uoi.gr/~zarras/mdw-ws/WebServicesConceptualArchitectu2.pdf (accessed 23 February 2009).

Lagoze, C., Arms, W., Gan, S., Hillman, D., Ingram, C., Krafft, D., Marisa, R., Phipps, J., Saylor, J., Terrizzi, C., Hoehn, W., Millman, D., Allan, J., Guzman-Lara, S. and Kalt, T. (2002). “Core services in the architecture of the National Science Digital Library (NSDL),” JCDL ‘02: Proceedings of the 2nd ACM/IEEE-CSJoint Conference on Digital Libraries, Portland, Oregon, pp. 201-209.

Le Bourgeois, F., Emptoz, H., Trinh, E. and Duong, J. (2001). “Networking digital document images,” Sixth International Conference on Document Analysis and Recognition, Seattle, Washington, pp. 379-383.

Mischo, W. H., Habing, T. G. and Cole, T. W. (2002). “Integration of simultaneous searching and reference linking across bibliographic resources on the Web,” JCDL ‘02: Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, Portland, Oregon, pp. 119-125.

Mothe, J., Chrisment, C., Dousset, B. and Alaux, J. (2003). “DocCube: multi-dimensional visualisation and exploration of large document sets,” Journal of the American Society for Information Science and Technology, vol. 54, no. 7, pp. 650-659, available at http://www.irit.fr/recherches/IRI/SIG/personnes/mothe/pub/JASIST03.pdf, accessed 30 March, 2009.

Nordstrom, K. E., Hartman, C. N. and Phillips, M. (2004). “The University of North Texas libraries’ portal to Texas history: archival challenges and solutions,” JCDL ‘04: Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries, Tucson, Arizona, p. 409.

OECD (2007). “Participative Web and User-Created Content: Web 2.0, Wikis and Social Networking,” available at http://213.253.134.43/oecd/pdfs/browseit/9307031E.PDF (accessed 23 February 2009).

Pierce M., Fox, G., Choonhan, Y., Mock, S., Mueller, K. and Balsoy, O. (2002). "Interoperable Web Services for Computational Portals," Supercomputing, Proceedings of the Conference on High Performance Networking and Computing, November 16-22, p. 39.

Smith, M.A. (2004). "Portals: toward an application framework for interoperability," Communications of the ACM, vol. 47, no. 10, pp. 93-97.

Teasdale, G. (2008). "La recherché intelligente se redécouvre à l'Université Laval: Un nouveau catalogue voit le jour," Argus, vol. 37, no. 2, pp. 7-10.

Tung-Ching L., Hsu-Ju, C. and Po-Hsun, C. (2008). Issue placement affecting factors for knowledge management systems initiation, International Journal of Innovation and Learning, vol. 5, no. 2, pp. 186-200.

Mohamed Ridda LAOUAR, Richard HACKEN, Mathew MILES, (2009) "The role of web services in portal design: approaches for an Algerian university library", Library Hi Tech, Vol. 27 Iss: 3, pp.460 – 479.