

San Jose State University

From the Selected Works of Anita S. Coleman

March, 2004

Communicating digital library services to scientific communities

Ying Liu, *University of Aberdeen*

Anita S. Coleman, *University of Arizona*



Available at: <https://works.bepress.com/anita-coleman/10/>

Communicating Digital Library Services to Scientific Communities

Ying Liu

Computing Science Department, University of Aberdeen
Kings College, Scotland, UK
yliu@csd.abdn.ac.uk

Anita S. Coleman

School of Information Resources and Library Science
University of Arizona at Tucson, USA
asc@u.arizona.edu

Abstract. To date, systematic approaches to the provision of and infrastructure of digital library knowledge services have not been able to model accurately the communication, protocols, and discourse practices in science disciplines. On one hand, there is a growing need for collocation and retrieval of scientific works online and not merely for services based on known-item identification. On the other, infrastructures for collection, such as the web service, ontologies, or the semantic web, need specialized knowledge. We argue that these infrastructures are not only limited to the web; they have also introduced new systematic contextures constraining communication in scientific communities. To assist the process of developing a better contexture for such types of knowledge, we report the lessons learned from the Alexandria Digital Library project and use water resources management as a given domain of geo-referenced knowledge to understand and develop library-user scenarios. This research is a pre-requisite for the development of knowledge services in a problem-solving environment.

1 Introduction

Recent advances in digital technologies, storages, and Web-driven infrastructures enable us to digitise, archive, and visualise informational materials on the Internet. These technologies have created many new opportunities as well as problems in organising digitized materials through gateways and repositories (all of which we are concerned with here in terms of digital libraries).^[1] Many researchers have conceded that the emphasis on digital library technologies and tools is out of step with the prevalent modes of user-communities for the processes of capturing, sharing, determining, and utilising information.^[2] According to a recent report (Lyman et al. 2002), the world currently produces between one and two billion-billion 8-bit bytes of information each year. Most of this information is in the form of images, sound, and numeric data. Digital textual documents account for only 0.003% of the total.

Indeed, for a scientific community of users, little of the information seems to be made available through digital library collections as most digital items are

file-based artifacts, i.e., packages of numerical models, large datasets, geo-referenced images, statistical data tables, etc. This has recently led some information scientists to shift the research focus to a new theme, i.e., to catalogue and index scientific items by treating scientific models as works, and, in turn, to classify and index the works as entities for knowledge retrieval.^[3] Historically there have been intensive studies in the construction of catalogs with the focus on known-item identification and retrieval. Little attention has been paid to the collocation and retrieval of like-items, related materials called *works*. A focus on *works* will help to advance understanding of the role *works* play in facilitating knowledge construction and in the importance of the work entity as key to the construction of bibliographic and other databases or internet search engines. New prototypes of digital library knowledge services are under research and development, e.g., the THREDDS (the THematic Real-time Environmental Distributed Data Services) in Benedict et al. (2003), or DLSI (Digital Library Service Integration) projects (DLSI homepage, Chen et al. 2003). For this line of research, we need to know how to collect materials in support of a user community, why the materials are collected, and how people actually use those materials.

This paper intends to show some contexts that critically affect digital libraries with services intended to facilitate the communication of scientific knowledge. We take water resources management as a given domain of geo-referenced knowledge for potential digital library users. The notion of “contexture” usually means “a weaving together of parts; structural character of a thing; system; constitution; texture” (Online Dictionary). Because the contexts in question are systematically inter-related, the notion now has its coherent place in internet-based infrastructures communicating to a scientific community in a given problem-solving environment.^[4] See Fig. 1 indicating our research methodology. The rest of this paper is organised as follows.

We first share the experiences and lessons learned from Alexandria Digital Library (ADL). ADL is a well-known geo-referenced digital library project funded by NSF, DARPA, and NASA since 1994. We show how an *infrastructural contexture* constrains the scalability of web-based information services and why ADL has been rapidly updating itself with new infrastructural technologies but reporting little for scientific usability of its large cartographic collection - now totalling more than 4 million individual items according to Chen (1998, pp 197). In designing several user scenarios for water resources management, we further discuss how a *knowledge contexture* determines the system's knowledge classification through different knowledge sub-domains in the users' disciplinary communication and information flow systems. Such a knowledge contexture can be seen in discussion about the hydrological system, one of the most vital elements in water resources management (Obasi 1996). Finally, we briefly consider the relationships between the revealed infrastructural and knowledge contextures and draw some conclusions.

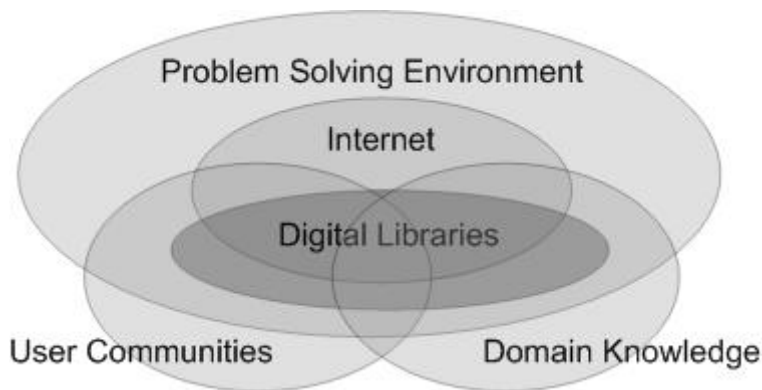


Fig. 1. Digital Libraries and Communities

2 Infrastructural Contextures: Lessons Learned from Alexandria Digital Library

In considering the flow of information and knowledge, the goals of developing a digital library are not different from the purposes of having a traditional library (Chen *op. cit.*, Dretske 1981, Kemp 1976):

- a) both provide information to users by collecting items;
- b) both serve as disciplinary and general human-communication systems by collecting and organizing the items of knowledge;
- c) both depend for their existence on the users' needs and uses for the communication, construction, and production of knowledge that take place in a social or other community setting.

The two, however, are clearly distinguished by completely different technological origins - one consists foremost of printed materials originating from print technology; the other is comprised foremost of items existing only in digital form, thereby originating from computer technology. This has determined two disparate sets of contextures: the printing items move around social communities as Figure 2 indicates; the digital items move through life cycles in internet computing environments where Web services,^[5] Web Ontology,^[6] or Metadata^[7] are based, to name but a few (see Figure 1). Such a Web system has become an increasingly integral element of our social infrastructure. Thus, the complexity of a digital library service stems not only from its infrastructure but also from the contexture from which its infrastructure augments, where we classify digital items we intend to collect and use. But before we illustrate this by ADL examples, it is necessary to define and clarify the new meanings that go beyond the ordinary

notion of infrastructure.

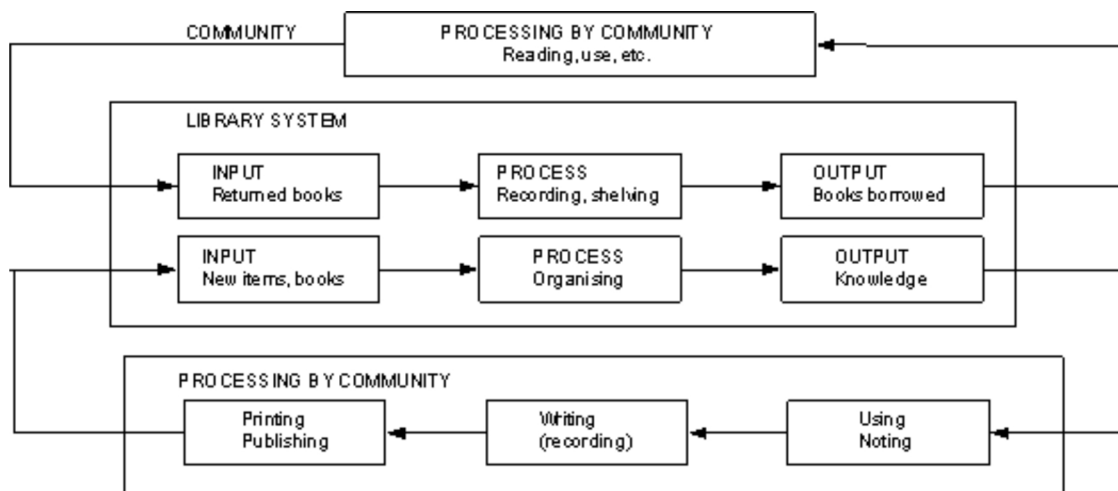


Fig. 2. Traditional Libraries and Communities

2.1 The Notion of Infrastructure

Star (2002) notes that infrastructure, often referred to as a list of technical specifications, black boxes, places, wires, plugs, roads, bridges, stations, etc., appears to be singularly boring as an object for scientists to study. Infrastructuring is usually seen as engineering work, the establishment of public services and utilities for societies and communities. Roads, railways, bridges, pipelines, and electricity are all instances of public, social infrastructures. Now people also use the term infrastructure to refer to any substructure or underlying system: most notably the information superhighway - the global information and communication infrastructure that includes the Internet, WWW, telephone networks, cable or satellite communication networks.

As we know, the Web is a collection of interlinked electronic items including documents, texts, images, music files, video, etc. hosted on servers all over the world, mostly on HTTP (Hypertext Transfer Protocol) servers (Schatz et al. 1994). The Web lives on the Internet by a set of protocols running over the net. Although the Web is a part of the net, the net is much larger than the Web. The net also hosts e-mail, FTP (file transfer protocol), peer-to-peer, VPNs (Virtual Private Network), telephony, automated sensors, wireless networks, and more.

In software engineering, the Web has driven Java-enabled technologies. In turn, large scale meta-computing such as the Grid^[8] embraces the net-centric technologies by adding more tiers on top of Java-based systems. Java can indeed “write it once, and run everywhere.” But without an infrastructure of Web-tiers and

net-layers, it may run all over the place for nothing or nowhere at all. This infrastructural impact has fundamentally influenced software life cycles^[9]: specification, design, coding, debugging, operation, and maintenance.

An electronic infrastructure is a set of Web-based software components and data network elements that are inter-operated and work together to support online events of data/information exchange. Typical infrastructural components are Web servers, application servers, APIs (Application Programming Interfaces), protocols, portals, repositories, databases, application-ware, middle-ware, PCs, work stations, electronic appliances, and programming languages. On top of these layers, a digital library infrastructure may include (Chen, *op. cit.*):

- a) systems and components for discovering, distributing, indexing, cataloguing, storing, retrieving, etc.,
- b) social and technical communication, and
- c) tiers of distributed software architectures that realise these functions.

2.2 Infrastructures and Tag Contextures in ADL

ADL is a prototype geo-referenced digital library intended primarily to serve a research audience (ADL homepage). It is a well-known system, a product of one of the six initial digital library projects funded by NSF, DARPA, and NASA. Its collection and services focus on geographical information: maps, images, geo-referenced data sets with text, and other information sources with links to geographic locations. It is a large collection of geo-referenced items that are organized and maintained in digital formats within complex Web-based infrastructures so that an end-user can access and explore the items. ADL has had three successive versions:

- a) a rapid prototype system comprised of a relational database of map and imagery metadata accessed through a desktop GIS (geographic information system, Frew et al. 1995);
- b) a Web prototype system that replaced the stand-alone GIS with an HTTP server, generating an HTML forms-based user interface accessible via the Web (Frew et al. 1997);
- c) the ADL integrated system that extended the HTTP server into full-fledged middleware, supporting HTTP interfaces to multiple clients and collections to multiple HTTP catalog databases (Frew et al. 1999).

In spite of these experiments with changing infrastructures, the research outcomes have a questionable impact on their research methodology and conceptual frameworks driving the design. According to the chief architect (Janée 1999 & 2003) as well as others, the development of the ADL and its services require substantial fundamental, infrastructural modifications (Boxall 2002). The problems are summarized in Tables 1 & 2. The two categories of problems are not only inter-related but also related externally to a service that is summarized in Table 3. Currently, ADL has evolved to become ADEPT (Alexandria Digital Earth Prototype) in the second phase of the development (from 1999 to 2004). ADEPT expands the use of ADL-features into new fields, e.g., classroom based geo-referencing e-learning applications. The focus of ADEPT has thereby broadened from a geospatial library to an integrated environment for managing, querying, and presenting geospatial information. This can be seen as a list in Table 4. ADEPT is involved in collaboration with new research partners for the development of new applications. For example, Kent State University is developing a structured database of scientific concepts for organizing, accessing, and using learning materials that ADEPT will provide (Coleman, et al 2001, Smith et al. 2002). ADEPT will be able to “plug-in” Iscapes (information landscapes) that the University of Georgia is developing, in which ontological environments are based on the Earth metaphor Iscapes (Isdis homepage). ADL, however, has proved to be very costly in terms of finance, human resources, and technology investment. More important, there are few references available to show how digital library research may tackle the fundamental problems inherent in tag-infrastructure contexts of digital libraries.

Table 1. Service Layer Problems

S-Category: a service layer is completely missing (S)	
S-Tags-Problem	ADL services are too simplistic. Indeed, a search is one “shot” followed by a stream of results back. Because of the next problem, a query is limited to the capacity of what one XML protocol tags.
S-Tags-Semantics-Problem	The semantics of all the metadata are both unknown and unknowable to ADL system.

Table 2. System Operational Problems

O-Category: no inter-operations between content holding resources (O)

Table 3. Integration Problems

I-Category: no integration with other tools and services (I)	
I-External-S-Plug-in-O-Problem	ADL' s single, monolithic components support no inter-operation with other tools and services. There is no new metadata standard adaptation, no service layer, and therefore no component to be plugged either in or out.

Table 4. More New Components

Solutions (S: service layer, O: operation, I: external integration, SYS: internal integration)	
S-O-Repository-Component	collection registry, content repository
O-Metadata-To-Metadata-Component	metadata mapper, harvest loader, item tracker
SYS-S-O-Component	collection aggregator

3 User Scenarios

The tag-infrastructureal contextures that constrain scalabilities of knowledge-retrieval services have to be understood in the knowledge contexture in which digital items are created, edited, described, indexed, or used. Furthermore, a knowledge contexture is not only interwoven within community-based communication among creators and other owners, but it is also embedded among individual or institutional users. (The following river authority board is an example that attempts to describe this). In this section, we outline a few user scenarios in order to explain the nature of knowledge that is sought or that underlines the use of information resources. In the next section, we further define and discuss the knowledge contexture.

What follows is a set of four related processes (described as user scenarios) enabling a final decision to be made in water resources management. The user scenario is designed to “identify the functions that the system should deliver, how these may be displayed to users, what parameters of the human-computer equation should be satisfied, and so on” (Crabtree et al. 1997).

Scenario I

Mark is a project engineer working for a local council. The council plans to build a huge holiday park near a river. Among many other environmental impacts to consider, Mark needs to initiate a set of conditions defining the amount of water the central park needs each year, including the amount, source, and times of discharge. Based on this information and prior to granting Mark a consent of water supply, the River Authority assesses if the requested discharge will have a significant detrimental effect on the river network downstream of the discharge point. To initiate this process, Mark logs into his laptop and gets connected to his computer station at his work. He opens the plan of the holiday park and begins to review the estimated figures and browse the maps.

Scenario II

Mark finally finishes his report and drags it to the River Authority's working folder. The Authority opens his report, browses the maps Mark has attached, and begins to retrieve the water-related objectives. These have been archived in a digital library. The Authority puts the information together and calls for a meeting with Mark and Kelly, another environmental engineer.

Scenario III

In the meeting, Kelly reviews the references provided by the Authority, logs into the digital library, and obtains an ecological model of the watercourse and downstream abstractions that have been used before in that region. She explains to Mark that an abstraction license may not be granted because the downstream area will likely be affected in summer seasons in the years to come.

Scenarios IV

Eventually the Authority, Mark, and Kelly agree that the plan has no potential breaches of legal restrictions on water quantity and quality and that all water-related objectives can be achieved. The Authority grants a consent to the city council and registers this case in the digital library.

4 Knowledge Contexture in Communication

The above scenarios describe a typical problem-solving process in a scientific community in which

- a) online archiving of scientific *works* is highly desirable;
- b) digital libraries have to cope with online raw data which are semi-structured

and file based;

- c) components of scientific *works* are often formatted datasets, file-based, and operated instrumentally;
- d) components of scientific *works* are usually acquired or systematically derived from a scientific simulation, modeling environment, or data instruments; and
- e) a considerable degree of infrastructural transparency between computing-system components and their integration with users' workplaces is required.

Examining the communication patterns in the scenarios above, we find that the layers of communications are based on levels of knowledge in terms of natural or modelled systems in which water resources management is primarily related to spatial water bodies, an atmospheric water cycle such as snow or rainfall, and the dynamic processes of human and social uses of water (Obasi et al. 1996).

The management involves a system called a relatively closed hydrological system (C). Hydrology is the science that deals with processes governing the depletion and replenishment of water resources of the land areas of the earth. The long journey of events marking the process of a particle of water from the atmosphere to the land masses and oceans and its return to the atmosphere is termed a hydrological cycle. If the cycle is solar powered and energy but not matter (water) is exchanged with the outside environment, then the cycle is called a relatively closed hydrological cycle. If the cycle involves the exchange of neither energy nor matter with the outside environment, it is called a closed hydrological cycle.

If we define the geologic formation of water storage as W and the set of the natural environmental elements and its effects as system M , the level of the knowledge about the cycle can be denoted by the interaction I_0 between W and N as $C: \{W, M\}$.

Management involves a water resource system (WR). The natural setting of the cycle is not simple. It is a result of interaction between man and the hydrologic environment: few river systems are unregulated by surface-storage facilities; atmospheric scientists are actively pursuing the goal of augmenting and controlling precipitation and other objectives. In short, considering only the natural hydrologic phenomena is no longer suitable. If we define the set of the artificial environmental

elements and its effects as a system AE, another level of the knowledge about water resources system can be defined by the interaction I1 between C and AE as WR: {C, AE}.

From a knowledge engineering perspective, WRM is concerned with discrepancies that occur when the information related to the major components of the system are compared with established criteria in order to predict and understand the impacts of any action taken to control, manage, and use water. We can then define the interaction (I2) between WR and the criteria system (CR). CR consists of its subsystems: the system that deals with human demands for water for survival and the effects on social systems; the system that deals with environmental demands, such as living plants and living organisms needing water, and the effects on environmental systems; the system that deals with fast-developing economic needs, such as industry and agriculture, demanding water, and the effects on economic systems, etc. Now we have the third level of knowledge about water resources system as WMR: {WR, CR}.

The most important point here is that the different levels of knowledge form varying contextures as Table 5 shows.

Table 5. Instances of Knowledge Contextures

C contexture (I0)	The knowledge that relates to the results of interaction between the geologic formation of water storage (W) system and the set of natural environmental elements (N).
WR contexture (I1)	The knowledge that relates to the results of interactions between C system and a set of artificial environmental elements (AE).
WM contexture (I2):	The knowledge that relates to the results of interaction between WR system and CR system.

Formal scientific communication is made up of the documentation and dissemination of concepts through works such as scientific models (Coleman 2002, Papazoglou et al. 1999). Models are well-known intellectual entities critical in the creation and transmission of scientific knowledge (if not all human knowledge). Communication of scientific knowledge is comprised of up to three levels of reasoning (Morrissey 2002):

- a) data and empirical observations (phenomenology);
- b) information, formulas, trends, and predictions generalized from data or derived from hypothesis and theory (description); and,
- c) paradigmatic axioms, models, metaphors (explanation).

5 Some Open Issues

The importance of a coherent infrastructure-knowledge contexture in digital libraries relies on the representation of knowledge organized as works. Appendix

III shows how this may be partially accomplished through metadata in which the levels of infrastructure detail are integrated in an online knowledge service. Similarly, cross-browsing and cross-searching are operated through subject gateways based on meta-data (Dublin Core homepage, Z39.50 Document, Koch 2000, Hydrological metadata).

If, however, we consider the scenarios presented in section 3, it is more likely for the planner to find a real-world entity (i.e., body of water) by matching the same concept in other related works. For this kind of integration, metadata (even with the use of controlled vocabularies) may be too broad. Metadata and controlled vocabularies such as thesauri are inadequate for representing and computationally enforcing explicit formalizations of the mental concepts that people have about the real world. For example, a body of water can be a lake that serves as recreation or as a habitat for specific species. Therefore, a special concept or name must refer to the lake. People perform such mental operations based on associations, roles, and relationships all the time.

Web ontology represents entities in hierarchies rather differently (see Appendix II). The choice of hierarchies as the representation of ontology, however, leaves us with new problems. How can ontology relate to a level of infrastructure as well as to knowledge contexture? For example, a lake can be an object at all the three levels described in Table 5: it can be an element in the water-storage category or it can be an element in a criteria system because, for the Parks and Recreation Department, the same entity is a lake. Thus, in a problem-solving environment, geo-referenced ontology must be designed with different views for the same geographic phenomenon. This is similar to an object that has an identity and can play different roles (Hornsby 1999, Pernici 1990, Albano et al. 1993, Wong et al. 1997, Fonseca et al. 2002).

6 Conclusion

Digital library knowledge service is not just a description of the content of a digital item by a single unadorned URL suggesting "click here." There is a growing need for collocation and retrieval of scientific *works* in a problem-solving environment, not merely for the services on known-item identification based on subjects of disciplines. Consequently, an internet-based infrastructure is not only about creating a new computing platform but is also about augmenting a systematic contexture for communication in a problem-solving environment. For scientific users, online archiving of referenced scientific data is highly desirable. Digital libraries, therefore, have to cope with online raw data that are semi-structured or file based. Referenced electronic data are formatted datasets and operated instrumentally. These datasets are usually acquired or systematically derived from a scientific simulation or modelling environment. Thus, a considerable degree of infrastructural transparency between computing system components and their integration with users' workplaces is required. Within these contextures, retrieval of a *work* on the Web requires digital library services to be able to cross levels of domain knowledge and to include observation or experiment abstractions and

measurements within a given problem-solving environment.

Appendix I Examples of Web Services

Web services are the latest software components and technologies designed to bring us online for “what we need, when we need it via any device we choose and access.” Using a Web service, one can run or interact with an application without the application’s being present on the user’s machine. This calls for an increase in infrastructural effort. Indeed, in providing Web services for geo-referenced digital libraries, since we not only have more sophisticated end-users, we need a set of more dedicated Web services. The two web service cases described below indicate how an infrastructure determines the scalability of a given application. In the IBM case, the classic client-server-DB approach is suitable to well-structured data services; the Sun situation, however, calls for standards and open services such as more platform portability versus front-end and back-end couplings, more inter-operations between applications versus databases mediation, and more resource transformation engines versus data-bits transaction servers.

IBM Phone Book Web Service

Service Definition	To bring the phone book service to the Web, IBM organizes materials in terms of service categories. Data-centric services are request-driven wrappers for the relational databases. This type of service would typically be implemented as a servlet and/or Enterprise JavaBean (EJB) in a J2EE environment, which maps names to phone numbers and provides get and set operations for accessing the data. Process-oriented services perform complex operations that are network-requests spanning or existing outside of individual user requests. Value-added services range from retrieving the phone number to searching for extraterrestrial information, e.g., through a phone number of a local restaurant to all the other entertaining programs around.
Main Infrastructural Components	Web servers, clients, relational databases, JavaScripts, API-clients, JavaScripted Objects, TCP/IP, HTTP, XML, SOAP.
Infrastructural Dependencies	Client-Server model, general purpose databases, hosting stores, additional owner stores, analytical engines, application logics.

Sun Microsystems Cost-effective Web Service

Service Definition	<p>Services on demand. The Brazilian National Healthcare System has developed a java-based solution to 20,000 clinics serving over 12 million patients. These applications run E10K servers using sophisticated Java-based security implementation. Information is fed through a hierarchy of regional servers from hospitals, clinics, and pharmacies. Every patient-clinician interaction is recorded in the system, including X-rays, EKGs, endoscopic video exams, and prescription information. Physicians have access to all patient information via patient medical ID card that will be replaced by JavaCard.</p> <p>Coordinated-Negotiated services. Services that cross preserve data providers, long term solution evolution, and equipment lines. Much data is streamed directly from instruments.</p>
Main Infrastructural Components	<p>Business systems, Web servers, application servers, browsers, portals, browser clients, workstations, laptops, cell phones, PDAs, repositories, XML/SOAP, RDF, J2EE, servers(XML, JAX-RPC, JAXM, JAXB, JAXP), JDBC, JMS, JAXR, DB servers.</p> <p>JAX-RPC Programming with SOAP</p> <p>JAXM Message oriented middleware, using JMS</p> <p>JAXP Process the low-level contents of an XML</p> <p>JAXR Access the UDI or ebXML registry to advertise or discover a service</p> <p>JDBS Java databse access APIs</p>
Infrastructural Dependencies	<p>Packaged Service Creation Tools With Standard Protocols: JavaCard, Jini, JXTA, other Grid Third Parties</p>

Appendix II: Web Ontology

Ontology has a long history in philosophy, in which ontology is the study of a systematic account of beings, or existence of things, or to being in the abstract as a “reality.” Science has shown a reality that is structured all the way down. At bottom it consists not of four types of gunk—earth, water, air, and fire—but rather of a finite number of definite particles, lawfully related one to the other (Zimmerman

1996). Thus it is believed that we might build up substantial information about our world from the elementary information we find at the bottom of reality. So, ontology, in terms of philosophy, is a theory of arguing and explaining.

The term "ontology" has been used for a number of years by the artificial intelligence and knowledge representation community. It is now becoming a part of the standard terminology of a much wider community that includes object modelling and XML (Nambiar et al. 2002). The key ingredients that make up ontologies are vocabularies of basic terms and a precise specification of what those terms mean so that they can be enforced using first order logic. Numerous researchers believe that the ontological approach can help develop useful tools (Fensel 2001):

- a) Ontology is more than a controlled vocabulary. It provides a set of well-founded constructs that can be leveraged to build meaningful higher-level knowledge. The terms in ontology are selected with great care, ensuring that the most basic (abstract) foundational concepts and distinctions are defined and specified. The terms chosen form a complete set, whose relationship one to another is defined using formal techniques. It is these formally defined relationships that provide the semantic basis for the terminology chosen.
- b) Ontology is more than a taxonomy or classification of terms. Although taxonomy contributes to the semantics of a term in a vocabulary, an ontology includes richer relationships between terms. It is these rich relationships that enable the expression of domain-specific knowledge without the need to include domain-specific terms.

Appendix III: Example of Metadata about Identification Information

Identification_Information:	Citation:	Citation_Information:	Originator: Publication_Date: Title: Geospatial_Data_Presentation_Form: Publication_Information: Publication_Place: Publisher: Other_Citation_Details: Online_Linkage:
	Description:	Abstract: Purpose: Supplemental_Information: Time_Period_of_Content: Time_Period_Information: Range_of_Dates/Times: Beginning_Date: Ending_Date: Currentness_Reference: Status:	

	Supplemental_Information: Time_Period_of_Content: Time_Period_Information: Range_of_Dates/Times: Beginning_Date: Ending_Date: Currentness_Reference: Status: Maintenance_and_Update_Frequency: Spatial_Domain: Bounding_Coordinates: West_Bounding_Coordinate: East_Bounding_Coordinate: North_Bounding_Coordinate: South_Bounding_Coordinate: Keywords: Theme: Theme_Keyword_Thesaurus: Theme_Keyword:
--	---

References

ADL homepage, Alexandria Digital Library Homepage: publications, research papers, current bibliography <http://www.alexandria.ucsb.edu>

Albano, A., Bergamini, R., Ghelli, G., & Orsini, R. (1993). An object data model with roles. In R. Agrawal, S. Baker & D. Bell (Eds.), Proc. of 19th international conference on very large data bases. Dublin, 39-51.

Arms, W.Y. (2000), Digital libraries. Cambridge: MIT Press.

Benedict, D., Caron, J., Davi, E., Kambic, R., & Nativi, S. (2003), Thematic real-time environmental distributed data services (THREDDS):

Incorporating real-time environmental data and interactive analysis tools into NSDL. *Journal of Digital Information (JoDI)*, 2 (4), <http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Domenico/>

Borgman, C. L. (1999), What are digital libraries? competing visions. *Information Processing and Management*, 35 (3), 227-243.

Boxall, J. (2002), Geolibraries, the global spatial data infrastructure and digital earth: A time for map librarians to reflect upon the moonshot. *INSPEL*, 36, 1-21. <http://www.ifla.org/VII/d2/inspel/02-lboja.pdf>

Campbell, A. (2000), Where are map libraries heading? Some route maps for the digital libraries. *LIBER Quarterly: J. of European Research Libraries*, 10, 489-498.

Carpenter, B. (2003), What is grid computing?, ISOC Member Briefings. (Grid Computing #11). 15 April, <http://www.isoc.org/briefings/011/index.html>

Chen, S-S. (1998), Digital libraries—the life cycle of information. BE (Better Earth) Publisher.

Chen, X., Kim, D-h, Nnadi, N., Shah, H., Shrivastava, P., Bieber, M., & Wu, Y-F. (2003), Digital library service integration. In Proc. of the 2003 Joint Conference on Digital Libraries. <http://csdl.computer.org/comp/proceedings/jcdl/2003/1939/00/19390384.pdf>

Chu-Carroll, M. C., & Sprenkle, S. C. (2000), Brewing better collaboration through software configuration management. *SIGSOFT*, 88-97.

Coleman, A. S. (2002), Scientific models as works. *Cataloging and Classification Quarterly*, 33 (¾), 129-159. Preprint available: <http://dlist.sir.arizona.edu/archive/00000060/>

Coleman, A.S. Smith, T.R., Mayer, R.E. and Buchel, O.A. (2001). Learning spaces in digital libraries. *Lecture Notes in Computer Science*, Vol. 2163. Berlin, Springer-Verlag.

Collier, M. (1997), Towards a general theory of the digital library. In Proc. of the International Symposium on Research, Development and Practice in Digital Libraries: ISDL'97, Japan. <http://www.dl.ulis.ac.jp/ISDL97/proceedings/>

Crabtree, A., Twidale, M., O'Brien, J., & Nichols, M. D. (1997), Talking in the library: implications for the design of digital libraries. In R. B. Allen (Eds.) [Proc. of the 2nd ACM International Conference on Digital Libraries](#), Rasmussen, E., Philadelphia, PA, USA, 221-228.

DELOS homepage, DELOS Network of Excellence on Digital Libraries. 12

- December (2001). <http://www.ercim.org/delos/>
- D-Lib homepage, Working Group, D-Lib Working Group on Digital Library Metrics. 13 January (2003). <http://www.dlib.org/metrics/public/metrics-documents.html>
- DLSI homepage, Digital Library Service Integration. <http://www.is.njit.edu/dlsi/>
- Does KM=IT? homepage, <http://www.brint.com/advisor/a091599.htm>
- Dretske, F. I. (1981), Knowledge and the flow of information. Basil Blackwell Publisher.
- Dublin Core homepage, Dublin Core Metadata Element Set, <http://www.dublincore.org/documents/dces/>
- eLib homepage, <http://www.ukoln.ac.uk/services/elib/>
- Fayad, M. E., & Johnson, R. E. (2000), Domain-specific application frameworks. John Wiley & Sons.
- Fensel, D. (2001), Ontologies: silver bullet for knowledge management and electronic commerce. Springer-Verlag.
- Fonseca, F., Egenhofer, M., Agouris, P., & Camara, C. (2002), Using ontologies for integration geographic information systems. Transaction in GIS, 6(3), 1-33.
- Frew, J., Carver, L., Fisher, C., Goodchild, M., Larsgaard, M., Smith, T., & Zheng, Q. (1995), The Alexandria rapid prototype: Building a digital library for spatial information. In ESRI International User Conference, Environmental Systems Research Institute, Inc, Palm Springs, 255-262. <http://gis.esri.com/library/userconf/proc95/to300/p255.html>
- Frew, J., Freeston, M., Freitas, N., Hill, L., Janee, G., Lovette, K., Nideffer, R., Smith, T., & Zheng, Q. (1999), The Alexandria digital library architecture. International Journal on Digital Libraries, 2 (4), 259-268.
- Frew, J., Freeston, M., Kemp, R., Simpson, J., Smith, T., Wells, A., & Zheng, Q. (1997), The Alexandria digital library testbed. D-Lib Magazine, July/August. <http://www.dlib.org/dlib/july96/alexandria/07frew.html>
- Greestain, D., & Thorin, S. E. (2002), The digital library: A biography. Washington: Digital Library Federation and Council on Library and Information Resources, 2nd Edition December. <http://www.clir.org/pubs/reports/pub109/pub109.pdf>
- Harter, S. P. (1997), Scholarly communication and the digital library: problems and

- issues. J. of Digital Information, 4 April. <http://jodi.ecs.soton.ac.uk/Articles/v01/i01/Harter/>
- Hornsby, K. (1999), Identity-based reasoning about spatial temporal change. PhD Thesis, University of Maine, USA.
- Houstis, E., Gallopoulos, E., Bramley, R., & Rice, J. (Eds.) (1997), Problem-solving environments for computational science. IEEE Computational Science & Engineering, special issue, July/September.
- Hydrological Metadata, Metadata for Hydrological Coverage, <http://edc.usgs.gov/sast/meta/fnl/hydro.html>
- IBM report, Next generation knowledge management: The complexity of humans. Executive Tek Report. IBM Global Services. 13 November 2002. http://www.ibm.com/services/files/etr_cynefin.pdf
- Janée, G. (1999), Current architectures & known limitations (sign). Presentation. <http://alexandria.sdc.ucsb.edu/~gjane/presentation-junkyard.html>
- Janée, G. (2003), Issues in georeferenced/geospatial digital libraries, 15 May. <http://alexandria.sdc.ucsb.edu/~gjane/archive/2003/issues.pdf>
- JERL (2002), Strategies for survival. Conference of the Journal of European Research Libraries.
- JISC homepage, Joint Information Systems Committee, <http://www.jisc.ac.uk/>
- Keller, C. P. (2001), The map library's future. Cartographic Perspectives, 38 , 73-77.
- Kemp, D. A (1976), The nature of knowledg - An introduction for librarians. Clive Bingley Ltd., London, UK.
- Kiczales, G., Hilsdale, E., Hugunin, J., Kersten, M. Palm, J., & Griswold, W. (2001), An overview of aspect. In Proc. of ECOOP, June, 327-353.
- Koch, T. (2000), Quality-controlled subject gateways: definitions, typologies, empirical Overview. The International Journal of Digital Information Research and Use, 24 (1). 26-45. <http://www.mcb.co.uk/oir.htm>
- Isdis homepage, ADEPT-UGA Prototype Process, <http://lsdis.cs.uga.edu/~adept/process.html>
- Lyman, P., & Varian, H. (2002), How much information. Report, School of Information Management Systems, University of California at Berkeley. <http://www.sims.berkeley.edu/how-much-info/>

- Metadata homepage, Information/Data/Metadata Management - General Resources, http://www.cens.ucla.edu/Education/Documents/Information_Data_Metadata%20Mgmt%20Resources.htm
- Morrissey, F. (2002), Introduction to a semiotic of scientific meaning, and its implications for access to scientific works on the web. *Cataloging and Classification Quarterly*, fall. <http://phoenix.liunet.edu/~smiragli/entities.html>
- Nalhotra, Y. (2002), Is knowledge management really an oxymoron? Unravelling the role of organizational controls in knowledge management. In D. White (Eds.), *Knowledge Mapping and Management*, Hershey, PA: Idea Group Publishing, 1-13. <http://www.brint.org/KMOxymoron.pdf>.
- Nambiar, U., Bressan, S., Lee, M. L., Li, Y-G. (2002), Current approaches to XML management. *IEEE Internet Computing*. July-August.
- Nardi, B. A., & O'Day, V. L. (1999), *Information ecologies: Using technologies with heart*. Cambridge, MA: MIT.
- NSDL homepage, Building a "memory" for the national science digital library. Online (7) 2, January. <http://www.npaci.edu/online/v7.2/index.html>
- NSF-EU homepage, <http://www.dli2.nsf.gov/internationalprojects/workgroups.html>
- Obasi, G. O. P. (1996), Hydrological knowledge: A vital element in water management. In *Proc. of National Conference on Water Resources Management*, Ryto, Poland October. http://www.wmo.ch/files/sg_statements/english/SG004E.pdf
- Online Dictionary, <http://www.thefreedictionary.com/Contexture>
- Ossher, H., & Tarr, P. (2001), Multi-dimensional separation of concerns and the hyperspace approach. In *Proc. of Symp. Software Architectures and Component Technology: The State of the Art in Software Development*. Kluwer.
- Papazoglou, M., & Hoppenbrouwers, J. (1999), Contextualizing the information space in federated digital libraries. *SIGMOD Record*, 28(1), March, 40-46. <http://infolab.uvt.nl/people/hoppie/papers/sigmod-record-1999.pdf>
- Parry, R. B. (2003), Who's saving the files? Towards a new role for local map collections?. *LIBER Quarterly, the Journal of European Research Libraries*, 13.
- Pernici, B. (1990), Objects with roles. In *Proc. of IEEE/ACM Conference on Office Information Systems*, Cambridge, MA., 205-215.

- Peterson, E. (2001), Building a digital library: With comments on cooperative grant projects and the goals of a digital library. *Library Philosophy and Practice*, 2 (3), <http://www.uidaho.edu/~mbolin/lppv3n2.htm>
- Poland, J. (2000), Cooperative development of the digital library: Identifying and working with potential partners. In 66th IFLA Council and General Conference, August , 13-18.
- Rusbridge, C. (1998), Towards the hybrid library. *DL-Magazine* July/August .
- Schatz, B. R., & Hardin, J. B. (1994), NCSA Mosaic and the World Wide Web: Global hypermedia protocols for the Internet. *Science Magazine*, 12 (265), August, 895-901. <http://www.dlib.org/dlib/july98/Rusbridge/07rusbridge.html>
- Schmidt, D., Stal, M., Rohnert, H., & Buschmann, F. (2000), Pattern-oriented software architecture. Volume 2, Patterns for Concurrent and Networked Objects. John Wiley & Sons.
- Smiraglia, R. (Eds.) (2002), Works as entities for information retrieval. Medford, Conn.: Scarecrow Press.
- Smith, J. (2000), Can a map be a geographic information retrieval tool?. *LIBER Quarterly*, *J. of European Research Libraries*, 10 (10).
- Smith, T. R., Zeng, M. L., & ADEPT Knowledge Team, Structured models of scientific concepts for organizing, accessing, and using learning materials. http://esm.ucsb.edu/fac_staff/fac/frew/cv/pubs/2002_concepts.pdf
- Star, S. L. (2002), Got infrastructure? How standards, categories and other aspects of infrastructure influence communication. The 2nd Social Study of IT Workshop at the LSE ICT and Globalization, London, April, 22-23.
- Voelter, M., Schmid, A., & Wolff, E. (2002), *Server Component Patterns*. Wiley & Sons.
- Wong, R., Chau, H., & Lochovsky, F. (1997), A data model and semantics of objects with dynamic roles. In M. Jackson & C. Pu (Eds.), *Proc. of 13th International Conference on Data Engineering*. Birmingham, UK, 402-411.
- Z39.50 Document, <http://lcweb.loc.gov/z3950/agency/document.html>
- Zimmerman, D. W. (1996), Could extended objects be made out of simple parts? An argument for "atomless gunk.". *Philosophy and Phenomenological Research*, 51, 1-29.

[1] Digital libraries preserve our knowledge, cultural experiences or treasures that we often find in art galleries, libraries, museums, or digital publications. A good conceptual debate on this subject can be found in “What are digital libraries?” (Borgman 1999). Also, see web sites for main international digital library initiatives and research programs: in UK (JISC homepage, eLib homepage), US (NSDL homepage), EU (DELOS homepage), Joint NSF-EU Working Groups (NSF-EU homepage), D-Lib Working Group (D-Lib homepage); an introduction book in general (Arms 2000); system technologies and integration (Chen 1998); and a biography (Greestein et al. 2002).

[2] This issue has been raised widely in internet computing research fields in general, e.g., (Nardi et al. 1999, IBM report 2002, Does KM=IT? Homepage, Nalhotra, 2002). For digital library development in particular, cases can be found in (Harter 1997, Collier 1997, Rusbridge 1998, Campbell 2000, Keller 2001, Smith 2000, JERL 2002, Parry 2003, Peterson 2001, Poland 2000).

[3] See special issue: “Works as entities for information retrieval,” J. of Cataloging and Classification Quarterly (Smiraglia eds. 2002).

[4] In Houstis et al. (1997), a PSE is described as a computer system that provides all the computational facilities needed to solve a target class of problems. These features include advanced solution methods, automatic and semiautomatic selection of solution methods, and ways to easily incorporate novel solution methods. Moreover, PSEs use the language of the target class of problems, so users can run them without specialized knowledge of the underlying computer hardware or software. By exploiting modern technologies such as interactive colour graphics, powerful processors, and networks of specialized services, PSEs can track extended problem-solving tasks and allow users to review them easily. Overall, they create a framework that is all things to all people: they solve simple or complex problems, support rapid prototyping or detailed analysis, and can be used in introductory education or at the frontiers of science.

[5] See Appendix I

[6] See Appendix II

[7] See Information/Data/Metadata Management - General Resources (Metadata homepage).

[8] See What is grid computing (Carpenter 2003)?

[9] See, e.g., Fayad (2001) shows how to build *domain specific application* frameworks; Schmidt (2000) deals with *networked objects* in meta-computing; Ossher, Kiczales, and Chu-Carroll separate *multi-dimensional concerns* with methods of hyperspace, hyper-aspect, and configurable services (see, e.g., Ossher et al. 2001, Kiczales et al. 2001, and Chu-Carroll et al. 2000 respectively; Voelter designs *server component patterns* that cross multi-application domains (Voelter et al. 2002).