UNITED NATIONS EDUCATIONAL,
SCIENTIFIC AND CULTURAL ORGANIZATION

DIGITAL LIBRARIES IN EDUCATION, SCIENCE AND CULTURE

ANALYTICAL SURVEY

UNESCO INSTITUTE
FOR INFORMATION TECHNOLOGIES IN EDUCATION

UNESCO
UNESCO Institute for Information Technologies in Education (IITE)

IITE PROJECT COORDINATOR
Azat Khannanov (IITE)

WORKING GROUP:
Digital Libraries in Education:
Leonid Kalinichenko – coordinating author (Institute of Informatics Problems of the Russian Academy of Sciences, Russian Federation)
Daniel Atkins (University of Michigan, USA)
Edward Fox (Virginia Polytechnic Institute and State University, USA)
David Fulker (University Corporation for Atmospheric Research, Boulder, Colorado, USA)
Yannis Ioannidis (University of Athens, Greece)
Stephan Kornig (Darmstadt University of Technology, Germany)
Bertram Ludascher (San Diego Supercomputing Center, University of California, USA)
Mary Marlino (University Corporation for Atmospheric Research, Boulder, Colorado, USA)
Pasquale Savino (Istituto di Elaborazione della Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy)
Narasimhiah Seshagiri (Supercomputer Education and Research Centre, Bangalore, India)
Tamara R. Sumner (University of Colorado, Boulder, USA)
Alexey Ushakov (University of California, Santa Barbara, USA)
Michael Wright (University Corporation for Atmospheric Research, Boulder, Colorado, USA)

Digital Libraries in Science:
Tarikere Basappa Rajashekar – coordinating author (National Centre for Science Information, Bangalore, India)
Suvarsha Walters (National Centre for Science Information, Indian Institute of Science, Bangalore, India)
Madhuresh Singhal (Aurigene Discovery Technologies, Bangalore, India)

Digital Libraries in Culture:
Simon Tanner (King's Digital Consultancy Services, King's College London, United Kingdom)

Editor for the Digital Libraries in Science and Digital Libraries in Culture parts:
Marilyn Deegan (Centre for Computing in the Humanities, King's College London, United Kingdom)

Digital Libraries in Education, Science and Culture. Analytical survey
Analytical survey *Digital Libraries in Education, Science and Culture* has been developed by the UNESCO Institute for Information Technologies in Education in the context of the UNESCO cross-cutting theme project *Methodologies for Digital Libraries*. The project aims to give an overview of current and future technologies and applications for digital libraries (DL) including ethical, social, pedagogical, organizational and economic aspects, as well as their impact on learning, cultural and scientific activities. The survey concentrates on the following topics: state-of-the-art in the field of DL usage for educational, scientific and cultural purposes, needs for DL with analysis of real demands for DL solutions from the educational, scientific and cultural communities, and the main perspectives and trends of further DL development.

FOR FURTHER INFORMATION PLEASE CONTACT:
UNESCO Institute for Information Technologies in Education
8 Kedrova St. (Bld. 3), Moscow, 117292, Russian Federation
Tel.: 7 495 129 2990
Fax: 7 495 129 1225
E-mail: info@iite.ru
Web: www.iite.ru

TABLE OF CONTENTS

PREFACE

UNESCO intends to foster new forms of networking between educational, scientific and cultural institutions, using digital libraries (DL), as well as production and deployment of digital educational materials. At the same time, there is a lack of understanding of how DL function, suspicion that DL promotes commercial approach in educational, scientific and cultural spheres, fear that developing countries are used as new markets, along with concerns regarding quality assurance of the content provided through DL. In addition, the policy, planning, development and management of digital libraries differ from the traditional libraries. Decision-makers need access to information that will assist them in appropriate and efficient performing of their tasks.

Within this framework UNESCO started a cross-cutting theme project *Methodologies for Digital Libraries*. The project aims to give an overview of current and future technologies and applications for digital libraries including ethical, social, pedagogical, organizational and economic aspects, as well as their impact on learning, cultural and scientific activities. This analysis will be used to develop methodologies for the establishment of digital libraries in UNESCO's fields of competence and to elaborate specialized training courses on the use of digital libraries for educational authorities, teacher trainers, educators, researchers and students. The specific objective is to provide capacity building for decision-makers and major regional DL developers in the target regions in forming a base for planning the forthcoming phases of future projects for DL development. For implementation of the project the UNESCO intersectoral team has been formed, consisting of UNESCO Institute for Information Technologies in Education (IITE), UNESCO Communication and Information Sector, UNESCO Education Sector and regional UNESCO Offices, with participation of recognized international experts and specialists.

The analytical survey *Digital Libraries in Education, Science and Culture* (along with specialized training course *Digital Libraries in Education*, pilot project for the creation of a digital library for education purposes in Earth sciences and set of face-to-face and online workshops in the target regions) is one of the main outcomes of the project. It is concentrated at the following topics:

- state-of-the-art in the field of DL usage for the educational, science and cultural purposes, in particular an overview of existing theoretical, methodological and technological approaches to DL development;
- needs for DL with analysis of real demands for DL solutions from educational, science and cultural communities and some explications of what can be done for this sphere through DL development;
- the main perspectives and trends of further DL development, including the problems, which should be solved, necessary resources for advancement, essential changes in state policies and plans and demands for international cooperation.

The first part of the survey – *Digital Libraries in Education* – has been already published by IITE in 2003. But taking into account the holistic and systematic approach, used for the entire UNESCO project *Methodologies for Digital Libraries*, close linkage between technological, methodological and organizational aspects of using DL in different fields and real demand for comparative cross-cutting analysis of DL development in education, science and culture, IITE considered it necessary to publish all three parts in one book.

Vladimir Kinelev
Director UNESCO IITE

GLOSSARY OF ACRONYMS

AAAS	American Association for the Advancement of Science
ACI	Autonomous Citation Indexing
ACM	Association for Computing Machinery
ADEPT	Alexandria Digital Earth Prototype
ADL	The US Department of Defense's Advanced Distributed Learning Network
ADN	ADEPT/DLESE/NASA Joint Metadata Framework
ADT	Australian Digital Theses
AHDS	Arts and Humanities Data Service
AJOL	African Journals OnLine
AMPS	Academic Media Production Services
APSR	Australian Partnership for Sustainable Repositories
ARIIC	Australian Research Information Infrastructure Committee
ARROW	Australian Research Repositories Online to the World
ASR-OSS	Australian Service for Knowledge of Open Source Software
BBSRC	Biotechnology and Biological Sciences Research Council
BOAI	Budapest Open Access Initiative
CAMiLEON	Creative Archiving at Michigan & Leeds: Emulating the Old on the New
CARET	Centre for Applied Research in Educational Technologies
CASI	Centre for Aerospace Information
CAUL	Council of Australian University Librarians
CBIR	Content Based Image Retrieval
CDSI	Committee on Dissemination of Scientific Information
Cedars	Curl Exemplars in Digital ARchiveS
CI	Core Integration (of NSDL)
CITIDEL	Computing and Information Technology Interactive Digital Educational Library
CLEF	Cross-Language Evaluation Forum
CMU	Carnegie-Mellon University
CODATA	Committee on Data for Science and Technology
COMET	Cooperative Program for Operational Meteorology, Education and Training
CO-ODE	Collaborative Open Ontology Development Environment
COVAX	Contemporary Culture Virtual Archives in XML
CRM	Conceptual Reference Model
CSPR	Committee on Scientific Planning and Review
CSTC	Computer Science Teaching Centre
CS-TR	Computer Science Technical Reports
CURL	Consortium of Research Libraries in the British Isles
DAEDALUS	Data-providers for Academic E-content and the Disclosure of Assets for Learning, Understanding and Scholarship
DAML	DARPA Agent Markup Language
DANA	Digital Archive Network for Anthropology
DARPA	Defense Advanced Research Projects Agency
DART	Dataset Acquisition, Accessibility and Annotation e-Research Technology
DC	Dublin Core
DCC	Digital Curation Centre
DDB	Die Deutsche Bibliothek
DERWeb	Dental Educational Resources on the Web
DFG	Deutsche Forschungsgemeinschaft
DIG35	Metadata Standard for Digital Images
DINI	Deutsche Initiative für Netzwerk Information
DIRECT	Decentralized Image Retrieval for Education
DL	Digital Library
DLE	Digital Library in Education
DLESE	Digital Library for Earth System Education

GLOSSARY OF ACRONYMS

DLI	Digital Libraries Initiative
DNER	Distributed National Electronic Resource
DOI	Digital Object Identifier
DPC	DLESE Program Center
DTD	Document Type Definition
DV	Digital Video standard
EAD	Encoded Archival Description
ECDL	European Conference on Digital Libraries
EPSRC	Engineering and Physical Sciences Research Council
ERCIM	European Research Consortium for Informatics and Mathematics
ESS	Earth System Science
ETD	Electronic Theses and Dissertations
ETRDL	ERCIM Technical Reference Digital Library
EZB	Elektronische Zeitschriftenbibliothek (Electronic Journals Library)
FAGS	Federation of Astronomical and Geophysical Data Analysis Services
FAIR	Focus on Access to Institutional Resources
FE	Further Education
FGDC	Federal Geographic Data Committee standard for digital geospatial metadata
FRODO	Federated Repositories Of Digital Objects
GILS	Global Information Locator Service
GOS	Global Observing Systems
HaIRST	Harvesting Institutional Resources in Scotland Testbed
HCII	Human Computer Interaction Institute
HE	Higher Education
ICADL	International Conference of Asian Digital Libraries
ICSU	International Council of Scientific Unions
ICTs	Information and Communication Technologies
IEEE	Institute for Electrical and Electronics Engineers
IITE	UNESCO Institute for Information Technologies in Education
IMS	Instructional Management Systems
INASP	International Network for the Availability of Scientific Publications
INEX	INitiative for the Evaluation of XML Retrieval
IPR	Intellectual Property Rights
IT	Information Technologies
ITU	International Telecommunication Union
JANET	Joint Academic NETwork
JCDL	Joint (IEEE/ACM) Conference on Digital Libraries
JDL	NASA's Joined Digital Library
JISC	Joint Information Systems Committee, UK
JISC IE	JISC Information Environment
JSTOR	Scholarly Journal Archive
K-MODDL	Kinematic Models for Design Digital Library
LANL	Los Alamos National Laboratory
LOCKSS	Lots of Copies Keep Stuff Safe
LOM	IEEE Learning Object Metadata
LTRS	Langley Technical Report Server
LTSC	IEEE Learning Technology Standards Committee
MAMS	Meta Access Management System
MAPS	Middleware Action Plan and Strategy
MBL	Multimedia-Based Learning
MC	Manchester Computing
MERRI	Managed Environment for Research Repository Infrastructure
METS	Metadata Encoding and Transmission Standard
MIT	Massachusetts Institute of Technology
MLE	Managed Learning Environments
MMIM	Molecular Medicine Informatics Model
MR	Metadata Repository

MXF	Material eXchange Format
NACA	National Advisory Committee for Aeronautics
NaCTeM	National Centre for Text Mining
NASA	National Aeronautics and Space Administration, USA
NCSA	National Center for Supercomputing Applications
NCSTRL	Networked Computer Science Technical Reference Library
NDLTD	Networked Digital Library of Theses and Dissertations
NLA	National Library of Australia
NLP	Natural Language Processing
NSDL	National Science Digital Library
NSF	National Science Foundation, USA
NSTA	National Science Teachers Association, USA
NTRS	NASA Technical Report Server
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OCLC	Online Computer Library Center
OIL	Ontology Inference Layer
OPAC	Online Public Access Catalogue
OSI	Open Society Institute
OSOSS	Online Self-Organizing Social Systems
OWL	Ontology Web Language
PAA	Priority Area Assessment
PANDAS	PANDORA Digital Archiving System
PANDORA	Preserving and Accessing Networked Documentary Resources of Australia
PC	Policy Committee
PERI	Programme for the Enhancement of Research Information
PERSIVAL	PErsonalized Retrieval and Summarization of Image, Video And Language Resource
PICats	Publishable Inventories and Catalogues
PORTAL	Presenting natiOnal Resources To Audiences Locally
QBIC	Query By Image Content
RBS	Ranking By Sources
RDF	Resource Description Framework
RDN	Resource Discovery Network
RLG-CMI	Research Libraries Group Cultural Materials Initiative
RoMEO	Rights Metadata for Open Archiving
RUBRIC	Regional Universities Building Research Infrastructure Collaboratively
SCALE	Services for a Customizable Authority Linking Environment
SCG	Semi-structured Conceptual Graphs
SCORM	Sharable Content Object Reference Model
SDLIP	Simple Digital Library Interoperability Protocol
SDSC	San Diego Supercomputer Center
SGML	Standard Generalized Markup Language
SHERPA	Securing a Hybrid Environment for Research Preservation and Access
SII	Systemic Infrastructure Initiative
SPARC	Scholarly Publishing and Academic Resources Coalition
SRB	Storage Resource Broker
SRU	Search Retrieve URL
STEM	Science, Technology, Engineering and Mathematics
STI	Scientific and Technical Information
SVG	Scalar Vector Graphics
TARDis	Targeting Academic Research for Dissemination and Disclosure
TDs	Theses and Dissertations
TEI	Text Encoding Initiative
THREDDS	Thematic Realtime Environmental Distributed Data Services
TTP	Trusted Third Parties
UCB	University of California at Berkeley
UKOLN	UK Office for Library and Information Networking

ViFas	Virtuelle Fachbibliotheken (Virtual Libraries)
VLE	Virtual Learning Environments
VRD	Virtual Reference Desk
VTLS	Visionary Technology in Library Solutions
W3C	World Wide Web Consortium
WATERS	Wide Area Technical Report Service
WDC	World Data Centers
WSDL	Web Services Description Language
WSIS	World Summit on the Information Society
XML	eXtensible Markup Language
XSCD	XML-based Secure Content Distribution

DIGITAL LIBRARIES IN EDUCATION

1. Introduction

The project aimed at Digital Libraries in Education (DLEs) is being developed in accordance with strategic goals of the UNESCO Institute for Information Technologies in Education (IITE). UNESCO intends to foster new forms of networking between teacher-training institutions and teachers, using digital libraries (DLs) as well as production and deployment of digital educational materials. This project started with the international expert meeting held in June 2001 in Moscow. The information materials *Digital Libraries in Education: State-of-the-Art Report* had been prepared as a basis for discussion during that meeting [IMEM01]. These materials contained a brief analysis of (1) current trends leading to quite rapid changes of learning environments in contemporary society and of (2) the most notable existing programmes for the development of DLEs. One of the recommendations of the expert meeting was to continue investigation of current experience with DLEs, under the IITE framework, and to prepare an analytical survey of the area, to be widely disseminated. This analytical survey is the current stage of the IITE project on DLEs.

The content of the analytical survey reflects the results of joint work and meetings with several US groups: the NSDL (National Science Digital Library) Policy Committee, UCAR (University Corporation for Atmospheric Research, Boulder, Colorado), SDSC (San Diego Supercomputer Center, University of California, San Diego), and the University of Michigan (Ann Arbor). The conversations involved Edward Fox, David Fulker, Mary Marlino, Timothy Spangler, Tamara Sumner, Michael Wright, Bertram Ludäscher, Michael Freeston, Alexey Ushakov, Daniel Atkins, and their colleagues. The analytical survey also incorporates information collected at these meetings and centres. Additionally the survey includes excerpts of [IMEM01] that were discussed in 2001 with Yannis Ioannidis (Greece, University of Athens), Stephan Körnig (Germany, Darmstadt University of Technology), Pasquale Savino (Italy, IEI CNR, Pisa), Narasimhiah Seshagiri (India, SERC, Bangalore), and other participants of the international expert meeting at the IITE in June 2001.

The analytical survey is a step in procedures for the promotion of international cooperation, globalization of products and methodologies, and their adoption in communities. It focuses on the application of information technologies in education. Through consideration of some carefully selected DLE projects, the survey provides a detailed analysis of current DLE technologies and their anticipated evolution, thus forming a base for planning the forthcoming phases of future projects for DLE development.

The survey does not emphasize hybrid libraries, which draw together online and physical collections and services, presenting them to the user in a seamless and integrated manner, supported by middleware that handles aspects, such as authentication and cross-searching. In universities and colleges, libraries offer their catalogues online, many through a standard Web browser. University libraries also offer access to increasing numbers of electronic journals and other online information sources, including those provided internally or from remote locations. Virtual versions of library services, such as reservations, registration, and reference enquiries, are also starting to be offered, particularly to distance learners.

The survey does not attempt to analyse every known DLE-related project. Instead, several DLE projects have been carefully selected to provide for identification of typical DLE features, characterization of the current state of DLE technologies, clarification of the social and organizational issues surrounding DLE development, and prediction of the DLE impact on education. Advanced frameworks and methodologies related to DLEs are described to forecast the further evolution of DLE technologies and applications.

The survey does not consider every educational discipline, instead concentrating on education in the natural sciences and engineering. Thus the specificity of DLEs for many other disciplines needs to be investigated further. Geographically, the report is based on information produced mostly in the USA and Europe. Collecting information about the state-of-the-art in other regions is important for UNESCO. Several significant issues (such as sustainability and economic issues, DLE globalization versus national or regional development) are not sufficiently analysed in this document and will require separate discussion.

As the area of DLEs is too broad a subject to cover exhaustively in one survey, and in view of the constraints on the survey mentioned in the previous few paragraphs, the survey also outlines subjects for further analysis.

The survey is the result of the joint effort of discussions with, and information provided by, the Working Group:

Daniel Atkins (USA, University of Michigan)
Edward Fox (USA, Virginia Polytechnic Institute and State University)
David Fulker (USA, University Corporation for Atmospheric Research, Boulder, Colorado)
Yannis Ioannidis (Greece, University of Athens)
Leonid Kalinichenko (Russian Federation, Institute of Informatics Problems, Russian Academy of Sciences, Moscow), coordinating author
Stephan Körnig (Germany, Darmstadt University of Technology)
Bertram Ludäscher (USA, San Diego Supercomputing Centre, University of California)
Mary Marlino (USA, University Corporation for Atmospheric Research, Boulder, Colorado)
Pasquale Savino (Italy, Istituto di Elaborazione della Informazione, Consiglio Nazionale delle Ricerche, Pisa)
Narasimhiah Seshagiri (India, Supercomputer Education and Research Centre, Bangalore)
Tamara R. Sumner (USA, University of Colorado)
Alexey Ushakov (USA, University of California, Santa Barbara)
Michael Wright (USA, University Corporation for Atmospheric Research, Boulder, Colorado)

The survey is structured as follows: Section 2 contains a general discussion of the role of digital libraries as an integral part of the rapidly changing educational environment. Section 3 presents the wide range of understanding of the digital library concept. Section 4 gives a general picture of DLEs as repositories of educational resources, with services. Section 5, *Integrated Learning Environments,* presents definitions and information about Managed Learning Environments (MLEs) and Virtual Learning Environments (VLEs) and interactions between them. The related pedagogic models for VLEs are also discussed there. Section 6 lists criteria for digital library quality within the learning environment. This section also contains a brief description of the proposal for a joint project between the UK Joint Information Systems Committee (JISC) and the US National Science Foundation (NSF) to incorporate VLEs and digital libraries into the learning process. Section 7 surveys activities to standardize educational metadata.

Sections 8 and 9 survey current DLE technologies. Several American projects (NSDL, DLESE, CITIDEL, and NDLTD) and several European projects (DNER, Scholnet, and Cyclades) were selected to demonstrate the current state-of-the-art and the planned evolution of technologies. (See those sections for the full names of these projects.) To show the anticipated evolution of DLE frameworks, from current ones based on the conventional library metaphor to more knowledge-based systems, Section 10 collects information on several advanced frameworks and methodologies related to DLEs. Issues in instructional course development with reuse of preexisting learning objects are discussed in Subsections 10.1 and 10.2. Work on extension of the information content of DLs for education and research with specialized educational resources of data (including real time data) is reflected in Subsections 10.3 and 10.4. The use of advanced infrastructures (cyberinfrastructure and data grid) as a possible basis for future DLEs is considered in Subsections 10.5 and 10.6. Research on knowledge-based approaches for DLE frameworks is included in Subsections 10.7 (navigation and search interface for NSDL based on science literacy benchmarks) and 10.8 (preparing a course of physical geography in an ADEPT concept-based architecture). Section 11 summarizes the survey. A reference list of publications used for compiling the survey is given at the end of the document.

2. Transforming the way to learn

2.1. Transformation of the educational system is inevitable

We are on the threshold of a revolution that is making the world's accumulated information and knowledge accessible to individuals everywhere. Digital technology will not only transform the intellectual activities of universities, it is likely to cause restructuring of the current higher-education enterprise into a global "knowledge and learning" industry.

A radical transformation of the educational system is coming under growing pressure in the school systems themselves, changing society into one in which knowledge work becomes ever more important, and developing information and communication technologies (ICTs) which are transforming the economies. The following trends indicate that the transformation is inevitable [DLTO01]:

- The number of students is still growing, as health and population increase.
- Different types of students are asking for education; different students bring different experiences with them.
- Increasingly, work and study are combined, and that leads to a need for more flexible learning arrangements in which the campus or school building is no longer central to the educational process.
- More generally, there is a trend towards lifelong learning.
- Lifelong learning leads to an emphasis on "learning to learn". Knowledge[1] becomes obsolete at an ever-increasing rate in a knowledge economy, and knowledge workers need to be able to refresh their knowledge on a regular basis.
- Because of the differences among students, there is a need to accommodate different learning styles, to provide customization and alternative learning routes. Courses have to take better account of the different experiences and backgrounds of students.
- Higher education institutions have long had a monopoly in providing education, but increasingly, companies and public bodies possess knowledge that can be reused for educational purposes, partly for in-house training (knowledge management) but also to offer to external markets.
- Education is under constant budget pressure; thus there is a need for more efficient and effective education.
- Students increasingly are behaving like consumers, and want to make informed choices about how and where they want to be educated, which implies that students are no longer committed to one institution.
- Teaching staff will exhibit more job-hopping behavior than they did in the past.
- There are too many dropouts in the current educational system.

A global learning environment of the future is envisaged as:

- being student-centred,
- being interactive and dynamic,
- enabling group work on real world problems,
- enabling students to determine their own learning routes,
- emphasizing competencies like information literacy to support lifelong learning.

Active learning implies that students do not limit themselves to resources supplied by their instructors, but also search for and organize new materials themselves in order to solve problems and to develop their competencies continuously.

2.2. Information users in the changing educational environment

It is important to analyse the characteristics of users who will make significant choices within this new educational environment. Starting with the students, one of the first and most important characteristics to note is that of preferring to be self-sufficient in their information gathering. It seems that users are beginning to perceive the traditional library as something used at the end, or at best the middle, of their information search. This has important implications for education programmes, as well as for understanding how those users who come to the library decide to do so. They do not come to the library first for the problem definition and information gathering phases. They prefer to dive into the problem alone first rather than coming to the common space.

Users want control of their own information environment. It is important to them to have some items owned for convenient personal consultation. They prefer to use a private good rather than a common good if they can. Users do

[1] The word "knowledge" is an overloaded term. This word is used here to emphasize that such phrases can often be found in the DL context. Of course, fundamental knowledge has not changed so rapidly.

not want to be dependent on anyone else if they can afford (in terms of both time and money) not to be. The convenience factor and the value added by the functionality of the service itself will be key in how choices about service providers are made.

The amount of information used in one's professional work that is available openly on the Internet is dramatically greater than it was in 1990. Users have the impression that they will find useful information free on the Web. If students do need to ask a person for information help, they go to a friend or co-worker because that person already has an understanding of the student's context for either the problem or their level of understanding.

Because of the characteristics described above, people are becoming ready to pay for ubiquitous, convenient, fast, and customized information access (this may not apply in developing countries). Faculty also wants control of their information environments, especially for teaching. Course management software packages are proliferating on campuses as part of a larger academic agenda to address the need to support information technology in both distance education and campus-based learning. Conventional libraries have an uncertain role in Web-based learning environments.

A second change in learning environments is that of a greater emphasis on the public scrutiny of teaching and learning. The quality of someone's teaching is no longer a personal matter or departmental matter. Faculty behavior in the classroom is brought out of the dimension of a personal contribution and is considered a service that is evaluated for its quality, just like other services.

2.3. Digital Libraries in Education as a way of restructuring the current higher–education enterprise into a global "knowledge and learning" industry

One of the natural responses to the above challenges consists in introducing the DLEs as a core of networks of learning environments and resources [SLRM01], that is:

- Designed to meet the needs of learners, in both individual and collaborative settings;
- Constructed to enable dynamic use of a broad array of materials for learning, primarily in digital formats;
- Managed actively to promote reliable anytime, anywhere access to quality collections and services, available both within and outside the network.

The digital library must not be seen as merely a digitized collection of information objects plus related management tools, but as an environment bringing together collections, services, and people to support the full cycle of creation, dissemination, discussion, collaboration, use, new authoring, and preservation of data, information, and knowledge. The challenges and opportunities that motivate advanced DL initiatives are associated with this view of the digital library environment. Work on digital libraries aims to help in generating, sharing, and using knowledge so that communities become more efficient and productive and the benefits of collaboration are maximized. It seeks to aid existing communities and to facilitate the emergence of new communities of research and education.

Introducing digital libraries into the education process has been underway in distance education for a number of years [DELO00]. With the Internet and the Web, distance education programmes can mount sets of materials on Web servers to support a course. The range of materials that currently are in digital form is great. In some disciplines enough materials are available with open access so that students already have access to broad collections. Digital libraries can provide adequately broad library services to local and remote students. One of the basic ideas is to join learning materials on various topics and written by many teachers in a digital library of courseware [DELI97]. Such a DL provides a basis for creating courses on specific topics.

Applying digital libraries in education has the potential to drastically change fundamental aspects of the classroom [WTSM99] in ways that could have an enormous impact on teaching and learning. The DLEs can be seen as an *information space* in which students are moving around intellectually, encountering new information, and working with the teacher and other students to make sense of what they encounter. A traditional library for education, a school library for example, typically includes textbooks, curriculum materials, artifacts (such as charts, physical samples, and equipment), enrichment books, and the teacher's own personal collection of teaching tools (as well as resources shared by other educators and learners). Taken generically, this space has been constituted over the decades to include the content required for teaching, for example, a sanctioned subject to high school students. The texts may be carefully designed and written, and the entire space is designed to bring students into the discipline

that the class represents. Thus students can be exposed to only the best products of the human mind. The content of education can be carefully evaluated and filtered to include only the most worthy. Such a point of view may lead to a controversy with the point of view that emphasizes expanding the number of educational resources and avoiding being prescriptive and confining.

Educational applications of digital libraries range from primary schools through graduate schools and across all disciplines. DLs change the possibilities for the education information space. Boundaries are expanded to include not only canonical versions of the subject, but other products put into DLs. However, the DL content can be controlled better than that of the Web. Although some would argue that content on the Web is motivating and interesting to students, it is varied and unpredictable in its design, rather than carefully designed to help students learn. Web sites can be complex and confusing or deep and significant. The quality, quantity, and substance of information available in the classroom teaching and learning space may become vastly different once the Web is included. Clearly, there are both positive and negative features of the new space opened up by the Web. In fact, positive and negative characteristics of the traditional text-based classroom and of the Web can be seen as mirror images of each other. DLE offers a middle ground between the overly constrained information space of traditional classroom resources and the overly open and undependable information space of the Web. Controls on content allow accountability and dependability while open submission provides opportunities for a wealth of materials beyond what an individual teacher or school system would envision and select.

New pedagogical methods should accompany DLs as an emerging technology to reach the compelling vision of education expressed in [ACRA98]:

“Any individual can participate in online education programmes regardless of geographic location, age, physical limitation, or personal schedule. Everyone can access repositories of educational materials, easily recalling past lessons, updating skills, or selecting from among different teaching methods in order to discover the most effective style for that individual. Educational programmes can be customized to each individual’s needs, so that our information revolution reaches everyone and no one gets left behind”.

Describing the situation in the USA, [ACRA98] reported that “in education, information technology is already changing how we teach, learn, and conduct research, but important research challenges remain. In addition to research to meet the scalability and reliability requirements for information infrastructure, improvements are needed in the software technologies to enable development of educational materials quickly and easily and to support their modification and maintenance. We know too little about how best to use computing and communications technology for effective teaching and learning. We need to better understand what aspects of learning can be effectively facilitated by technology and which aspects require traditional classroom interactions with the accompanying social and interactive contexts. We also need to determine how best to teach our citizens the powers and limitations of the new technologies and how to use these technologies effectively in their personal and professional lives”.

“Access to and use of IT, particularly in educational settings (K-12 [primary and secondary] as well as higher education), is a prerequisite to building the skills base that will allow our citizens to function productively in the information society of the next century. It is also a critical stepping stone for instilling interest and developing the skills of the budding IT researchers who will be essential to sustaining our national research capabilities”.

It was predicted that “the Nation [is] facing an impending crisis in preparing workers to be productive in an economy that is increasingly dependent on IT. Although the use of computers in education is increasing at all levels, and computer literacy is increasing dramatically across the country, too little percentage of the population are entering or receiving necessary re-training in the computing, information, and communications professions. Market forces alone will not correct the problem. The government must do more to help educate and re-train people in these crucial fields and to bolster the academic pipeline from elementary school to post-graduate study”.

This and other analyses gave rise to various research and development programmes for DL technologies in education around the world and resulted in planning specific research areas, including [NSF996]:

- preservation and archiving of digital scholarly information, including technology and procedures for long-term information asset management;
- utilization of digital libraries in educational technology at all levels of instruction, electronic publishing, and scholarly communication technology, including
- collaboratories, online repositories, and new methods of organizing scientific knowledge distribution.

3. The digital library concept: A common term with many interpretations

Digital libraries are becoming an integral part of digital learning environments. At the same time, the notion of "digital library" is subject to a broad range of definitions. Different audiences associated with a digital library have different interpretations; they evaluate a digital library differently and use different terminologies. On one end of the range, DLs are considered to be related to physical libraries performing similar functions, thus creating a *hybrid library* (combining traditional and electronic resources). On the other end, DLs are considered to be knowledge repositories, and services, organized as complex information systems. For example, global information repository projects are devoted to the accumulation of digital forms of information related to the Earth, universe, art, environment, or humans.

Such diversity of definitions is reflected in the different communities related to DLs. One such community is the research community formed in the United States by the Digital Libraries Initiatives (DLI) of NSF. DLI did not attempt to define "digital library" strictly. In order to incorporate a wide range of possible approaches and domains, the concept is treated broadly and vaguely. Thus the projects cover a wide range of topics, stretching the possible meaning of "digital library". This is perfectly acceptable for research – frontiers need to be stretched. On the other hand, the practicing community, the majority of whom reside in operational libraries, concentrate on building operational digital libraries, maintaining and operating them, and providing services to users. As a result, thousands of digital libraries have emerged worldwide. The efforts are diverse. Numerous types of collections and media are included and processed in many different ways. Some of them are located in physical libraries, creating a hybrid library, while others are not bound to a library at all. JISC in the UK (in contrast to the US NSFs DLI effort) promotes the innovative application and use of information systems and information technology, and emphasizes the development of content and new technologies that would be widely applicable and not just of benefit to the participating institutions. In 1999 NSF and JISC established a joint initiative to bring together the best elements of the two funding bodies.

To show the difference of understanding of the notion of DLs, several definitions of "digital library" follow. The broad (unrestricted) definition may be considered closest to the approach taken by the research community [DLCO00]:

"Digital libraries are organized collections of digital information. They combine the structure and gathering of information, which libraries and archives have always done, with the digital representation that computers have made possible" [PDLB97].

The following definition may serve as a bridge between the research and practicing communities:

"Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information; they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium. The content of digital libraries includes data and metadata. Digital libraries are constructed, collected, and organized by (and for) a community of users, and their functional capabilities support the information needs and uses of that community" [COVB99].

The Digital Library Federation (USA) agreed in 1999 on the following working definition of a digital library representing a definition of the practicing community. In this definition the emphasis is on an organizational or institutional setting for the collection of digital works and aspects related to its functioning in the larger context of service:

"Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities" [DLFS99].

The United Nations Task Force on Digital Libraries gives the following definition of digital libraries:

"Digital libraries are organized collections of information resources in digital or electronic format along with the services designed to help users identify and use those collections. Digital libraries promise to provide more effective

information services than has been possible in the past, by offering the following advantages: faster delivery, a wider audience, greater availability, more timely information, more comprehensive".

The Technical Committee on Digital Libraries of the IEEE (Institute of Electrical and Electronics Engineers) Computer Society (IEEE-CS) uses the more general term "(digital) collective memory" to emphasize the convergence of libraries, museums, archives, and collections of all kinds including those of private citizens. Collective memory development faces challenges in several areas, including storage, classification, and indexing; user interfaces; information retrieval; content delivery; presentation, administration; and preservation.

Diversity of understanding of what digital libraries are leads to a wide range of possible visions for DL frameworks and methodologies of use: from the conventional library metaphor to knowledge-based systems. This survey focuses on evaluation of digital libraries in learning environments.

4. Digital libraries of educational resources and services

"The network is the library" [SLRM01]: In a library, be it digital or analog, the essential transaction is the same: a user interacts with content. But richer interaction is possible within the digital environment, not only as more content is put within reach of the user, but also as more tools and services are put directly in the hands of the user. These include the abilities to search, refer, validate, integrate, create, customize, publish, share, notify, and collaborate, to name but a few. Students, teachers, faculty, and those pursuing continuing education will "connect to learn"; but they will also "learn to connect", as they leverage their participation with other users of the library and its resources.

By networking users and content with tools, the digital library enables three chains of support. First, *users supported by profiles* are able to form *learning communities*. These can be communities of one or they may be communities of thousands; they may be short-lived communities born of immediate needs, or they may grow into persistent communities. However, an important concern to acknowledge is the potential loss of privacy, which must be balanced against the potential gain in personalization of a user's experience. A second chain of support closely related to the first is that *content supported by metadata enables the formation of customizable collections of educational objects and learning materials*. These collections may target an individual or they may target a community; and they may learn and adapt to the behavior of their users. Finally, *tools supported by common protocols or standards enable* the development of varied application services that enhance the value of the library's content for the learner.

The following long-range objectives for DLEs were formulated by Tom Kalil (The White House) and further articulated in an article by Zia [SLRM01]:
- Lifelong learning.
- Learning anytime anywhere.
- Distance learning demonstration programme.
- Government as "model user" of technology-based training.

For these objectives, a number of intermediate goals are formulated, such as:
- Improve student performance.
- Get more students excited about science.
- Increase the quantity, quality, and comprehensiveness of Internet-based science educational resources.
- Make these resources easy to discover and retrieve for students, parents, and teachers.
- Ensure that these resources are available over time.

Studies show [SLWSJ98] that the Internet has the potential to transform the highest level of education, but only a fraction of that potential is now being realized. Some of this gap lies in the maturation process that is part of any transition, but a larger part is the result of fragmentation. Resources of great value are not being used because students and faculty do not know about them, or do not know how to use them.

While great efforts have been placed on creating materials, less attention has been given to organizing them, maintaining them in the long term, helping people find them, and training people how to use them. For example, a faculty member who is planning a course has only the most rudimentary tools to discover what materials are available or whether they have proved effective in other courses. A student who is researching a topic is forced to choose between general-purpose Web search services and commercial databases designed for scientific and technical research. Neither faculty nor students can safely rely on resources that might be withdrawn without notice, or change subtly overnight.

A DLE is envisaged as a *comprehensive library of the digital resources and services* that are available for education in science, mathematics, engineering, technology, and other disciplines. The key word here is "comprehensive". Faculty are very specific in wanting a single place where they and their students can discover, use, and possibly contribute a wide range of materials.

A DLE is considered to be a *federation of library services and collections* that function together to create a digital learning community. Organizationally, a DLE will consist of a small central operation with a wide range of partners. Some of the services and collections are already well organized; for these, the DLE will act as a gateway. Others exist but are poorly organized; for these, the DLE will stimulate the creation of specific services. Some materials are fragmented,

unorganized, or hard to find; in these cases the DLEs will build library services and may even manage specific collections. Across all these areas, the DLEs will provide tools to help faculty and students find and use materials, with services to assist them in evaluating quality and appropriateness.

DLEs will take a broad view of science and technology, and of scientific education. The primary audience is faculty and undergraduate students, but there is no hard distinction between the needs of high school students, undergraduates, and graduate students, nor between students in formal programmes, independent learners, and the general public.

DLEs should have a variety of *financial models* for access to the materials; some content will be free of charge while other materials will be available on a fee basis.

The range includes curricula and courseware materials, lectures, lesson plans, computer programmes, models and simulations, intelligent tutoring systems, access to remote scientific instruments, project-based learning, tools, the results of educational research, scientific research reported both formally in journals and informally in web sites, raw data for student activities, and multimedia (image, audio, or video) banks. DLEs should provide services for authors and instructors, such as annotation, evaluation, and peer review of donated materials. For students and faculty, they will offer the capability to search for desired information by subject area, to have access to scientific data sets, to interact with peers, and to provide archiving, location-independent naming, recommender systems, selective dissemination of information, and copyright management. Faculty, students, and other clients, such as independent learners, will be able to participate in forums. Interdisciplinary activities, lifelong learning, and the process of education will all benefit. In this way, the DLE will be much more than the sum of its parts, and will promote change and innovation in scientific and technical education at all levels.

The following are guiding principles that DLEs should follow:
1. Be driven by educational and scientific needs.
2. Facilitate educational innovations.
3. Be stable, reliable, and permanent.
4. Be accessible to all (though not all materials will be free).
5. Build on, and leverage, past and current work in courseware libraries, digital library research, and successful commercial sites.
6. Be adaptable to new technologies.
7. Support the decentralized creation of services.
8. Provide tools and organizational background for the integration of resources.

DLEs are intended to encourage *the dissemination of research in educational methods.* They will also facilitate the involvement of industry and government laboratories in the educational process. Whereas some universities benefit from guest speakers from industry or government in the classroom, not all schools are able to arrange such visits. The digital library, enabled by new information technologies, would provide a forum for real time video or voice communication to a wider range of learners. These *virtual lectures and discussions* could be captured and then added to the library for later access.

DLEs will also facilitate *cross-institutional sharing* of educational resources, including all types of courseware, as well as materials for distance and self-learning. The ultimate goal is the development of a community of science and technology educators who use the library for cross-disciplinary and cross-institutional collaboration. Access and discussions with authors and prior users would be possible, along with an archive of past reviews and discussion of materials in the DLE. The collections could be annotated and linked to these discussions and reviews.

The digital library also opens the opportunity for students at different institutions to work on *joint projects or experiments*, perhaps sharing and adding to the same data set and its analysis. This would also promote *physical resource sharing*, as students and instructors may have varying access to high-end instrumentation, computational capabilities, data collections, and technology.

The following rationale for DLEs in science and math has been expressed [SLWSJ98]:
1. Student performance in math and science is poor and needs to be improved.
2. Today's Internet lacks the cataloguing, organization, archiving, collections management, etc., of a library.
3. The effort to connect every classroom to the Internet will be of limited value without high-quality content.
4. A digital library can be a resource for the entire population (marginal cost of dissemination is almost zero).

One of the methods of determining the success of digital libraries in improvement of student learning is to examine whether they are helping to achieve pedagogical objectives. Development of scientific thinking in students might be one of the criteria. Examples of skills that are to be developed in students by educators are asking questions, acquiring information, organizing information, analyzing information, and answering questions in certain scientific disciplines.

Accessibility [SLWSJ98] is an important property of DLEs that requires a two-part strategy. The first is that the library should be realistic in its technical expectations. Since a range of factors, including network bandwidth, availability of computers, and costs, can limit accessibility, the library must be designed to accommodate a wide range of users and be realistic about the technology that they use. However, not all DLE services need be limited to the lowest common denominator of the current capabilities of computers, networks, students, and faculty. Technology is improving rapidly, and the library must grow with it.

The second part of the strategy is that the DLE should work vigorously with concerned individuals and organizations, including federal and local agencies, to ensure that all students and faculty have good Internet access. Modern scientific and technical education requires that all faculty and students have computers and telecommunications, with the training to use them effectively.

Sustainability [SLRM01] is another important property of DLEs. There are strong arguments for the national DLE to be considered a "national treasure" and supported as a public good; indeed the frequent calls for open, free access to content are rooted in this view. An attractive scenario for the long-term management of the digital library is to place responsibility in the hands of a non-profit organization.

This vision of DLEs still begs the question of how creators will be compensated for their efforts. For contributions of "fine-grained" content (e.g. short Applet tutorials or simulators) the digital library can offer recognition from peers, which would be suitable and important "compensation". Digital rights management technologies also hold promise for identifying usage of, and then appropriately providing compensation for, content. This would allow the creators and purveyors of content to differentially price and/or repackage portions of "coarse-grained" material that has been disaggregated (some publishers have begun to offer custom runs of selected textbook chapters to professors). It has been observed that reconceptualizing information as a service rather than a good offers the opportunity for new revenue streams that can be directed back towards content creators. This view suggests interesting possibilities for the development of new services for users that could be available, for example, individually or through affiliation with existing organizations, such as professional societies. More generally, these considerations may lead to rethinking of reward systems, such as promotion and tenure, to reflect the importance of developing, sharing, and using educational resources in DLs.

The role of digital libraries in the learning environment will be clarified further. First, models of learning environments will be characterized.

5. Integrated learning environments

5.1. The Virtual Learning Environment and its interaction with the Managed Learning Environment

Major changes are happening across the educational community, affecting all aspects of learning, teaching, and administration activities. These present opportunities for huge benefits to colleges and universities from new learning environments and management technologies.

To discuss the involvement of digital libraries in the process of education, a model of learning environments is required. A JISC model of a Managed Learning Environment (MLE) and a Virtual Learning Environment (VLE) [MLEB02, MLEJIS, MLEVLE, VLEREQ, VLEGRE, VLEPED] will be used here. JISC has described an MLE as follows:

"MLE is the taking advantage of the potential of new technology based learning environments to integrate information systems around the learner. These learners may be working in different modes at different times, on campus, off campus, part time or full time. To support this, information systems will need to be student centred and fully accessible from multiple locations. They will need to be integrated at multiple levels, ensuring interoperability between administrative and financial systems, learning support and learning environments, and between collaborating institutions".

Using such a definition, educational institutions need to, first, rethink their educational and organizational processes as an essential part of MLE development, and second, make student-centred approaches central to that development. All MLE development should combine two processes: the design and implementation of appropriate secure and robust technical systems, and the enabling of effective educational and organizational integration towards an improved student experience. Of course, rather than being automatically or necessarily aligned, these two processes often operate with varying degrees of separation and lack of integration.

A VLE [MLEVLE, JISCMLE, VLEREQ] consists of the components through which learners and tutors participate in online interactions of various kinds, including online learning. The principal functions that the complete VLE needs to deliver are:

- Controlled access to curriculum that has been mapped to elements (or "chunks") that can be separately assessed and recorded. (Computer Curricula 2001, http://www.acm.org/sigcse/cc2001/, the new curriculum for computing developed by the Association for Computing Machinery, ACM, and IEEE-CS, is a fresh document waiting for the application of such an approach.)
- Tracking student activity and achievement against these elements using simple processes for course administration and student tracking that make it possible for tutors to define and set up a course with accompanying materials and activities to direct, guide, and monitor learner progress.
- Support of online learning, including access to learning resources, assessment, and guidance. The learning resources may vary from self-developed to professionally authored and purchased materials that can be imported and made available for use by learners.
- Communication between the learner, the tutor, and other learning support specialists to provide direct support and feedback for learners, as well as peer-group communications that build a sense of group identity and community of interest.
- Links to other administrative systems, both in-house and externally.

As shown in Figure 1, the VLE will act as a "portal" to online curriculum mapping, assessment, communication, delivery, tutor support, and tracking facilities.

The VLE makes up only one part of a college's overall systems (both computerized and noncomputerized). Interfacing between these systems is possible by "connecting up" the constituent parts through the use of interoperability standards such as the IMS (Instructional Management System). Examples of these interfaces are ones between the student record system and the VLE, and between learning resources (or content) and the VLE.

VLEs are learning management software systems that synthesize the functionality of computer-mediated communications software (email, bulletin boards, newsgroups, etc.) and online methods of delivering course materials (e.g. the Web). Several systems are emerging for the management of online learning, but none is currently able to deliver

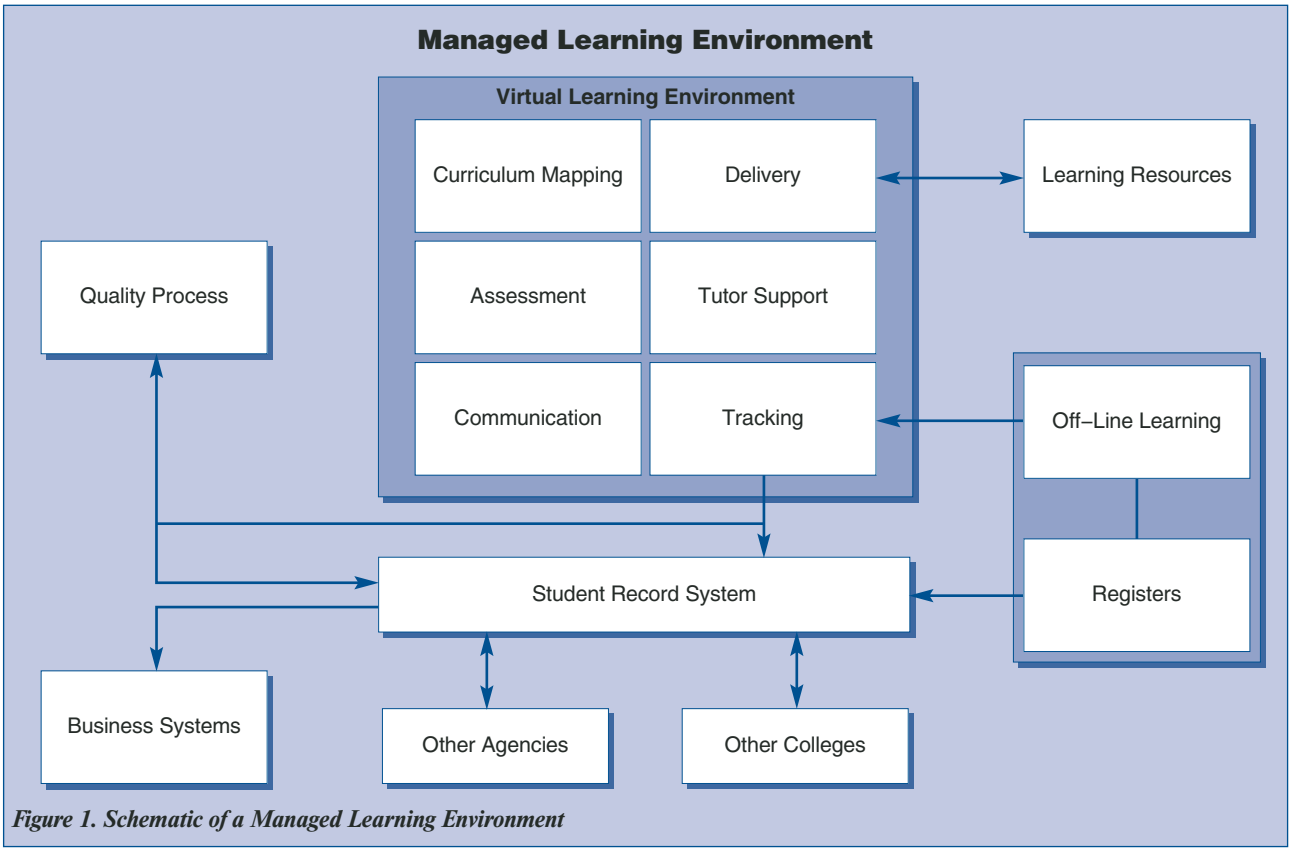

Figure 1. Schematic of a Managed Learning Environment

the full set of functions and linkages shown in Figure 1. Reports of evaluations carried out on particular VLEs, indicating the various pedagogical assumptions that developers may make in implementing VLEs, can be found at http://www.jisc.ac.uk/jciel/mlesg/.

Most VLE systems are intended not simply to reproduce the classroom environment online, but to use the technology to provide learners with new tools to facilitate their learning. They aim to accommodate a wider range of learning styles and goals, to encourage collaborative and resource-based learning, and to allow greater sharing and reuse of resources. A summary of the features of a number of current systems is provided in [VLETLS,VLEGRE]. In the MLE/VLE particular attention is given to the issues of interoperability, both between VLEs and various administrative systems and between VLEs and various providers of learning content.

5.2. Integrated learning environments in different countries

The e-learning situation worldwide is driven by similar factors, namely, technological developments increasing and changing the expectations of learners, changes in society resulting in changes in the nature of the student population (including globalization of learning), and new developments and understandings of what learning is and how it can best be accomplished.

The US higher education system is not as driven by central government as it is in other countries (e.g. in the UK). In addition to the traditional, privately run, large universities, there is growing use of commercial contracting out of instructional responsibilities rather than using tenured faculty, and increasingly private firms, such as Microsoft[2], are collaborating in higher education or running their own universities. These last developments are already becoming factors in the UK and are likely to increase in coming years, once again driven by the globalization of the educational marketplace.

[2] See: http://research.microsoft.com/programs/. The Jones Education Company aims to "get the cost of real estate out of education" and uses cable television to deliver six certificates and 11 degrees in conjunction with 14 institutions. For these and more see: Michael Thorne. *Universities of the Future.* PowerPoint Presentation at SeSDL videoconference seminar series 2000/2001; scroll down page at: http://www.sesdl.scotcit.ac.uk:8082/seminars/index.html for link.

The terms "VLE" and "MLE" are not used in the US; distance learning is frequently referred to as remote learning, and the terms "virtual classroom" and "Web-based instruction" are examples of terms used in a broader sense than VLE. There is no generic term for a networked learning environment encompassing both campus- and distance-based online instruction. This is in spite of the fact that the systems or platforms referred to as VLEs and MLEs are widely developed and used in the US, by both commercial firms and educational institutions. The difference in language probably indicates a difference in conception or priorities.

5.3. Alternative pedagogic models for evaluating Virtual Learning Environments

Two different models (one from education, the other from systems modeling) have been explored in [VLEEVA] as a basis for constructing a pedagogical evaluation methodology for VLEs. The educational model was developed and applied to the use of learning technology in higher education by Laurillard [CONV93] as a Conversational Framework. The Organizational Model is drawn from the Viable Systems Model for modeling organizational systems proposed by Beer [ORGM81]. It has previously been suggested that this organizational systems approach may be applicable in a pedagogical context [VLEEVA]. Basic ideas of these frameworks are presented in Figures 2 and 3.

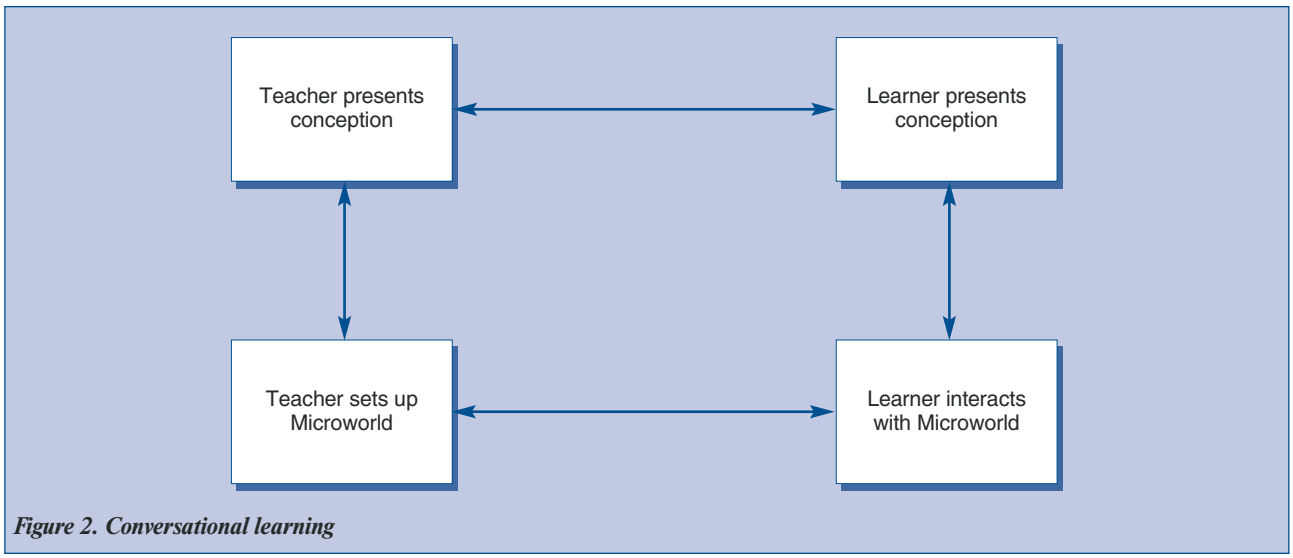

Figure 2. Conversational learning

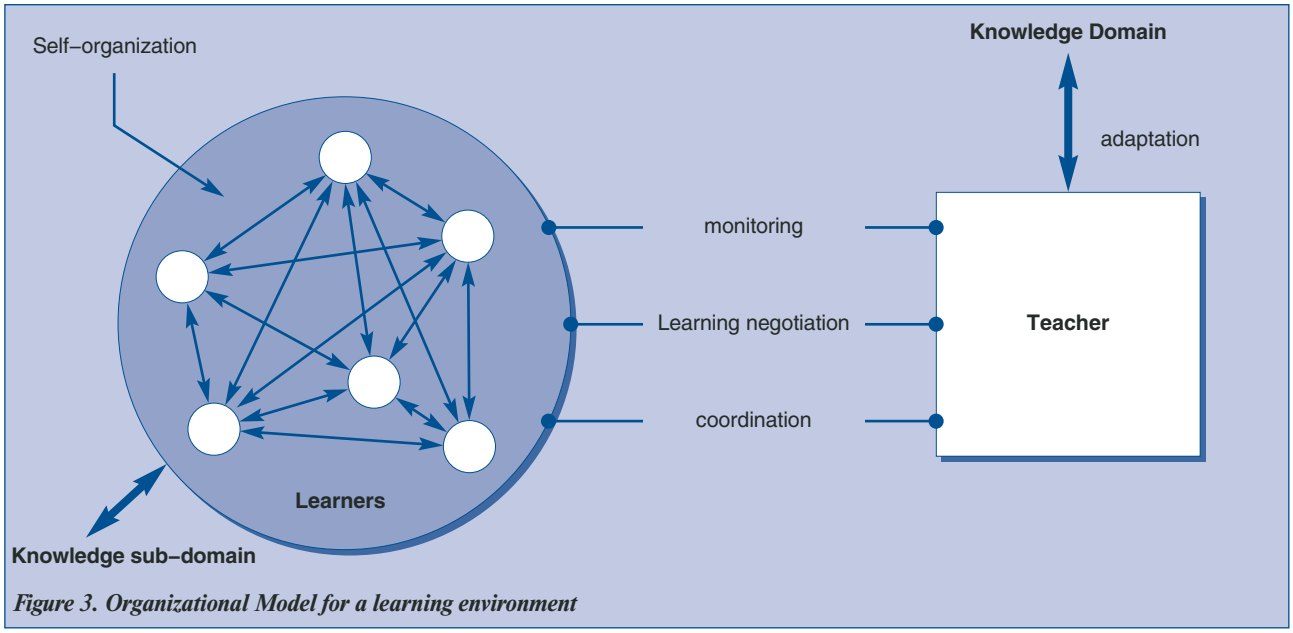

Figure 3. Organizational Model for a learning environment

The US Department of Defense's Advanced Distributed Learning Initiative (ADL) summarizes many approaches to solving the scalability or "teacher bandwidth" problem in a model close to the conversational one:

- A one-to-one instructional model in which a teacher tailors instruction to individual student needs is preferable to other instructional models.
- Human (teacher-student) interaction in large scale learning environments is not economically feasible; therefore,
- Automating feedback and other learning support via intelligent instructional systems is the only viable solution to providing scalable online learning.

But automated instructional systems completely lack human interaction and social negotiation, which learning theorists are increasingly stressing as crucial to supporting meaningful learning. Highly decontextualized learning objects are reusable in the greatest number of learning contexts, but they are also the most expensive and difficult for instructional designers to reuse, creating a "reusability paradox". Fortunately, however, when educators have been trained and motivated, they can easily identify and share very small knowledge resources that can be readily reused by others, once discovered. This process can be facilitated by tools that support the synthesis and construction of larger resources out of a number of small resources. Examples of such tools include the Walden's Path software from the digital library research group at Texas A&M University, USA, the instructional architect from Utah State University, USA, and colleagues, and the VIADUCT tool that is part of the CITIDEL (www.citidel.org) effort at the Virginia Polytechnic Institute and State University, USA.

Some ideas of learners' organization and interaction with a knowledge subdomain of the Organizational Model are reflected in the online self-organizing social systems (OSOSS) [OSOSS]. The most significant departure of the OSOSS approach from conventional learning objects approaches is that it relies on human beings to locate, assemble, and contextualize the resources. The OSOSS provides a conceptual framework for a new method of indexing, discovering, combining, using, and evaluating digital educational resources:

Indexing and discovery: Learning objects are not catalogued with metadata and submitted to a central repository. Community members know of existing resources and local resource collections. Learners gather information from a variety of sources. Individual resources are discovered through "community queries" in which community members respond with pointers to resources they know personally.

Combination: Learning objects are not automatically populated into one of many instructional templates. Without the direction of any single grand architect, peers contribute relevant resources and descriptions of how they might be employed within the context of the initiator's problem.

Use: Learners do not sit through a temporal sequencing of resources and assessments linked to decontextualized instructional objectives. They employ resources provided by peers as mediational means in the solution of a self-selected problem or accomplishment of another self-selected goal.

Evaluation: Learning objects are not critiqued out of an instructional context with a summative quality rating. Learners evaluate the relevance and suitability of resources within a specific learning context.

In an OSOSS learners are provided with meaningful learning support "anytime anywhere", yet the support is reached with human-to-human interaction. Learning objects are successfully embedded in a meaningful learning context, but the objects are discovered and contextualized by humans – again without scalability's becoming an issue.

Potential problems with the OSOSS approach:

- A standard curriculum may be difficult to impose on individuals in an OSOSS.
- Assessment of individuals may be difficult to carry out in an OSOSS.
- Required feedback may not be immediate in an OSOSS.

6. Digital libraries and Virtual Learning Environments

6.1. Tasks of digital libraries in the learning environments

Within the context of changes in society, technology, and education in recent years, there have been two key developments relating to e-learning infrastructure in UK universities and colleges:

- The adoption of virtual learning environments and managed learning environments.
- The implementation of digital and hybrid libraries.

VLEs are tools which support e-learning through the provision and integration of Web-based materials, including learning materials, links to other resources, online communication tools (such as electronic bulletin boards), and assessment tools. When such VLEs are integrated with other information systems and processes of the institution, e.g. student records, the resultant system is generally referred to as an MLE.

For VLEs, truly digital libraries are required with all resources and services available online. Some of the candidate tasks that DLEs could support include [EDLL00]:

1. Highly directed uses, such as lab exercises to reinforce a specific disciplinary concept.
2. Instructional modules that introduce concepts in an incremental manner and can be customized and extended by faculty for use in lectures.
3. Free form exploration conducted by students preparing term papers or faculty putting together a lecture that might include personal manipulation of data sets, information visualization, and the integration of new information or data sets to augment existing content.
4. Collaborative applications that might be used by students doing team projects or faculty and teaching assistants who are team teaching.
5. Discipline- or domain-specific methods of building knowledge that support specific information seeking and use processes.

The key characteristic of learners with regard to the linkage of VLEs with digital libraries is their diversity. More and more learners are learning from home, from their workplace, part-time, or from a geographical distance to their course. They are coming from all age groups, and are learning throughout their lives. They are coming to the university expecting more, based on their experiences with the Internet and other information and communication technologies. There is no longer a typical "higher education" learner. Where library and information resource support to teaching was once comfortably housed in a library building, that support must now be provided to all students regardless of the medium or location of their learning.

6.2. General criteria for digital library quality within the learning environment

Quality of the resources to be discovered in the library: There is a great deal of discussion and divergence — some libraries focus on quantity as in the public library model; some focus on quality, as in specialized collections that might be found in a public library.

Seamless access: This includes seamless integration between the learning environment and the library or information resources at any point in the VLE and within one user's portal across different courses, departments, or even institutions. The most important aspect of this was the single sign-on; one authentication procedure, regardless of whether the user is accessing the VLE from on- or off-campus.

Warning notes that were sounded included potential problems with seamless cross-searching of different databases, indexes, and other information resources.[3] Lack of interoperability of search vocabularies, and a lack of awareness of and strategies to deal with this in course design, could lead to confusing, ineffective resource discovery experiences for learners.

All library functions online: Concerns about this include the potential diminishment of two important educational functions of traditional libraries: serendipitous browsing (finding the book you need right next to the one you were actually searching for); and their social function as a place to meet fellow students and discuss sources of information, etc.

[3] See JISC/RSLP funded project HILT (High Level Thesaurus) for a full analysis of this issue, at: http://hilt.cdlr.strath.ac.uk/

Individualization for the learner: This concept includes such ideas as the student portal, which could cross institutions and be available throughout a learner's life; the Amazon.com idea of tailoring resources and notifying the user about relevant resources; the ability to save and share searches; the ability to take and embed notes with information resources, and to share resources; and settings for "level", such as undergraduate, third-year, etc., with options to adjust upwards if the user wishes.

Flexibility for the teacher: Teachers would like to be able to adapt or update courses easily, including the information resources embedded in or linked to them, from anywhere. Flexibility in terms of being able to design the course according to their own pedagogical approach, rather than having it dictated by the system, was also seen as extremely important, and vital for bringing academics on board with e-learning. Finally, the system should have the capability to feed back data to the teacher about what information resources and services are being used.

Universal accessibility: Universal accessibility includes accessibility for users with differing physical abilities, adaptability to differing learning styles; availability on- and off-campus (an issue with regard to certain subscription library materials), equitable access for distance learners abroad (usually the biggest problems are access to hard copy resources and time zone problems with communications), equitable access for the economically disadvantaged (those who have to wait in line at a computer centre versus those with a PC or laptop of their own), and usability on any platform or hardware.

6.3. Example of a project incorporating a VLE and digital libraries into the learning process

Incorporation of VLEs into the learning process and interaction of VLEs with digital libraries can be illustrated by the framework planned according to a recent JISC/NSF proposal [GEOFRA]. The framework shows how the courses, content, and delivery mechanisms at the participating institutions (University of Southampton, University of Leeds, University of California at Santa Barbara, Pennsylvania State University) are linked to a range of digital resources through the application of VLE and DL technologies. As courses and modules cover particular topics, students will use links to digital resources that include a geospatial classification, within each of the following areas: human geography, geomorphology, geographical information science, and Earth observation.

Layer 1 in Figure 4 outlines the existing courses offered at partner institutions. Within each course, smaller components of learning can be identified that might involve one lecture and seminar in a week, or one practical class or one field visit, which are denoted as student learning nuggets, and which form *Layer 2* in the framework. *Layers 3* and 4 represent the integration of VLE and DL technology, where *Layer 4* shows a DL middleware technology for managing collections of resources, such as those shown in *Layer 5*. The instances of the middleware can interoperate, so that all the resources in the distributed collection are available to the VLE users.

The Alexandria Digital Earth Prototype (ADEPT) [ALEXBI, ADEPRO, ADEDLE, ADEPTS, ADEPTA, EGEODL] project has developed distributed digital libraries for heterogeneous geo-referenced information. In ADEPT, *libraries* are sets of collections. Libraries expose a single standard set of interfaces to all their collections, making it possible to issue a single query against multiple collections. (By contrast, the interfaces to collections are not standardized; instead, a library has standard mechanisms for adapting itself to whatever interfaces the collection exposes.). A library is, in effect, a "collection broker", mediating standardized access to its.

ADEPT incorporates a *bucket framework*: a canonical, simplified representation of the source metadata of heterogeneous collections, allowing uniform querying across all the collections.

The ADEPT architecture has recently been extended [ADECOA] to support an innovative form of VLE, based on the hypothesis that learning should proceed from a formal presentation of concepts and their relationships within a domain of knowledge. The architecture supports a representation of a domain ontology, or concept space, linked to a collection of learning objects. The granularity of these objects is more appropriate to their use in VLEs than traditional "information containers", such as books.

The proposal [GEOFRA] that has been accepted aims at the resource-based learning that involves active participation with multiple resources. Students are motivated to learn about a topic by trying to search for and evaluate authentic information. This learning experience mimics real life in targeting the learner as the routine information hunter and interpreter who constructs knowledge by problem solving with information tools. The advantages to this approach include:

- a student-centred approach to learning;
- promotion of the development of thinking skills (such as problem solving, reasoning, and critical evaluation);
- improving student research skills, which supports the research-led mission of all four partners;
- flexible and adaptable resources and materials for different learning styles and strategies;
- integration of key skills and competences within the academic framework.

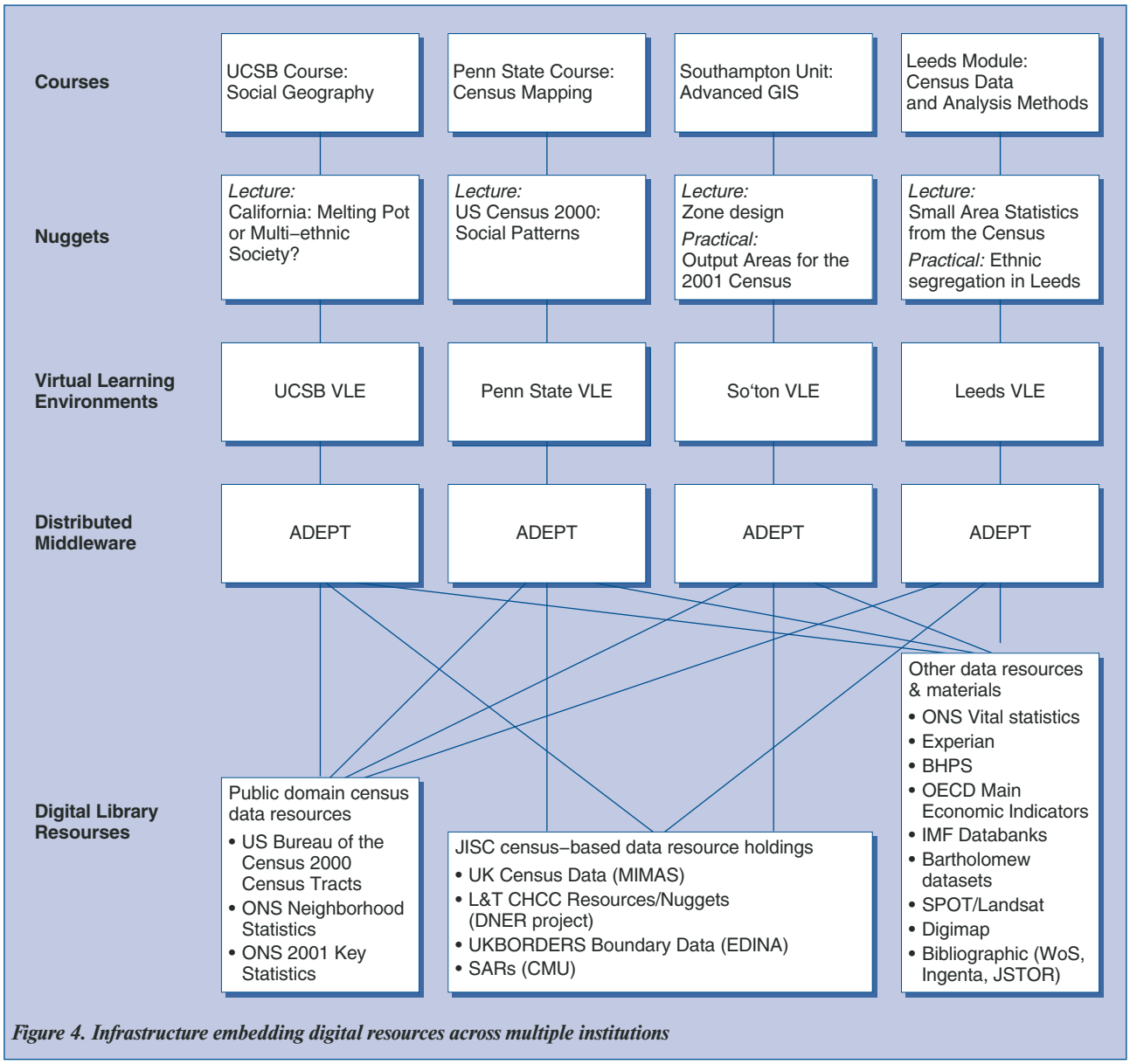

Figure 4. Infrastructure embedding digital resources across multiple institutions

The rest of this survey is devoted to the current state and anticipated evolution of DLEs that eventually should meet the requirements discussed above. The survey shows that the existing technology and services constitute a step in the evolution to true DLEs.

7. Taking a common view of educational metadata

Instructional Management Systems (IMS) Project: Designers and developers of online learning materials have an enormous variety of software tools at their disposal for creating learning resources. These tools range from simple presentation software packages to more complex authoring environments. They can be very useful in allowing developers the opportunity to create learning resources that might otherwise require extensive programming skills. Unfortunately, the wide variety of software tools available from a wide variety of vendors produce instructional materials that do not share a common mechanism for finding and using these resources. Descriptive labels can be used to index learning resources to make them easier to find and use. Such metadata specification makes the process of finding and using a resource more efficient by providing a structure of defined elements that describe, or catalogue, the learning resource, along with requirements about how the elements are to be used and represented.

In 1997, the IMS Project, part of the nonprofit EDUCOM consortium (now EDUCAUSE) of US institutions of higher education and their vendor partners, established an effort to develop open, market-based standards for online learning, including specifications for learning content metadata [LRBP11]. Also in 1997, groups within the US National Institute for Standards and Technology (NIST) and the IEEE P.1484 study group (now the IEEE Learning Technology Standards Committee, LTSC [LTSA01]) began similar efforts. The NIST effort merged with the IMS effort, and the IMS began collaborating with the ARIADNE Project (www.ukoln.ac.uk/services/elib/projects/ariadne), a European project with an active metadata definition effort. In 1998, IMS and ARIADNE submitted a joint proposal specification for standardization to IEEE, which formed the basis for the current IEEE learning object metadata (LOM) standard. IMS publicized the IEEE work through the IMS community in the US, UK, Europe, Australia, and Singapore during 1999 and brought the resulting feedback into the ongoing specification development process.

The IEEE LOM base document [LOMN00] defines a set of metadata elements that can be used to describe learning resources. This includes the element names, definitions, data types, and field lengths. The specification also defines a conceptual structure for the metadata. The specification includes conformance statements for how metadata documents must be organized and how applications must behave in order to be considered IEEE-conforming. The IEEE base document is intended to support consistent definition of metadata elements across multiple implementations. The number of items defined within the IEEE base document was large and many participating organizations within the IMS community recommended that a select core of elements be identified to simplify initial implementation efforts. The IMS developed a representation of the metadata in Extensible Markup Language (XML) [LRXB11] and surveyed its member institutions around the world to identify the core elements. The IMS *Metadata Best Practice and Implementation Guide* [LRBP11] provides general guidance about how an application may use the core and extended metadata elements.

Many metadata implementers were initially optimistic that their participation in the IMS consortium would help produce a relatively small but well defined and agreed upon set of metadata elements. This optimism soured as the set of proposed metadata elements grew increasingly larger.

Learning object metadata [LOMN00]: This standard specifies a conceptual data schema that defines the structure of a metadata instance for a learning object. For this standard, a learning object is defined as any entity, digital or nondigital, that may be used for learning, education, or training. For this standard, a metadata instance for a learning object describes relevant characteristics of the learning object to which it applies. Such characteristics can be regrouped in general, educational, technical, and classification categories. The conceptual data schema specified in this standard will allow for linguistic diversity of both learning objects and the metadata instances that describe them. This standard will be referenced by other standards that will define the implementation descriptions of the data schema so that a metadata instance for a learning object can be used by a learning technology system to manage, locate, evaluate, or exchange learning objects.

The purpose of this standard is to facilitate search, evaluation, acquisition, use, sharing, and exchange of learning objects, for instance by learners or instructors. Catalogues and inventories may take into account the diversity of cultural and linguistic contexts in which the learning objects and their metadata will be exploited. By specifying a common conceptual data schema, bindings of learning object metadata will have a high degree of semantic interoperability.

Basic metadata structure: A description of a learning object consists of data elements. The latter are grouped into *categories*. The base scheme consists of nine such categories:

a) The *General* category groups the general information that describes the resource as a whole.
b) The *Lifecycle* category groups the features related to the history and current state of this resource and those who have affected this resource during its evolution.
c) The *Meta-metadata* category groups information about this metadata record itself (rather than the resource that this record describes).
d) The *Technical* category groups the technical requirements and characteristics of the resource.
e) The *Educational* category groups the educational and pedagogic characteristics of the resource.
f) The *Rights* category groups the intellectual property rights and conditions of use for the resource.
g) The *Relation* category groups features that define the relationship between this resource and other targeted resources.
h) The *Annotation* category provides comments on the educational use of the resource and information on when and by whom the comments were created.
i) The *Classification* category describes where this resource falls within a particular subject classification system.

Data elements: Categories contain data elements. For each element, the base scheme defines:

- *Name:* the name by which the data element shall be referenced.
- *Explanation:* the definition of the element.
- *Size:* the number of values allowed.
- *Order:* whether the order of the values is significant (only applicable for elements with multiple values).
- *Value space:* the set of allowed values for the data element — typically in the form of a vocabulary or a reference to another standard.
- *Data type:* a set of distinct values.

The following data types are included: list values, vocabularies, minimum-maximum values, and character sets. For each of the data elements, the specification includes the data type from which it derives its values, such as LongString or Date, etc. These will be defined separately, and will be implemented in a particular way in a particular system. In order to maximize interoperability, future work may define a common representation for these data types. In the absence of such a common representation, an exchange format, such as XML, would allow systems with different representations to achieve interoperability through a conversion process.

DELOS Working Group on Metadata Registries: Thomas Baker (Fraunhofer Institute of the German National Research Center for Information Technology - GMD), who formed the DELOS Working Group on Metadata Registries, identified problems of harmonization of various metadata standards in the following manner.

Standards organizations and metadata implementers could in principle link their standards (as well as any application-specific extensions based on them) by cross-referencing their schemas over the Internet. However, there are currently a variety of models for defining the nature and attributes of metadata entities ("elements"). One important model is the ISO 11179 set of standards for defining the attributes of data elements and the architecture of registries. Another is the somewhat simpler model being used by the Dublin Core Metadata Initiative in its prototype schema registry, which is based on RDF schemas. Other terms are used for the IEEE learning objects metadata model and in XML Schemas. Yet others are used for web services. An appropriate level of harmonization between these various approaches would help ensure a degree of interoperability as such initiatives deploy networked registries.

8. Current Digital Libraries for Education: American projects

8.1. National Science Digital Library (USA)

NSDL is a large and comprehensive DLE initiative. It now involves 118 projects, funded by the US National Science Foundation, but additional organizations are engaging as partners as well — as the organization takes shape, it may evolve into a nonprofit institution, NSDL, Inc., or become an official sub-branch of the federal government. Current operations relate to four funding tracks, connecting with a parallel volunteer structure for governance.

The core integration track aims to provide coherence and production-level operational support, yielding a central portal to services and collections. The collections track focuses on content, with different groups working on varied horizontal or vertical slices of the STEM field. The services track emphasizes both general and specialized services, including those needed in any digital library and those appropriate to support teaching and learning. Examples discussed below include the strand map service built on the strand maps developed by the American Association for the Advancement of Science's (AAAS) Project 2061 and the concept map tool in the GetSmart project. The research track encourages innovation, new approaches, and evaluation/assessment.

The following subsections explore various aspects of NSDL, which had its first glimmerings over a decade ago, and which now is in a crucial phase of initial testing and launch. Later in this report are discussions of DLESE (Section 8.2), which is connected with NSDL, and CITIDEL (Section 8.3), one of the NSDL collection projects, to provide more details regarding how the NSDL deals with particular areas in science, technology, engineering, and mathematics, in particular Earth systems (DLESE), and computing and information technology (CITIDEL).

8.1.1. NSDL features

We concentrate first on a macro way of looking at digital library support for learning environments and on learning resources in general. The most visible representative of this approach is the US NSF National Science Digital Library (NSDL) programme [SLRM01, NSDLAR, ICADL02]. This programme seeks to bring together a vast collection of learning resources supporting all possible kinds of education, ranging from primary to graduate and lifelong learning, into one big library for the USA — and even beyond. As such, the NSDL approach is consistent with the large scale of many of the other NSF digital libraries projects.

The goal of the NSDL programme is to enhance all aspects of scientific and technical education in America. The NSDL library will be a distributed information environment for accessing quality-assured digital resources from many sources. NSDL resources span a nearly unlimited range of materials with educational value, including web pages of all sorts, digital objects such as geospatial images, proxies for physical objects, such as specimens, and threaded discussions.

Beyond its digital resources, the NSDL environment will include an extensible set of services to enhance the experience of library use. These will offer, for example: interfaces for browsing and discovering NSDL resources; tailored views of NSDL; means to annotate resources (augmenting owner-supplied metadata); support for social interactions among NSDL users; and managed access to resources by various groupings of end-users (i.e. by enforcement of usage policies). NSDL services and content eventually may alter basic pedagogic and academic practices in science, technology, engineering, and mathematics education.

NSDL is envisioned as an integrated information environment, constructed in a highly distributed effort. The goal is for end-users to interact with NSDL mainly as a coherent whole rather than as a set of individual collections and services. These ideas for coherence parallel, in several ways, the technical architecture for the Distributed National Electronic Resource (DNER) under development in the United Kingdom [DNERLP, DNERAV]. Planned NSDL features are characterized below.

End-Users: NSDL end-users are viewed primarily in their roles as educators and learners, though many simultaneously are library builders or content providers. The NSDL information environment can be characterized as supporting four high-level activities by such end-users:

- *Discovery:* NSDL facilitates discovery of content and services corresponding to end-user needs and interests.
- *Access:* NSDL enables and manages access to (discovered) content and services, potentially resulting in use for educational purposes. Use often results in further cycles of discovery, access, and use.
- *Tailoring:* Educators and learners tailor NSDL for personal purposes and for use by specified groups, such as classes of students.
- *Social interaction:* End-users may enrich their NSDL information environment through social interaction and community discourse.

Content: NSDL content is typically made available in the form of collections, defined to be any aggregation of one or more items. There will be collections of metadata about other collections. NSDL content will be characterized (and discovered) mainly via "metadata records" that describe content at the collection or item level. The accessible resources include:

- digital collections, managed by dispersed institutions;
- primary and derived information;
- data and metadata;
- diverse services, some embedded with human roles;
- policy-controlled and non-policy-controlled items.

The content available through NSDL includes (though not exclusively) the following types, with or without policy-controlled access:

- Web pages — such as lesson plans, teacher guides, monographs, abstracts, manuscripts, scholarly journals, and still images — that are accessible and usable via conventional browsers.
- Digital items used outside the browser environment or with special plug-ins (usually after downloading), or requiring specialized access protocols. Examples include numeric data, geospatial images, moving pictures, sound collections, music scores, learning objects, and computer simulations (e.g. simulations of real-world objects and processes).
- Discussions on special topics archived from community discourse (e.g. email threads).
- Digital proxies for physical items, such as textbooks, lab supplies, and specimens.
- Thematically organized collections of the above (potentially nested).

Most content-specific aspects of the library are addressed via library services. For example, the initial release of NSDL will include a content-based search service restricted to textual documents in common formats.[4]

Services: Services for displaying, processing, and analyzing images, maps, and other scientific data, as well as specialized portals, may be characterized and "discovered" somewhat like other library content. Non content-based services (such as help desks or community forums) will focus on some form of social interaction, though recorded discourse from such interaction may well become NSDL content. Some services are compound, i.e. they depend on other services and therefore must interoperate via matching or brokered protocols. Some services may provide real physical objects (e.g. books or specimens) that correspond to digital proxies. As with content, some NSDL services will have policies that constrain usage; typically, such policies will be part of a metadata record describing the service.

Discovery: The act of discovering NSDL content (or services) generally entails using a general-purpose Web browser to call upon one or more services designed for searching, browsing, or querying NSDL collections. Nonbrowser contexts for using NSDL discovery capabilities also are envisioned. For example, the user of a geographic information system (GIS) may wish to search NSDL for maps, images, or gazetteer services without ever leaving the GIS environment.

Access: Access to (discovered) content within NSDL is often very simple, such as when the desired resource is openly available on the Web. Access becomes more complex when use of the resource is controlled by policy or when the resource requires more than a browser for effective use.

Whenever access entails the use of services, or when the desired resource is itself a service, then it becomes necessary to interface such services (applying protocol matching or brokering) to the end-user's information environment. A minimal analysis shows use of the resource to include the following:

[4] NSDL is being developed by a large federation of groups working on collections and services well integrated. NSDL now is up to 118 projects. One key track is for collections and one track is for services. Examples of collections and services will be provided later.

- unpacking the resource (as may be typical with learning objects and software packages);
- viewing it (e.g. visualizing a large data set) or listening to it;
- processing it (e.g. loading the resource into a spreadsheet or computer model);
- incorporating or assembling it into other (new) resources;
- storing, sharing, or publishing it for use by others.

Tailoring: User adaptations of NSDL (e.g. to match personal needs or to support a specified group of students) generally take one of two forms. The first is the creation of constrained views of NSDL, such as portals that are designed for specific education levels. A second form of tailoring is the creation of individual or group "profiles" that trigger automatic setting of user-interface parameters.

Social interaction: NSDL will continue the long historical tradition of libraries as centres of scientific discourse. NSDL supports the following activities by users:
- launching and joining electronic discussion groups;
- reviewing, editing, or annotating content developed by others;
- posing or answering questions at a human-mediated "help desk";
- posting messages and announcements that reach a target audience;
- participating in collaborative educational or library development.

NSDL cannot be characterized as a one-way flow of information from providers to users. Users are both recipients and creators of primary content, secondary content, and metadata, especially as they give shape to the NSDL social context.

NSDL will create and manage a registry of services. Service descriptions are expected to include the nature of the service (human, programmatic, or both); a broad categorization (e.g. marking, discovery, content manipulation, as above); interfacing technology (e.g. SOAP, Z39.50); and perhaps semantic metadata, such as subject area and educational level.

8.1.2. NSDL core integration

The core integration effort has concentrated on building the central framework for the NSDL, with key services. Work to integrate the first NSDL collections into this framework has been done for the initial release in December 2002. In the next phase, the focus will be on maximizing the impact of the library on education. This will require enormous growth in the scale of the library, and partnerships with all facets of the science education community.

The core integration team provides the organizational and technical glue that binds the NSDL projects into a single coherent whole. Computers and networking can significantly enhance learning, when combined with pedagogy that is informed by current understandings about cognition, knowledge creation (i.e. constructivism), and the dynamics of collaborative learning [LTCOGN]. Thus a central NSDL challenge is to create not just a rich repository but also an *intellectual commons*, where students and educators interact and are stimulated to change the way they teach and learn.

Networks of institutions and individuals (educators, librarians, learners, publishers, parents, etc.) are to be stimulated to become stakeholders in NSDL, utilizing its resources, enhancing its quality, and finding it to be a place for fruitful discourse on education and learning.

The library is considered to be a cognitive tool, fostering the active creation of knowledge by both teachers and students. In the case of teachers, the knowledge gained will pertain to effective modes of pedagogy, in addition to subject knowledge about the worlds of science, technology, engineering, and mathematics.

Fitting NSDL to the needs of different user groups is made possible by the technical strategy of *one library, many portals.*

Over the next five years NSDL is expected to serve millions of users and provide access to tens of millions of digital resources [NSDLCO].

The architectural design (see Figure 5) for the initial NSDL phase is based on sharing of human and machine-generated metadata and exploitation of that metadata for the deployment of core services (e.g. search and discovery).

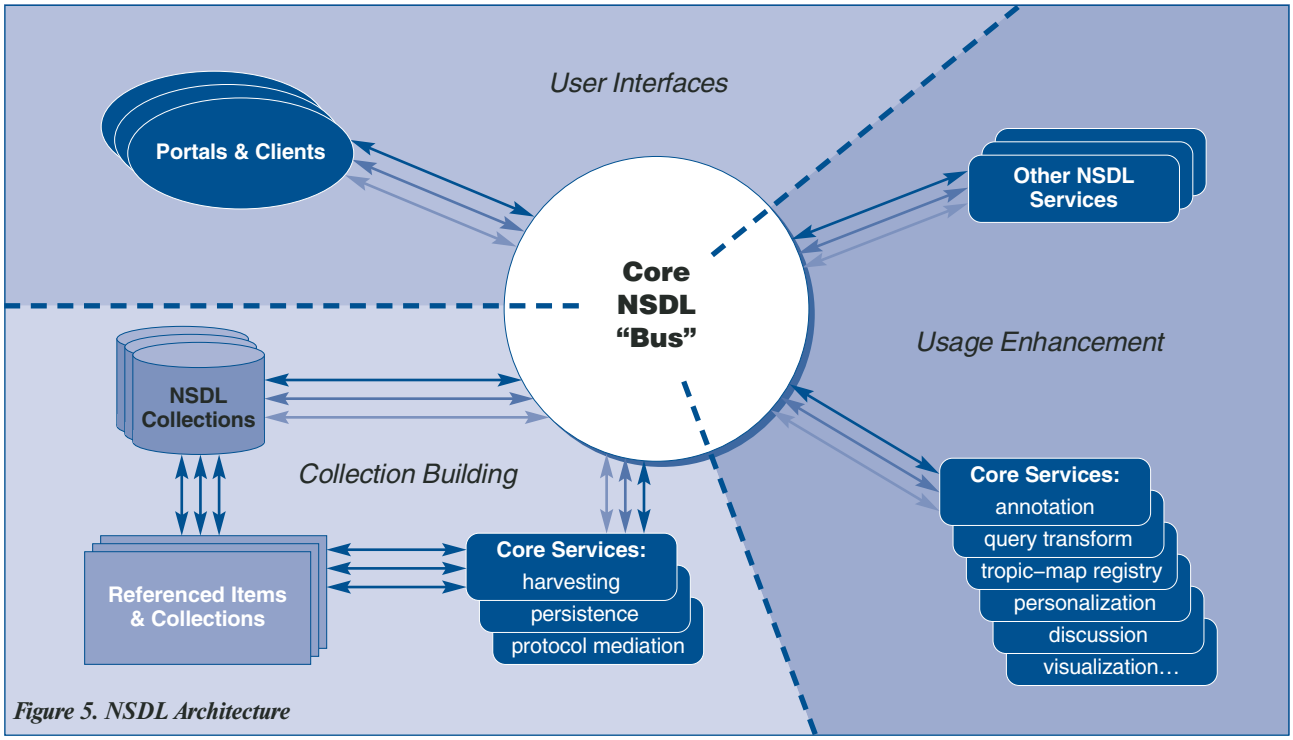

Figure 5. NSDL Architecture

8.1.3. Components of NSDL core architecture

As is illustrated in Figure 5, the core architecture of NSDL involves a number of integrated components.

Metadata repository

By combining metadata from many collections, the metadata repository can be considered to be a generalization of the concept of a union catalog, i.e. a catalog that combines records from many libraries. For metadata NSDL is adopting the following strategy:

- Item metadata from cooperating collections in any of eight supported "native" formats [NSDLME].
- When appropriate, automatically crosswalk native metadata to qualified Dublin Core [DCQUAL], which will provide a lingua franca for interoperability.
- When item-level metadata does not exist and where possible, process content and generate metadata automatically [METAGE].
- Accept that item-level metadata will not always exist but mandate that collection-level metadata always exists. Concentrate limited human effort on the creation of this collection-level metadata.
- Assemble all metadata, core and native, item and collection, in the central metadata repository.
- Expose metadata records in the repository for service providers to harvest.

The NSDL Standards and Metadata Workgroup, whose members represent all NSDL projects, identified the following list of preferred metadata element sets:

- Dublin Core,
- Dublin Core with DC-Ed extensions,
- LTSC (IMS),
- ADL (SCORM),
- MARC 21,
- Content Standard for Digital Geospatial Metadata (FGDC),
- Global Information Locator Service (GILS),
- Encoded Archival Description (EAD).

Three categories of metadata are available and can be stored in a metadata repository:

1. *Collection-level metadata:* The repository contains a registry of the collections that are known to the NSDL. There is a metadata record for each collection. It uses qualified Dublin Core, including descriptive information about the

material in the collection (e.g. courseware for high school biology) and, where available, technical information about protocols, formats, etc., and access restrictions and authentication requirements.
2. *Original item-level records:* Whenever a collection is able to supply item-level metadata that follows one of the preferred standards, the original records as supplied by the collection are stored in the metadata repository.
3. *Normalized item-level records:* The preferred metadata standards differ considerably in their syntax and semantics. Many collections implement only part of a standard and others have no standardized metadata. To enable the delivery of services that require consistent information about items from such diverse collections, all item-level metadata records are normalized by metadata crosswalks to Dublin Core.

By providing item-level metadata both in its original form and also as normalized metadata, the repository offers service providers a choice. Services that are able to make use of the original metadata can use it. Others can use the simpler, normalized records.

Qualified Dublin Core (with DC-ED extensions [DCEDEX]) is the normalized format for both item and collection metadata. Eight native metadata formats are supported. The metadata repository stores both the native metadata and the DC metadata.

The following approaches for adding metadata to the metadata repository are supported: metadata ingest via OAI (Open Archive Initiative), metadata ingest via FTP, email, or Web-upload (batch); metadata ingest by direct entry or metadata ingest by gathering.

Both the Dublin Core and native metadata records in the repository are made available to services through the OAI protocol. No restriction is placed on access to the Dublin Core records, making it possible for any party, NSDL-funded or external, to create a service building on the data in those records. In most cases, the contents of the native metadata records are also open-access.

Search services

Generally speaking, any item that has metadata in the repository is accessible via the search and discovery component. Where possible, the search and discovery component also allows search by the actual content of the resources corresponding to a record in the repository. The content is accessed using open network protocols (e.g. HTTP or FTP) linked via the identifier in the metadata record. In the first phase of NSDL, for content to be available via search, the content must be freely accessible over the Internet and it must be stored in one of a small set of textual formats: open formats such as ASCII text and HTML and a handful of proprietary formats, such as PostScript, PDF, and Microsoft Word.

Content-based search initially supports only textual queries, meaning that nontext items (e.g. images, sound recordings) are accessible via metadata only. Thus, the discoverability of nontextual items depends on adequate descriptive metadata: e.g. in the DC title or description elements. Search and discovery services are available using metadata, content, and any combination of the both. The query language that supports these services is independent of the architecture. However, the service provides a language modeled after Z39.50 type 102 ranked list queries [Z203RLQ]. The search engine interacts with its clients — portals, not people — using the SDLIP protocol [SDLIP].

Authentication and authorization

The core access management system relies on standard (e.g. Kerberos [KERBER], LDAP [LDAP05]) or emerging protocols (e.g. Shibboleth [SHIB01]) to distribute identity verification (authentication) and cohort membership (authentication) to the administrators of distinct communities of users.

User profile server: The NSDL core architecture includes a profile server, which holds attributes associated with a user. Portal interfaces, or other services, may use the Profile Server to store and retrieve information to customize a user's experience. For example, a portal may store search preferences and histories, disability information, grade level, etc., in the profile Server and adjust its interface behavior when the user next visits the portal.

Rights management broker: The core architecture also includes a rights management broker service. The rights broker enforces access decisions for items in the library based on the characteristics of the user and of the item.

8.1.4. Levels of interoperability in NSDL

Three levels of digital library interoperability are identified [NSDINT]: federation, harvesting, and gathering.

Federation: Federation can be considered the conventional approach to interoperability. In a federation, a group of organizations agree that their services will conform to certain specifications (which are often selected from formal standards). The libraries that share online catalog records using Z39.50 are an example of a federation. Another federation is the ADEPT project for geospatial materials, led by the University of California at Santa Barbara, USA; one of the NSDL partners in the core integration production team [ALEXBI].

Harvesting: The Open Archives Initiative is based on the concept of metadata harvesting [OAI001]. The metadata about collection items can be harvested by service providers and built into services such as information discovery or reference linking. While services built by metadata harvesting may be less powerful than those provided by federations, the burden of participating is much less. As a result, many more organizations are likely to join and keep their systems current.

Gathering: Even if the various organizations do not cooperate in any formal manner, a base level of interoperability is still possible by gathering openly accessible information using a Web crawler. The premier examples of this approach are the Web search engines. Because there is no cost to the collections, gathering can provide services that embrace large numbers of digital libraries, but the services are of poorer quality than can be achieved by partners who cooperate directly.

Some of the most interesting Web research at present can be thought of as adding extra function to the base level, which will lead to better interoperability, even among totally noncooperating organizations. Even though the concept of a fully semantic Web is still only a vision, it is reasonable to expect that the level of services that can be provided by gathering will improve steadily. ResearchIndex (also known as CiteSeer) is an example of a digital library built automatically by gathering publicly accessible information. Thanks to the assistance of NEC, both the data and software used in ResearchIndex are being integrated into CITIDEL (www.citidel.org), one of the NSDL collection projects.

8.1.5. Initial release of NSDL

The initial release of NSDL will be in December 2002, involving mostly the results of funding provided to over 60 projects. The discussion below sketches key aspects of that release, and then proceeds to consider future growth plans, which will involve over 30 more additional projects funded in the fall of 2002. Section 8.1.6 outlines the plan for continued evolution. Section 8.1.7 contains a discussion of the NSDL community, which will expand both in terms of size and engagement as NSDL content and services expand. Section 8.1.8 sketches the organization of NSDL, with support by NSF, guidance by the NSF-selected National Visiting Committee, steering efforts of the Policy Committee, and focused efforts by Standing Committees.

Providing technology: An architectural framework has been established [NSDLNSF] to support the core technology and to integrate resources created by others, including:

- Building a metadata repository, which holds native and standardized metadata records for each collection and item known to the NSDL;
- Implementing interfaces (primarily OAI protocols) by which the metadata repository is populated and by which its contents may be accessed to construct various library services;
- Interfacing and testing fundamental library services;
- Building the primary portal from user-interface components that can be reused in a broad array of additional portals.

Operating core services: The first library-wide services include the main NSDL portal, a comprehensive search and discovery service, and an initial authentication and authorization service.

NSDL emphasis in the next phase will be on growth: expanding the collections, adding new services and partners, and above all encouraging use for education.

Growing the collections: A major goal is to make the NSDL a comprehensive library, covering all areas of scientific, technical, engineering, and mathematics education. This requires a vast expansion in the number of collections, far beyond those funded by the NSF. NSDL strategy has three collection-building strategies: working with other digital libraries, partnering with publishers, and building collections automatically from the Web.

Ultimately, the success of the NSDL will be judged by its impact on STEM education. A variety of reforms is required, including the adoption of inquiry-based science learning, and emphasis on the process of doing science and the integration of scientific research into education. Digital libraries promise to be a powerful tool in realizing the goals of these reform initiatives.

The NSDL goal is to be the resource of choice for disciplinary, curricular, and pedagogical issues. Hence, NSDL should be a partner and major asset for every significant initiative in education for science, technology, engineering, and mathematics.

NSDL focuses its efforts on developing strategic partnerships with a key set of professional organizations, including the National Science Teachers Association and the American Library Association (ALA). Within the ALA, NSDL aims to develop partnerships with the American Association of School Librarians, the Association for Library Services to Children, and the Public Library Association, as well as the Association for Library Collections and Technical Services. The outcome of these partnerships for NSDL will be an intimate knowledge, through consultation and feedback, of the pedagogical and resource needs of teachers and librarians as they face the daily realities of American classrooms and schools.[5]

Many of the collection track projects of NSDL are actively engaged in collaborations with professional societies related to their specific subject areas. These collaborations lead to increased access to extensive collections of materials. An example is the CITIDEL (Computer and Information Technology Interactive Digital Education Library) project, a collections project that covers the entire realm of computing materials. CITIDEL has entered into an accord with the Association for Computing Machinery (ACM) by which the metadata of the entire ACM digital library is accessible for search through CITIDEL. In addition, ACM has agreed to provide free access to specific entries in the ACM DL that are particularly relevant for educational purposes. NSDL's aim is to engage a group of such partners, who cover all disciplines and all levels of science education. The societies can help in understanding how adequately NSDL resources represent the most current resources and research in their fields.[6]

Project 2061 [ATLAS] (see discussion later in this survey) supports "the development of new tools for teachers, curriculum developers, and textbook authors and publishers". As the national library for scientific education, NSDL intends to become the distributor of these educational resources.

8.1.6. Planned NSDL evolution

Extending the architecture

Describing and integrating services: A multi-service framework requires standards for describing services, to permit interactions among them and access to them by users and agents [NSDLNSF]. It is planned to leverage recent developments in Web services, including the Web Services Description Language (WSDL) work within the W3C and perhaps the .net work from Microsoft.

Creation and management of federated ontologies: The NSDL will encompass resources from many disciplinary domains. Providing a coherent end-user view will require developing and establishing conceptual relationships among distributed resources.

Creating, packaging, and accessing complex content: The resources available in the NSDL will be rich and complex. Dynamic and multimedia content requires access methods and digital object structuring standards that extend far beyond those now available via Web standards.

Scientific data: Of particular value in science education are numerical data sets and other digital objects that are of little use except in conjunction with appropriate tools.

Preservation of digital resources: The utility of the NSDL depends on the integrity of the resources, notably their longevity and stability. There is no one solution to the problems of digital preservation and longevity.

[5] This orientation to get a bottom-up feedback may provide an average pattern of requirements. How to lead to excellence based on the average is an open issue.

[6] To make such partnership efficient, the resource representation in DLE should be more focused, more structured, more curriculum or subject oriented.

Data provenance: Another factor related to the integrity of digital resources is their record of origin. This is especially true for scientific data, where origin and derivation are critical to determining the veracity and utility of the data.

Annotation and review frameworks: The metadata repository has flexible means for annotating resources. This is intended to support educational services by linking reviews of resource quality and appropriateness, and encouraging collaborative reviews.

Authentication of digital objects: Communities of authors, publishers, librarians, and consumers have consistently emphasized the importance of document authenticity. In the digital realm, comprehensive document authentication is in only the early stages of development, but considerable progress can be made with simple tools.

Accounting: For authors, an important value in NSDL will be to understand how their work is being used (e.g. by whom and how often), so that they receive the same level of academic credit for digital citations as for print citations. The planned NSDL access management infrastructure is uniquely positioned to supply this information to an author without compromising the privacy of the user. Work on logging standards [GONCAL] across NSDL may also have broad impact.

Collaborative learning: Supporting collaborative learning is of central importance to NSDL. NSDL plans to work with the collaborative learning community to begin the process of identifying interoperability standards for integrating collaborative learning systems with digital library substrates.

Large-scale growth

Version 2 of NSDL is scheduled for summer 2004 [NSDLNSF]. By the end of this phase, NSDL should be a very large, production-quality library, with substantial richness of educational resources, and a rapidly growing community of users and contributors. The community support and evaluation activities will be continued.

In parallel with this growth in resources, a major emphasis on educational outreach is planned, building partnerships with groups of educational users, including discipline-specific groups, sharing resources with them, providing services via NSDL, understanding their needs, and seeking resources that meet them.

Fundamental to this growth, in Version 3 and beyond, is an increased emphasis on creating communities of educators and learners, and feedback mechanisms to facilitate parent-teacher-student interaction. One of the important directions is a technical framework, for sharing and managing semantic information, to support the sharing of learning concepts across diverse communities and the creation of new learning communities within the NSDL rubric.

8.1.7. NSDL education community

NSDL provides opportunities to develop a new science, technology, engineering, and mathematics education community [PATHWAY] that is interdisciplinary in nature. There is intrinsic value in recognizing and accentuating the connections among the knowledge bases, skills, and methodologies employed by the contributing disciplinary communities. NSDL can make a substantive contribution towards bridging current disciplinary boundaries. An NSDL that effectively integrates concepts, knowledge, and methods across the STEM disciplines will be a much greater resource than a collection of discipline-specific libraries.

The NSDL community is an aggregate of many existing and intersecting communities, including:
- disciplinary groups,
- educational group,
- technology and information science group,
- special interest groups (e.g. policy makers, journalists, commercial sector),
- learners of all kinds — students and citizens-at-large.

8.1.8. NSDL governance

The NSDL governance model [NSDGOS, NSDLGOI] must:
- create a public sphere of influence that promotes partnering and shared vision and balances interests which sometimes differ,

- establish a framework for accountability,
- reflect the special nature of key players.

To achieve its purpose, the NSDL governance model (Figure 6) is built around committees, subcommittees, task forces, and interest groups that provide several domains of guidance. In decreasing order of authority (and formality), the advisory entities in this model are as shown in Table 1.

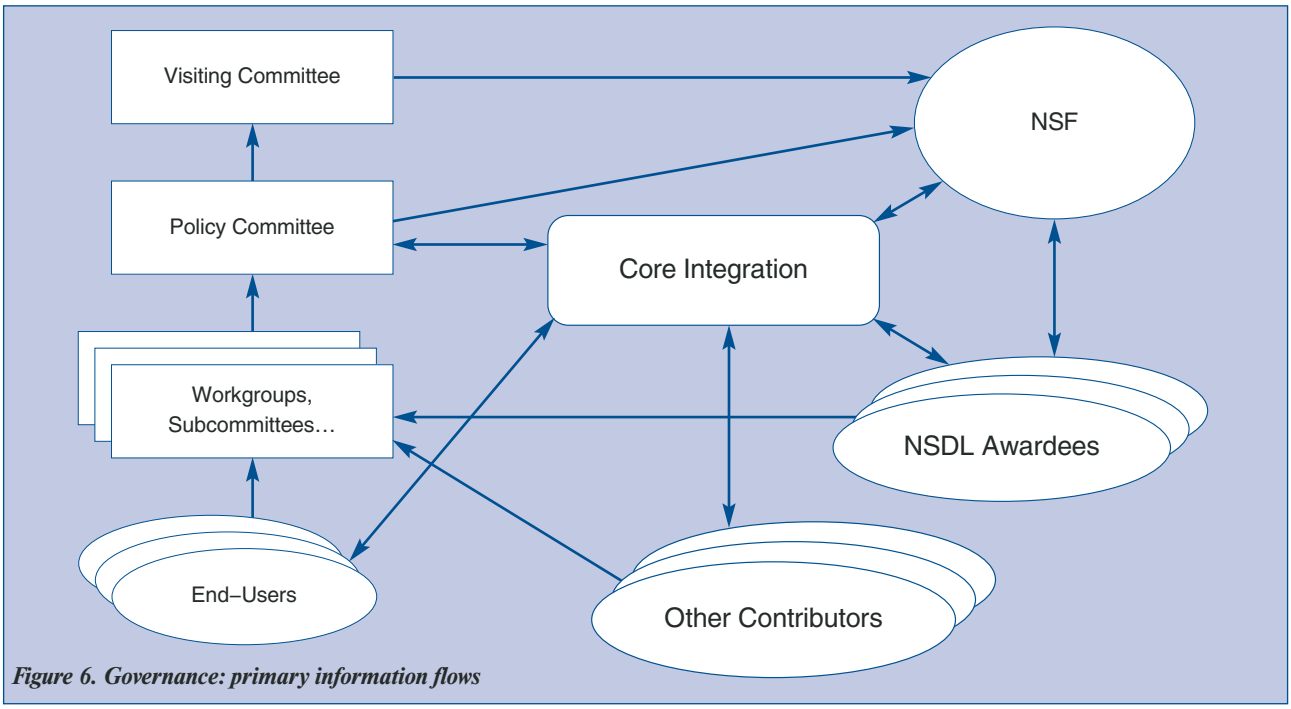

Figure 6. Governance: primary information flows

This table summarizes the key elements of the governance model, organized by the types of guidance and decisions that pertain:

Entity	Area of Decision Making and Guidance	Composition
National Visiting Committee	Accountability and top level strategy and feedback (to NSF and the CI team)	NSF appointed advisory group
Full Assembly	Adopts the governance model, elects the Policy Committee, and responds to polls when appropriate	One representative from each project (PI or designee)
Policy Committee	Empowered to act on behalf of the Assembly, articulating policies, strategies, and priorities	Elected by the Assembly from a slate prepared by a Nominating Committee
Topical Standing Committees	Each reports to the Policy Committee, which charters it to reflect a subdomain of responsibility	Defined by the Policy Committee
Subcommittees and Task Forces	Ad hoc groups may be chartered to address specialized needs of the NSDL such as a long–term governance, incorporation, and communications. Each reports to a specified committee	Defined by Policy or Standing Committees as needed
Spontaneous Interest Groups	Address ad hoc common interests from individual project members	Composed of individuals within and beyond the NSDL projects as appropriate
Project–specific	Operations, management, and budgets	Individual awardees

Table 1. Elements of NSDL governance

Topical Standing Committees set by the NSDL Policy Committee in spring 2002 include: Community and Services, Content, Educational Impact and Evaluation, Sustainability, and Technology.

The Community and Services Standing Committee is focused on library users and learners, and also on supporting other communities and views, such as teachers, the primary and secondary education communities, and other groups involved in education. Relative to earlier governance documents, it serves as a users committee. The Content Standing Committee serves as the primary group to recommend or adopt policies associated with the creation, development, and maintenance of individual collections within NSDL. It addresses issues of quality, in conjunction with other groups, such as Educational Impact and Evaluation. The purpose of the Educational Impact and Evaluation Standing Committee is to ensure that participatory and stakeholder evaluation principles are integrated into the development and implementation of NSDL. The Sustainability Standing Committee's mission is to facilitate the development of an NSDL entity for long-term sustainability of the NSDL collective through a diversified funding stream. Intellectual property issues and publisher relations are among the tasks of the Sustainability Subcommittee. The Technology Standing Committee is a forum for those involved in building the technical infrastructure of the NSDL. As such, the Committee serves as the voice of the Assembly in matters ranging from technical standards (such as metadata) through the technical integration of resources and services into the NSDL architecture.

8.2. Digital Library for Earth System Education (USA)

8.2.1. DLESE version 1.0

The Digital Library for Earth System Education (DLESE) [DLESNSF, DLESEC, DLESME] mission is to "improve the quality, quantity, and efficiency of teaching and learning about the Earth System, by developing, managing, and providing access to high-quality educational resources and supporting services through a community-based, distributed digital library". Over the past two years, DLESE has emerged to support the specific educational needs of the geoscience community within the larger NSDL network. In the tradition of community libraries, DLESE can fundamentally change how students learn, instructors teach, and researchers interact, by providing new ways of sharing information, tools, and services. DLESE serves as a vehicle for the geoscience community to respond to the challenges of systemic educational reform and the changing technological landscape.

Version 1.0 of DLESE is the operational library that was released in August 2001, with approximately 1,000 carefully selected resources in its initial collection. Version 1.0 provides educational discovery features that enable users to search by grade level, educational resource type, and keyword. This version also contains a resource cataloguer, and community oriented services, such as discussion forums for working groups and a community-posting tool. Version 1.0 creates an initial three-tier library architecture in which requests made in the user interface are mapped to operations over centrally held metadata through middleware services. To ensure interoperability with the NSDL, support for the Open Archives Initiative harvesting protocol has been implemented. This protocol is being used by DLESE to support transport of metadata between it's distributed collections.

The DLESE collections grow through community contributions from individuals or institutions. Contributions consist of individual resources or entire collections. The effectiveness of this collections model depends on having community members who are able to create and share their efforts to further common intellectual goals. The DLESE Program Center (DPC) enables the community to consciously and actively shape the intellectual framework of the DLESE collection by providing tools, components, and services that reflect DLESE policy, assure collection quality, and promote pedagogical innovation. Much of the DPC's prior work has focused on creating resource characterization frameworks (a metadata framework implemented in an XML schema), resource characterization tools, and services necessary to grow the collections and support resource discovery.

Version 1.0 supports resource discovery using keywords, grade level, and educational resource type descriptors (map, visualization, activity, lab, etc.). Discovery is provided through both basic and advanced search interfaces and a graphical browsing interface.

Library awareness and effective use of library services by educators and students have been promoted by DPC through a strong presence at geoscience professional meetings, community Working Groups (e.g. Diversity, Data Access, K-12), and through the DLESE annual meeting. These efforts have been targeted towards stimulating library development projects.

To evaluate user experiences, library collections and services, and educational effectiveness, the DPC has been focused on: (1) formative studies of current and potential library users to inform library design, (2) ethnographic studies of library building processes, (3) participant surveys to gauge the usefulness of each annual meeting and to provide broad input into library planning, and (4) online survey instruments for the beta test programme.

DLESE is a community-owned and governed digital library offering high-quality electronic resources that foster learning about the Earth at all educational levels. The DLESE Strategic Plan, which outlines the five-year vision for the next stage of library development, was formally approved in November 2001. When fully operational, DLESE will offer peer-reviewed teaching and learning resources, interfaces, and tools to allow exploration of Earth data; services to help users effectively create and use educational resources; and an "intellectual commons" facilitating sharing, collaboration, and excellence in Earth system education. DLESE users include learners and instructors in all venues, many of whom are also resource contributors, developing educational materials, providing scientific knowledge, and evaluating DLESE holdings.

8.2.2. DLESE evolution plan

During the next five-year performance period [DLESNSF], the DLESE Program Center (DPC), in conjunction with the larger DLESE community, will engage in a programme of work to achieve the following three goals:
1. Establishing DLESE as the premier, trusted source for high-quality geoscience educational resources.
2. Promoting sustainable library growth through community capacity-building and community participation in library governance, development, and operations.
3. Enabling the library to serve as a catalyst for geoscience educational reform.

These goals will be realized in stages and measured against two-year and five-year benchmarks outlined in the Strategic Plan. These benchmarks are articulated in two major library versions, Version 2.0 and Version 3.0 (Table 2).

DLESE version 2.0

In Version 2.0 (due for release in the summer of 2003), collections development services and resource discovery efforts will be extended to support science literacy standards and the Earth system science (ESS) perspective. Users will be able to search across multiple peer-reviewed collections, according to an Earth system perspective and a variety of science education benchmarks and standards. Community forums and library services supporting the effective use of resources and professional development will be available. Version 3.0 will represent a significant step towards supporting data use in geoscience education. Users will be able to search across spatially and temporally indexed resources, such as data, maps, images, and field guides. Gazetteer services will support user-centred geo-referenced discovery interfaces using place names and Earth system events. Integrated tools and services to assist with age-appropriate exploration of data will be available. Users will create and share a variety of personalized collections.

For Version 2.0 [DLESNSF], the current DLESE architecture will be extended to support distributed community collections and to interoperate with NSDL collections and services. Some of these services are already in place, in particular, a repository management system and Open Archives Initiative (OAI) data and service provider interfaces. These OAI interfaces support both DLESE's native ADEPT/DLESE/NASA (ADN) metadata framework, and the Dublin Core (DC) framework as recommended by NSDL. The ADEPT middleware, with its innovative search bucket architecture and mechanisms to work with geo-spatial searching, will be incorporated into the DLESE discovery system. The ADEPT bucket architecture enables collections to map their metadata framework onto common semantics (buckets) that can then be used by shared library services, such as discovery systems.

Thus, for Version 2.0, DPC efforts will focus on extending the resource metadata framework and resource characterization tools to enable contributors and collections developers to characterize their holdings according to a variety of national science literacy standards and an Earth system science perspective. In Version 2.0, DLESE discovery services will be extended to enable users to search and browse DLESE collections using the US national science education standards, the geography standards, and the Earth system science (ESS) vocabulary.

DLESE version 3.0

In Version 3.0 [DLESNSF] (the version to be delivered at the end of the five-year performance period), the architecture will be extended to support resource discovery using geo-referenced information, i.e. geospatial footprints (e.g. regions

	Version 1.0	Version 2.0	Version 3.0
Web Dimension			
Discovery	Grade level, educational resource type, key words	Educational discovery: standards, Benchmarks, ESS	Geo–referencing and Earth system events
Collections	Modest (about 1,000 resources)	Reviewed collections with broad topical breadth for primary, secondary, and higher education	Thematic and personalized collections
Community	Working groups, community postings	Community Forum, Prototype teaching and learning centre	Robust teaching and learning centre
Services & Support	Search tips available; support for resource contributors provided	Online end–user support (FAQs, Tutorials); community support personnel available	Robust community input mechanisms; email support provided
Data Dimension			
	Data sets and tools present in broad collection	Specially constructed thematic resources combining data, tools, and learning aids for primary, secondary, and higher education	Metadata extensions for integrated access to Earth data and learning resources
	DAWG convened; demonstrator projects under way	Specifications for data access and delivery developed with DAWG	Ability to create personalized inventories of data and tools
Use Dimension			
Target Audience	Contributors, faculty early adopters	Extended to mainstream primary, secondary, and higher education teachers/faculty	Extended to students and informal learners
Volume Capacity	1,000s of sessions/month	10,000s of sessions/month	100,000s of sessions/month
Building Dimension			
Governance	Committees established; some policies in place	All major library policies developed and in place	Mature community–based governance
Collections Development	Focus on object–level growth	Collection–level growth; collections developers toolkit	Brokering for persistent collections in place
Processes	Manual growth and maintenance processes	Collections management and harvesting processes automated	Automatic resource/collection cataloguing
Evaluation	Web metrics collected; beta testing is routine	Collection metrics collected; formative evaluations are routine	Summative evaluations of DLESE under way

Table 2. DLESE versions

on the surface of the Earth specified by latitude and longitude) and temporal footprints (e.g. a point or slice in time). The architecture also will be extended to support existing and emerging federated protocols necessary for interoperating with distributed DLESE and NSDL partners. For example, supporting the widely used Z39.50 protocol would enable DLESE to interoperate with existing libraries and bibliographic services.

Version 3.0 will also include support for geo-referenced discovery. Inquiry-oriented Earth system education makes use of many nontextual resources, such as maps, images, and data, to bring hands-on activities into the classroom. Indexing by temporal and spatial dimensions will be supported. Users will be able to locate such geo-referenced resources using a direct manipulation interface (e.g. selecting a region on a map) or via textual search terms, such as "New York". To support these discovery features, the place name gazetteer developed by ADEPT and the Map and Imaging Laboratory at UCSB, USA, will be incorporated into the discovery system.

In Version 3.0, the discovery system will be extended to support the educational use of Earth data and data analysis tools. Educational use of real time and archived data is one of the foundations of the geoscience education reform movement and is critical for integrating geoscience research into educational practice. Tools appropriate to primary, secondary, and higher education learning contexts are needed for scientific data mining, analysis, and visualization.

Throughout the performance period 2003-2007, the DPC will continue to design library interfaces and services for broad accessibility. In future versions, partnerships with key professional societies that serve underrepresented populations will be established.

Metadata collaboration

ADEPT, DLESE, and NASA's Joined Digital Library are taking their current frameworks as the basis for developing a common metadata framework, the ADEPT/DLESE/NASA (ADN) Joint Metadata Framework. ADEPT, DLESE, and NASA are using XSL to transform the ADN XML catalogue records to Dublin Core records for use in NSDL. For the ADN Joint Metadata Framework, DLESE is responsible for schema development, maintenance, and documentation for required metadata and an overarching framework schema; ADEPT is responsible for geospatial and temporal components and model description schemas. All three organizations are responsible for collection level type information. A collection level metadata framework based on, using the Alexandria Digital Library collection level metadata is being investigated.

Since DLESE will be a federated system, search buckets might be used to facilitate searching [ADEPTA]. Search buckets are a small set of high-level metadata attributes that can be used to make collections appear to have the same searchable metadata. They are also designed explicitly for querying. The underlying metadata fields of a record do not map one-to-one to a bucket. Rather, groups of fields map to one bucket and fields may map to more than one bucket.

Collections policy

The scope of the DLESE collection is Earth system education, with particular emphasis on interdisciplinary areas. The collection will favour materials that bring the Earth into the classroom or other learning site, connecting the general with the specific, theory with evidence, and the global with the local. The collection will favour materials which are well documented, easy to use, bug-free, motivational for learners, pedagogically effective, scientifically accurate, and which foster mastery of significant understandings or skills. The types of resources and tools to be collected include research and education materials and sources of content, tutorials and learning resources, field trips, and tools.

Community collections

Community collections that collaborate with DLESE include NASA's Joined Digital Library (JDL), the Digital Water Education Library (DWEL) for primary and secondary education, the Electronic Encyclopedia of Earthquakes, the Geoscience Digital Image Library (GeoDIL), the Atmospheric Visualization Collection (AVC), and the COMET Multimedia Database.

Relationship to geoscience, science education, and information technology partners

DLESE has established relationships with science and science education professional societies, including the American Association for the Advancement of Science (AAAS), the American Geological Institute (AGI), the American Geophysical Union (AGU), the Incorporated Research Institutions for Seismology (IRIS), the National Science Teachers Association (NSTA), and the emerging Center for Ocean Sciences Education Excellence (COSEE) and Earthscope efforts. These partners provide outreach opportunities for DLESE through their conferences and workshops. AAAS library interfaces are under development to enable users to discover resources via the AAAS strand maps [ATLAS]. This would enable resources indexed via either the benchmarks or the standards to be queried and retrieved.[7] AAAS intends to integrate DLESE into their professional development offering for educators. DLESE has a codevelopment partnership with NASA and plans to extend its partnerships to other federal and state agencies during the performance period 2003-2007.

In the data arena, DLESE has two important partnerships, with the US NSF funded Unidata Program Center and the emerging Geoinformatics effort [SINHA00]. Both are developing mechanisms for describing Earth system data for discovery and use in educational and research settings. In the THREDDS (Thematic Realtime Environmental Distributed Data Services) project the Unidata Program Center is developing services to enable users to create personalized and sharable collections of data and data analysis tools.

[7] This is an absolutely different model and interface. Serious work on resources registration is required to achieve the indexing of resources via benchmarks.

8.3. Computer and Information Technology Interactive Digital Education Library (USA)

NSDL has many projects funded through the collection track. CITIDEL is a representative example, employing the latest in digital library technology. As such it is a microcosm of NSDL, with aspects related to content, services, and community building.

8.3.1. CITIDEL foundation

CITIDEL is built on a strong foundation of prior work and engages a dynamic and active community of computing educators. The CITIDEL goals are to:

- provide the greatest possible access to high-quality resources that contribute to learning about computing at any level;
- provide the greatest possible visibility to the products of talented educators and developers who produce materials that support student learning about computing;
- provide access to materials that may have value within specific contexts, without undue barriers;
- support open exchange of materials and comments on materials discovered;
- support community building and mutual support of teachers and learners with common interests;
- support the use of discovered materials in learning activities.

Foundation work to address these goals includes the use of the Computer Science Teaching Center (CSTC, http://www.cstc.org), a resource submission, review, and sharing resource funded by NSF in 1997. CSTC allows easy submission of materials for wide dissemination by simple upload procedures. Volunteers register to review materials, thus providing a level of assurance to resource users that the materials have been evaluated and found to be working and sound. Materials are categorized to support searching by subject area and type of resource, as well as by author name or resource title and keywords. CSTC was extended with multimedia-related content as a result of the US NSF-funded Curriculum Resources in Interactive Multimedia project [CRIM] that emphasized knowledge modules to aid in education [MCCKM].

In addition to CSTC, CITIDEL incorporates access to resources collected by other projects, including JERIC (the ACM *Journal of Educational Resources in Computing*, http://www.acm.org/pubs/jeric/ [CASSEL]), ResearchIndex (http://www.researchindex.org), the CS Virtual History Museum, the Networked Computer Science Technical Reference Library (NCSTRL, www.ncstrl.org), PlanetMath (a collaborative mathematics encyclopedia, http://planetmath.org/), and the ACM Digital Library (http://www.acm.org/dl). Efforts are under way to include metadata from the IEEE Computer Society, DBLP (the Computer Science Bibliography at http://dblp.uni-trier.de/), and relevant parts of the Networked Digital Library of Theses and Dissertations (NDLTD) collection, so that by 2003 there may be roughly a million records. The number of partners participating in CITIDEL grows regularly, just as the number of collections projects within NSDL grows. One of the challenges, at both the level of individual collections and the level of NSDL overall, is to assure the continuing accessibility and the status of maintenance of member collections. As is evident from the list of partners, CITIDEL includes resources of many types. Some are created specifically for supporting computing education. They include demonstration modules and simulators that display the behavior of computing components or algorithm performance, for example. Other resources are general-purpose tools or publications including refereed journal articles. Each of these has its place in supporting learning.

CITIDEL permits direct submission of materials to the library, but provides access to resources supported by others, too. CITIDEL also obtains materials by crawling the Web in search of candidate materials; with the permission of the author these materials are submitted to CSTC for evaluation and thus become part of CITIDEL. CITIDEL is really an example of NSDL on a smaller scale, bringing together a diverse set of resources and presenting them to the learning community as an integrated collection.

Digital libraries have differing policies regarding the evaluation of resources presented to its users. The approach of CITIDEL is to have a variety of types of materials, but to have them clearly distinguished for the user's convenience. Thus, CITIDEL searches will yield refereed resources from ACM journals and casual products of an individual's creativity. The searcher will know the source of the resource and can apply judgment with regard to the amount of confidence to place in the resource.

CITIDEL intends to serve the entire community of computing educators in the United States, with the understanding that access to the rest of the world is available as well. To reach the largest possible population, the resources of

CITIDEL must be available in both English and Spanish immediately, and in other languages in time. Spanish language access is a central component of the initial design effort.

Overall, the CITIDEL project presents materials with characteristics in three dimensions as shown in Figure 7. Note that quality is dependent upon source and corresponding approaches to add value.

Figure 7. The three dimensions of CITIDEL resources

8.3.2. CITIDEL growth

A key aspect of building a comprehensive digital library by integrating diverse resources is a plan for effective metadata harvesting. CITIDEL is a participant in the Open Archives Initiative. All member collections of CITIDEL are OAI sites and CITIDEL regularly harvests metadata from these sites in order to keep search results current with the collection revisions.

From Figure 8, we see that CITIDEL is both a harvester and a provider of OAI data. CITIDEL provides its metadata to the overall NSDL OAI harvester, contributing to the umbrella library, while serving the specialized needs of the computing community.

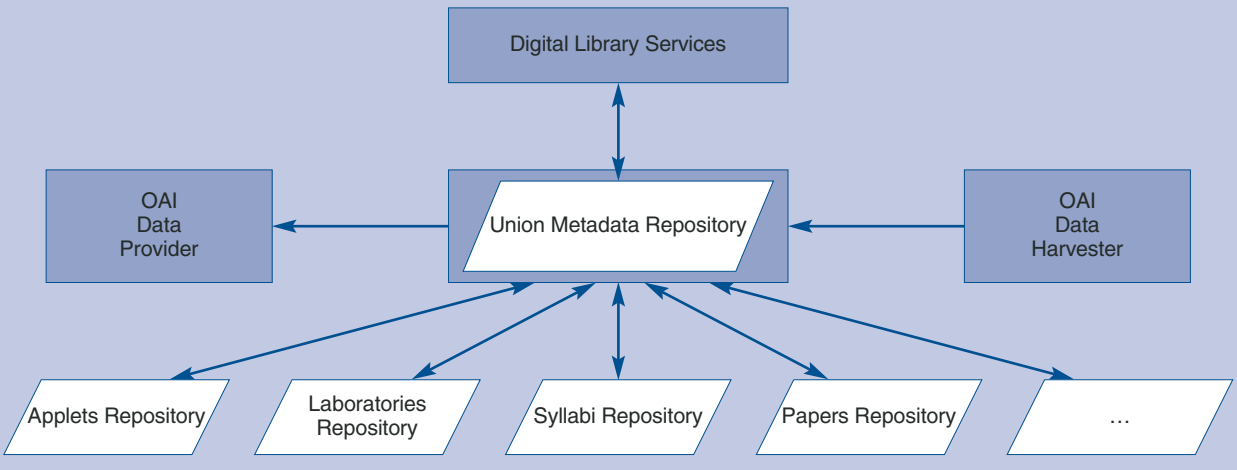

Figure 8. Metadata collection and provision in CITIDEL

CITIDEL provides a separate view for each of its user communities, including educators, learners, and administrators of the resource. Services are tailored to the needs of each group as shown in Figure 9. Filtering profiles allow a user to specify qualifications for the resources to be returned in response to searches. For example, some users may choose not to see materials that have not been peer-reviewed; other users may choose to limit the resources they see based on their education level or amount of familiarity with the topic. These profile characteristics will certainly change over time.

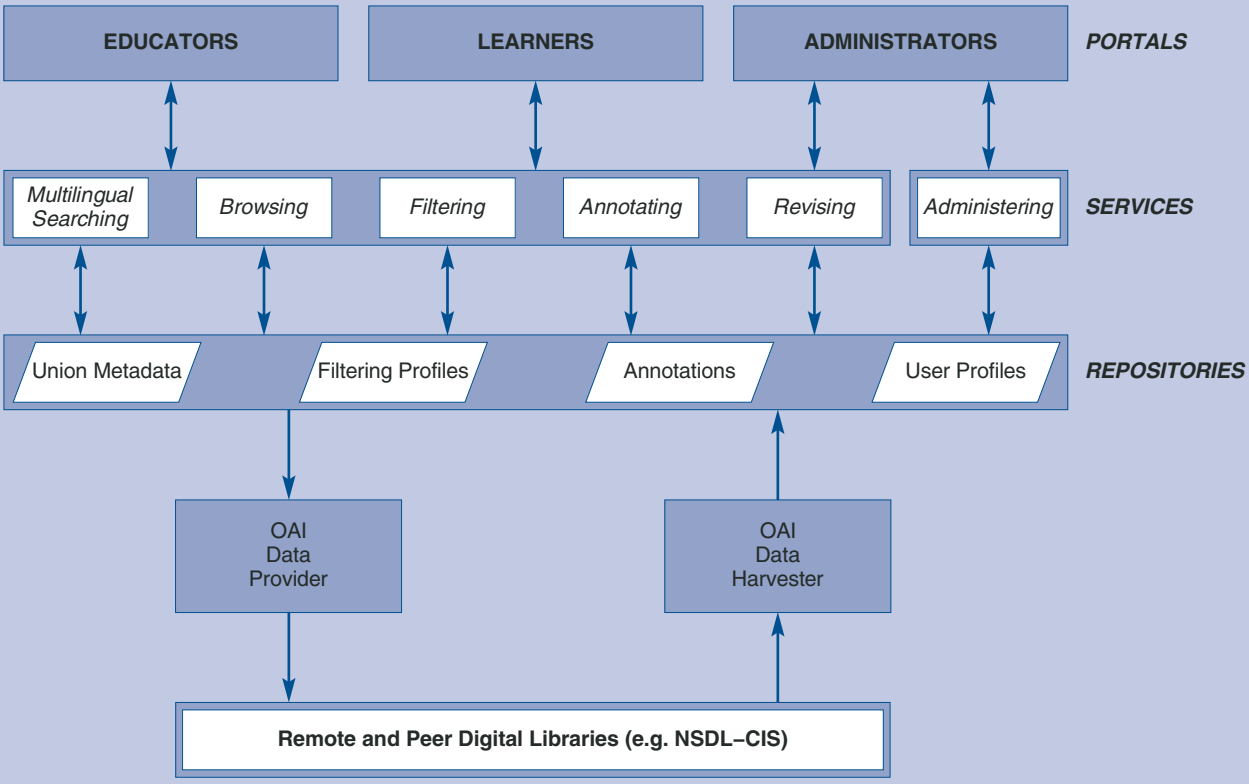

Figure 9. Architecture for local and interoperable CITIDEL services

CITIDEL additions

CITIDEL is more than a collection of resources and more than a collection of collections of resources. In addition to the usual DL services, such as browsing, filtering, annotating, revising, and administration, and the more unusual multilingual support, CITIDEL includes a number of innovative service activities. An example is the VIADUCT (Virginia Instructional Architect for Digital Undergraduate Computing Teaching) project.

VIADUCT is a supplement to the CITIDEL search and browse facilities that allows a user to group a subcollection of resources into a meaningful and useful teaching facility. The premise is that gathering resources is only the first step in using them effectively. The resulting collection of materials must be organized and presented in the context of overall learning goals. VIADUCT allows a CITIDEL user to gather materials and insert them into a syllabus, which defines a unit of a course as developed by an individual teacher. The resulting syllabus is a vehicle for the teacher to interact with students using resources discovered in CITIDEL, a concrete representation of that course unit for future use and revision, and a form of that course unit design that can be shared with other teachers for their use and their further development. Syllabus components constructed and populated with VIADUCT can easily be published in CITIDEL, becoming a new generation of resources for sharing in the computing educational community.

8.3.3. CITIDEL connections

CITIDEL exists to serve the community of computing educators. Thus, its connections to that community are critical. One of the foundation stones of CITIDEL, CSTC was created with funding from the US National Science Foundation,

with support from the ACM Education Board. CSTC has benefited by close collaboration with the ACM Special Interest Group in Computer Science Education (SIGCSE, http://www.acm.org/sigcse/). Through the generosity of ACM, the contents of the *Journal on Educational Resources in Computing* and the publications of SIGCSE are freely available through CITIDEL. Other ACM digital library content is searchable through CITIDEL, with access to the full content limited by the usual subscription requirements. Like the overall NSDL, CITIDEL must address the issues of recognizing a subscriber to a member service in order to support access and to provide fee-based access to nonsubscribers.

Because CITIDEL is both a product of and a service to the computing community, its design and implementation provide illustrations for many of the issues that other DLs will have to address as they mature. CITIDEL is accessible at www.citidel.org.

Related projects

The CITIDEL project is connected with two NSDL services projects. One, led by Hsinchun Chen and Ann Lally at the University of Arizona, has developed a system supporting concept maps, integrated with searching, summarization, and visualization. Thus, CITIDEL and other collections can be searched and the results can be used to build concept maps whose nodes link to resources or other lower-level maps. The other project, GetSmart (http://collab.dlib.vt.edu/runwiki/wiki.pl?GetSmart), has been deployed in autumn 2002 in computing and information technology courses at Virginia Tech and the University of Arizona. While related to efforts described below (e.g. strand maps in Project 2061), GetSmart focuses on empowering students to create their own concept maps, as an aid to understanding readings and course materials, and to developing their term projects. In addition to supporting study and documentation efforts, GetSmart helps with class presentations and group discussion (e.g. with group and class-level maps that result from comparison of the maps of individuals).

In addition, the "DL in a box" effort (http://dlbox.nudl.org/), led by Su-Shing Chen at University of Florida, and involving Joe Futrelle at the National Center for Supercomputing Applications (NCSA), builds upon the Open Digital Library (ODL) effort launched at Virginia Tech (http://oai.dlib.vt.edu/odl/). The idea behind ODL [SULEMA1, SULEMA2] is to have a collection of components that conform to a set of lightweight protocols that minimally extend the protocol used in the Open Archives Initiative. Some of these components are key elements of the CITIDEL software. Recently, the ODL component for annotation was integrated into CITIDEL, demonstrating how easy it is to add services in the ODL framework. The DL in a box effort aims to extend, package, and document the current set of components so that NSDL and other efforts that require digital library software support can expand upon this framework, building their core technology around well-tested and modular routines that can be easily extended.

8.4. Networked Digital Library of Theses and Dissertations (USA)

NDLTD, www.ndltd.org, is an international digital library initiative supporting education [NDLTD1, NDLTD2]. It represents a rather distinctive approach to DLE development, which fits well with UNESCO interests, and has had strong UNESCO support in recent years. A number of developing nations have become involved, since there are immediate benefits, and since there is a clear and simple sustainability model. NDLTD has benefited from support of the latest in digital library technologies, and so also is interesting since it is based upon open source software.

8.4.1. Education focus

NDLTD focuses on graduate education and research. It deals with a key phase of graduate studies, namely the preparation of a thesis or dissertation (TD), but also can cover undergraduate theses or graduate project reports. Though historically TDs were submitted in paper form, NDLTD aims to move the community of scholars rapidly toward an emerging genre of electronic theses and dissertations (ETDs).

Given the emerging collection of ETDs, students as well as other researchers and even educators derive significant benefit. By and large, universities make ETDs freely available, and NDLTD services support centralized discovery. There is intense interest in this content from around the world, since it serves many educational purposes.

First, students preparing their own thesis or dissertation can learn from the works of others. Of course, they can learn specific content, related to their research interest. But also they can learn about style and organization, presentation, and even about use of multimedia and hypermedia technologies. Since TDs generally have long bibliographies, as well

as extensive literature reviews, students may be aided in carrying out their own literature review, and may have an easier time getting oriented to the related work in their field. Indeed, the world's collection of TDs constitutes one of the largest and most comprehensive collections of overviews of research in varied specialty areas. Further, since TDs provide details on the methods used by researchers, they also are a rich mine regarding research methods and particular methodologies. By mining the TDs of the world, students should also be able to avoid wasting time to reinvent the work already carried out by others, when that related work is not easily discovered.

Second, since ETDs are electronic documents, it is possible to work with small portions that fit in with a specialized need. If the ETD is a traditional document, formed simply as "electronic paper" (e.g. as a PDF version of a Word document), then even if the entire work is in a single file, tools will allow a single page or small span of pages to be isolated. If the ETD is broken into files, such as for each chapter, then these smaller segments can be readily accessed alone. Further, if there is multimedia content, a single image, video, audio segment, data file, spreadsheet, or other component can be utilized. Teachers can show these components in class, can assign them to students in connection with an online assignment, or can encourage their study in connection with project work.

Further, ETDs may help students identify topics that they can explore in conjunction with their own research. Many theses have sections that summarize open problems. Others suggest additional work that can be done to extend the recently completed study. According to one report, a student who prepared an ETD in South Africa was contacted by a student at the University of California, Berkeley, USA, who wanted to extend the first student's efforts.

Finally, working with the NDLTD collections and services, students gain familiarity with modern digital library systems. Since 1998, NDLTD has provided federated search access to a number of sites maintained by members, and more sophisticated technology can now help in that regard [MARIAN1]. Today, the MARIAN (Multiple Access Retrieval of Library Information with Annotations) system, a modern research digital library system that is being enhanced with multilingual support, can help enhance access to ETDs [MARIAN3]. Indeed, current research on generating digital libraries based on formal specifications is allowing MARIAN to support content from a variety of sites, through a variety of search and harvesting mechanisms, with views governing how the internal semantic network representation is managed to support user information needs [MARIAN2].

8.4.2. Learning by doing

Further educational benefit derives with regard to NDLTD when students learn by doing. Many students learn best this way, rather than by simply reading or listening to lectures. Since a thesis or dissertation may take months or years to develop, and often is an object of pride to a young scholar, learning while preparing an ETD may have particularly lasting effect. Part of the learning process deals with exposure to the latest document processing and electronic publishing technologies.

Thus, as time moves forward, it is more and more common for students to prepare their thesis or dissertation with a computer. Then it is a small matter, if proper prior planning was undertaken at the university, for the student to prepare a version suitable for long-term preservation (e.g. using PDF or XML, as well as suitable standards for multimedia, like JPEG and MPEG). Finally, it is quite simple for students to fill in a form (that thus captures needed metadata), and to upload files, thus getting familiar with electronic submission processes.

Such exposure to electronic publishing and author-submission procedures is particularly important for modern scholars. Many funding agencies, including the US NSF, require proposals to be prepared and submitted electronically. Many conferences and journals require papers to be managed in a similar fashion. Those reviewing papers or proposals often must access others' works from a DL-like system, and must submit their reviews in similar fashion. In short, the activities of preparing and submitting an ETD are good preparation for a number of commonly required scholarly tasks.

Clearly, however, technologies related to electronic publishing, as well as to use of digital libraries, rapidly change. Fortunately, there is a good tracking of students learning through work on their ETDs with young researchers engaging in common activities. In other words, if universities always ensure that their students who work on ETDs learn about state-of-the-art methods in electronic publishing and in digital library access, then all will be prepared to work as scholars after graduation.

This learning process actually has an even broader effect. First, students who will make their works widely available have greater need to learn about matters of digital preservation, patent, copyright, and intellectual property than they might

if only a paper document were prepared. Thus, they may be exposed to these important issues, which unfortunately are often are poorly understood by many senior members of academia. Second, since students often turn to their mentors when preparing their theses or dissertations, there is increased exposure of the faculty to many of the above-mentioned issues and skills. Thus, ETD activities often have a stimulating effect on campus scholarly publishing and library budget discussions, and may facilitate institutional progress on these matters.

8.4.3. A cost-effective packaging

While many DLE-type efforts discussed in this survey require considerable investment, and have not yet led to sustainable programmes, work with ETDs clearly illustrates how sensible packaging of efforts can lead to beneficial and cost-effective practices. The following illustrates how this may work. Though there are many variants, we focus here on the simplest approach, which was adopted at Virginia Tech, USA, after 1996, and has since been adapted in a variety of other contexts.

Since the beginning of 1997, students preparing a TD at Virginia Tech have conformed to the ETD requirement then put in place (after pilot efforts dating back to 1987). In particular, as part of their graduate activities, all who prepare a thesis or dissertation must do so electronically, and must submit their ETD for review by the graduate school. Once the work is approved, the automated system that handles submission and subsequent processing then allows library handling, including cataloguing. Works are made accessible in accordance with the approval form that has been signed by the student author as well as the student's graduate or supervising committee. The university library accepts no paper submission, so the entire workflow is managed electronically. This leads to so much saving in terms of paper handling and transport, avoidance of binding and shelving, and even student copying, that the overall programme costs less than the earlier paper-based approach.

These savings result from the connection of training of graduate students with the skills they need as scholars, and then placing responsibility on their shoulders, through a requirement. It is facilitated by the use of automation software that is made available as open source by Virginia Tech. That software even includes allowing a university to become a data provider as part of the Open Archives Initiative, facilitating development of a global union catalog of ETD metadata.

Benefits are considerable, even beyond the above-mentioned savings and learning. First, student research becomes much more available. Whereas paper theses and dissertations historically have been read by very few beyond the set of faculty involved in advising and examining the student, ETDs are typically read by hundreds or thousands. Given the investment of time and effort by the student, and the institutional support that they receive over several years of study, it is clearly sensible to leverage that investment by making student work accessible. Indeed, based on experience at a number of NDLTD sites, a university may have hundreds of thousands of accesses per year to its ETD collection, thus expanding its visibility in the research community. This system may be of particular value in developing countries, where research is now rarely disseminated on the international scene; in the world of ETDs, every participating university can benefit from a truly level playing field.

On the global scale, the benefits are even clearer. If there is full participation in NDLTD worldwide, there should be more than 200,000 ETDs made available each year. If production follows the awarding of graduate degrees, this will level the playing field regarding research access. Furthermore, whereas there are efforts under way to make journals more widely available globally, due to economic barriers it is likely that it will take a number of years for smaller universities, especially in developing countries, to gain access to 200,000 journal articles per year.

8.4.4. Current status

There are national ETD efforts in Australia, Brazil, and Germany in addition to expanding efforts in India, Korea, and USA, as well as support by national institutions like the British Library and the national libraries of Canada and Portugal. All told, there are over 150 members of NDLTD, with roughly half in the USA, but with fastest growth outside the US. Some members, like OhioLink, which represents the 79 colleges and universities in Ohio, demonstrate that the number of universities engaged in ETD production is well over the 150 count of members mentioned. Now that NDLTD, which was launched in 1996, is maturing as a federation, more and more of its members are shifting to require ETDs, which will ensure a rapid increase in ETD production.

Many NDLTD members are leading universities, and many are known for their innovative practices. California Institute of Technology, Johns Hopkins University, Louisiana State University, Michigan Tech, Massachusetts Institute

of Technology, Pennsylvania State University, University of Texas at Austin, University of Virginia,Virginia Tech, and West Virginia University are among the more than 60 institutions that have joined in the USA. International members come from many countries, including Australia, Brazil, Canada, China, Finland, France, Germany, Hong Kong, India, Korea, Mexico, Norway, Portugal, Russian Federation, South Africa, Singapore, Spain, Sweden, and United Kingdom. In addition, NDLTD benefits from members that represent disciplinary interests such as for physics (PhysDiss) and mathematics (MathDiss).

UNESCO support is focusing on increasing ETD efforts in South America, the Caribbean, Eastern Europe, Africa, and other regions. UNESCO funded preparation of online and CD-ROM versions of an ETD Guide (http://etdguide.org), with over 400 pages, encouraging and supporting new initiatives with guidance for students, faculty, administrators, and trainers. The guide is available in English, French, and Spanish. UNESCO offices in Uruguay and France (headquarters) are coordinating efforts, working in concert with NDLTD. Some of the support will allow interested parties from key areas to attend a special training programme undertaken in conjunction with ETD 2003, the latest in a series of annual international conferences, scheduled for Berlin in May 2003.

8.4.5. Advanced technology

NDLTD has been actively engaged in the development and deployment of the latest electronic publishing and digital library technologies. In 1988, a document type definition for electronic dissertations was developed to facilitate authoring using SGML, and in subsequent years various tools were tested to help in such preparation. In recent years this has led to document type definitions as well as schemas that relate to authoring using XML. In the early 1990s, as soon as PDF was available in test form, ETDs were prepared using that format. And many ETDs have made use of standards like JPEG and MPEG as soon as they were established.

In 1998, access to the distributed collections of ETDs was facilitated by a federated search system set up by NDLTD at its collection site, www.theses.org. In 1999, NDLTD was represented in the Santa Fe, New Mexico, USA, meeting that led to the Open Archives Initiative, and NDLTD has been one of the lead groups involved in the OAI steering and technical committees ever since. As a result of over two years of international meetings and discussion, the ETD metadata standard, ETD-MS, was devised, and along with MARC 21, is the recommended representation format, which, along with Dublin Core, all NDLTD OAI data providers are encouraged to support. OCLC is working to make available more than four million TD records through OAI, and Virginia Tech maintains a union catalogue for ETDs that was among the first union collections harvestable as part of OAI.

Digital library support for NDLTD continues to be state-of-the-art. In the mid-1990s, studies were undertaken with the test versions of the IBM digital library software, and full-text content searches were demonstrated. The above-mentioned Open Digital Library framework was tested early on for applicability to NDLTD, and an experimental service using that component-based approach has been available for over a year. The Virtua system, one of the most comprehensive commercial library automation packages, now extended to be a full digital library system, has been deployed by VTLS Inc. in connection with its support of the union catalogue, with interfaces supporting 14 languages and content available already in six languages.

MARIAN, the research digital library system developed at Virginia Tech, has worked with NDLTD data since it first became available. MARIAN has demonstrated a variety of types of support for distributed collections, including through OAI, Harvest, and Z39.50 [MARIAN1]. MARIAN also supports a modular approach to DL development, but in a more principled way than most, building as it does upon a theoretical framework for digital libraries (5S, which builds upon five key constructs: societies, scenarios, spaces, structures, and streams). NDLTD support is now available as driven by a description of NDLTD collections and services in the 5S specification language, 5SL [MARIAN2]. MARIAN supports Unicode and searching in multiple languages, incorporates a view mechanism similar in concept to that in database systems, supports varied ontologies, and is implemented in Java in versions available as open source software, scalable to run on laptops or as a distributed system running on a network of computers [MARIAN3].

Because of its educational focus, its international scope, its simple sustainability model, and its incorporation of the latest DL technology, NDLTD is particularly amenable to integration with other UNESCO efforts, such as the move toward DLEs.

9. Current Digital Libraries for Education: European projects

9.1. The Distributed National Electronic Resource and the hybrid library (UK)

DNER [DNERLP, DNERAV] is a generic term used to describe the wide range of information-related activities and services supported by the Joint Information Systems Committee (JISC). The DNER coordinates informational activity in three broad areas:

1. The creation of a strategic national resource of educational and learning materials: this consists of collections of materials made available on a national level via networked systems designed to deliver them to the end-user.
2. The creation of a framework for community resources: institutions and individuals within institutions also manage or create resources that may be articulated as part of a wider "community" resource.
3. The creation of a resource discovery framework for a global resource collection.

The collections themselves consist of a wide range of resource types: scholarly journals, monographs, textbooks, abstracts, manuscripts, maps, music scores, still images, geospatial images, vector and numeric data, moving picture collections, and sound collections.

For content and service delivery the predominant approach is plain HTTP, but this limits how the services can be used because of the limited semantic structures in the protocol. The aspiration here is to make more services available through more structured protocols such as Z39.50, LDAP, and the Open Archives Initiative. In that way, their data is more directly accessible and manipulable for others to use. This aspiration is broadly in line with the vision of the "semantic Web" espoused by the World Wide Web Consortium. The most important difference between the DNER and hybrid library projects is the most obvious one. Whereas the hybrid library projects are based in individual institutions, the DNER is a national service.

9.2. Scholnet and Cyclades: Extending the role of digital libraries (EU)

Scholnet (IST-1000-20664) and Cyclades (IST-2000-25456) are two digital library projects funded by the European Union 5th Framework Programme and coordinated scientifically by the IEI – CNR [SCCA01]. Both projects aim at extending the role of a digital library by providing services that support remote communication and collaboration among scholars. In particular, the goal of Scholnet is to develop a digital library providing an enhanced set of specialized services, while Cyclades is focused on the need to develop a service environment on top of large heterogeneous and multidisciplinary interoperable archives.

Scholnet (http://www.ercim.org/scholnet) aims at enabling the immediate dissemination and accessibility of technical documentation within a globally distributed multilingual community. Scholnet also aims at developing a digital library infrastructure to support communication and the collaboration within networked scholarly communities.

In order to achieve this objective, Scholnet will provide:

- *Traditional digital library services on multimedia documents:* These services enable scholars to communicate through the publication of and access to not only textual documentation, such as technical reports, project deliverables, or workshop proceedings, but also videos of tutorials or seminars (possibly synchronized with corresponding textual slides), training sessions, project presentations, demos, etc.
- *Handling of document annotations*: Annotations can be textual notes, ratings, links, etc., associated with either the entire document or its parts. Annotations can be authored by different people and will have public or group restricted access privileges.
- *Monolingual and multilingual search and retrieval services:* Monolingual search is provided in all of the project languages; if a user specifies the search language, the system searches only those documents that contain information in that language. In addition, a cross-language search facility allows users to query in their own language and retrieve documents matching the query in other languages.
- *Automatic personalized information dissemination service:* When a new document arrives in the digital library, a proactive facility sends messages to those users who, on the basis of their system-maintained profiles, are potentially interested in its contents.

From the technical point of view the Scholnet infrastructure will be built by extending, and partially rethinking, the basic services provided by the ERCIM Technical Reference Digital Library (ETRDL) (http://www.iei.pi.cnr.it/DELOS/ETRDL).

Cyclades (http://www.iei.pi.cnr.it/cyclades) will develop an open, collaborative virtual archive service environment supporting both single scholars and scholarly communities in carrying out their work. In particular, it will provide functionality to access large, heterogeneous, multidisciplinary archives distributed over the Web and to support remote collaboration among the members of communities of interest.

Cyclades will run in the data environment composed by the archives that adhere to the Open Archives Initiative's harvesting protocol specifications (http://www.openarchives.org). From the technical point of view, Cyclades will consist of the following federation of independent but interoperable services:

- *Access:* supports harvest-based information gathering, plus indexing and storage of gathered information in a local database.
- *Query and browse:* supports the users in formulating queries and develops plans for their evaluation. In particular, it provides an advanced multilevel browse facility, completely integrated with the search facility, that allows one to browse at the levels of schema, attributes, and documents.
- *Collection*: provides mechanisms for dynamically structuring the overall information space into meaningful (from some community's perspective) collections.
- *Personalization:* supports information personalization on the basis of individual user profiles, and profiles of the working communities the user belongs to. User and community profiles are automatically inferred by monitoring the user behavior.
- *Recommendation:* provides recommendations about new published articles within a working community. The choice about what recommendations to send, and to whom, is based on the profiles of both the user and the working community.
- *Collaborative work:* supports collaboration between members of remotely distributed working groups by providing functionality for creating shared working spaces referencing users' own documents, collections, recommendations, related links, textual annotations, ratings, etc.

10. Advanced frameworks and methodologies related to DLEs

10.1. Instructional course development in the presence of learning module repositories

A gap exists between current DLEs and their use in course development. DLE evolution in this direction can be envisaged as applying the successful experiences of instructional course development in the presence of teaching module repositories. Impressive experience in this respect has been accumulated in the Cooperative Program for Operational Meteorology, Education and Training (COMET) that was formed in 1989. The COMET programme [COMET] was originally envisioned as a broad effort to affect meteorology education and training in the United States. However, the programme has recently been involved in activities to enhance meteorology education in universities and meteorological services throughout the world. The COMET mission is to serve as a premier resource to support, enhance, convey, and stimulate scientific knowledge about the weather for the benefit of weather-information providers, educators, and users.

A typical COMET multimedia-based learning (MBL) module contains one to 12 hours of highly interactive instruction and incorporates case studies, graphics, animations, and video to provide an effective educational experience. Concepts are introduced via both computer text and spoken dialogue and are reinforced by displays of such graphic materials as time-sequenced satellite data, radar data, or videos demonstrating laboratory experiments. At various points in each module, students have the opportunity to test their understanding of the concepts presented. If the student answers incorrectly or would like more detailed information, she/he can access additional material that is typically presented by an expert on the topic.

COMET modules provide interactive, multimedia training on satellite meteorology, marine meteorology, hydrology, fire weather, and much more for the meteorologist, educator, and weather enthusiast. The COMET programme has instructional designers, author-programmers, staff meteorologists, graphic designers, and animators who produce computer-based training modules and related learning materials.

Currently COMET does not espouse the concept of a digital library. At the same time, to make the graphic images created in COMET projects available to the sponsors and the public, to enhance media reuse by the sponsors, and to enhance educational materials available to the meteorological community at large, the COMET programme has developed a Hydrometeorological Multimedia Database. Besides the images, instructional components can be found by text search applying keywords in a query. These examples show how a digital library of instructional materials can be organized to make materials customizable and reusable in new courses. Finally, information on various weather cases (e.g. on severe weather, hurricanes, winter weather, fire weather, floods, cyclones) is arranged into a simple classification system. They include radar and satellite data, model data, text data, upper-air and profiler data, and other contributed data sets. These case studies are being used in National Weather Service offices and university classrooms across the United States.

Some of the instructional modules contain concept maps in their definitions showing interrelationships of concepts used (Figure 10). For example, Mesoscale Convective Systems (MCSs) appear in many forms, ranging from a relatively disorganized mass of convective cells, to highly organized convective lines. At times the systems appear like small synoptic-type cyclones with bands of convective cells spiraling around the cyclone centre. The instructional modules and exercises provide the opportunity to apply many of the concepts presented in both the Conceptual Models and the Physical Processes portions of the Web site (http://www.comet.ucar.edu/modules/MCSMatrix.htm).

Such concept maps may serve as hints to define an ontology and to structure the information relevant to the instructional course on MCSs. This would provide for relating to those definitions of DL resources that contain any relevant information. Such a DL mediation interface is considered to be an advanced DL feature.

10.2. Learning objects for reuse

Learning objects are elements of a new type of computer-based instruction grounded in the object-oriented paradigm of computer science. Instructional designers can build small (relative to the size of an entire course) instructional

components that can be reused a number of times in different learning contexts. Learning objects include multimedia content, instructional content, learning objectives, instructional software and software tools, and persons, organizations, or events referenced during technology-supported learning [LOMN00].

The Learning Objects Metadata Working Group (a working group of the LTSC of the IEEE) describes the purpose of learning objects as being: "to enable computer agents to automatically and dynamically compose personalized lessons for an individual learner". Regretfully, this intention has not succeeded in any instantiation.

In fact, diverse definitions have arisen for learning objects and similar terms used with the same meaning [CLOIDP]. People have coined the term "instructional design theory", which is in its infancy. The online book [IULOOB] that contains a collection of philosophical papers (e.g. [CLOIDP, NOIOBJ, RBLPSS]) discusses the nature of learning objects and instructional design theory as if no information systems design with reuse existed before.

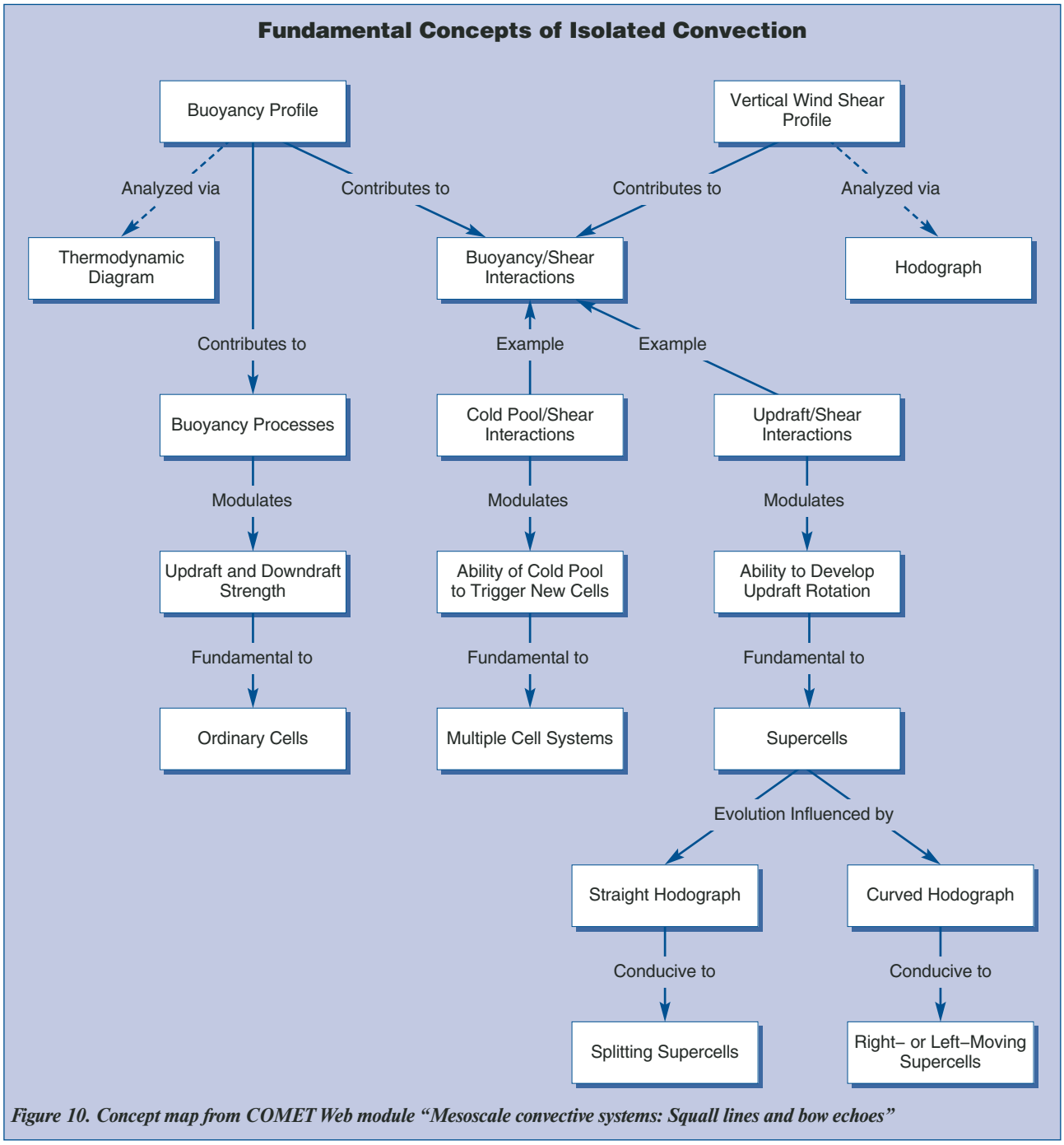

Figure 10. Concept map from COMET Web module "Mesoscale convective systems: Squall lines and bow echoes"

At the same time, a note on the COMET experience [COMIDP] shows that an engineering approach does exist for at least the top-down process of instructional development. The process includes the following phases: Initial Requirements Definition, Analysis and Planning, Design, Content Development, Component Development, Product Development, Publication, Implementation, Evaluation. Deep modifications are required to make the process a mixture of top-down and bottom-up (reuse of preexisting learning objects) approaches. The component definitions (learning objects) should appear alongside the requirements definitions. Both kinds of definitions should be provided with proper modeling allowing users to accomplish three things: (1) Users should be able to discover learning objects (stored in DLEs) and fragments of learning objects that are potentially relevant to the requirements definitions. "Relevant" means that learning objects are ontologically (contextually) suitable and fragments of requirements definitions can be substituted for learning object definitions so that users will not notice the substitution (frequently, it might be necessary to adapt learning objects to match requirements definitions). It should also be possible to specify generic learning objects that could be further customized. (2) Users should be able to compose adapted learning objects into larger definitions that would refine the requirement definition. (3) Finally, users should be able to justify that the resulting construct can serve as an implementation of the requirements definition.

For software components, an approach similar to the above has been experienced in a specific tool [CBISDT]. For instructional development, however, this remains an open issue. The state of current research only underscores this observation [DLREUSE, COMPLO].

10.3. Community organization around vast specialized educational resources and data

Development of specialized educational resources and data requires adequate organizational efforts in the relevant community. MeteoForum [METFOR], a pilot programme developed jointly by the COMET and Unidata programmes of the University Corporation for Atmospheric Research in Boulder, Colorado, USA, may serve as an example of such efforts. Under the MeteoForum concept, the international network of Regional Meteorological Training Centres (RMTCs) of the World Meteorological Organization (WMO) will work collaboratively with universities to enhance their roles of training and education through information technologies and multilingual collections of resources. MeteoForum will build a stronger sense of community among the RMTCs through Internet-based interactions and through the process of sharing educational concepts, educational materials, and hydrometeorological data with one another and with universities.

COMET has established a multilingual MeteoForum Web site to foster interaction and communication among MeteoForum members. The site, http://www.meteoforum.ucar.edu, will be coordinated with the WMO "Virtual Library".

The COMET programme brings to MeteoForum its recognized excellence in (1) distance learning (lectures, student interaction, and a large collection of educational graphics that are accessible via the Web or on CD-ROMs); (2) an outreach programme which involves partnerships between the academic research community and forecasters; (3) residence courses in advanced meteorology and hydrology that include lectures and hands-on laboratory activities; and (4) case studies, a collection of meteorological data sets for specific events (available online). Some of its vast resources have been translated into Spanish. COMET has the expertise to help MeteoForum create new materials or translate existing materials (e.g. case studies that focus on regional phenomena, such as Amazon squall lines or El Niño related floods).

Unidata [UNIDAT] activities that are of potential use to MeteoForum include: (1) providing tools to visualize, analyse, organize, receive, and share data; (2) facilitating data access to a broad spectrum of observations and forecasts (most in real time); (3) supporting faculty who use Unidata systems at colleges and universities; and (4) building a community where data, tools, and best practices in education and research are shared.

10.4. Information content for science and research more broadly

Important resources for research and education include not only archives of data but also streams of real time data, as well as software services. Such resources should become part of DLE content. We now consider DLEs from this perspective, taking the Unidata Program Center (UPC) [UNIDATI, UNIDATP] as an example.

UPC offers software and services that enable universities to acquire and use atmospheric and related data on their own computers, often in real time or "near-real time" — that is, the data are sent to participants almost as soon as the

observations are made. The UPC's software and services are available to any college or university at no cost. Member institutions provide their own computers, network connections, human resources, and other requirements for participation, including access fees for certain data. Located in Boulder, Colorado, the UPC serves more than 150 universities. Through computer networking, Unidata participants are members of a mutually supportive "virtual community" — a nationwide group of electronically linked individuals who hold common academic interests in the atmospheric and related sciences and who share similar needs for data and software.

The nationwide Unidata community has established a cooperative network of computers running the Local Data Manager package and exchanging data in real time via the Internet. Known as the Internet Data Distribution (IDD) system, this network is linked to data sources at various locations and provides universities with various types of data at low cost. Participants in the IDD system can access numerous data streams, including information from surface and upper-air observing stations worldwide, radar systems nationwide, forecasting models run by the National Weather Service (NWS), national lightning detection networks, wind profilers, and satellite-borne images. In addition, Unidata does provide mechanisms for accessing some archived data sets and case studies, and some Unidata sites do archive data streams in raw, encoded form.

Unidata was founded in the atmospheric science domain; however many universities employ Unidata systems or data to support education and research that falls outside that domain. This reflects a national trend toward interdisciplinary education and research that Unidata will support by continuing to emphasize generality and broad utility in its systems, especially the data management components.

Descriptive information (metadata) about core data stream contents and about other data accessible through Unidata will be made accessible through special Unidata Web pages. Part of this objective will be to advertise and link to Web pages created by the NWS and others, such as those describing the new Geostationary Operational Environmental Satellite (GOES) sensors and the quantities they measure.

Just as the Web and digital library technologies have simplified the process of publishing and accessing multimedia documents, the Thematic Realtime Environmental Distributed Data Services [THREDDS] will provide needed infrastructure for publishing and accessing scientific data in a similarly convenient fashion. These services are intended for students, educators, and researchers to publish, contribute, find, and interact with data relating to the Earth system in a convenient, effective, and integrated fashion.

THREDDS will establish both an organizational infrastructure and a software infrastructure. A team of data providers, software tool developers, and metadata experts will work together to develop a software framework that allows users to publish, find, analyse, and display data residing on remote servers. The heart of THREDDS, however, is metadata contained in the publishable inventories and catalogues (PICats). Based on XML, PICats can be created in many different ways. Educators will incorporate PICats of illustrative data sets into educational modules that also include the tools for data analysis and visualization. Indeed students will eventually be able to use PICats to point to data sets related to their research projects, just as they now use URLs to point to relevant documents. Since they are text-based, PICats can be "harvested" and indexed in digital libraries using specialized tools that make use of the internal structure and semantic content, as well as tools similar to those used by current document search engines.

Data collections are a cornerstone of environmental research and education. New levels of accessing and using data are now achievable because of evolving technologies, even as the amount and variety of Earth system data are increasing daily. Recent parallel progress in the worlds of scientific data management and education-oriented digital libraries is highlighting common needs to discover widely distributed data sets and to use unfamiliar data meaningfully with a comprehensive set of analysis tools for:

- visualizing complex, multidimensional data;
- integrating and overlaying data from multiple sources;
- gracefully handling coordinate systems, measurable quantities, units of measure, and sampling variations.

The THREDDS system has four main components (Figure 11 — compare with Figure 5). First, data-access protocols, such as DODS, HTTP, and FTP, provide Internet access to scientific data sets, using URLs to name the data sets. These existing protocols are already in wide use by data servers in the scientific community. Second, the proposed PICats provide lists of available data sets and a framework for specifying the semantics of data sets, sometimes called "use metadata". The PICats will contain data set inventory and data set description components. Third, the proposed PICat servers are distributed processes that monitor a set of PICats and provide integrated discovery services. Fourth, existing

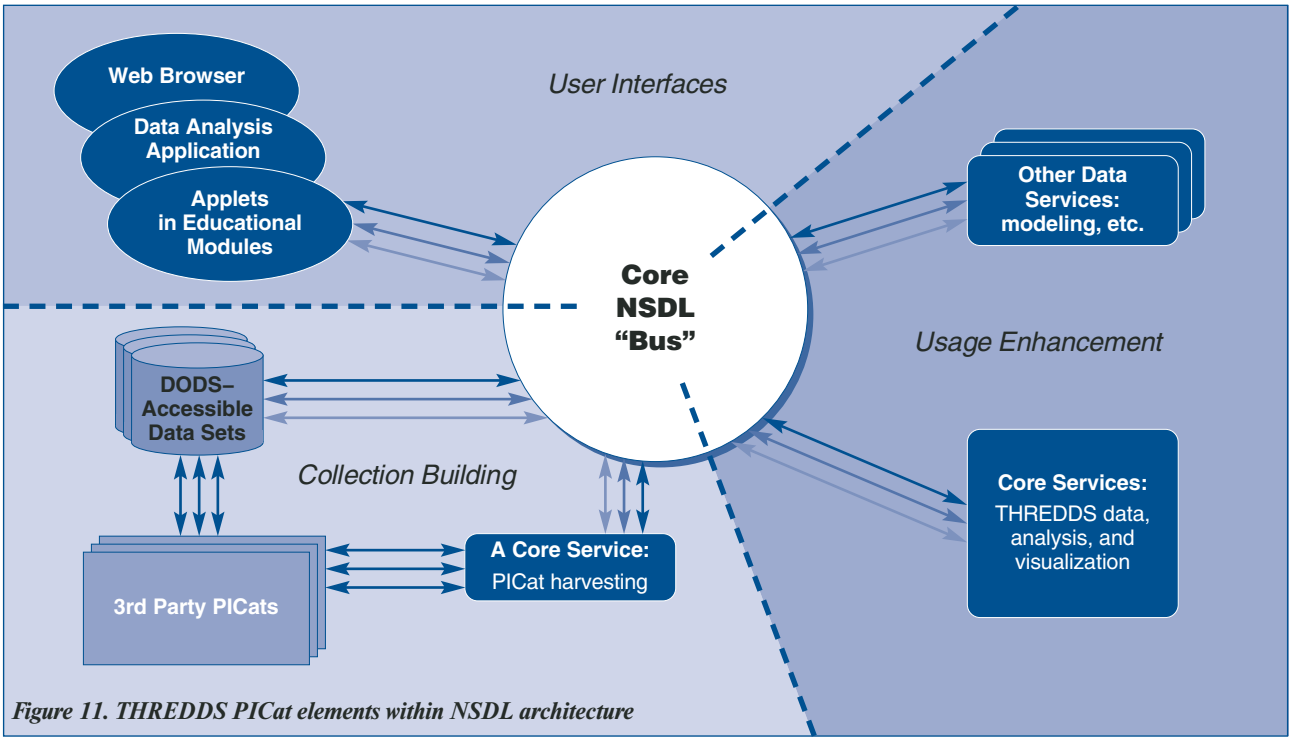

Figure 11. THREDDS PICat elements within NSDL architecture

visualization and analysis clients will be extended to connect to the PICats, PICat servers, and the data servers themselves.

The PICat server is a digital library service that enables the discovery of scientific data-set inventories, and allows searching by category, standard quantity, and possibly space and time regions. An online digital library can include a user interface (or portal) to the THREDDS PICat Server, allowing users to search, discover, and download scientific data sets from a browser.

10.5. Cyberinfrastructure

The term "cyberinfrastructure" [CYBERI] was recently coined by the US NSF management to connote not only advanced scientific computing but also a more comprehensive infrastructure for research and education based upon distributed but federated networks of computers, information resources, online instruments, and human interfaces. It provides a convenient way to talk about infrastructure based on information technology (IT), in contrast to more traditional science infrastructure.

In combination, IT capabilities will in combination allow access to complex services as well as raw computing resources through the network, enabling both collaboration and sharing over distance and time.

This picture is consistent with the vision of the Grid,[8] the modern Internet, distributed computing, and collaboratories. It is the basis for what some are calling *e-science*. A schematic of such cyberinfrastructure services is shown in Figure 12.

An environment is envisioned in which raw data and recent results are easily shared, not just within a research group or institution but also between scientific disciplines and locations. There is an exciting opportunity to share insights, software, and knowledge, and thus to reduce wasteful re-creation and repetition. Key applications and software that are used to analyse and simulate phenomena in one field can be utilized broadly in other fields. However, this will only take place if the standards and underlying technical infrastructure are accessible by all.

[8] More recently the term Grid has evolved to mean a more comprehensive structure linking people, information, and tools/facilities as indicated in Figure 12. The term xGrid, where x can be a discipline or a place, is now coming into use, e.g. the BioGrid or the MGrid (Michigan Grid). In this sense the terms "Grid" and collaboratory represent similar ideas.

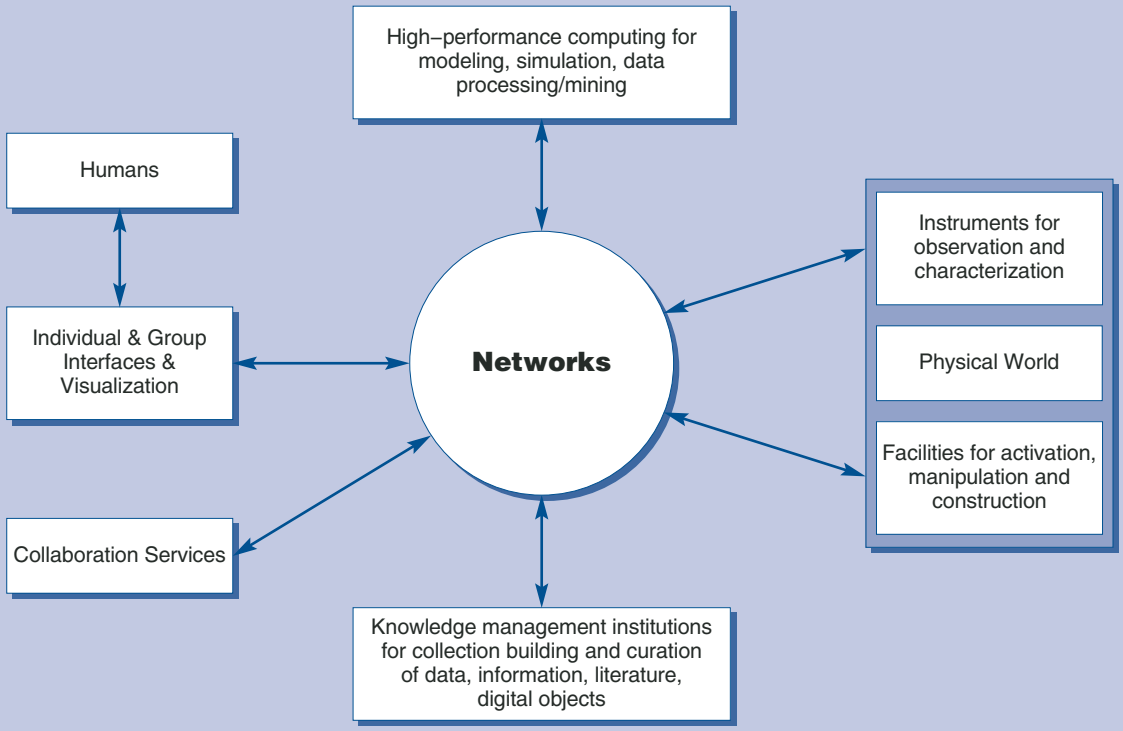

Figure 12. Schematic of services for cyberinfrastructure to support research

The Grid, built around the Internet and Web, is an infrastructure designed to provide scalable, secure, high-performance mechanisms for discovering and negotiating access to remote resources. Ultimately, it should allow scientific collaborators to share resources on an unprecedented scale, and allow geographically distributed groups to work together very effectively.

The new infrastructure is to be applied to educate the next generations of scientists using the best techniques, and to ensure broader participation without respect to field boundaries, institutional wealth, personal origin or bodily ability. It will maximize international collaboration and resource sharing through standardization and networking.

There are also significant risks and costs:

- Researchers in different fields and at different sites will adopt different formats and representations of key information, which will make it forever difficult or impossible to combine or reconcile their information.
- If no decision is made to curate and store raw and intermediate research results indefinitely along with polished and final publications, irreplaceable data will be lost.
- Effective use of cyberinfrastructure can break down artificial field boundaries, while differing tools and structures can isolate scientific communities for years.

Collaboration among disciplines is growing at an unprecedented pace and now includes, in some cases, hundreds of scientists working on a single project across the globe. Cyberinfrastructure must support this type of collaboration in a reliable, flexible, and cost-effective manner.

A significant need exists in many disciplines for long-term, distributed, and stable repositories for data and metadata that institutionalize public-domain data holdings. These repositories must provide tutorials and documentation on data format, quality control, interchange formatting, and translation, as well as tools for data preparation, fusion, mining and knowledge discovery, and visualization. A key element associated with filling this need is the development of middleware and related data storage strategies.

More and more disciplines are expressing a compelling need for nearly instantaneous access to selected databases (both local and distributed) and related services, particularly because such access often drives the actual collection of the data.

Users are expressing the need for nearly instantaneous access to real time data streams from observing platforms or computations, where such information feeds prediction models and decision support tools used in time-critical decision-making. Users also require an ability to control complex instruments remotely with exceptional network quality of service. For example, the National Virtual Observatory (http://www.us-vo.org/nvo-proj.html) links astronomy with cyberinfrastructure in the forms of Grid computing and federated access to massive data collections. A similar Grid-based approach enables neuroscientists to remotely control high-energy electron microscopes that are located a continent away: cf. the Telescience project at http://www.npaci.edu/envision/v16.2/telescience.html.

10.6. Data grids

One of the aims of the Grid is to promote the open publication of scientific data. If this is realized then it is expected that many of the Grid advances will come from applications that can combine different information about a single entity to gain a more complete picture of that entity, and to aggregate similar information about different entities. Achieving this will require support for integrating data from multiple data sources.

A good example of a data grid involves the Storage Resource Broker (SRB) [SRBDLE] that has been developed at the San Diego Supercomputer Center (SDSC) to provide "uniform access to distributed storage" across a range of storage devices, via a well-defined API. The SRB supports file replication, and this can occur either off-line or on-the-fly. Interaction with the SRB is via one of several APIs: the end-user can use a client-side GUI[9] or a standard Web browser.[10] SRB servers are managed by an administrator, with authority to create user groups. A key feature of the SRB is that it supports metadata associated with a distributed file system, such as location, size, and creation date information. It also supports the notion of application-level (or domain-dependent) metadata, specific to the content and not generalizable across all data sets. In contrast with traditional network file systems, SRB is attractive for Grid applications in that it can handle very large volumes of data, which can transcend individual storage devices, because of its metadata awareness and capabilities, and many other features, such as file replication, authentication, etc.

Features of SRB include:

- A metadata server (MCAT) that holds information on the data, users, and resources managed by the SRB: it can also be used to hold application-specific metadata. However, a limitation is that there appears to be no general mechanism for connecting MCATs into a hierarchy (for example, to allow the scalable federation of servers).
- A logical naming scheme for data sets: the mapping from logical name to the physical file is done automatically when a data set is accessed.
- A federation facility that allows a set of SRB servers to offer a single interface to clients.
- Authentication can be through the Grid security infrastructure.

As well as files, the SRB can manage data stored in relational databases. SRB looks to be a possible candidate to support THREDDS functionality [THREDDS].

10.7. Curriculum–based interfaces for DLEs

10.7.1. Project 2061: Atlas of science literacy

Project 2061 is a long-term effort of the American Association for the Advancement of Science (AAAS) to reform primary and secondary education so that all citizens attain science literacy — that is a basic understanding of the natural and social sciences, mathematics, technology, and their interactions [ATLAS]. The project is creating a coordinated set of tools and services — books, CD-ROMs, online resources, and professional development workshops — that educators, parents and families, and community leaders can use to make meaningful and lasting improvements in teaching and learning for all students. Thus the project supports education standards.

Science literacy should be approached not as a collection of isolated abilities and bits of information, but as a rich fabric of mutually supporting ideas and skills that must develop over time. From primary school to high school, what students learn should build on what they learned before, make sense in terms of what else they are learning, and prepare them for what they will learn next. To help students achieve science literacy, educators need to see how the ideas and skills that students learn in different grades and topics — and even disciplines — depend on and support one another. Atlas

[9] http://www.npaci.edu/dice/srb/inQ/inQ.html

[10] http://www.npaci.edu/dice/srb/mySRB/mySRB.html

[ATLAS] depicts this pattern of connections in a set of "strand maps" that provide a graphic representation of students' growth of understanding. Each map displays the ideas, skills, and connections among them that are part of achieving literacy on a particular topic, showing where each step along the way comes from and where it leads. "Strand" denotes identifiable concepts or stories that are developed in groups of benchmarks across different grade levels.

The maps in Atlas do not prescribe a particular curriculum or instructional strategy. Instead, they present a framework meant to inspire a variety of different ways to design and organize learning experiences suited to local circumstances. The maps in Atlas are built from benchmarks – the learning goals presented in the AAAS's Benchmarks for Science Literacy.

The horizontal lines in each map (see Figure 13 for an example) delineate the US grade ranges (where K-12 covers US primary and secondary education levels) in which most students should be able to achieve particular benchmarks. In each map, a few principal strands are pointed out to help the reader find things in the map and get a sense of its content. These strands are labeled along the bottom of the map (Figure 13). A line between boxes implies that understanding one benchmark contributes to understanding the other.

This notation is emphasized in Atlas to show that the structuring and definition of any discipline can be standardized. Each benchmark may serve as a template denoting resources in digital collections relevant to the benchmark's educational content.

10.7.2. Science literacy benchmarks as possible navigation and search interface for NSDL

A formal coding of the benchmarks as semi-structured conceptual graphs (SCGs) is proposed in [CURINT] to be used in NSDL. Each registered collection and digital object within a collection is mapped back to such SCGs. The system admits a set of mapping predicates, such as *example_of, discovery_of, description_of, explanation_of, anecdote_about*, etc. When a library holding $H is placed in the collection, the author may fill in a set of predicates, e.g. *example_of* ($H,

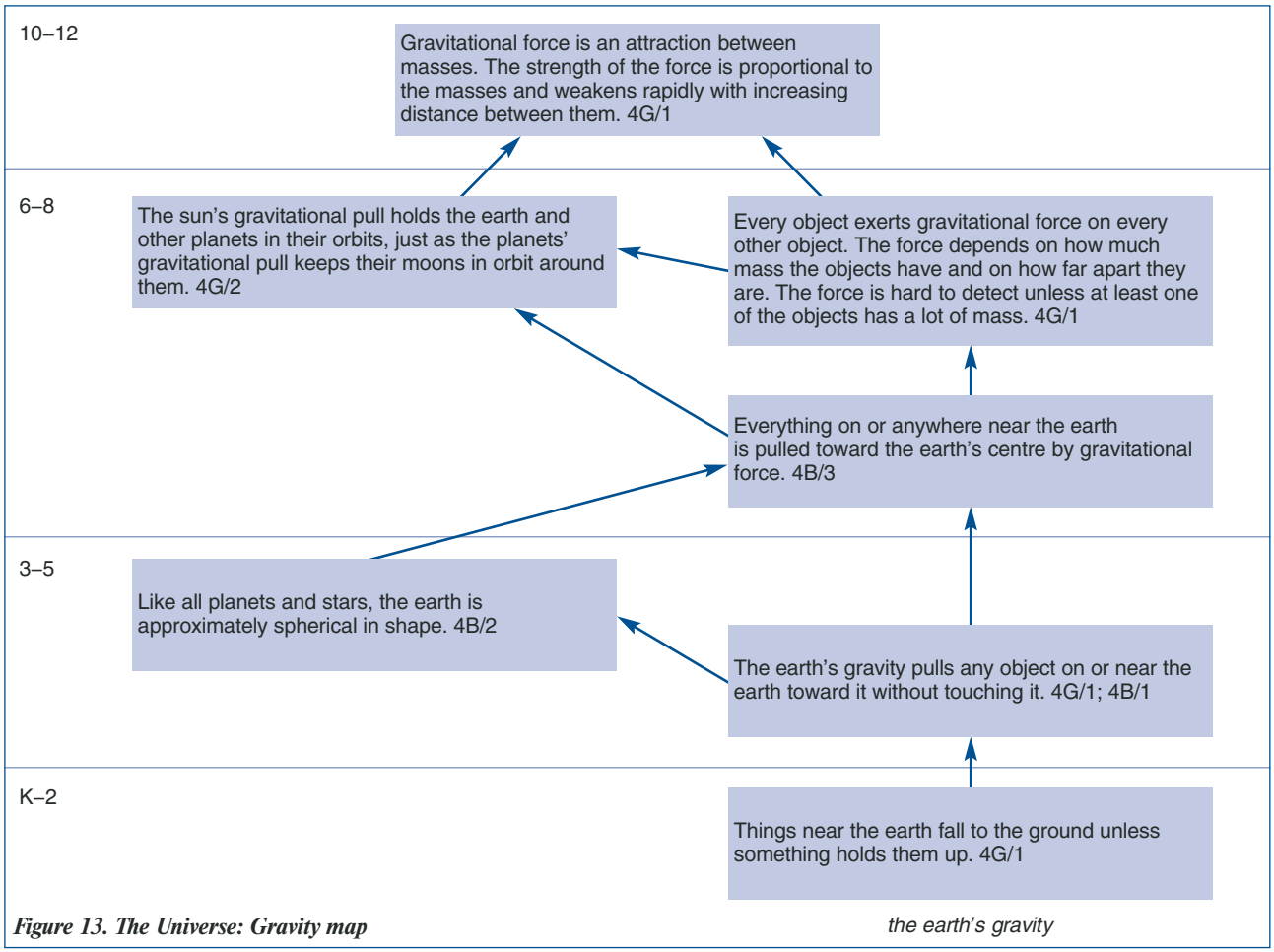

Figure 13. The Universe: Gravity map

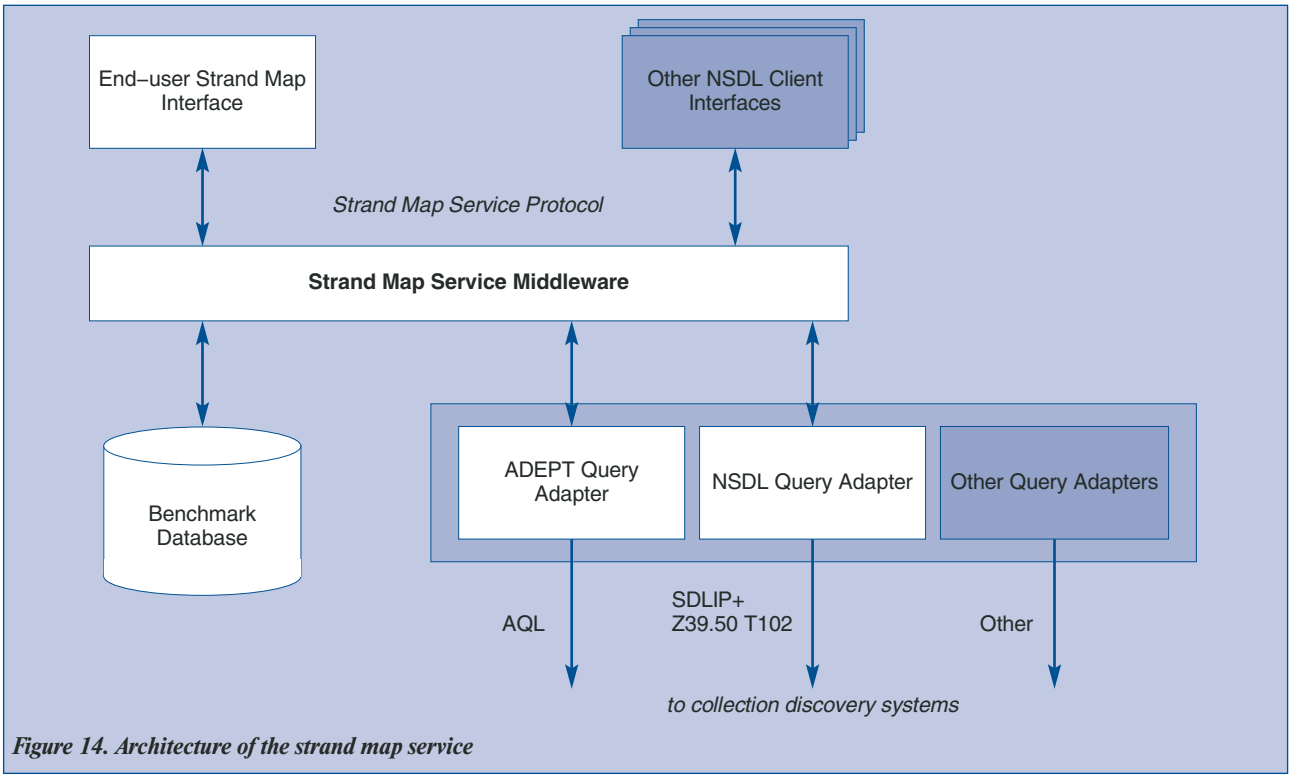

Figure 14. Architecture of the strand map service

atom), *description_of* ($H, hydrogen_atom) where the first term is a reference to the holding and the second term denotes a concept that the digital object "is about". When such SCGs are applied, it becomes possible to navigate and search in a digital library based on concept and discipline-oriented structuring of information. Note the difference between such "curriculum"-oriented structuring of a library and the conventional bibliography-oriented structuring.

Advanced applications based on SCGs are indicated in [CURINT]. Suppose an educator wants to evaluate a textbook, represented as a linear graph of chapters, where each unit is a graph of concepts. The system employs graph-matching techniques to determine whether the textbook contains certain concepts by first searching for the units touched by each chapter, and then by computing the difference between the concept set within the unit and the sets covered by the chapter. Similarly, the system can also determine whether the order of concept coverage in the textbook preserves the same dependency structure as the SCGs.

Another project Strand Maps as an Interactive Interface to NSDL Resources [STMINT], is also oriented toward discovery of educational resources that support the learning goals, or benchmarks, articulated in the strand maps. During the two-year performance period, investigators from the University of Colorado, the American Association for the Advancement of Science, the Digital Library for Earth System Education Program Center, and the University of California at Santa Barbara will work together to develop the service middleware and protocol that enable the strand map service to be integrated into library discovery systems and other NSDL services. The project includes development of other related services. One goal of this project is to provide educators with a tool to locate resources that are aligned with benchmarks through well-designed interactive interfaces and educationally tailored query mechanisms. The strand map service will be analogous to the UCSB place name gazetteer. The architecture of the service is shown in Figure 14. The basic function of the middleware (along with the other components) would be to find resources (of type X) "aligned" to a particular benchmark.

10.8. Concept–based organization of learning materials and courses

10.8.1. Concept space specifying a domain of interest

Issues of proper definition and structuring of learning materials in learning environments are analysed in [ADECOA]. The idea of access to digital collections by content is based on the premise that "scientific concepts" and "relationships between concepts" provide a powerful level of granularity with which to support effective access and use for learning.

The sets of concepts and interrelationships developed over many centuries by scientists and mathematicians constitute nothing less than the fundamental building blocks from which useful representations of reality are created, applied, evaluated, and modified.

The structure of the educational experience should reflect the structure of the concept space specifying the domain of interest. Understanding the knowledge in some domain of science requires students to understand how sets of concepts and their interrelationships are developed and applied in representing the phenomena of the domain. It looks both feasible and valuable to organize, access, and use scientific knowledge explicitly in terms of the sets of concepts that underlie some domain of scientific knowledge rather than in terms of the information objects contained in the domain.

Within the framework of existing digital library technology, it is proposed to develop:
- a concept model for representing concepts and their interrelationships;
- domain-specific knowledge bases of such representations;
- associated DL collections of "illustrative materials" concerning different aspects and attributes of the concepts;
- services supporting the creation, modification, viewing, and use of concepts for various purposes in learning contexts.

It is assumed that such an explicitly concept-based support for learning leads students to a deeper understanding of: (1) the nature, structure, and classes of the concepts that, together with the interrelationships between the concepts, provide a basis for scientific development in specific domains of knowledge; (2) the scientific roles of various classes of concepts across the spectrum of scientific activities; and (3) the global structure of some domain of scientific knowledge in terms of the underlying framework of concepts. Advantages that may accrue to instructors include the efficient reuse and repurposing of the knowledge bases of concepts and the associated collections and services in creating instructional support materials.

There is a growing consensus that science education should be a meaningful learning activity in which students learn to think like scientists rather than solely remembering information. In [ADECOA] concepts are considered to belong to one of three classes:
- *Abstract concepts:* have operational semantics defined in terms of syntactic (or computational) manipulations of symbolic representations. Three possible subclasses are syntactic (linguistic) concepts, logical concepts, and mathematical concepts.
- *Methodological concepts:* have a semantics defined in terms of the various classes of scientifically well-defined operations that may be carried out in relation to them. Possible subclasses of this class include concepts relating to procedures for scientific observation, scientific information seeking, scientific data gathering and interpretation, scientific hypothesis construction, theory construction, and scientific communication.
- *Concrete concepts:* have a semantics defined in terms of scientifically well-defined operations that provide the concept with an interpretation of defined phenomena in the domain of scientific knowledge. The concept of river discharge, for example, has a characterizing set of operations that defines the concept in terms of various sets of measurement procedures that may be carried out in real-world contexts. An example of such a procedure is one to determine the amount of water passing across some cross section of a river during a given interval of time. These concepts are, by and large, the class of concepts used in model and theory construction.

10.8.2. ADEPT concept–based architecture

The Alexandria Digital Earth Prototype (ADEPT) is being extended with a knowledge representation system that includes a knowledge base of concepts, a collection of associated materials, and a set of services for use in supporting an introductory course in physical geography at UCSB in fall 2002. The ADEPT concept-based architecture [ADEVLE] is shown in Figure 15.

The instantiation of this architecture involves:
- *Constructing a knowledge base of concepts:* Given the existence of a set of concepts for some domain, a knowledge base of these concepts is constructed by taking the abstract concept model as the basis for an XML schema from this model, and creating an XML record for each concept.
- *Constructing a collection of information* illustrating specific facets of the concepts: creating a DL collection of information illustrating the facets of some concept (such as representations of the measurement operations for some measurable concept) involves cataloguing the item as part of an ADEPT collection. This involves using the ADN metadata standard for learning objects, with an extension that involves a field for the concept name(s) and facet(s).

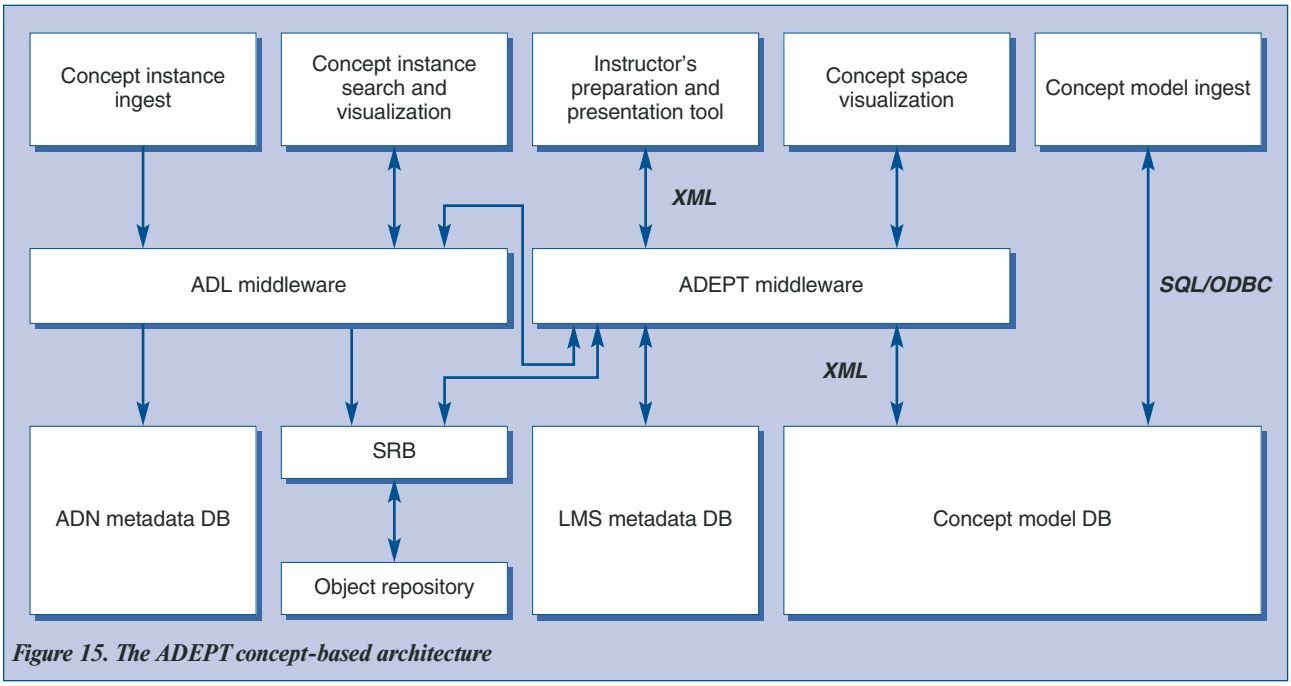

Figure 15. The ADEPT concept-based architecture

- *Creating services* that provide access to the knowledge bases and collections: Three sets of services over the concept knowledge bases and associated collections are planned: services for manipulation of entries in the concept knowledge base and the collection of illustrative items; services for searching the concept knowledge base and the collection of illustrative items; services providing both textual and visual views of the concept knowledge base and collections.
- *Developing a client that provides access to the knowledge base, collections, and services* from the perspective of the general user: This client includes functionality supporting the following operations on both the knowledge base and the collection: (1) creation and modification; (2) access; and (3) the creation of views, and the use of the material for various learning purposes.

10.8.3. Physical geography course in a concept–based architecture

The course on physical geography being developed may be viewed as a set of subtrajectories relating to each of the major topics covered in the course and corresponding to the chapter organization of a typical textbook. All of the highest-priority concepts are being modeled with the use of an XML schema. The collections of associated materials that illustrate the various facets of each concept model are being created as regular ADEPT (multimedia) collections, accessible by the operational search middleware services of ADEPT. The metadata created for each item follows the ADN metadata content standard for instructional materials augmented with an additional field for the relevant concept(s) and their facets.

The services associated with the knowledge base and collection include services for creating a trajectory through concept space that may, for example, form the backbone of a session constructed by the instructor in relation to the following high-level lecture plan:

- The instructor motivates the session with collection items containing (typically) implicit representations of the phenomena encompassed by the concepts of hydraulic geometry. These may include, for example, aerial photographs of rivers taken over one significant portion of their length and representing (implicitly) aspects of the concept of downstream hydraulic geometry. This concept refers to the changes in the discharge, width, average depth, velocity, etc., of a river as it flows downstream. Items may also include videos taken at a cross-section of a river that show the changes in the discharge, width, average depth, velocity, etc., of the river over time in response to a rainfall event.
- The instructor begins a discussion of the concepts that might be useful in constructing empirical and theoretical representations of hydraulic geometry. All such concepts may be found by constructing from the knowledge base a view of all the concepts that relate to the topic of hydraulic geometry, including the concepts of width (w), average depth (d), average velocity (v), and total discharge (Q) at a cross-section of a river channel. The students might be

led to develop specific models of concepts not previously specified in any detail. Using items from the collection, for example, the instructor may lead the students to a representation of a river as a sequence of cross-sections. The instructor may also encourage development of explicit representations for cross-sections in terms of the product of width and average depth.

- The instructor may then interact with the students in establishing hypothetical relationships between concepts (i.e. hypotheses) that may be considered not only to be important characterizations of hydraulic geometry but, if verified, to provide (partial or full) objective representations of the concept of hydraulic geometry. The instructor, may, for example, provide the students with data sets from the collection that can be used for constructing and testing hypotheses about the mathematical relationships of concepts.

A prototype interface is being implemented as a three-screen (three-window) system, including the lecture window containing a trajectory through the concept space, the knowledge window reflecting the concept base, and the collection window of selected resources that illustrate presented concepts.

11. Summary

1. Digital libraries have become a core ingredient, a collective memory of the educational environments of today and of the future. Hybrid libraries have already become widely used components of many universities around the world.

2. In several countries (USA, UK, Germany) a national digital library for education in science, engineering, and technology is being developed as an important ingredient of the national educational infrastructure. While the development of a DLE is a continuous process of collecting, classifying, conceptualizing, and using information, the process is paralleled by rapid technological advancements. Together, these two developments lead to the evolution of DLE frameworks and methodologies for their application.

DLEs in each country are dependent on the language used by it's educational community, the national culture, and the national traditions in education. This should be taken into account for the globalization of DLEs.

3. The development of national DLEs requires the involvement of various groups in society. Besides educators and learners, the community includes members of professional societies, information providers, researchers, and representatives of industries. Interrelationship of the interests of the community members should be addressed, in order to develop a reasonable strategy leading to DLE sustainable development and gradual evolution. This process leads to formation of a wide community around the DLE, providing for development, governance, collecting of information, and use in education. Large DLEs should function as nonprofit organizations. Experience from NSDL as a national project is expected to show how such challenges can be met.

Building sustainable and scalable communities where education-oriented best practices and information are shared (national and global communities, domain- or discourse-oriented communities) requires specific attention and organizational measures during DLE development and evolution.

If DLEs are to succeed, there must be serious social change, at least, in two important aspects: community building and development of a sharing culture. In particular, such change should result in getting educators working with researchers and with professional societies, educators who are willing to share education resources as well as their time (through such activities as development of reusable instruction modules, mentoring, reviewing, and advising). Educators must feel empowered to reuse resources, with community support, if required. New schemes for rewards and policies for promotion and tenure must be developed. New outlets, like the ACM *Journal of Educational Resources in Computing*, which is one of various related efforts to motivate people to share high-quality resources, must be created.

4. Organizational frameworks and interdependencies of components of the DLE infrastructure are quite complicated. These interdependencies impose serious challenges on the decisions taken during DL development. For example, users of existing specialized course modules may rely on software products developed by industries or military organizations. If such courses are to be included in the national DLE, the issues of availability of such products for the academic environment must be resolved. MeteoForum (see sections 10.3, 10.4) is an example of an organizational structure required to solve such problems.

A reasonable compromise must be found between the approach of selecting only high-quality education-oriented objects for DLEs (as in DLESE) and the approach of collecting a broad class of materials including millions of objects (the NSDL strategy).

5. DLEs in science include not only textual and multimedia information, but also real time data — the results of measurements provided by specialized organizations on the national level or globally. DLEs also provide access to expensive scientific instruments (e.g. electronic microscopes, telescopes) and to specific services. Such possibilities create efficient conditions to involve learners in research in the early stages of their education.

6. DLEs should provide various services, such as cataloguing, archiving, selective dissemination of courseware and other instructional materials developed internationally, annotation, evaluation, cross-lingual search and retrieval, personalization, recommendation, instructor support, and copyright management.

7. A gap exists between the technological and informational possibilities of existing DLEs and the availability of systematic reports by teachers and learners about actual experience using them. During the work on this survey, we found some success stories that have been collected at the University of Michigan, USA, showing how "collective memories" can be successfully used in a classroom. Analysis and evaluation of the real practice of DLE use require specific attention from DLE governance bodies and from the respective funding organizations. Systematic evaluations are still required to analyse the effectiveness of the new models of learning environments and new pedagogical methods built on DLEs (e.g. being student-centric, enabling group work on real world problems, customization of educational programmes for different students, interaction of learners with the required subject domain, curriculum-based access to learning objects, course design with reuse of existing educational modules, etc.).

8. Alternative approaches (visions) for DLE frameworks.

Existing DLE architectures, technologies, and methodologies for using them have not yet become mature. DLE frameworks start simply, being based in metadata and organized according to the metaphor of a conventional library. With time and experience, these frameworks are expected to evolve into more knowledge-based systems using conceptual definition of subjects, curricula, and ontologies introduced (in parallel with the metadata layer) for the subject definitions in various areas of teaching (e.g. see sections 10.7, 10.8).

According to the framework based on the metaphor of a conventional library, a DLE as a collective memory can be considered a container extending the conventional library (cataloguing) practice. In this case the granularity of the memory is at the level of "information entities" — electronic versions of books, journal articles, images, and videos. Metadata schemas support retrieval focused on information entities (as in the conventional tradition of library bibliographic cards), not on subject structuring and the respective granularity of retrieved items. Such an approach looks reasonable at least because of the large heritage of traditional information entities and the significant difficulty in getting access to proper information items by content.

On the other hand, analysis of DLEs and specifically of DLEs for different branches of science shows that information in such libraries (acting as collective memories) should be structured differently. Textbooks and courses are not good information entities any longer. "Bibliographic cards" are not suitable for information discovery. Educational domains in different branches of science should be properly structured. More suitable entities would be concept spaces, theories, models, hypotheses, experimental results and measurements, curricula, and educational modules. Scientists have spent centuries to reach well-defined structures, concepts and theories in the various branches of science. These definitions cannot be used following the conventional library metaphor, but are more suitable as a guiding principle for information structuring and search in DLEs (see sections 6.3, 10.7, 10.8).

For this reason, the gradual evolution of DLEs from the current framework based on the metaphor of a conventional library to more knowledge-based organization is expected. With time and experience, these frameworks will be upgraded with conceptual definitions (ontologies) of subject domains and curricula along with the conventional metadata so that information resources can be registered in accordance with the proper subject definition and granularity. This trend will also lead to a higher level of coherency of the information collected in a specific subject domain, by contrast with metadata use, where collected materials are more diverse though less relevant to the subject.

There is a discussion that DLEs need to be constructed to facilitate access by both humans and computer, i.e. must be semantic and support cross-resource and cross-domain searching and harvesting. To allow for best dissemination practices, the reusability of resources, and harvesting at semantic levels beyond OAI, current research, such as the work related to the semantic Web (e.g. DAML/OIL and OWL), should not be overlooked.

Several directions of research and development surveyed in this report indicate that such evolution has already started.

9. Current DL technologies constitute a combination of well known database, information retrieval, and web techniques. Some promoters of DL technologies often prefer not to notice that they really reuse well known techniques for DL purposes. Just a few examples:

- Metadata registries are often just databases with schemas defining structures of the respective metadata.
- OAI is similar to some well known data-warehousing techniques.
- "Subject mediator" technologies aim at capturing community agreements on data structures, thesauri, and ontologies for specific subject domains in a meta-information database. Such definitions are in a sense analogous to the registries of metadata elements, though they are much more advanced. Subject mediators are designed to

convert and reconcile a diversity of structures, terminology, and ontology of multiple data sources relevant to the subject domain with those of the mediator, as well as to provide a uniform query interface to these data sources registered at the mediator.
According to this approach, the subject domain model is to be defined by the experts in the field independently of relevant information sources. This model may include specifications of data structures, terminologies (thesauri), concepts (ontologies), methods applicable to data, and processes (workflows) characteristic of the domain. These definitions constitute the specification of a subject mediator. The process of developing a specification for a subject domain in the respective community is called mediator consolidation. After the definitions are consolidated, an operational phase of the mediator starts. During the operational phase, information providers can disseminate their information for integration into the subject domain independently of each other and at any time. To disseminate, they should register their information sources at the mediator. The registration assumes contextualization of sources in the mediator (reconciling of terminological and ontological differences) and defining source structures and behaviors in terms of the mediator schema.

If we think about what DC is or what application profiles are or what Z39.50 profiles are or what the Alexandria metadata schema is (many examples of various metadata schemas can be given), we easily see that they specify "schemas" of sets of attributes characterizing information resources in respective domains. These schemas were the result of consolidation in special working groups, conferences, and meetings held in the respective communities. In the process of registration of a source at the mediator, the differences in mediator and source terminology, ontology, and structuring are to be reconciled. Converting, harmonizing, and maintaining element sets of different metadata registries constitute a simplified instance of this process.

This observation is important: mediators are a more general technique than metadata registries. Mediated schemas are more general than schemas of metadata registries. Mediator technology can support a range of new DL technologies that are emerging (see above); metadata registry technologies cannot.

- The search bucket mechanism [ADEPTA] is just a specialized mediator designed for a concrete schema defining "collection-neutral search buckets". More can be reached in the framework of a Local-as-View (LAV) mediator technique. In LAV a mediator's schema is a schema of a subject domain. The subject domain schema is agreed to by the domain community independently of the existing collections. Each relevant collection can be registered at the LAV mediator as a materialized view. Query rewriting algorithms exist to transform mediator queries into query plans above the registered collections. Some benefits a general mediator architecture can bring, when compared to buckets, are:
  a) More flexibility in developing geo-referenced mediators (instead of rigid bucket schema, any geo-referencing mediator can be defined and implemented).
  b) The possibility of introducing data mediators (not only a search mediator) over various geo-referenced collections.
  c) Buckets in their current form probably should be treated as personalization over some mediators.
  d) General collection registration tools can be reused. An object-oriented data model can be introduced into the ADEPT architecture. Query rewriting algorithms of the mediator are more powerful and sound than those of the current bucket search facilities. Extensions for semantic mediation are playing an increasingly important role in federations of scientific data sets, and it can be expected that they might do so in DLs as well. Essentially, semantic data mediation can be seen as an extension of database mediators and query rewriting, supported by ontologies and other knowledge representation techniques.

The observations considered in item 9 lead to a conclusion that tighter collaboration with specialists in knowledge systems and mediation systems is required for developing DLEs and related technologies.

10. The subject of Digital Libraries for Education is too broad to cover exhaustively in one survey. Through consideration of several carefully chosen projects, this survey report attempts to concentrate on advanced topics, to analyse the current state of the technology, and to foresee the probable directions of its forthcoming evolution. The report could not consider every educational discipline, mostly concentrating on education in the natural sciences and engineering. Thus the specificity of DLEs for many other disciplines needs to be investigated further. Geographically, the report is based on information produced mostly in the USA and Europe. For completeness, the DLE development programmes in the rest of the world need additional serious analysis. Even collecting information about the state-of-the-art in different countries is difficult, because of the insufficient level of information available and the diversity of the presentation languages.

Several important issues were not sufficiently analysed in this survey and will require separate discussion:

- Sustainability and economic issues are crucial for DLE development. What are the visions here? What are the approaches? What may be funded by governments, and what by private companies or organizations? What will result from volunteer efforts undertaken in connection with educational institutions, or in connection with professional societies? How can the numeric dominance of students be leveraged to have their efforts lead to benefits (through sharing of theses, dissertations, technical reports, portfolios, etc.)?
- Preserving the national language as well as a cultural and historical identity in the education and globalization of DLEs. For example, is a move toward an international DLE reasonable? Or is it better to have national DLEs that somehow interoperate? If so, who would help guide and ensure interoperability? The argument that each nation needs a DLE is not well justified. Can every nation afford one? Why not have regional DLEs? Why not have DLEs for language groups? Are differences in language really so important that we must have different DLEs when content may be in different languages? What about countries like India where there are many languages? What of the fact that mathematics education is less tied to natural language because the language of mathematics is used?

11. Digital libraries are becoming a core ingredient, a collective memory, of the educational environments (global, national, university or domain-oriented) of today and of the future. A main conclusion of the survey is that to provide a competitive education, different countries should establish their own DLEs (e.g. as a national DLE, collaboratively with other DLEs, or as a regional DLE). They cannot passively wait until suitable global digital educational content is formed. The digital content of DLEs remains dependent on the language (or language groups) used by the educational community in each country, as well as the culture and national traditions in education. A significant amount of time is required to form the national community around DLEs, collect the DLE content, and educate specialists to develop, maintain, and govern DLE.

On the other hand, a DLE is distinguished from other ICTs applicable to education (e.g. multimedia, distance learning) by several important features:

- To establish a DLE (after technology is installed as software components), serious efforts are required to collect (harvest, integrate, gather, register) the digital resources, and to maintain and continuously extend them. If the digital content is not completely borrowed from another DLE, this process requires specific organizational efforts and investments. It is not a task that can be done by a separate individual (the way an educator can individually establish and use multimedia technology preparing courses). Governance, maintenance, and a community must be arranged around the DLE to make it sustainable.
- To make a DLE useful, additional efforts are required to provide for preserving the proper quality of the digital content. This is also not an individual effort. Various organizations in society must be involved in the process of creating digital content of the required quality.
- To make information in digital form widely available requires supporting rights of access and use, including copyright, preservation of the integrity of the document, licensing, and payment for use.
- In DLEs with digital content a wide set of interrelated services require administration and development.

DLEs are complex technologies that will require specific considerations and planning to provide for their deployment (compared to other, more compact technologies).

Acknowledgements

This analytical survey is devoted to recent research and development in the area of Digital Libraries in Education. Because this is very active area of research, it was not possible to reflect all the related works in one survey. The cited works are listed in the References. We would like to thank the community of authors of these publications and apologize to them in the likely event that we have misinterpreted some of their ideas.

The authors express their deepest gratitude to the experts in the field who read the survey draft and provided detailed remarks and suggestions that helped to improve the survey: Terry Smith (UC Santa Barbara, USA), Lillian N. Cassel (Department of Computing Sciences, Villanova University, USA), Bernd Wegner (TU Berlin, Germany), David Casal (University of East Anglia, UK), and Efim Kudashev (SRI RAS, Russian Federation).

Specific gratitude the analytical survey Working Group expresses to Mary Marlino for the arrangement of the technical editing of the text of the analytical survey in Boulder. We appreciate the contributions of the National Science Digital Library Core Integration Office, the Digital Library for Earth System Education, and especially the services of Marie Boyko of the University of Colorado in the preparation of this final manuscript. The editing contributed a lot to the quality of the survey text.

References

[ACRA98] President's Information Technology Advisory Committee Interim Report to the President, August 1998, National Coordination Office for Computing, Information, and Communications 4201 Wilson Boulevard, Suite 690, Arlington, VA 22230.

[ADECOA] Structured models of scientific concepts as a basis for organizing, accessing, and using learning materials. Terence R. Smith, Marcia L. Zeng, ADEPT Knowledge Organization Team. JCDL, 2002.

[ADEDLE] ADEPT/DLESE/NASA. ADN joint metadata content model. http://www.dlese.org/Metadata/, 2001.

[ADEPRO] The Alexandria Digital Earth Modeling System (ADEPT). NSF Proposal. http://www.alexandria.ucsb.edu/adept/proposal.pdf, 1999.

[ADEPTA] Greg Janée, James Frew. The ADEPT Digital Library Architecture. JCDL, 2002.

[ADEPTS] T.R.Smith, G.Janée, J.Frew, and A.Coleman. The Alexandria digital earth prototype system. In Proceedings of the Joint Conference on Digital Libraries. June 2001.

[ADEVLE] Michael Freeston. The ADEPT Virtual Learning Environment. Talk at the SDSC Seminar on DLE: *Bringing Scientific Databases to the Classroom*. http://kbi.sdsc.edu/events/DLE-06-02

[ALEXBI] Alexandria digital library: Publications, research papers, current bibliography. http://www.alexandria.ucsb.edu

[ATLAS] Atlas of Science Literacy. Project 2061. AAAS. Washington, DC, 2001.

[CASSEL] L. Cassel and E. Fox. Introducing the ACM Journal of Educational Resources in Computing (JERIC), editor-in-chiefs' introduction. 1(1), March 2001, http://doi.acm.org/10.1145/376697.382399

[CBISDT] Dmitry O. Briukhov, Leonid A. Kalinichenko. Component-based information systems development tool supporting the SYNTHESIS design method. Proceedings of the conference on Advances in Databases and Information Systems (ADBIS), LNCS, Springer Verlag, 1998.

[CLOIDP] David A. Wiley. Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. The Instructional Use of Learning Objects, online book, http://www.reusability.org/read/

[COMET] Cooperative Program for Operational Meteorology, Education and Training, http://www.comet.ucar.edu/

[COMIDP] COMET Instructional Development Process Outline, http://www.comet.ucar.edu/

[COMPLO] Jean R. Laleuf, Anne Morgan Spalter. A Component Repository for Learning Objects. A Progress Report JCDL '01, June 24-28, 2001.

[CONV93] Laurillard, D. Rethinking University Teaching - a framework for the effective use of educational technology, London, Routledge, 1993.

[COVB99] Borgman C.I. What are digital libraries? Competing visions. Information processing & management, 35(3), 1999, 227 – 243.

[CRIM] Edward A. Fox, Rachelle S. Heller, Anna Long, and David Watkins. *CRIM: Curricular Resources in Interactive Multimedia*, in Proceedings ACM Multimedia '99, Orlando, FL, Nov. 1-5, 1999.

[CURINT] Ontology Services for Curriculum Development in NSDL. Amarnath Gupta, Bertram Ludäscher, Reagan Moore. Talk at the SDSC Seminar on DLE: *Bringing Scientific Databases to the Classroom* http://kbi.sdsc.edu/events/DLE-06-02, based on short paper at JCDL'2002 with the same title.

[CYBERI] Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. January 2003, Daniel E. Atkins, Chair, http://www.cise.nsf.gov/evnt/reports/atkins annc 020303.htm

[DCQUAL] Dublin Core Qualifiers, 2000, http://purl.org/DC/documents/rec/dcmes-qualifiers-20000711.htm

[DCEDEX] Dublin Core Education Working Group, http://purl.org/DC/groups/education.htm

[DELI97] Matija Maroli, Marko Privosnik. A distance learning application based on a digital library of courseware 1997 IFIP, Chapman and Hall.

[DELO00] Digital Libraries for Distance Education (Editorial). D-Lib Magazine, October 2000, Volume 6, Number 10.

[DLCO00] Tefko Saracevic. Digital library evaluation: Towards an evolution of concepts. Library Trends, vol. 49, N 2, Fall 2000.

[DLESEC] Mary Marlino, Tamara Sumner. Research Issues for Educational Community-based Digital Libraries: The Digital Library for Earth System Education. Paper presented at the Annual Meeting of the American Educational Research Association New Orleans, LA, USA, April 1-5, 2002.

[DLESME] Michael Wright, Mary Marlino, Tamara Sumner. Meta-Design of a Community Digital Library. *D-Lib Magazine*, May 2002, Volume 8, Number 5.

[DLESNSF] The DLESE Program Center: Providing Infrastructure for a Distributed Community Library, NSF Proposal, http://www.dlese.org/documents/proposals/DPC_Infrastructure_dist2.pdf, 2002.

[DLFS99] Digital Library Policies, Organizations, and Practices. Digital Library Federation Survey, 1999.

[DLREUSE] Tamara Sumner and Melissa Dawe. Looking at Digital Library Usability from a Reuse Perspective. JCDL '01, June 24-28, 2001.

[DLTO01] Hans Roes. Digital Libraries and Education. Trends and Opportunities. *D-Lib Magazine*, July/August 2001, Volume 7, Number 7/8.

[DNERLP] DNER Learning and Teaching Programme, http://www.jisc.ac.uk/dner/development/programmes/learnteach.html

[DNERAV] Adding Value to the UK's Learning, Teaching and Research Resources: the Distributed National Electronic Resource (DNER), www.jisc.ac.uk/pub/index.html#dner

[EDLL00] Christine Borgmann, et al. Evaluating digital libraries for teaching and learning in undergraduate education: a case study of the Alexandria Digital Earth prototype. Library Trends, vol. 49, N 2, Fall 2000.

[EGEODL] G. H. Leazer, A.J. Gilliland-Swetland, and C.L. Borgman. Evaluating the Use of a Geographic Digital Library in Undergraduate Classrooms: ADEPT. ACM Conference on Digital Libraries, 2000.

[GEOFRA] JISC/NSF Proposal: Digital Libraries in Support of Innovative Approaches to Teaching and Learning in Geography. JISC/NSF Digital Libraries in the Classroom, JISC 07/01.

[GONCAL] Marcos André Gonçalves, Ming Luo, Rao Shen, Mir Farooq Ali, and Edward A. Fox. An XML Log Standard and Tool for Digital Library Logging Analysis. In Research and Advanced Technology for Digital Libraries, 6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002, Proceedings, eds. Maristella Agosti and Constantino Thanos, pp. 129-143.

[ICADL02] E. Hoffman and E. Fox, *Building Policy, Building Community: An Example from the US National Science, Technology, Engineering, and Mathematics Education Library (NSDL)*, short paper for ICADL '2002, Singapore, Dec. 2002.

[IMEM01] *Digital Libraries in Education: State-of-the-Art Report*. Information materials prepared for the IITE expert meeting on DLE by Leonid Kalinichenko and Liang Zhang, May 2001.

[IULOOB] The Instructional Use of Learning Objects, online book, http://www.reusability.org/read/

[JISCMLE] Managed Learning Environments (MLEs) in Further Education: Progress Report, JISC Circular 7/00, July 2000. http://www.jisc.ac.uk/pub00/c07_00.html

[KERBER] J. G. Steiner, B. C. Neuman, and J. I. Shiller, *Kerberos: An Authentication Service for Open Network Systems,* presented at Winter 1998 Usenix Conference, 1998.

[LDAP05] W. Yeong, T. Howes, and S. Kille, *Lightweight Directory Access Protocol*, Internet Engineering Task Force, Request for Comments 1777, March 1995, http://www.ietf.org/rfc/rfc1777.txt

[LOMN00] Draft Standard for Learning Object Metadata, IEEE P1484.12/D5.0, 11 November 2000.

[LRBP11] IMS Meta-Data Best Practice and Implementation Guide Final Specification Version 1.1. Available at http://www.imsproject.org

[LRXB11] IMS Learning Resource XML Binding Specification Final Specification Version 1.1. Available at http://www.imsproject.org

[LTCOGN] Jonassen, David H. and Thos. C. Reeves. Learning with Technology: Using Computers as Cognitive Tools. In Handbook of Research for Educational Communications and Technology. New York: Macmillan Library Reference USA, 1996.

[LTSA01] Draft Standard for Learning Technology — Learning Technology Systems Architecture. IEEE P1484.1/D8, 2001-04-06.

[MARIAN1] Marcos André Gonçalves, Robert K. France, and Edward A. Fox, MARIAN: Flexible Interoperability for Federated Digital Libraries. In Proceedings 5th European Conference on Research and Advanced Technology for Digital Libraries, ECDL'2001, September 4-8 2001, Darmstadt, Germany.

[MARIAN2] Marcos André Gonçalves and Edward A. Fox. 5SL – A Language for Declarative Specification and Generation of Digital Libraries. Long paper in Proc. JCDL'2002, Second Joint ACM / IEEE-CS Joint Conference on Digital Libraries, July 14-18, 2002, Portland, pp. 263-272.

[MARIAN3] Marcos André Gonçalves, Paul Mather, Jun Wang, Ye Zhou, Ming Luo, Ryan Richardson, Rao Shen, Liang Xu, and Edward A. Fox, Java MARIAN: From an OPAC to a Modern Digital Library System. In Proceedings of the 9th String Processing and Information Retrieval Symposium (SPIRE 2002).

[MCCKM] E. Fox and L. Kieffer. Multimedia Curricula, Courses and Knowledge Modules, ACM Computing Surveys, Dec. 1995, 27(4): 549-551.

[METAGE] E. D. Liddy and S. Sutton. Breaking the Metadata Generation Bottleneck, 2000, http://cnlp.org/research/project.asp?recid=6.

[METFOR] Arlene G. Laing, Linda Miller, Timothy C. Spangler, Dave Fulker. MeteoForum: A pilot project for advanced international meteorological training centers, 18th International Conference on IIPS, 2002, http://ams.confex.com/ams/annual2002/18IIPS/abstracts/28315.htm

[MLEB02] Jos Boys. Managed Learning Environments, joined up systems and the problems of organizational change. A JISC Report. March 2002.

[MLEJIS] Managed Learning Environments. http://www.jisc.ac.uk/mle/reps/infopack.html

[MLEVLE] MLEs and VLEs explained. JISC, http://www.jisc.ac.uk/mle/

[NDLTD1] E. Fox, M. Gonçalves, G. McMillan, J. Eaton, A. Atkins, and N. Kipp. The Networked Digital Library of Theses and Dissertations: Changes in the University Community. In special issue on *Information Technology and Educational Change* of The Journal of Computing in Higher Education, 13(2): 3-24, Spring 2002.

[NDLTD2] Hussein Suleman, Anthony Atkins, Marcos A. Gonçalves, Robert K. France, and Edward A. Fox, Virginia Tech; Vinod Chachra and Murray Crowder, VTLS, Inc.; and Jeff Young, OCLC. Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access - Part 1: Mission and Progress, Part 2: Services and Research. D-Lib Magazine, 7(9), Sept. 2001, http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html, http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html

[NOIOBJ] Andrew S. Gibbons, Jon Nelson, Robert Richards. The Nature and Origin of Instructional Objects. The Instructional Use of Learning Objects, online book, http://www.reusability.org/read/

[NSDINT] William Y. Arms, et al A Spectrum of Interoperability. The Site for Science Prototype for the NSDL. D-Lib Magazine, January 2002, Volume 8, Number 1.

[NSDLAR] David Fulker and Greg Janée. Components of an NSDL Architecture: Technical Scope and Functional Model, JCDL'2002.

[NSDLCO] Carl Lagoze (ed.). Core Services in the Architecture of the National Digital Library for Science Education (NSDL), JCDL, 2002.

[NSDLGOI] Interim NSDL Governance Model, Adopted 12/4/2001, http://doclib.comm.nsdlib.org/InterimGovern-final-3.doc

[NSDGOS] Dave Fulker. A Short Term NSDL Governance Model. October 1, 2001, PPT file, http://www.smete.org/nsdl/workgroups/governance/ NSDLgovernance32.ppt

[NSDLME] NSDL Standards Working Group, Metadata Resources Page, 2001, http://128.253.121.110/NSDLMetaWG/IntroPage.html

[NSDLNSF] Collaborative Project: Core Integration – Leading NSDL toward Long-Term Success. Narrative, References, and Subcontractors' (Budgetary) Statements of Work for a Proposal to NSF, Submitted 6 May 2002.

[NSF996] International Digital Libraries Collaborative Research. Announcement Number NSF 99-6.

[OAI001] C. Lagoze and H. Van de Sompel, The Open Archives Initiative: Building a low-barrier interoperability framework, presented at Joint Conference on Digital Libraries, Roanoke, VA, 2001.

[ORGM81] Beer, S. (1981) The Brain of the Firm, 2nd Edn, Chichester: John Wiley, 1981.

[OSOSS] David A. Wiley, Erin K. Edwards. online self-organizing social systems: the decentralized future of online learning, 2002, http://wiley.ed.usu.edu/docs/ososs.pdf

[PATHWAY] Cathryn A. Manduca, et al. (Eds.) Pathways to Progress: Vision and Plans for Developing the NSDL. March 20, 2001, http://nsdl.comm.nsdlib.org/Comm.Center_Archive/workgroups/whitepaper.pdf

[PDLB97] Lesk M. Practical digital libraries, Books, bytes, and bucks. Morgan Kaufmann, 1997.

[RBLPSS] Michael J. Hannafin, Janette R. Hill, James E. McCarthy. Designing Resource-Based Learning and Performance Support Systems. The Instructional Use of Learning Objects, online book. http://www.reusability.org/read/

[SCCA01] Scholnet and Cyclades: Extending the Role of Digital Libraries. D-Lib Magazine, April 2001, Volume 7, Number 4.

[SDLIP] A. Paepcke, R. Brandriff, G. Janee, R. Larson, B. Ludäscher, S. Melnik, and S. Raghavan, *Search Middleware and the Simple Digital Library Interoperability Protocol*, D-Lib Magazine, 5 (3), 2000.

[SHIB01] Shibboleth, 2001, http://middleware.internet2.edu/shibboleth/

[SINHA00] A. K. Sinha, *Geoinformatics: A defining opportunity for earth science research. A white paper* http://www.geoinformaticsnetwork.org/whitepaper.html, 2000.

[SLRM01] Lee L. Zia. Growing a National Learning Environment and Resources Network for Science, Mathematics, Engineering, and Technology Education: Current Issues and Opportunities for the NSDL Program. D-Lib Magazine, March 2001, Volume 7, Number 3.

[SLWSJ98] REPORT OF THE SMETE LIBRARY WORKSHOP. Held at the National Science Foundation, July 21 to 23, 1998.

[SRBDLE] Arcot Rajasekar. Managing Digital Library Collections with the SDSC Storage Resource Broker. Talk at the SDSC Seminar on DLE: *Bringing Scientific Databases to the Classroom*, http://kbi.sdsc.edu/events/DLE-06-02

[STMINT] Strand Maps as an Interactive Interface to NSDL Resources. NSF Proposal Number DUE 0226286, 2002.

[SULEMA1] Hussein Suleman and Edward A. Fox. A Framework for Building Open Digital Libraries. D-Lib Magazine, 7(12), Dec. 2001, http://www.dlib.org/dlib/december01/suleman/12suleman.html

[SULEMA2] Hussein Suleman and Edward A. Fox. Designing Protocols in Support of Digital Library Componentization. In *Research and Advanced Technology for Digital Libraries, 6th European Conference, ECDL 2002, Rome, Italy, September 16-18, 2002, Proceedings*, eds. Maristella Agosti and Constantino Thanos, pp. 568-582.

[THREDDS] Ben Domenico. Thematic Realtime Environmental Data Distributed Services THREDDS. http://my.unidata.ucar.edu/content/software/thredds/index.html

[UNIDAT] Overview of the Unidata Program. http://www.unidata.ucar.edu/factsheets/overview.html

[UNIDATI] David Fulker, Sally Bates, Clifford Jacobs. Unidata: A Virtual Community Sharing Resources via Technological Infrastructure. Bulletin of the American Meteorological Society, 78(3).

[UNIDATP] Unidata 2003: Proposal to the National Science Foundation. Principal Investigator: David W. Fulker, http://unidata.ucar.edu/proposals/NSF2003/

[VLEEVA] Britain, Sandy and Liber, Oleg, *A Framework for Pedagogical Evaluation of Virtual Learning Environments*, http://www.jisc.ac.uk/jtap/htm/jtap-041.html

[VLEGRE] Archer, R. Virtual learning environments for distance learning in HEIs: a guide to resources. Unpublished report for IMRI, University of Nottingham, 2000.

[VLEPED] Britain, S.; Liber, O. A framework for pedagogical evaluation of virtual learning environments. JTAP, 1999, http://www.jisc.ac.uk/jtap/htm/jtap-041.html

[VLEREQ] Requirements for a Virtual Learning Environment. JISC, August 2000, http://www.jisc.ac.uk/vle

[VLETLS] Links to the VLE and MLE tools on the INSPIRAL Web site, http://inspiral.cdlr.strath.ac.uk/resources/toolssystems.html

[WTSM99] Raven Wallace. Teaching with the Web: Challenges in a Complex Information Space. Dissertation. March 1, 1999, http://www-personal.umich.edu/~ravenmw/disssumm.html

[Z203RLQ] P. Ryall, Type 203 Ranked List Query Spec, 1996, http://www.itl.nist.gov/iaui/894.02/projects/rlq/RLQ_spec_Feb96_ZIG.html

The Internet resources as at the end of 2002.

DIGITAL LIBRARIES IN SCIENCE

1. Introduction

While developments in science span hundreds of years, developments in digital libraries (DL) as we know them today, in comparison, are very recent. Many consider the DL initiatives launched in the USA in the mid 1990's to have given the major impetus for the development of DL as a field of study and application, and to have engaged worldwide interest in R&D into and the application of DL. DL is therefore just over a decade in age, even younger than the Internet and the Web. Science and DL are interdependent. Within its short period of development, DL has demonstrated its value to scientific progress by enabling improved creation, capture, preservation, organization and dissemination of scientific information. Along with the advent of the Internet, the constant desire of scientists to improve distribution and access to scientific information led many of them to develop path-breaking applications and services, for instance the WWW at CERN, the Mosaic web browser at NCSA and the pre-print distribution system in physics and mathematics at LANL. These developments, combined with initiatives, such as DLI, are showing today the way forward the development of more powerful and advanced DL systems. These systems, many still in demonstration and proof-of-concept stage, are expected to bring about significant improvements in science communication. More importantly, developments in DL are giving rise to new forms of science like e-Science: collaborative research based on shared data not possible through traditional methods of disseminating science information. DL is thus not just improving the way science communication is done, but is also causing a shift in the way science itself is done.

An analytical survey of digital libraries in science at this juncture is of significant interest and use. This report presents such a survey. It presents the state-of-the-art in the field of DL usage for scientific purposes, in particular an overview of existing theoretical, methodological and technological approaches, and it shows several examples of DL for science in different countries, especially those which can be considered leaders in this field. The survey identifies trends, challenges and problems in the development of DL for science. It also identifies recent and current policies, plans and initiatives for international cooperation that are contributing to the successful development and deployment of DLs. Lessons learnt from such plans and initiatives will be of significant use during the development and implementation of similar programmes elsewhere.

The survey has made an attempt to cover most leading science DLs. The survey also covers several major DL initiatives and collections that are not specific to science but are interdisciplinary in scope, since the theories, methods and technologies used in these DLs would be of significance across a broad range of discipline areas. These DLs have been studied by visiting them on the web and also from related published literature. In identifying technological trends and challenges, the survey has leaned heavily on select recent literature, particularly relevant international conferences. Planning and policy related surveys have been made using documents available on relevant project web sites and in related published literature.

Regarding example DLs studied in the survey, the focus has been predominantly on publicly accessible, open DLs, rather than commercial services. Further, the education and research community has been the user domain of interest in identification and selection of DLs for study. There is significant interest in national level licensing and provision of access to commercial digital library services, including e-journals, databases and e-books. We have excluded such digital libraries from this survey since such DLs have limited access and information about methods and technological approaches used in them is not easily available.

The survey has a few shortcomings. It does not aim to be exhaustive in covering all DLs in science, as such coverage would be extremely difficult and time consuming. Instead, in selecting a few leading DLs in science, the survey focuses on the theoretical, methodological and technological approaches used in these, with the assumption that this would be representative of most current developments. The survey also does not pretend to cover all relevant literature related to DLs in science, for the same reasons. Our interest has been far more pragmatic in limiting ourselves to selective relevant literature. We do hope that the report is successful in presenting a fairly substantial and purposeful coverage of current developments with pointers to the future.

In first part of the report we present several examples of DLs in science, engineering and related areas, organized in terms of the country of their origination/location. It also provides an overview of theoretical, methodological and technological approaches used in the development, deployment and usage of such digital libraries. The second part of

the report presents a review of current trends and challenges in the development of DL in science, based on example DLs covered in the first part and also on current literature. In the third part of the report, we review national and international plans, policies and programmes that have formed the basis for several successful DL projects, with a pointer to their relevance for replicating DL development in other situations.

2. Science Digital Libraries – Examples, Theories, Methods and Technologies

In this part of the survey we discuss several examples of digital libraries in science, engineering and related areas. We also provide an overview of existing theoretical, methodological and technological approaches used in the development, deployment and usage of these digital libraries. We present and discuss the digital library examples taken mostly from developed countries. These examples include both operational digital libraries and those that have been developed for the demonstration of new approaches and technologies.

2.1. United States of America

2.1.1. National Science Digital Library (NSDL)

(http://www.nsdl.org/)

2.1.1.1. Introduction

NSDL is probably the largest distributed science digital library in the world today. It is supported by the National Science Foundation (NSF) and builds on the work supported under the NSF multi-agency Digital Libraries Initiative (DLI) Phase I and Phase II (http://www.dli2.nsf.gov/).

NSDL is the National Science Foundation's online library of resources for science, technology, engineering and mathematics (STEM) education and research. It is a free online library of exemplary STEM related resources. NSDL aims to provide an organized point of access to collections and services from resource contributors that represent the best of both public and private institutions, including universities, museums, commercial publishers, government agencies and professional societies. NSDL supports teaching and learning at all levels, from preschool through adult, with materials ranging from journal articles and lesson plans to interactive animations, and from real-time data sets and technology-based tools to ask-an-expert services.

The NSDL mission and vision statements clearly spell out the aim, purpose, nature and scope of the digital library. The NSDL mission is to "...enhance science, technology, engineering and mathematics education for all Americans by developing and maintaining a library of high quality multimedia resources with many portals of online access. The NSDL embodies long-standing library traditions of service, stewardship, democratization of knowledge, rich discourse, equal access, fair use and privacy, as well as innovations that foster a spirit of inquiry." The vision of the NSDL programme is to "...enable the collection, organization, discovery and delivery of quality learning and teaching resources appropriate for educators and learners at all levels. The resulting network of learning environments and resources should provide reusable, shareable and interoperable resources that enable learners at all levels to access and use materials suited to their needs, both within and across traditional STEM disciplinary boundaries. Such materials should also include assessment and evaluation tools and findings, and should harness new understandings about pedagogy and the processes of learning that are founded on a solid research base. The resource collections, services and core infrastructure of NSDL will facilitate the development and dissemination of both new and tested materials and methods, thereby promoting continual improvements in STEM education at all levels."

2.1.1.2. NSDL collections, users and collection building strategy

NSDL primarily includes collections of resources and learning environments developed in projects funded by NSDL. It also includes non-NSDL funded collections and services, including materials from publishers and other governmental agencies. NSDL is intended to meet the needs of students and teachers at all levels – from pre-K to K-12, undergraduate, graduate and lifelong learning. It will serve both the individual learner seeking understanding and groups of learners engaged in collaborative exploration of concepts, and will support formal and informal modes of learning.

NSDL collection is a ***collection of sets of resources.*** These ***sets of resources*** are referred to as ***collections.*** Collections that form part of the NSDL Collection are not actually held within NSDL-owned computers or storage systems. Instead, individual ***collections*** typically are held and managed by their owners or providers. However, NSF asks each

site to display a prominent NSF logo and link to the nsdl.org portal site. NSDL does not aggregate and store collection resources, but instead uses metadata to describe a given library resource. The NSDL metadata repository manages two types of metadata – collection metadata, describing the overall collection, and item metadata, describing each resource. NSDL accepts simple and qualified DC (Dublin Core) and also native metadata types. Certain item-level metadata terms, like "Audience" and "Resource Type", are important for discovering and reusing a collection's resources.

A collection provider has to follow a few processes before the collection becomes accessible through nsdl.org site. These processes include:

- review of metadata design with NSDL metadata expert;
- collection registration through an online form providing the collection level metadata;
- testing the collection OAI server for compatibility with the OAI-PMH protocol for metadata harvesting; and
- after the OAI server passes, working with NSDL to get in the harvest queue.

After these processes are successfully completed, the search service creates an index using both metadata and text crawled from item's URL at publicly available sites. Ranking of results is based on several factors including the number of times the query term appears and uncommon terms that are considered "more important" which receive higher scores.

Through its distributed digital library building strategy, the NSDL programme aims to create and develop a comprehensive cyberlearning infrastructure. Features of this strategy include:

- Practices and policies for community-based review and other mechanisms that assure the quality and usability of resources;
- Practices and policies for collections management issues, such as archiving, preservation and de-accessioning;
- Articulation of standards that promote stability, interoperability and reusability of a wide variety of learning objects;
- Effective ways to handle intellectual property issues that focus on maximising the value of content;
- Login and authentication systems;
- Archiving services;
- Digital rights management systems.

NSDL provides access to a large number of STEM related digital libraries. In Table 1 we provide a sample of these collections and their scope. Most of these collections are developed and/or made available from NSDL funded projects.

Table 2. Sample NSDL supported science digital library collections and services

Collection/Service name	Collection Scope
Biology Education Online – An Interactive Electronic Journal	Creating a peer–reviewed online journal on the subject of teaching and learning biology, and a collection of reviewed digital resources aimed at biology educators at many different levels.
Atmospheric Visualization Collection	Aggregating visualizations of atmospheric data from the Atmospheric Radiation Measurement program's Southern Great Plains site, which is the world's largest group of remote sensing atmospheric instruments.
Math DL – A Library of Online Learning Materials in Mathematics and Its Applications	Developing a collection of independently reviewed learning materials for learning and teaching undergraduate mathematical sciences.
Collection and Distribution of Geoscience (Solid Earth) Data Sets	Two goals define the focus of this project: 1) collecting, organising and maintaining geoscience data sets for use in a national digital library for STEM education, and 2) developing advanced user tools to manipulate, map, model, visualisize and analyse the collected data sets. The proposed activities include development of a series of learning modules that college and secondary school teachers can integrate into pre–existing earth science curricula.
The Alsos Digital Library	This web–based collection of references to resources offers a broad, balanced perspective of topics relating to the origins, functions and legacies of the Manhattan Project. The central task of this project is to integrate these references into a structured architecture that allows users to examine this important period of history from many scientific, engineering, technological and social perspectives.

Collection/Service name	Collection Scope
Analytical Sciences Digital Library	The aim is to classify, catalogue, link and reference information or discovery material pertinent to innovations in curricular development and supporting resources in the analytical sciences.
A Digital Library of Ceramic Microstructures	The goal of this project is to create a digital library of microstructures for functional ceramics with emphasis on materials used for structural, electronic and thermal applications in undergraduate and graduate science and engineering education.
Mathematical Software Collection	This project is aggregating high–quality numerical software for science and engineering education to support a rich, highly interactive and inquiry–based learning environment.
Gender and Science Digital Library (GSDL)	A high–quality, interactive library of K–16 gender and science resources, the GSDL assists educators in promoting and implementing gender–equitable science education in both formal and informal settings.
Computing and Information Technology Interactive Digital Educational Library (CITIDEL)	CITIDEL serves the computing education community in all its diversity and at all levels. Domain areas of coverage include computer engineering, computer science, information science, information systems, information technology and software engineering.
Geotechnical, Rock and Water Resources Library	This project is promoting widespread access to quality information, resources and activities in support of learning, teaching and research in the areas of geotechnical engineering, rock engineering and water and its use.
Reciprocal Net – A Distributed Molecular Database	The Reciprocal Net project is constructing and deploying an extensive distributed and open digital collection of molecular structures. Associated with the collection are software tools for visualising, interacting with, and rendering printable images of the contents; software for the automated conversion of local database representations into standard formats which can be globally shared; tools and components for constructing educational modules based on the collection; and examples of such modules as the beginning of a public repository for educational materials based on the collection.
Electronic Encyclopaedia of Earthquakes	This project is expanding a pilot version of the web–based Electronic Encyclopaedia of Earthquakes into a major portal for students, educators and others seeking information about the science of earthquakes, earthquake engineering and the practical aspects of hazard characterization and loss reduction.
The Digital Archive Network for Anthropology (DANA)	This project is supporting a network infrastructure to link distributed databases with content of relevance to the domain of anthropology.
Virtual Telescopes in Education (TIE)	This project is seamlessly integrating telescopes equipped with remote access and control capabilities into one virtual observatory by providing the services required to operate this facility, including a scheduling service, tools for data manipulation, an online proposal review environment, an online “Virtual TIE Student Astrophysical Journal” for publication of results, and access to related educational materials provided by the TIE community.
www.eSkeletons.org: An Interactive Digital Library of Human and Primate Anatomy	This project is expanding the range of content at www.eskeletons.org, both in terms of the degree of learner interactivity with the materials, and the amount of interaction among users. New high resolution X–ray computer tomography technologies allow the inclusion of species of a much smaller body size than those initially posted at the site, and much faster completion of scans. The collection provides students with a more complete understanding of the range of primate diversity and facilitates a great diversity of lab exercises.
The Moving Image Gateway	This project is developing a web portal for moving images that combines an archives directory database with a union catalogue to provide access to the world's moving image collections.
Kinematic Models for Design Digital Library (K–MODDL)	A team of faculty and librarians are aggregating educational materials associated with the 220 late 19th–century model machine elements designed for research and teaching by the founder of modern kinematics, Franz Reuleaux (1829–1905).
A Digital Library Collection for Computer Vision Education	This project is gathering high–quality material into a comprehensive digital library collection for computer vision education. The resource links to assignments at a variety of institutions and hosts a set of vetted assignments, complete with data sets and solutions.
Math Tools Project	The Math Forum is aggregating mathematical software critical to the learning of school mathematics, including software for handheld devices, small interactive web–based tools, such as applets, and other small modules based on software application packages.

Collection/Service name	Collection Scope
Ceph School: A Pedagogic Portal for Teaching Biological Principles with Cephalopod Molluscs	Using cephalopods as model organisms, Ceph School strives to help students understand basic principles in biology, observe the methodology of scientific research, and become familiar with cephalopods.
CaREN: Career Resources Education Network for STEM	The Gender and Diversities Institute is developing and maintaining a science, technology, engineering and mathematics career development digital library aimed at engaging middle and early high school students.
Materials Digital Library: MatDL.org	Identifies and collects materials science educational resources. Submission, editing and composing tools enable experts and novices to characterize contributions to the MatDL using Dublin Core and IEEE Learning Object Metadata schemes, as well as domain–specific markup languages (ML), such as Materials ML.
Grid based Bioinformatics Digital Library	This project is applying advances in digital library technologies to the emerging domain of bioinformatics, and developing interaction tools that support learning based on identifying and visualising associations among key dimensions of bioinformatics resources.
CAUSEweb: A Digital Library of Undergraduate Statistics Education	CAUSE – Consortium for the Advancement of Undergraduate Statistics Education – is a national organization whose mission is to support and advance undergraduate statistics education in four target areas: resources, professional development, outreach and research.
The Digital Chemistry Library	This project establishes a digital collection of tightly integrated chemistry learning objects. A major focus of the project is the development of descriptive chemistry learning object vocabularies to identify analysis methods, models and context machinery.
The Tree of Life Project (ToL): A Digital Library of Biodiversity Information	The ToL is a collaborative effort of biologists from around the world. On more than 3000 World Wide Web pages, the project provides information about the diversity of organisms on Earth, their evolutionary history (phylogeny) and characteristics.

2.1.1.3. NSDL architecture

NSDL core infrastructure has been designed to contain millions of records; to offer a wide range of services; and to support long-term evolution and growth. Given the heterogeneous community of participants and technologies, the library has been developed with two key notions: *a spectrum of interoperability* and *one library, many portals*. A spectrum of inter-operability provides a low cost of entry into the library in order to support the broadest acceptance by participants. To support this notion, the initial architectural design for this library is based on sharing of human and machine-generated metadata and exploitation of that metadata for the deployment of core services (e.g. search and discovery, and archiving). NSDL users will be very diverse, and will include students, instructors, the public at all levels, librarians, NSDL federated partners and community interest groups. To best serve such a diverse audience, the library is designed to support the notion of one library, many portals. The goal is to provide many different views of the library but with user interfaces that convey the sense of a single library.

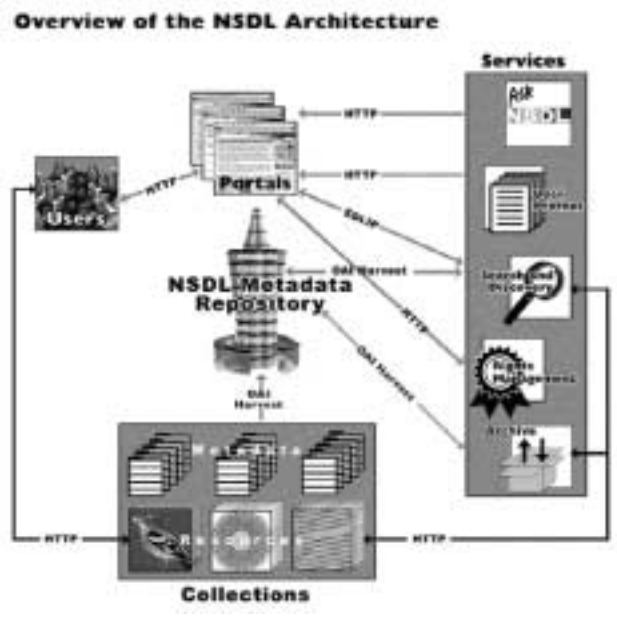

Figure 16. NSDL architecture

Figure 16 shows the principal features of the core NSDL architecture. A description of the major components follows.

Metadata Repository: The Metadata Repository (MR) is the central record of all resources known to the NSDL, and provides storage of the metadata ingested from participating projects, as well as metadata gathered from open-access web

resources. The MR was conceived as a generalization of the concept of a union catalogue, i.e. a catalogue that combines records from many libraries. The MR exports the metadata to services, such as search and browse, in both raw and normalized forms, via the Open Archives Initiative (OAI) protocol for metadata harvesting (http://www.openarchives.org).

The MR is implemented using a relational database system, wrapped with standard interfaces and protocols that hide the underlying database implementation. Metadata is ingested via an automated ingest process, which performs services, such as crosswalks to Dublin Core (DC), and any pre-processing necessary to ingest and normalize collection metadata. The OAI metadata harvesting protocol is the primary method of ingest and the XML format is used for providing and ingesting metadata.

Search and Discovery Service: The search and discovery component of the NSDL provides fundamental capabilities for locating resources and collections within the library. Search services allow any item represented in the MR to be found, but in reality the metadata provided by collections vary dramatically in formats, quality and comprehensiveness.

To address this challenge, the search service combines indexing of metadata from the MR with indexing of full text content using open network protocols (e.g. HTTP or FTP) where the content is linked via the identifier in the metadata record and is freely available. The underlying technology uses the Jakarta Lucene search engine, an open source text search engine (information at http://jakarta.apache.org/), and the SDLIP search protocol (http://www.dlib.org/dlib/march00/paepcke/03paepcke.html). Search services are directly accessible to both library portals and service providers who wish to search the contents of the MR directly.

Access Management Service: NSDL core access management must accommodate a widely varying set of requirements from the users of the library and from the providers of intellectual property, i.e. the NSDL collections and the providers of other services to the library. While many items in the library will be freely available and anonymous user access is permitted, access to some materials is restricted. Access management services include: a user Profile Server, which holds attributes associated with a user allowing portal interfaces, or other services to store and retrieve information to customize the user's experience; and a Rights Management Broker service that enforces access decisions for items in the library based on the characteristics of the user and of the item.

The core access management system relies on the Shibboleth protocol (http://middleware.internet2.edu/shibboleth/) to distribute identity verification (authentication) and cohort membership (authorization) to the administrators of distinct communities of users. In other words, the user's "home" institution performs user identity and capability management. Federated communities performing user identity and capability management can easily tie-in to this system using standard protocols (e.g. Kerberos and LDAP).

User Interface and Portals: Users of the NSDL access collections and services through portals. Because library users will be very diverse, the NSDL will eventually offer several portals to support the unique needs of each user community (e.g. the main portal, specialized portals or other personalized portals). The Core Integration (CI) team (see under "Governance" section in 4.1.1) maintains the main library portal at nsdl.org using PHP, MySQL and the Internet Scout Portal Toolkit running on Apache Servers. CI will be working with partners to build specialized NSDL portals that highlight tools and functions available to a specific audience.

Archive Service: A basic requirement of national libraries is stewardship of the materials assembled within those libraries. The CI team is building persistent archive services to retrieve materials represented in the MR from public sites (with a crawl depth of 10 levels) and archive both the metadata and content for future retrieval. Web materials that are deleted or "lost" will be recoverable through archive services, and users will be provided options in NSDL search results to retrieve prior versions of resources.

User Help: Text-based user help is available to users through nsdl.org. CI provides Virtual Reference Desk (VRD) capabilities using AskNSDL (http://asknsdl.askvrd.org/) as a central component of the online NSDL help system. This system aids users in gaining assistance from one another, e.g. from more experienced colleagues who have expertise in using a specific collection or can answer discipline-specific pedagogy questions, for example.

The VRD service is e-mail based, and patron questions are distributed to a large group of experts. Responses are not intended to be immediate in this model, as they are in the traditional face-to-face reference interaction. However, this approach is believed to help establish a culture of sharing expert knowledge and a true sense of community among NSDL users and contributors.

The choice of technologies, standards and mechanisms for the core infrastructure is a complex process in which the needs of the NSDL, the availability of the technology, the costs, the adoption rate of end library users, and the wishes of NSDL partners are all important. The challenge is to interpret trends and adopt the technical approaches that will best stimulate the growth of the NSDL while keeping it broadly accessible.

2.1.1.4. Theories, methods and technologies

NSDL employs various technologies, standards and protocols upon which the components of the library are built and interoperate. Here are some starting points to understand how NSDL is being built:

- Future infrastructure developments include Fedora;
- OAI – Open Archives Initiative, used for sharing collection metadata;
- Dublin Core – metadata standard used in the core library;
- XML – schema used by OAI for metadata harvesting;
- WSDL, REST and SOAP – Internet services definition and exchange protocols;
- NSDL services:
  - search and discovery service;
  - authentication and login;
  - AskNSDL – NSDL virtual reference desk;
  - archiving.

Table 3. NSDL research track projects

Theories, methods and technologies	Scope of work
Content Creation in Digital Libraries Using Competitive Intelligence Systems	This project is using high–end multi–agent collaborative competitive intelligence systems to identify and acquire relevant content for digital libraries. By analysing experts' information seeking and problem solving strategies, and translating these into tools to guide novice learners, the project intends to develop a service that complements human–centred knowledge discovery and transfer.
Peer Review of Digital Learning Materials	Developing tools and processes for quality assurance of digital learning materials.
Metadata Generation	This project is developing a system for automating the classification and assignment of metadata through use of natural language processing and machine learning.
Recommendation System	This project is providing instructional support by automated recommendation of learning objects to enhance a national library of interactive web–based mathematics learning manipulatives, with a primary focus on K–12 education.
Enhancing the Interoperability of Collections and Services	Faculty at a number of institutions are collaborating to develop and implement solutions aimed at enhancing the interoperability of both collections and services for the NSDL. A particular emphasis is on exploring the requirements for supporting "tightly federated" collections, featuring close adherence to particular metadata frameworks that enable federated search services to be built.
The Internet Scout Project's Personalized Content Delivery System	This project is developing a Personalized Content Delivery System (PCDS) to promote and showcase NSDL materials and services, while at the same time collecting and disseminating information about the best new online SMET resources from outside the NSDL.
Developing Virtual Reference Desk Capabilities for the NSDL	The goal of this project is to integrate human expertise with virtual reference desk capabilities to support the library. A variety of evaluation methodologies are being used to help develop a better understanding of the digital reference process.
Data Discovery Toolkit and Foundry	This NSDL Services project is developing an initial set of "data discovery" tools to enable students to manipulate real (and real–time if needed) data sets for visualization purposes.
Textual – Geospatial Integration Services for NSDL	Operational services are being constructed that make it easy for users of the NSDL to find and integrate semantically–related information in heterogeneously represented items.
Decentralized Image Retrieval for Education (DIRECT)	A peer–to–peer content based image retrieval (CBIR) service is being developed to enable a user to designate a query image so that the library may be searched for images of similar content. Image matching is done not by text metadata but by the colour, texture and shape of the image objects.

Theories, methods and technologies	Scope of work
An OAI–Compliant Federated Physics Digital Library for the NSDL	This project is designing and building a cross–archive searching service, based on the Open Archives Initiative, that federates heterogeneous collections having metadata that differ in richness, syntax and semantics.
Developing a Learner–Centered Metathesaurus for Science, Mathematics, Engineering and Technology Education	This collaborative project is investigating the creation of a learner–centred metathesaurus based on the analysis of transactions between learners and different types of learning resources.
Managing Authority Lists for Customized Linking and Visualization	The Services for a Customizable Authority Linking Environment (SCALE) project supports two broad classes of service for the NSDL. First are linking services to bind key words and phrases automatically to supplementary information to help students, professionals outside a particular discipline and the interested public to read documents with unfamiliar technical terms and concepts. A second class of service bases automatic linking on authority control of names and terms and on links among different authority lists, such as thesauri, glossaries, encyclopaedias, subject hierarchies and object catalogues.
ReMarkable Texts: A Digital Notepad for the NSDL	Faculty and students are investigating capabilities for an innovative pen–based digital notebook to enable users, particularly students, to work and interact with NSDL's digital materials in a personalized manner. The main features include viewing, notetaking, annotating (e.g. freehand ink, post–it notes and bidirectional fine–grained hyperlinks), organising, and collaborating on multimedia documents, all with the ability to replay the temporal sequence of one's notes in the contexts in which they were made. Whiteboarding and audio facilities are also supported.
MetaTest: Evaluating the Quality and Utility of Metadata	Researchers are evaluating the use and utility of metadata from multiple perspectives. These include the comparison of the subjective quality of metadata that is assigned both manually and automatically to learning resources; the comparison of the retrieval effectiveness of metadata that is assigned manually versus automatically to learning resources; determination of searching and browsing behaviours of users when engaged in information seeking in the digital library; and an analysis of the relative contribution of individual elements of the GEM + Dublin Core metadata scheme to users' searching and browsing behaviour.
Web Lecture Archiving System for Professional Society Meetings	This project is developing a device and associated software to capture lectures from professional society meetings and permit their cataloguing and playback on the World Wide Web.
Topic Maps–based Digital Libraries	This project is studying the application of the new ISO standard, Topic Maps, within the framework of NSDL. Such an approach supports efficient context–based retrieval of learning resources and understanding of the ontological structure of specific subject domains.
Quality Analysis of Metadata Records in the NSDL Metadata Repository	This project is performing a quality assessment and analysis of metadata records in the NSDL Metadata Repository. The project is expected: 1) to classify the evident quality problems and recommend quality assurance solutions; 2) to explicate the relationships between metadata quality and the design principles that underlie the metadata services of the NSDL Metadata Repository; and 3) to provide an empirical basis for an NSDL metadata quality assurance program.
General Recommendation Engine (GRE) for NSDL	This project is developing a general recommendation engine (GRE) that any NSDL system can use to provide recommendation services to its users. GRE selects the relevant materials for users: students doing assignments, teachers preparing for classes, or researchers trying to understand a new topic area. GRE integrates the three most dominant recommendation technologies – collaborative filtering (CF), content–based filtering (CB) and knowledge–based recommendation (KB).
Personal Collections: Enhancing the Utility of the NSDL	This project is providing the equivalent of a personal digital library, allowing patrons to store digital library holdings together with the associated identifying metadata. The service enables users to classify resources using personal taxonomies or organizational schemes, add annotations (such as the suggested citation), share collections with colleagues in collaborative learning and research settings, and search colleagues' personal collections using metadata. Users can also create “packages” of resources and learning objects for use by learning content authoring tools and learning management systems. The design is based on Internet services architecture and makes maximal use of existing open source technologies.

2.1.2. Other Major Science Digital Libraries in the USA

2.1.2.1. ArXiv
(http://lanl.arxiv.org/)

Started in August 1991, arXiv.org (formerly xxx.lanl.gov), also known as Los Alamos National Laboratory (LANL) e-print service, is a fully automated electronic archive and distribution server for research papers. ArXiv is owned, operated and funded by Cornell University and partially funded by the National Science Foundation. It covers unrefereed articles self-archived by the authors. The areas covered include physics and related disciplines like mathematics, nonlinear sciences, computer science and quantitative biology. The contents of arXiv conform to Cornell University academic standards.

Authors can submit their papers to the archive using the World Wide Web interface. They may also update their submissions, though previous versions remain available. Users can retrieve papers from the archive either through an online interface, or by sending commands to the system via e-mail. Users can also register to automatically receive an e-mail listing of newly submitted papers in areas of interest to them. Facilities to view recent submissions and to search old submissions are also provided via the World Wide Web interface.

Evolution

August 1991	Physics e-print archive started: hep-th archive with e-mail interface.
1992	FTP interface added; hep-ph and hep-lat added locally; alg-geom, astro-ph and cond-mat added remotely.
December 1993	Web interface added.
November 1994	Data at some remote archives (using the same software) moved to main site, the remote sites become mirrors.
June 1995	Automatic PostScript generation from TEX Source.
April 1996	PDF generation added.
June 1996	Web upload facility added.
From 1996	World wide mirror network grows.
From 1999	ArXiv involved in the OAI.

ArXiv Software: The software running arXiv comprises of the order of 30,000 lines of Perl running under Linux with numerous other programs (TEX, ghostscript, tar, gnuzip, etc.). The Perl code has evolved through many incremental changes on a system that has run continuously for about a decade, punctuated only by moves to newer hardware and a few short outages.

Involvement of arXiv in the OAI: The Open Archives Initiative (OAI) developed from a meeting held in Santa Fe in 1999 which was initiated by Paul Ginsparg (arXiv, Los Alamos National Lab.), Rick Luce (Los Alamos National Lab.) and Herbert Van de Sompel (University of Ghent, Los Alamos National Lab.). ArXiv has continued to be actively involved in both management of the initiative and technical development of the protocol. The subset Dienst protocol resulting from the Santa Fe meeting was implemented at arXiv.

2.1.2.2. CiteSeer (ResearchIndex)
(http://citeseer.ist.psu.edu/)

CiteSeer (also known as the ResearchIndex) is a scientific literature digital library and search engine that focuses primarily on the literature in computer and information science. It contains freely available, full-text research articles (journal pre-prints and papers where available, conference proceedings, technical reports) downloaded from the web. It indexes PostScript and PDF research articles. CiteSeer uses search engines, crawling and document submissions to harvest papers.

The articles are indexed by an Autonomous Citation Indexing (ACI) system which links the records together through references cited within, and citations made to, an article. It provides links to related articles and can identify the context of a citation, allowing researchers to see what their peers have said about the cited work.

CiteSeer computes citation statistics and related documents for all articles cited in the database, not just the indexed articles. CiteSeer locates related documents using citation and word based measures and displays an active and

continuously updated bibliography for each document. It shows the percentage of matching sentences between documents. CiteSeer provides the context of how query terms are used in articles instead of a generic summary, improving the efficiency of search.

Other services include full-text, Boolean, phrase and proximity search. It provides automatic notification of new citations to given papers, and new papers matching a user profile.

CiteSeer aims to improve the dissemination and feedback of scientific literature and to provide improvements in functionality, usability, availability, cost, comprehensiveness, efficiency and timeliness in the access of scientific and scholarly knowledge. Rather than creating just another digital library, CiteSeer provides algorithms, metadata, services, techniques and software that can be used in other digital libraries.

CiteSeer was developed at the NEC Research Institute by Steve Lawrence, Lee Giles and Kurt Bollacker. It is currently hosted at Penn State's School of Information Sciences and Technology under the direction of Professor Lee Giles.

2.1.2.3. Networked Computer Science Technical Reference Library (NCSTRL)

(http://ncstrl.mit.edu/)

The Networked Computer Science Technical Reference Library, or NCSTRL (pronounced “ancestral”), is an international collection of computer science technical reports from Computer Science departments and industrial and government research laboratories, made available for non-commercial and educational use. Mostly NCSTRL institutions are universities that grant PhDs in Computer Science or Engineering, with some industrial or government research laboratories.

The NCSTRL project is a follow-on to the DARPA-funded CS-TR project and the NSF-funded WATERS project. The project is an international consortium with participation at many levels.

- The NCSTRL Steering Committee is responsible for determining the policy direction of the project and providing long-term oversight. Steering Committee membership includes representatives from the funding agencies, library community, and the North American and European computer science community.
- The NCSTRL Working Group is responsible for the operational oversight of the project. Working Group membership includes representatives from the CS-TR and WATERS projects, and the working group is one of the D-Lib Working Groups.
- The Cornell Digital Library Research Group is responsible for support and maintenance of the current NCSTRL technology.

Technology

NCSTRL is based on two previous technologies for technical report libraries. Dienst is a protocol and implementation for distributed digital library servers. WATERS is a system that links distributed FTP report repositories via a centralized index. NCSTRL uses an extension of the original Dienst architecture to combine the distributed Dienst servers (known as NCSTRL-standard sites), the FTP repositories (known as NCSTRL-lite sites) and a central Dienst indexer for the L (see below).

The NCSTRL architecture combines the power and flexibility of Dienst with the ease of installation of WATERS. The technology underlying NCSTRL is a network of interoperating digital library servers. The digital library servers provide three services: repository services that store and provide access to documents; index services that allow searches over bibliographic records; and user interface services that provide the human front-end for the other services. Search requests from users generate parallel protocol requests to the distributed index servers. A distinguished central index server acts as the search engine for the documents that are stored in FTP repositories. Document access requests generate a protocol request to the repository in the network that contains the document.

The services interoperate using an open protocol, so that other software systems can use them as well. The power of NCSTRL comes from the architecture, while the ease of installation comes from having two levels, Lite and Standard. The Lite version is intended for sites with few resources and will have a lower startup investment, while the Standard version will offer greater functionality. Sites participating in NCSTRL will be able to install either. No matter which they install, the complete technical report collection will be available to all parties. NCSTRL will have a uniform user interface,

hiding almost all the underlying diversity. Users should not need to know which level of software a site is running, and departments will have a smooth upgrade path from the basic to the advanced should they desire additional capability.

2.1.2.4. NASA Technical Report Server (NTRS)

(http://ntrs.nasa.gov/)

NASA's technical information is available via the NASA Technical Report Server (NTRS) to provide students, educators and the public with the access to NASA's technical literature. It is a part of the NASA Scientific and Technical Information (STI) Program, whose mission is to collect, archive and disseminate NASA aerospace information, and locate domestic and international STI pertinent to NASA's missions and Strategic Enterprises. Examples of NASA's STI include research reports, journal articles, conference and meeting papers, technical videos, mission-related operational documents and preliminary data.

NTRS also collects scientific and technical information from sites external to NASA to broaden the scope of information available to users. NTRS's Simple Search searches for NASA information only and its Advanced Search can search for NASA and non-NASA information. It also facilitates browsing and weekly updates. Most of the NASA information does not have full-text document images. However documents can be ordered by contacting the NASA Center for AeroSpace Information.

The scientific and technical information is gathered from NASA's technical report servers and non-NASA sites using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The major archives harvested by NTRS are arXiv Physics E-print Server, BioMed Central, Energy Citation Database (OSTI), NASA Center for AeroSpace Information (CASI), NASA Jet Propulsion Laboratory, National Advisory Committee for Aeronautics (NACA) etc.

History of NTRS

NASA's history with web-based DLs dates back to 1993, when a WWW interface was provided for the Langley Technical Report Server (LTRS). Prior to this, LTRS was simply an anonymous FTP server that distributed technical reports authored and sponsored by NASA Langley Research Center. However, LTRS provided access to only reports from NASA Langley Research Center, and not other NASA centres and institutes. Beginning in 1994, software used to create LTRS was shared with other NASA installations and "LTRS-like" DLs were setup. However, it was not until 1995 that the NASA Technical Report Server (NTRS) was set up to provide integrated searching between the various NASA web-based DLs. Since then NTRS has proven extremely successful with the general public and averages approximately 30,000 search sessions per month.

Technology/Methodology

NTRS is implemented as a specialized bucket and uses a variety of technologies: the Virginia Tech OAI-PMH harvester, an OAI-PMH repository (thus making NTRS an OAI-PMH aggregator), a MySQL database, the AWStats http log analysis facility, and a variety of scripts to integrate the various aspects. Buckets are object-oriented, intelligent web storage units that contain metadata, data and methods to operate on the metadata and data. Buckets already implement approximately 30 methods (or "verbs" in OAI-PMH parlance), so adding the six OAI-PMH verbs as bucket methods was easily accomplished. Buckets were originally built to hold full-text content in DLs, but they have proved useful in other scenarios as well, including building OAI-PMH repositories. The bucket architecture includes advanced facilities for capturing and sharing logs, which is useful for NTRS Recommendation Service. The Recommendation server was added to NTRS in September 2003.

2.1.2.5. OAIster

(http://www.oaister.org/o/oaister/)

OAIster is a project of the University of Michigan Digital Library Production Service. The goal is to create a collection of freely available, previously difficult-to-access, academically-oriented digital resources that are easily searchable by anyone. OAIster provides a one-stop service for users interested in useful digital resources. Users retrieve not only information (metadata) about resources; they have access to the actual resource. For instance, instead of just the catalogue records of a slide collection of Van Gogh's works, users are able to view images of the actual works. Actually harvested metadata that has no corresponding digital resource is not indexed in OAIster. The service encompasses as broad a collection of resources as possible (i.e. with no subject parameters) made freely accessible on the Internet.

This service was built through collaboration between the University of Illinois at Urbana-Champaign (UIUC) and University of Michigan. The metadata harvester built by UIUC was used for the first two years of the project, while University of Michigan developed mechanisms to regularly export and transform the harvested data.

Currently it is harvesting the records (using OAI Protocol) from more than 527 institutions (a.k.a. data providers), it contains almost 5,800,000 records and is being updated weekly.

The site provides the facility to browse the records alphabetically by institution, where it gives the detail about the institution and URL of the data provider. A search can be made through a search box in all the fields of all the participating libraries. 10 results are displayed at a time and again are sorted institution wise on the left hand side of the page. Also a search can be done within a particular field, such as Title, Author, and Subject, and can be limited to Resource Type like text, image, audio, video and dataset.

2.1.2.6. Networked Digital Library of Theses and Dissertations (NDLTD)
(http://www.ndltd.org/)

The Networked Digital Library of Theses and Dissertations (NDLTD) is an international organization dedicated to promoting the adoption, creation, use, dissemination and preservation of electronic analogues to traditional paper-based theses and dissertations. Fundamentally, NDLTD is an effort to improve education while building a digital library and expanding current library services and resources. While many education efforts have focused on how students can learn through accessing a digital library, NDLTD also deals with how students learn by preparing an electronic document and submitting it to a digital library. Further, and key to solving various concerns raised, NDLTD strives to ensure that students are prepared for work in the world of publication.

One can search and browse the Electronic Theses and Dissertations (ETDs) across multiple institutions at once. The various options are:

- Using harvested collection from OCLC – ETD search and discovery system based at Virginia Tech, using data from OCLC's central collection;
- Using older smaller collection harvested by Virginia Tech – ETD search and discovery system based at Virginia Tech, using an older snapshot of data collected locally;
- VTLS search – ETD search and discovery system based at VTLS, using data collected locally, and exposed through the Virtua system;
- Search SRU Provided by OCLC – Web Service machine interface to perform remote searches through OCLC's central metadata collection. External services and portals may directly connect to this service for seamless integration of ETD searching into any other system.

Users can search the full-text of metadata (including abstract) for the full collection or parts thereof, i.e. issue one query to search all sites in parallel. More than 56 universities participate in this programme. Universities can keep their own archive or have it managed by an archival service, and search engines can be at each university or run by state, regional, national, or other services. For performance reasons, backup and regional replication systems can be included in the overall architecture.

2.2. United Kingdom

Much of the digital library related services in the UK have emerged out of projects supported under JISC, eLib and UKOLN programmes. We discuss a few significant DL examples and projects here.

2.2.1. Example Digital Libraries and Projects

2.2.1.1. The Resource Discovery Network (RDN)
(http://www.rdn.ac.in/)

RDN is the UK's free national gateway to Internet resources for the learning, teaching and research community. The service currently links to more than 100,000 resources via a series of subject-based information gateways (or hubs). The RDN is primarily aimed at Internet users in UK further and higher education but is freely available to all. In contrast to search engines, the RDN gathers resources, which are carefully selected by subject specialists in RDN partner

institutions. Users can search and browse through the resources and be confident that the results will connect them to web sites relevant to learning, teaching and research in their subject area. The RDN is a collaboration of over seventy educational and research organizations, and builds upon the foundations of the subject gateway activity carried out under the JISC's (Joint Information Systems Committee) eLib Programme. Table 4 lists the various RDN hubs.

Table 4. RDN hubs

RDN hub	Subject coverage	Website
Altis	Hospitality, Leisure, Sport & Tourism	http://www.altis.ac.uk/
Artifact	Arts and Creative Industries	http://www.artifact.ac.uk/
BIOME	Health, Medicine and Life Sciences	http://biome.ac.uk/
EEVL	Engineering, Mathematics and Computing	http://www.eevl.ac.uk/
GEsource	Geography and Environment	http://www.gesource.ac.uk/
HUMBUL	Humanities	http://www.humbul.ac.uk/
PSIgate	Physical Sciences	http://www.psigate.ac.uk/
SOSIG	Social Science, Business and Law	http://www.sosig.ac.uk/

UKOLN provides technical and interoperability advice role for the Resource Discovery Network (RDN), as well as hosting and developing the central RDN web services. An interdisciplinary search service of RDN, called **ResourceFinder**, has now been developed using the OAI-PMH version 2.0 protocol for metadata harvesting, enabling cross-searching of metadata records harvested from different RDN hubs. This activity has also provided an opportunity to bring greater consistency across the RDN in terms of the records being harvested.

2.2.1.2. JISC IE Metadata Schema Registry

(http://www.ukoln.ac.uk/projects/iemsr/)

The overall objective of the project is to provide access to information on metadata vocabularies and application profiles used for resource description in the UK learning, teaching and research communities. The project will demonstrate how a metadata schema registry might act as the primary source for authoritative information about metadata schemas recommended by the JISC Information Environment (IE) Standards framework. The Registry will be targeted at the spectrum of education communities, aiming to provide a service sufficiently generic to handle schema based on both the Dublin Core and IEEE LOM (Learning Objects Metadata).

2.2.1.3. eBank UK

(http://www.ukoln.ac.uk/projects/ebank-uk/)

eBank UK is a JISC-funded project which is a part of the Semantic Grid Programme. The project is being led by UKOLN in partnership with the Combechem project at the University of Southampton (http://www.combechem.org/) and the PSIgate Physical Sciences Information Gateway (http://www.psigate.ac.uk/newsite/) at the University of Manchester. This initiative is set in the context of the JISC Information Environment which supports end-users to discover, access, use and publish resources as part of their teaching, learning and research activities. The project will build on the technical architecture currently being deployed within the context of the ePrints UK project. The architecture supports the harvesting of metadata from e-print archives in UK academic institutions and elsewhere using the OAI Protocol for Metadata Harvesting (OAI-PMH). The eBank UK project will augment this work by also harvesting metadata about research data from institutional "e-data repositories". Initially this will encompass data made available by Combechem, but will include data from other sources in the longer term. Metadata records harvested from e-data repositories will be stored in the central database alongside the e-print metadata records gathered as part of the ePrints UK project. Within the first year of funding, eBank UK has made progress on a number of fronts. An OAI-PMH repository has been set up at the University of Southampton. The repository extends the use of Open Archives to the dissemination of experimental research data. It is specialized to support the deposition of data from crystallographic experiments. The eprints.org software was modified to deliver the repository. Curation of data must include support for discovery services. eBank UK has developed a demonstrator that uses the OAI-PMH to aggregate

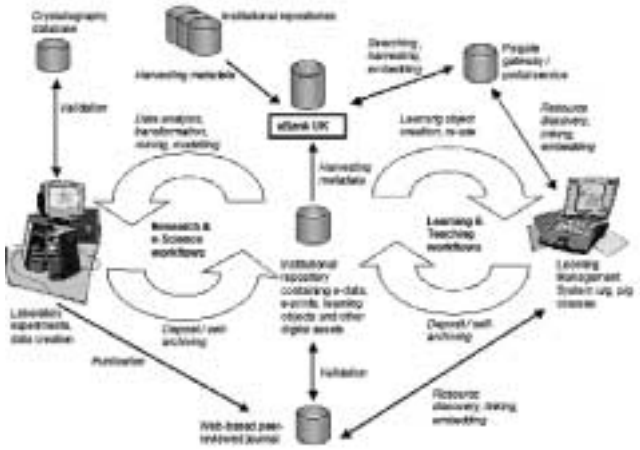

Figure 17. The Scholarly Knowledge Cycle

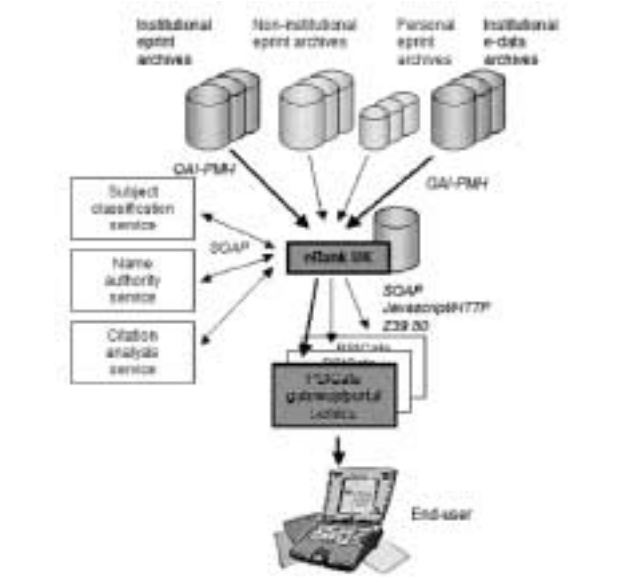

Figure 18. eBank UK information architecture framework under development

metadata, and supports discovery both of datasets and related publications. Embedding techniques have been used to integrate the search interface into PSIgate, which has an established user base hailing from scientific disciplines. These demonstrators are useful for learning about different ways in which data can be discovered and re-used.

The Knowledge Cycle visualized in the eBank project, showing the integration of creation, distribution and use of research data and other types of research output, in the larger context of learning and research, is shown in Figure 17.

Figure 18 outlines the eBank UK information architecture framework under development.

2.2.1.4. FAIR (Focus on Access to Institutional Resources)

(http://www.jisc.ac.uk/programme_fair.html)

The JISC Information Environment is envisaged as a virtual place where members of colleges and universities can deposit and share useful content (e.g. research outputs). The current collection of JISC funded content has the potential to grow to embrace both externally generated content from publishers and aggregators and community-generated resources. To achieve the latter, staff and students will need a *place* or *places* in which to lodge suitable content and products and a means for exchanging and adding to it. This is an important part of developing a relevant and useful portfolio of resources. The FAIR programme has been developed to create the mechanisms and supporting services to allow this process to prosper and these *places* to be built. The work of this programme has been inspired by the success of the Open Archives Initiative (OAI), which has proved to be a simple mechanism that allows metadata about resources to be harvested into services that can be searched by staff and students. FAIR programme aims to evaluate and explore different mechanisms for the disclosure and sharing of content (and the related challenges) to fulfil the vision of a web of resources built by groups with a long term stake in the future of those resources, but made available to the whole community of learning. Specific objectives are to:

- Explore the OAI protocol as a mechanism for disclosure and sharing a range of resource types, e.g. images, museum content, e-prints, e-theses;
- Explore, in addition, possible alternative mechanisms for disclosure;
- Explore the challenges associated with disclosure and sharing, including Intellectual Property Rights (IPR) and the role of institutional repositories; test out the delivery of disclosed information through established JISC services and investigate other means of delivery;
- Inform work towards developing the JISC Information Environment.

FAIR projects have been arranged into three clusters (see box below). These clusters will aid in the management of the programme but will also serve as focal points for other organizations and interested parties to become involved in FAIR and to contribute to the success of the programme.

E-prints and e-theses cluster: These projects are examining the use of OAI as a means of sharing and disclosing information about materials held in institutional repositories, and the issues that relate to this.

Museums and Images cluster: Projects in this cluster are examining the disclosure of assets from institutional collections, particularly museum and image materials, through established JISC services using OAI-disclosed metadata.

Institutional Portals: Exploring how disclosed assets and other externally-sourced JISC collections and services can be made available within institutions alongside locally managed assets. This is to ensure that users do not have to go to many locations to access the information they need. OAI is one of the mechanisms through which such collections might be delivered.

FAIR Cluster 1: E–prints and E–theses

DAEDALUS – University of Glasgow. http://www.lib.gla.ac.uk/daedalus/

Investigating the use of different repository software to store different types of materials within the University, from which they can then be disclosed to the wider community. These materials include pre–prints, post–prints (deliberately kept separate), electronic theses and dissertations (ETDs), administrative documents and grey literature.

HaIRST – University of Strathclyde. http://hairst.cdlr.strath.ac.uk

HaIRST involves a consortium of both HE and FE institutions, and is investigating the ways in which different materials can be disclosed and shared in different situations. The project is developing a model of how different levels of metadata can be incorporated in an overall search for materials across the range.

SHERPA – University of Nottingham. http://www.sherpa.ac.uk

SHERPA concentrates on the building up of a body of content of self–archived articles in order to increase the range of research outputs available on open access within institutional repositories. It includes both pre–prints and post–prints.

TARDis – University of Southampton. http://tardis.eprints.org

This project concentrates on the technical and cultural aspects of setting up an institutional repository and making this an integral part of the research and publication process within the university. It is also exploring the role of the library in providing such a service.

ePrints UK – UKOLN, University of Bath. http://www.rdn.ac.uk/projects/eprints–uk/

Within the context of OAI, the four projects above are developing data providers, repositories of content and associated metadata. ePrints UK is building an OAI service provider. Although individual repositories will be accessible at institutions, ePrints UK is envisaged as the route to search across all or a selection of them.

RoMEO – University of Loughborough. http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html

This one–year project is now complete. Surveys of authors, journal publishers, OAI data providers and OAI service providers were undertaken and built up a body of data on the rights these various parties are willing to allow and/or would like to see. A directory of publisher self–archive policies has been created. Solutions to allow all e–prints to include an IPR statement in associated metadata have also been created.

Theses Alive! – University of Edinburgh. http://www.thesesalive.ac.uk

Developing a pilot distributed system for the management of e–theses. This will involve the development of an OAI–compliant thesis submission system and the testing of an infrastructure to enable e–theses to be published on the web by institutions.

Electronic Theses – Robert Gordon University, Aberdeen. http://www.rgu.ac.uk/library/e–theses.htm

To complement the work of the DAEDALUS (storage) and Theses Alive! (submission and dissemination) projects, this project is examining the practice and methods of e–theses production, management and use, including theses that have been digitized and those born digital.

FAIR Cluster 2: Museums and images

Harvesting the FitzWilliam – FitzWilliam Museum, University of Cambridge.
http://www.fitzmuseum.cam.ac.uk/ htf/index.html

This project is preparing and digitising items from the FitzWilliam collections for disclosure using OAI and delivery through the Archaeology Data Service and Arts & Humanities Data Service. Issues being addressed include the metadata requirements for museum objects and the disclosure of images using OAI.

Accessing the Virtual Museum – Petrie Museum, University College London. http://www.petrie.ucl.ac.uk/randd/avm.html

This project is preparing and digitising items from the Petrie Museum of Egyptian Archaeology. Disclosure will be through the Archaeology Data Service. The major issue being addressed is the creation of metadata for objects which require transliteration from original scripts and which do not naturally fit with the Dublin Core requirements of OAI.

Hybrid Archives – Arts & Humanities Data Service (AHDS), King's College London.
http://www.ahds.ac.uk/about/ projects/hybrid–archives/index.htm

Hybrid Archives is examining the preservation requirements for institutional collections disclosed for delivery through a service, such as the AHDS. The project is developing a model of partial deposit, whereby a preservation copy of the actual content is deposited at the AHDS, while the OAI–disclosed metadata still connects users to the original collections.

BioMed Image Archive – Institute for Learning & Research Technology, University of Bristol. http://www.brisbio.ac.uk/

The BioMed Image Archive itself is an established source of biomedical images that has been available for a number of years. This project is investigating ways that would allow institutions to self–archive images to the archive remotely, in order to share these with the wider community.

FAIR Cluster 3: Institutional portals

PORTAL – University of Hull. http://www.fair–portal.hull.ac.uk/

PORTAL is examining how JISC collections and services can be embedded within an institutional portal. It is taking advantage of Hull's current work in developing an institutional portal and the project will be assessing user needs for such a portal and what information should be delivered through it.

FAIR Enough – Western Colleges Consortium, Keynsham, Somerset. http://www.fairenough.ac.uk/

The Western Colleges Consortium is a collection of five further education colleges in the West Country. This project is examining the needs of further education for access to JISC collections and services within a virtual learning environment, and how these sit alongside local assets.

2.2.1.5. ePrints UK

(http://www.rdn.ac.uk/projects/eprints-uk/)

The principal objective of the project is to develop a demonstrator of a national service through which the UK Higher and Further Education communities could access the collective output of e-prints from UK repositories. This two-year project was funded by the JISC as part of its Focus on Access to Institutional Resources (FAIR) Programme. The project builds upon RDN's experience in implementing the Open Archives Initiative Protocol for Metadata Harvesting

to share metadata between the RDN partners to create the aggregated search facility, ResourceFinder (see above). A Service Demo is available at: http://eprints-uk.rdn.ac.uk/. This is the basic version of the e-Prints UK Portal. Discipline-focused views of available e-prints will be provided through the use of an automatic subject-classification web service offered by OCLC. Furthermore, the project will use "name authority" and "citation analysis" web services (offered by OCLC and the University of Southampton respectively) to enhance the metadata harvested from available archives.

2.2.1.6. Digital Curation Centre

(http://www.dcc.ac.uk/)

Scientists and researchers across the UK generate increasingly vast amounts of digital data, with further investment in digitization and purchase of digital content and information. This volume is increasing dramatically with the advent of Grid-enabled applications. The scientific record and the documentary heritage created in digital form are at risk, because of technology obsolescence and the fragility of digital media. Working with other practitioners, the Digital Curation Centre will support UK institutions to store, manage and preserve these data to ensure their enhancement and their continuing long-term use. The purpose of the Centre is to provide a national focus for research into curation issues and to promote expertise and good practice, both national and international, for the management of all research outputs in digital format. The establishment of the DCC follows a recommendation in the JISC Continuing Access and Digital Preservation Strategy (October 2002) to establish a UK Digital Curation Centre to help solve challenges which could not be solved by any single institution or discipline, including generic services, development activity and research.

2.2.1.7. The National Centre for Text Mining

(http://www.nactem.ac.uk/)

National Centre for Text Mining (NaCTeM) is operated by a consortium of three Universities: the University of Manchester which leads the consortium, the University of Liverpool and the University of Salford. The service activity is run by the National Centre for Dataset Services (MIMAS), based within Manchester Computing (MC). As part of previous and ongoing collaboration, NaCTeM involves, as self-funded partners, world-leading groups at San Diego Supercomputer Center (SDSC), the University of California at Berkeley (UCB), the University of Geneva and the University of Tokyo. NaCTeM's initial focus is on bioscience and biomedical texts as there is an increasing need for bio-text mining and automated methods to search, access, extract, integrate and manage textual information from large-scale bio-resources. NaCTeM was established in Summer 2004 with funding from the Joint Information Systems Committee (JISC), the Biotechnology and Biological Sciences Research Council (BBSRC) and the Engineering and Physical Sciences Research Council (EPSRC), with the consortium itself investing almost the same amount as it received in funding. The paramount responsibility of NaCTeM is to establish high-quality service provision in text mining for the UK academic community, with particular focus on biological and biomedical science. Initial activity will establish the framework to enable a quality service, and to identify "best of breed" tools. Evaluation and choice of appropriate tools is ongoing, and tools will be customized in cooperation with partners and customers, bearing in mind existing competition and advantages to be gained from cooperation with technology providers.

The overall aims of NaCTeM are:

- To provide a one-stop resource and focus primarily for the UK text mining community for user support and advice, service provision, dissemination, training, data access, expertise and knowledge in related emerging technologies, consultative services, tutorials, courses and materials and demonstrator projects;
- To drive the international and national research agenda in text mining informed by the collected experiences of the user community, allied to existing and developing knowledge and evaluation of the state of the art;
- To consolidate, evolve and promulgate best practice from bio-related text mining into other domains;
- To widen awareness of and participation in text mining to all science and engineering disciplines, and further to social sciences and humanities, including business and management;
- To maintain and develop links with industry and tool suppliers, to establish best practice and provision.

2.2.1.8. DSpace@Cambridge

Dspace@Cambridge: http://www.dspace.cam.ac.uk/
Project Site: http://www.lib.cam.ac.uk/dspace/index.htm

The DSpace@Cambridge Project is a three-year collaboration started in 2003 between Cambridge University Library, Cambridge University Computing Service and MIT Libraries. Funded by a grant from the Cambridge-MIT Institute, its objective is developing an institutional repository that will preserve and disseminate digital materials created by or associated with the University of Cambridge. It utilizes DSpace™, an open source digital repository software platform developed jointly by the Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard Laboratories.

The project has focus on two areas of interest. Firstly, DSpace will be developed as a means of preserving digital content. The expertise that the University Library has acquired during its four years of participation in the Cedars and CAMiLEON digital preservation projects will be utilized for this purpose. Secondly, DSpace's ability to support learning management systems will be developed in collaboration with CARET and other University departments.

Strategies will also be established regarding deployment costs, intellectual property rights issues, organizational change, institutional governance and policy issues, based on the proposed implementations, and these lessons will be applied in support of the shared goals of MIT and Cambridge University.

The specific goals of the project are:

- To identify needs within the institution for digital asset management services, and a deployment strategy to satisfy these needs;
- To establish DSpace at Cambridge to preserve: learning materials, internally produced digital objects and externally published digital objects, including the ability to enable appropriate use;
- To demonstrate applicable Cedars and CAMiLEON approaches within an operational DSpace context. Share technologies between institutions;
- To work with other institutional entities to agree on standards applicable for the deposit of educational materials in a digital repository (AMPS and CARET, for example). Collaboratively develop interoperability standards with Learning Management Systems using emerging standards from the Open Knowledge Initiative work, or other similar initiatives;
- To position the institutions to deliver archived educational digital assets through emerging Learning Management Systems;
- To leverage the business model created at MIT to produce the Cambridge University business model, including costs, funding streams and institutional structures, in co-operation with the business schools at both institutions.
- To inform national digital repository strategies;
- To produce pragmatic approaches that enable practical solutions to be applied to rights management issues.

Material for populating DSpace in the context of Cambridge University will potentially come from a variety of both internal and external sources, including, for example, learning materials created by CARET, by teaching departments and by individuals, papers written by teaching and research staff, technical reports, theses, databanks, administrative documents (the official University archive), personal files, working notes, purchased digital texts, bibliographic datasets, hosting of e-journals published by Cambridge academics, legal deposit of digital publications, etc.

DSpace@Cambridge is implemented in three phases: planning (phase 1), implementation (phase 2) and pilot project (phase 3).

Early Adopters and Fast Followers: In the first stages the project team identified a small number of "early adopter" communities to provide content and test the system before it is made available to the University at large. Each user community is able to customize the DSpace system to meet its individual needs and manage its own data submission process.

Together they represent a cross-section of academic activities and offer a variety of file formats. Those who are already actively contributing, or have agreed to provide, material are:

- Archaeology;
- Cambridge University Library;
- CamRAD (Cambridge Rock Art Database);
- CARET (Centre for Applied Research in Education Technology);
- Chemistry – Unilever Centre for Molecular Informatics;
- CRASSH (Centre for Research in the Arts, Social Sciences and Humanities);

- CUMIS (Cambridge University Moving Image Studio);
- Economics – Depertment of Applied Economics;
- Fitzwilliam Museum;
- Music;
- Pathology;
- Philosophy;
- Department of Social Anthropology.

In the third phase the communities will expand to include "Fast Followers" identified during a University-wide survey in August 2004, adding a broader range both of subjects and of file types to those already present. The project team will explore how the repository can also support core University activities, such as electronic theses and dissertations, and administrative records. This phase will also see the project develop its business plan, leading to the launch of DSpace@Cambridge as a sustainable repository service.

Note: In parallel with the Cambridge-MIT Libraries partnership, another six universities (Columbia, Cornell, Rochester, Ohio and Washington in the United States, plus Toronto in Canada) are developing collaborative arrangements with MIT to implement DSpace for themselves, thus creating the first federation of DSpace institutions.

2.2.1.9. Semantic Grid
(http://www.jisc.ac.uk/index.cfm?name=programme_semantic_grid)

The JISC (http://www.jisc.ac.uk/) and the Engineering and Physical Science Research Council (http://www.epsrc.ac.uk/) have both undertaken to fund projects to progress the development of a Semantic Grid. Lying at the core of scientific development is the discovery of new knowledge. The rapid increase in the volume and variety of data inherent within e-Science and mirrored by e-Commerce and e-Government means that any supporting infrastructure must provide a set of semantic services. These core services must be able to equip data with meaning and generate a surrounding semantic context in which data can be meaningfully interpreted. Fundamental research on knowledge systems and services is needed to allow us to move from the current data centric view supporting the grid to a semantic grid with a powerful set of knowledge services. Two projects initiated under the Semantic Grid are Collaborative Open Ontology Development Environment (CO-ODE), lead by the University of Manchester, and eBank UK, lead by UKOLN at the University of Bath.

2.2.1.10. BioMed Image Archive
(http://www.brisbio.ac.uk/index.html)

The BioMed Image Archive is an online collection of about 8000 medical, dental and veterinary images for use in learning, teaching and research. All the images have been donated by academics working in the biomedical fields in different countries. Images can be used free of charge for non-profit educational purposes. There are several ways of finding images in the archive. Options include basic search, advanced search, browsing by keywords, browsing by terms derived from the image descriptions, and thesaurus search. To aid searching, the images have been catalogued using Medical Subject Headings (MeSH) (http://www.nlm.nih.gov/mesh/).

2.2.1.11. Dental Images
(http://www.derweb.co.uk/)

2,500 online dental images from the DERWeb (Dental Educational Resources on the Web) Image Library. DERWeb aims to provide dental academics and practitioners with the definitive collection of information to aid in the advancement of knowledge and technology in the field of dentistry around the world.

2.2.1.12. CogPrints: CogSci – Cognitive Science Disciplinary Research Archive
(http://cogprints.org/)

Launched in 1998 and edited by Professor Stevan Harnad, CogPrints is an electronic archive for self-archiving papers in any area pertinent to the study of cognition. These include Psychology, neuroscience, Linguistics, many areas of Computer Science, Philosophy, Biology, Medicine, Anthropology, as well as any other portions of the physical, social and mathematical sciences. It contains more than 2400 articles.

Authors can deposit their pre-prints or post-prints of paper in areas relevant to cognitive sciences which is freely available for all from the server.

CogPrints runs on *EPrints2*, the archive-creating software, which generates e-prints archives that are compliant with the Open Archives Protocol for Metadata Harvesting OAI 2.0 (http://www.openarchives.org/). It is often held up, along with the physics archive arXiv, as proof that the author self-archiving model of publishing can work.

2.2.2. DL Initiatives and Programmes

Digital library related initiatives in the UK have mainly come from JISC, UKOLN and eLib programmes.

2.2.2.1. The Joint Information Systems Committee (JISC)

(http://www.jisc.ac.uk/)

The Joint Information Systems Committee (JISC) is a committee of the UK further and higher education funding bodies that, through funding of services and the management of development programmes, aims to help and facilitate institutions in their use of the Internet and other Information and Communications Technology (ICT) applications. The JISC's mission is "to help further and higher education institutions and the research community organize their ambitions in exploiting the opportunities of information and communications technology by exercising vision and leadership, encouraging collaboration and co-operation and by funding and managing national development programmes and services of the highest quality."

Among several objectives JISC has formulated three key related objectives to meet over a period of 2002-2006:

- Build an online information environment providing secure and convenient access to a comprehensive collection of scholarly and educational material;
- Help institutions create and maintain Managed Learning Environments (MLE) to support students;
- Ensure the continued provision of, and wide access to, a world-leading network to support education and research in the UK.

These activities will be designed to support learning and teaching, research, the management and administration of institutions and those sector outreach activities where a UK focus can add value.

The key driving factor in formulating and working towards these objectives has been the recognition that the Internet has become vital to further and higher education, not just to enhance distance and flexible learning but as a useful tool in traditional learning and teaching, administration and management, and in research. In meeting its objectives JISC aims to provide:

- An information environment that is easy to use and incorporates necessary access controls;
- Online information resources;
- The underlying network (JANET);
- Advisory and consultancy services;
- An innovative development programme.

The JISC works with other organizations in related areas, such as the creation of learning materials, development of international standards, provision of access to content outside further and higher education and training, and staff development.

DL related development strategy of JISC

Developing an "Online Information Environment (IE)" is the key development strategy of JISC with several related projects and programmes. Goal of the IE strategy is to "build an on-line information environment providing secure and convenient access to a comprehensive collection of scholarly and educational material". The Information Environment is the key to achieving an interoperable distributed national electronic resource. A major element of this strategy is to build on existing partnerships and forge new ones to contribute to a vision of a single, worldwide information environment. This environment will eventually provide access to the majority of information needed by scholars and researchers for their day-to-day work.

JISC IE vision and architecture
(http://www.jisc.ac.uk/ie/)

An Information Environment can be characterized as the set of network or online services that support publishing and use of information and learning resources. At the moment online services providing digital resources tend to operate in a stand-alone manner. The user is therefore required to navigate a complex set of different websites with different search interfaces in order to locate relevant resources. Similarly, the resources offered tend to be characterized by a lack of mediation to provide vital signposts to explain context and relevance to the user. It has been recognized that this is one of the key features limiting take up of digital resources. The key challenge is therefore to achieve a coherent and shared information environment that will overcome these obstacles. It is an environment that will need to be compatible with the developments that are taking place within colleges and universities, such as the growth of institutional web sites, portals, intranets and virtual learning environments (VLE) and managed learning environments (MLE). The Information Environment is proposed to offer the user a more seamless and less complex journey to relevant information and learning resources.

The IE technical architecture specifies a set of standards and protocols that support the development and delivery of an integrated set of networked services that allow the end-user to **discover, access, use** and **publish** digital and physical resources as part of their learning and research activities. Examples of the kind of activities supported by the architecture include:

- Integration of local and remote information resources with a variety of "discovery" services (for example, the RDN subject portals, institutional and commercial portals and personal reference managers) allowing students, lecturers and researchers to find quality assured resources from a wide range of content providers including commercial content providers and those within the higher and further education community and elsewhere;
- Seamless linking from "discovery" services to appropriate "delivery" services;
- Integration of information resources and learning object repositories with Virtual Learning Environments (for example, allowing seamless, persistent links from a course reading list or other learning objects to the most appropriate copy of an information resource);
- Open access to e-print archives and other systems for managing the intellectual output of institutions.

Examples of the kinds of content that are available through the JISC IE include scholarly journals, monographs, textbooks, learning objects, abstracts, manuscripts, maps, music scores, Internet resource descriptions, still images, geospatial images, and other kinds of vector and numeric data, as well as moving picture and sound collections.

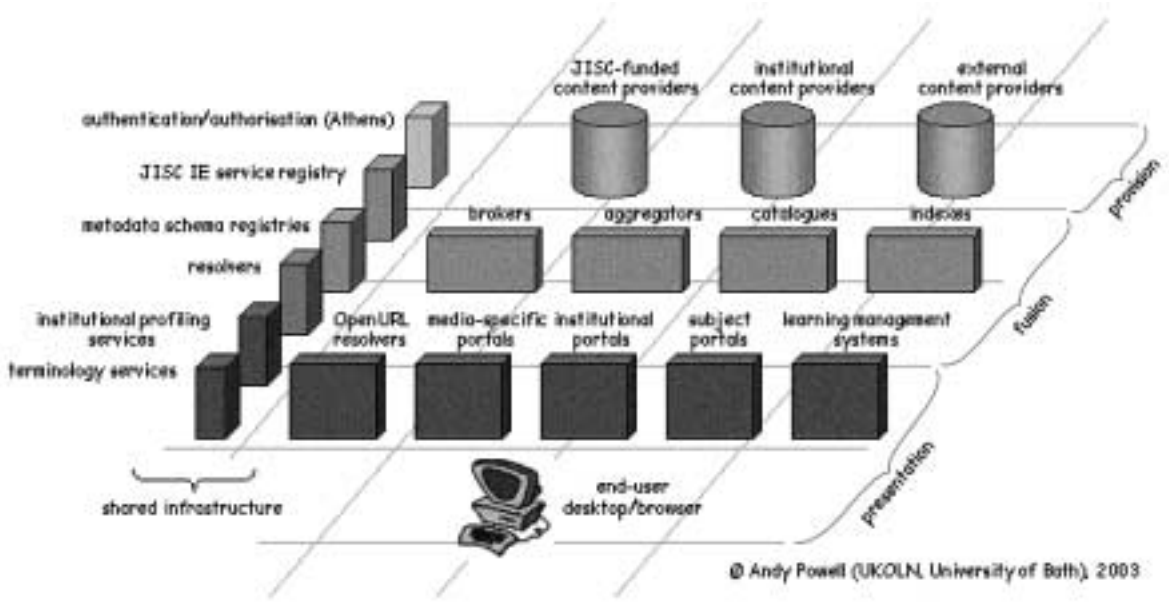

Figure 19. JISC Online Information Environment (IE) architecture

The JISC IE is being realized through the Distributed National Electronic Resource (DNER). One key aspect of DNER is the implementation of a range of commonly-agreed technical standards and protocols.

2.2.2.2. UKOLN

(http://www.ukoln.ac.uk/)

UKOLN is a centre of expertise in digital information management, providing advice and services to the library, information, education and cultural heritage communities by:

- Influencing policy and informing practice;
- Promoting community-building and consensus-making by actively raising awareness;
- Advancing knowledge through research and development;
- Building innovative systems and services based on web technologies; and
- Acting as an agent for knowledge transfer.

UKOLN is funded mainly by the JISC and the European Community. UKOLN is involved in several projects and programmes of JISC. Two key activities of UKOLN are Policy and Advice; and Research and Development.

Policy and Advice

UKOLN's Policy and Advice team members seek to disseminate advice and support to UK user communities on best practices in the development and use of digital library services. The areas of advice include web accessibility, interoperability, digitization, collection description (metadata schemas and tools) and quality assurance.

Research and Development

Some of the current R&D projects UKOLN is involved in include: the Digital Curation Centre; the Delos Network of Excellence; emerging metadata standards; open access to data and journal article e-print repositories; digital preservation and the management of metadata schemas.

2.2.2.3. eLib

(http://www.ukoln.ac.uk/services/elib/)

In the early 1990s it became clear that electronic information would play a significantly increased role within higher education (HE). What was not clear was how this would occur and what the impacts would be, particularly on libraries. At the same time, the growth in student numbers and increasing costs were both major concerns for higher education libraries. These issues were linked in the report of the Libraries Review Group report on library provision (The Follett Report), published in December, 1993. The Electronic Libraries Programme (eLib) was an initiative of the Higher Education Funding Councils' Joint Information Systems Committee (JISC), established to implement the IT recommendations from the Follett Report by addressing the issue of change within higher education libraries. The eLib Programme consisted of three phases. Phases 1 and 2 together formed a successful £15M programme over a period of 3 years from 1994 to 1997. eLib Phase 3 was a £4.1M three-year programme which sought to consolidate this work in a practical context and to extend Phase 1 and 2 benefits by helping to achieve “critical mass” in key areas.

In its first two phases, eLib addressed uncertainties by taking a broad approach with 59 projects. An important overall aim was the inclusion of a wide range of stakeholders. eLib used mature technologies within a managed framework covering a limited number of key development areas. These areas are given below:

- Phase 1:
  - Electronic document delivery;
  - Electronic storage of books and journals (digitization);
  - Electronic journals;
  - On-demand publishing and the electronic book;
  - Awareness and training;
  - Navigational tools (Access to network resources);
- Phase 2:
  - Pre-prints;
  - Quality Assurance;

  - Electronic short loans;
  - Images;
- Phase 3:
  - Hybrid libraries;
  - Large scale resource discovery (clumps);
  - Digital preservation;
- Supporting studies.

The eLib Programme was judged to have been successful and good value for the investment. It provided good coverage of the issues involved in the development of electronic libraries. A number of services were established as a result of eLib projects. Within the HE organizations it involved, eLib has improved awareness and practical experience of the opportunities offered by electronic resources, allowing them to adapt more effectively to rapid changes in technologies and evolving user expectations. Overall, eLib was evaluated to have had positive impact on libraries, HE community, publishers and rights holders, commercial, public and international.

2.2.3. Theories, Methods and Technologies: Focus Areas

DL related R&D focus areas in the UK

Research focus	DL project	Remarks
National level Internet resources gateway service	RDN	Metadata disclosure and cross–searching using disclosed metadata. Data uniformity across RDN hubs. Interoperability.
Metadata vocabularies (schemas) and application profiles	JISC IE Metadata Schema Registry	For describing resources of relevance to teaching, learning and research.
Research data archiving in e–print archives, harvesting and resource discovery using OAI–PMH	eBank UK	
Institutional repositories – for archiving and disseminating institutional research output (publications)	FAIR (Focus on Access to Institutional Resources)	OAI protocol as a mechanism for disclosure and sharing a range of resource types – publications, theses, museum ent. IPR, technical and cultural issues. Open access to research output.
Institutional portals	FAIR (Focus on Access to Institutional Resources)	Integration of access to external and internal assets through OAI–based metadata disclosure.
National level cross–archive search services	ePrints UK	Interoperability.
Digital data Curation – long term storage, preservation and access to digital data	Digital Curation Centre	
Text mining	The National Centre for Text Mining	Search, access, extract, integrate and manage textual information from large–scale biological and biomedical resources.
Ontology development	Semantic Grid	To equip data with meaning and generate a surrounding semantic context in which data can be meaningfully interpreted.
Image search and retrieval	BioMed Image Archive. Dental Images	
Information Environment	JISC Information Environment Architecture	Technical Architecture and standards.

2.3. Germany

2.3.1. DL Examples and Projects

2.3.1.1. Vascoda

(http://www.vascoda.de/)

Started in August 2003, Vascoda is an integrated interdisciplinary Internet portal for scientific and scholarly information in Germany. It is sponsored by the Bundesministerium für Bildung und Forschung (Federal Ministry for Education and Research – http://www.bmbf.de/) and the Deutsche Forschungsgemeinschaft (German Research Foundation – http://www.dfg.de/lis).

The following types of information resources can be found through vascoda:

- Electronic full-text access to journal articles, working papers, etc;
- Subject gateways, offering reliable and quality controlled access to subject-specific websites containing scientific and other high-standard resources;
- Bibliographic data of other types of media (printed books and journals, microfilms, etc.), often combined with document delivery systems to enable access to these materials;
- Other types of information resources, such as calendars, statistical data, videos, etc.

A uniform user interface provides opportunities for subject-specific and interdisciplinary searches. In addition to the central search engine, it also offers four subject-based portals. You can search on each of these gateways or make an interdisciplinary search on two or more areas:

- Humanities and Area Studies;
- Economics and Social Sciences;
- Life Sciences;
- Engineering and Physical Sciences.

It links the search results to electronic full text if the access is free or if the institute from where the end users searches has access to the particular resource. Otherwise it links it to document delivery services and pay-per-view options.

Partners: Around 37 institutions with almost 30 different individual services are integrated within vascoda. The partners are libraries, as well as information centres and other institutions providing quality academic information. The aim is to bundle resources in a central gateway site. Vascoda is structured in a flexible manner in order to ensure that the circle of partner institutions and subject areas can always expand. The first release of vascoda is realized through a joint effort of three major partners.

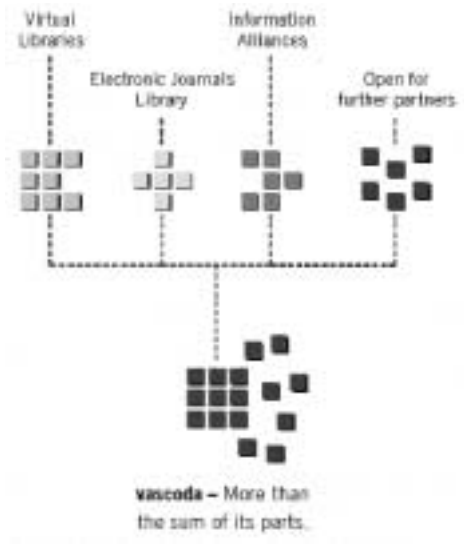

Figure 20. Structure of the Vascoda portal

The *Electronic Journals Library* (Elektronische Zeitschriften-bibliothek or EZB) offers access to scientific and academic full-text journals. The services of the EZB is integrated into the vascoda portal to give users the license information of the journal articles retrieved through the portal. The user will be able to find out immediately if the document can be obtained free of charge or which institution has a subscription to the journal and how it can be used. If the user accesses vascoda through a subscribing institution, he or she can also directly access the full-text of the journal.

The *Information Alliance* is another partner. It focuses on providing the digital full-texts of published materials and grey literature. Partners of the information alliance are EconDoc (http://www.econdoc.de/), GetInfo (http://www.getinfo-doc.de/), infoconnex (http://www.infoconnex.de/) and MedPilot (http://www.medpilot.de/). Together they make an effort to

combine the search results of databases with the electronic full-text provided either by publishers, grey-literature- or preprint-servers, etc. This is now accessible through vascoda.

The third partner is the Subject-Based *Virtual Libraries (ViFas)*. Twenty-three subject gateways that offer access to information and documents relevant for research in a specific field of study are integrated into vascoda.

Vascoda will be a central access point for all fields of research covering science, life sciences and the humanities, as well as social sciences. The use of the services of the portal is free of charge and open to everyone while some of the services of the partners will be charged for.

Vascoda is now the basic module for a "Digital Library Germany". In the future it will provide Germany, as a location for research and scholarly activity, with a convenient information and verification tool. At the start of 2005, the Federal Ministry of Education and Research (BMBF) (http://www.bmbf.de/) approved funding for the second development phase. The second release will offer complex search and browsing options enabled by metasearch engines, standardization of metadata and possibly a joint cataloguing database for Internet resources. Focus is also to integrate local and regional libraries more fully into the vascoda portal.

2.3.1.2. Dissertations Online at Die Deutsche Bibliothek (DDB – The German Library)

The German Library has collected electronic theses and dissertations from the Universities since 1998. The service at Die Deutsche Bibliogthek was promoted by the DFG as a project "Dissertationen online". The objective was to evolve a uniform and legally correct process of handling post-doctoral theses and dissertations produced by the universities in Germany, for electronic archiving and retrieval. This system was developed by DDB in co-operation with the university libraries.

About 87 university libraries are currently using this procedure. The metadata can be input by the authors through an interactive online form at the web server of the DDB which then sends the data to the appropriate registering institute for processing. It can be also sent by e-mail containing metadata or referring to metadata in conformity with the MetaDiss standard. The full-text documents of the dissertations and post-doctoral theses are registered by the university libraries with DDB, which then downloads them to its Archive Server DEPOSIT.DDB.DE (http://deposit.ddb.de/) using standard transfer protocols.

Users can access online dissertations and post-doctoral theses held by Die Deutsche Bibliothek from the OPAC of die Deutsche Bibliothek Frankfurt am Main via Z39.50-Gateway or directly from the OPAC (http://dbf-opac.ddb.de/).

For each online publication display, a front page is automatically generated in HTML on the basis of the descriptive data provided by the submitting institution. The authentication code and any viewers that may be required are also embedded here. The online documents can be called up directly, provided the applicable legal restrictions permit. Dissertations and post-doctoral theses for which an access fee is charged are available for internal use only.

So far approximately 34,500 theses and around 450 dissertations were announced (as on: 01.06.2005).

Coordination Agency DissOnline

The Coordination Agency DissOnline (http://www.dissonline.de) was established at Die Deutsche Bibliothek in February 2001. This agency serves as a point of contact in matters relating to online dissertations and post-doctoral theses and is responsible for promoting and co-ordinating further development of the results of the "Dissertationen online" project.

DissOnline informs libraries, doctoral candidates, faculties and computer centres about tools, online support, contacts and preferred formats and tools recommended by Die Deutsche Bibliothek. The Coordination Agency advises university libraries when introducing or adapting the workflow for online dissertations and provides the workflow for the transfer of online dissertations and their metadata to DDB.

It also recommends technological innovations and advancements that are corresponding to the current state-of-the-art technology and national and international efforts for standardization.

Methods and Technologies used in DDB

Documents must be made available for downloading in one of the "preferred data formats" specified by Die Deutsche Bibliothek. The DDB also collects the metadata in the MetaDiss format which is based on the Dublin Core Metadata element set.

To ensure permanent access to the electronic resources, the DDB assigns a Persistent Identifier for dissertations since September 2001. The namespace "National Bibliography Number (NBN)" within the URN system, agreed upon by the Conference of Directors of National Libraries, is also being used.

In order to guarantee the integrity of documents consisting of multiple individual files, a standard for the transmission of multiple-file documents (container model) is applied.

Once a copy of the document is downloaded to the Archive Server and checked for viruses, an MD-5 hash code is computed. The hash code is used to identify document copies in relation to the archived reference version. Long-term preservation is ensured through the application of several different security technologies.

2.3.2. Theories Methods and Technologies: Focus Areas

DL related R&D focus areas in Germany

Research focus	DL project	Remarks
National level Internet resources gateway service	ViFas	Twenty three subject–based virtual libraries or portals on specific subjects.
Metadata vocabularies (schemas) and application profiles	Dissertations Online	MetaDiss provides metadata standards for the Dissertations online.
Institutional repositories – for archiving and disseminating institutional research output (publications)	Dissertations Online	Centralized archiving of the ETDs from universities across the nation.
National level cross–archive search services	Vascoda	Integration of varied DLs into a single portal that will form the basis for DL of Germany.
Digital data curation – long term storage, preservation and access to digital data	Dissertations Online	Using namespaces, persistent identifiers and MD–5 hash techniques for long–term preservation of documents.

2.4. Australia

2.4.1. DL Examples and Projects

2.4.1.1. PANDORA – Australia's Web Archive

(http://pandora.nla.gov.au/)

The digital collections of the National Library of Australia include full-text databases, online Australian government publications, archived websites, and online copies of significant Australian material in traditional formats – photographs, paintings, cartoons, transparencies, negatives, postcards, maps, printed music, manuscripts, books and journals. PANDORA is probably the most ambitions and challenging DL project.

PANDORA, Australia's Web Archive, is a growing collection of copies of Australian online publications, established initially by the National Library of Australia in 1996, and now built in collaboration with nine other Australian libraries and other cultural collecting organizations. The purpose of the PANDORA Archive is to collect and provide long-term access to selected online publications and web sites that are about Australia, are by an Australian author on a subject of social, political, cultural, religious, scientific or economic significance and relevance to Australia, or are by an Australian author of recognized authority and make a contribution to international knowledge.

Access: There are a number of different ways to find resources in PANDORA. All archived titles are catalogued and a record is included in the National Bibliographic Database (Kinetica), as well as in partners' online catalogues. Access is provided via hotlinks in the catalogue record. Access is also provided via subject and title lists on the PANDORA Home Page, where there is also a search facility for the Archive.

Legal deposit: While the National Library relies on the legal deposit provisions of the *Copyright Act 1968* for deposit of one copy of each edition of a print work with the Library. Since these provisions do not yet cover electronic publications, PANDORA partners seek the permission of publishers before copying a title into the Archive.

PANDORA Digital Archiving System (PANDAS): To support the acquisition and management of increasing volumes of data, as well as to support more efficient distributed archive building among partners, the Library developed the PANDORA Digital Archiving System (PANDAS). PANDAS supports following workflows defined by the staff of the National Library's Digital Archiving Section:

- Identifying, selecting and registering candidate titles;
- Seeking and recording permission to archive;
- Setting harvest regimes;
- Gathering (harvesting) files;
- Undertaking quality assurance checking;
- Initiating archiving processes; and
- Organising access, display and discovery routes to, and metadata for, the archived resources.

These workflows are supported by means of the following functions:

- The management of administrative metadata about titles that have been either selected for archiving, rejected, or are being monitored pending a decision;
- The management of access restrictions;
- The scheduling and initiation of the harvesting of titles selected for archiving;
- The management of the quality checking and assurance process and associated problem fixing;
- The preparation and organization of archived instances for public display through title entry pages, and title and subject listings; and
- The provision of defined management reports.

Persistent identifiers: PANDAS assigns a system generated running number to each title when it is registered. This number becomes part of the persistent URL applicable to each archived title's entry page. The PANDAS persistent URL is generated according to a schema developed by the National Library for its digital collections.

PANDORA archive size and growth: As on March 2005, the archive consisted of over 8,000 titles stored in over 23 million files, occupying a storage space of over 818 GB.

Policies and procedures: PANDORA operations are guided by a set of policy and procedure documents. These include: collection development policy, selection guidelines, persistent identification and resolution system, digital preservation policy, preservation metadata specifications, business process model, and logical data model.

2.4.1.2. Australian Digital Theses system (ADT)

(http://adt.caul.edu.au/)

The aim of the ADT Program is to establish a distributed database of digital versions of theses produced by the postgraduate research students at Australian universities. The theses will be available worldwide via the web. The ideal behind the Program is to provide access to, and promote Australian research in the international community. Initially funded by the Australian Research Council, ADT is now an initiative of the Council of Australian University Librarians (CAUL), participated by a large number of Australian university libraries. ADT partners with the NDLTD programme in becoming part of a growing international network of universities providing access to theses.

Motivation: Approximately 4,000 degrees are awarded each year in Australia. Postgraduate theses represent a significant proportion of Australia's research activity. However, lack of easy access to this information means, that other researchers can wait months or years before papers or books describing aspects of the research are published.

These publications do not always comprehensively cover the valuable information in a thesis, information in many cases is then effectively lost. The ADT programme was initiated with a view to overcome some of these problems.

How does the ADT model work? The local institution uses ADT software to process theses in digital format. The theses are mounted on the institutions' own server/s. The local institution can also integrate access to theses via the local infrastructure, e.g. Web catalogue, provide local search interface, etc. The ADT software automatically generates DC metadata. This metadata is gathered on a regular basis into a central metadata repository/database. The metadata database is searchable using an Australian-developed proprietary search engine, with links provided back to the local institution housing the digitized theses files. The ADT software is based on the deposit process software first developed at Virginia Polytechnic Institute.

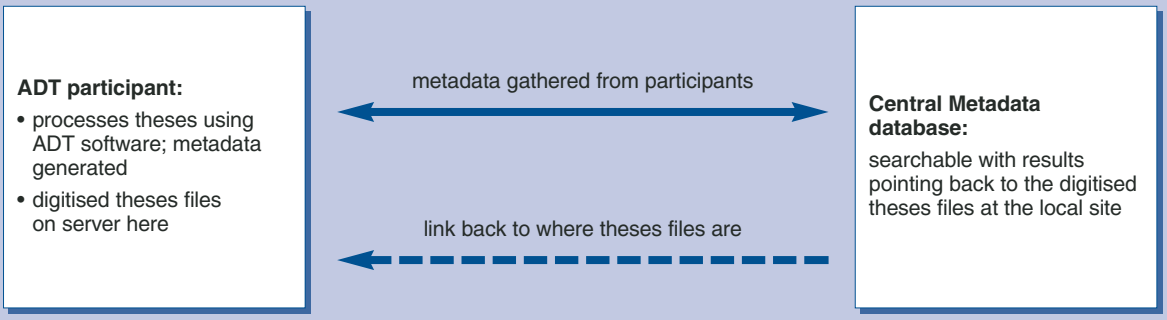

Figure 21. Australian Digital Theses system model

2.4.1.3. ARROW (Australian Research Repositories Online to the World)

(http://arrow.edu.au/)

ARROW (Australian Research Repositories Online to the World) is a three-year consortium project for the establishment of institutional repositories and associated resource discovery mechanisms. Monash University is the lead institution, and the other partner members are the University of New South Wales, Swinburne University of Technology and the National Library of Australia. It is intended that, once the technology infrastructure and workflows are in place, other universities will be invited to participate.

ARROW repositories will include a wide range of digital formats, including pre- and post-publication journal articles and conference proceedings, electronic versions of higher degree theses, and digital collections (such as image databases). The ARROW solution is also intended to support self-published open-access journals. If successful, it is anticipated that these repositories will evolve to include research data, and teaching and learning objects, supported by associated authentication and rights management software.

ARROW's objectives are to create a service framework that achieves national relevance by using interoperable solutions that will have wide applicability, and, by working with national and international standards, projects and initiatives, achieve the following aims:

- Gain high levels of stakeholder engagement with the project's aims, activities and developments;
- Establish affordable, sustainable and cost-effective solutions supported by institutional commitment and appropriate business models;
- Implement hardware and software that is, in its elements and the way it is put together: scalable, modular, interoperable, open-standards based, referencing a common services layer, supporting value-added services, implementing digital rights management and taking into account archiving and preservation issues;
- Achieve a critical mass of content;
- Enhance access to and increase the use of the scholarly research output of participating institutions;
- Improve scholarly communication by encouraging open or low-cost access to research output in a copyright environment that protects the individual's intellectual property and enhances the impact of their research.

The ARROW project is funded by the Australian Commonwealth Department of Education, Science and Training, under the Research Information Infrastructure Framework for Australian Higher Education. The ARROW Project is sponsored as part of the Commonwealth Government's Backing Australia's Ability.

The ARROW Consortium comprises Monash University (lead institution), the National Library of Australia, the University of New South Wales, and the Swinburne University of Technology.

Under the ARROW Project, a few repositories are already operational (http://arrow.edu.au/repositories) and an ARROW search of Australian repositories is available (http://search.arrow.edu.au/).

2.4.2. DL Initiatives and Programmes

The Systemic Infrastructure Initiative (SII) is providing $246 million over 5 years to upgrade research infrastructure at Australian universities. As part of SII the Australian Research Information Infrastructure Committee (ARIIC) was established in August 2003.

ARIIC's Vision

The Australian Research Information Infrastructure Committee is the principal body advising the Australian Government on research information issues. Its goals are:

- To improve the access of Australian researchers to the information they need to carry out their research;
- To make the results of Australian research widely available and easily accessible.

Key areas of interest for the Committee are:

- Development of principles to increase the accessibility of research information;
- Promotion of standards-based approaches to information discovery, storage and sharing;
- Fostering practical demonstrations of integrated information management solutions;
- Promoting cross-institutional, cross-sectoral and transnational cooperation in research information management;
- Securing improved access to key information resources, including major research data sets and databases of research publications;
- Advocacy in assisting researchers embrace the potential offered by an improved information infrastructure.

The ARIIC advises the Government on the information infrastructure requirements of the Australian higher education sector and their intersection with the wider information and technical infrastructures used by the scholarly and research community. A number of projects were announced in 2003 and 2005 (see boxes) in the area of research information at various universities.

2003 Federated Repositories of Digital Objects (FRODO) Projects

Meta Access Management System Project (MAMS)
Lead Institution – Macquarie University

The Australian Research Repositories Online to the World (ARROW)
Lead Institution – Monash University

Australian Digital Theses Program Expansion (ADT)
Lead Institution – University of New South Wales

Australian Partnership for Sustainable Repositories (APSR)
Lead Institution – Australian National University

Many of the funded projects comply with more than one of the areas identified and support many of the themes and problems that are emerging in the areas of access to scholarly literature.

2005 Managed Environment for Research Repository Infrastructure (MERRI) Projects

A. MANAGING AND INTEGRATING LARGE DATA SETS

Time Sync: Mapping the Global Financial System
Lead Institution – University of New South Wales

BlueNet: The Australian Marine Science Data Network
Lead Institution – University of Tasmania

Molecular Medicine Informatics Model (MMIM): A Multi–Institutional, Multi–disciplinary Research and Training Platform for Clinical Research
Lead Institution – University of Melbourne

B. TECHNICAL DEVELOPMENT AND DEPLOYMENT PROJECTS

Australian Service for Knowledge of Open Source Software (ASK–OSS)
Lead Institution – Macquarie University

Middleware Action Plan and Strategy (MAPS)
Lead Institution – University of Queensland

Legal Protocols for Copyright Management: Facilitating Open Access to Research at the National and International Levels
Lead Institution – Queensland University of Technology

C. TECHNICAL INTEROPERABILITY AND ACCESSIBILITY PROJECTS

Dataset Acquisition, Accessibility and Annotation e–Research Technology Project (DART)
Lead Institution – Monash University

E–Security Framework for Research
Lead Institution – University of Queensland

Regional Universities Building Research Infrastructure Collaboratively (RUBRIC)
Lead Institution – University of Southern Queensland

2.4.3. Theories, Methods and Technologies: Focus Areas

DL related R&D focus areas in Australia

Research focus	DL project	Remarks
Distributed and interoperable data archiving	BlueNet	Highly distributed archiving facility to support the long term data curation requirements of Australia's marine science researchers.
Preservation and sustainable DLs	Australian Partnership for Sustainable Repositories (APSR)	APSR will focus on accessing and sustaining digital collections.
Networking of Information Systems	Meta Access Management System Project (MAMS)	This project is developing the software for creating better linkages between university information technology systems.
Institutional repositories – for archiving and disseminating institutional research output (publications)	ADT	Redevelop existing central repository of the Australian Digital Theses Program (ADT) to increase its coverage and utility.
Open Access archiving and publishing	ARROW	Development of repository and the enabling metadata to support independent scholars, as well as those associated with institutions.

Research focus	DL project	Remarks
Collaboration between University institutional repositories	Regional Universities Building Research Infrastructure Collaboratively (RUBRIC)	RUBRIC takes on the best practice developed from the existing FRODO projects and develop similar institutional repositories in thirteen other universities.
Secure Information access	Secure access and the authentication and authorization of researchers	Secure access and the authentication and authorization of researchers, who access services and infrastructure across global networks.
Grid Computing	Dataset Acquisition, Accessibility and Annotation e–Research Technology (DART)	The DART project is a significant and strategically important project of interest to many institutions who are taking up the challenge of e–Research.
Intellectual Property Rights	Legal Protocols for Copyright Management	Develop a set of legal protocols and generic licences that can be used across universities to facilitate and break down barriers to open access to copyright material.

2.5. New Zealand

2.5.1. DL Examples and Projects

2.5.1.1. New Zealand Digital Library Project

(http://www.nzdl.org)

The New Zealand Digital Library project is a research programme at The University of Waikato whose aim is to develop the underlying technology for digital libraries and make it available publicly so that others can use it to create their own collections.

The collections at the New Zealand Digital Library Project (http://www.nzdl.org) consist of a number of humanitarian and United Nations collections like "Food and Nutrition library", "WHO health library for disasters", "Agricultural information modules", etc. Most of these are also available on CD-ROMs. There is also a set of demonstration collections, showing the use of the Greenstone digital library software for various types and formats of collections. This includes image and video collections, music library, collections in various languages like Arabic, Chinese and Maori, newspapers and journal collections and unique collections like "acronym extraction demo" and "language extraction demo". Collections that have been documented are grouped separately.

The digital library site provides several collections, including historical documents, humanitarian and development information, computer science technical reports and bibliographies, literary works and magazines. All are available over the web, and can be accessed through searching and browsing interfaces provided by the Greenstone digital library software. Greenstone documentation collection is also available separately.

The Greenstone software (http://www.greenstone.org/) is Digital Library software for organising information and making it available over the Internet or on CD-ROM. It allows collection building, indexing and searching by browsing of document collections. It is open-source software, available under the terms of the GNU public license.

2.6. Other Examples

2.6.1. Scirus

(http://www.scirus.com/)

Scirus is the most comprehensive science-specific search engine on the Internet. Driven by the latest search engine technology, Scirus offers unique functionalities designed for scientists and researchers. It searches over 200 million science-specific web pages, enabling users to quickly pinpoint scientific, scholarly, technical and medical data on the web, and to find the latest reports, peer-reviewed articles, patents, pre-prints and journals that other search engines miss.

Scirus is different from other search engines as it focuses only on web pages containing scientific content. It searches more than 200 million science-related pages and filters out non-scientific sites. It provides 2 search options – Basic and Advanced.

Basic Search is a simple keyword search where one can also choose to display the results only from Journals or only from web sources. By default, it combines the results from both.

Advanced Search is having the option of searching in the Complete Document; Article Title; Journal Title; Author(s) Name; Author Affiliation(s); Keyword; ISSN; and (part of a) URL. Further one can restrict the results to a specified date range and narrow it down by Information Types, such as Abstracts; Articles; Books, Company Homepages; Conferences; Patents; and Preprints, etc., or by File Formats (HTML and PDF). One can also do a search in selective Content Sources, such as ScienceDirect, BioMed Central, Medline, ArXiv, Patent Offices, NDLTD, etc. – individually or combined. One can also select to search in a range of subject areas including health, life, physical and social sciences.

Scirus automatically rewrites the query to improve the results and also supports clustering. Users have the option of setting the search preferences and of saving the results. Search results in Scirus are, by default, ranked according to relevance. It is also possible to rank results by date. Scirus uses an algorithm to calculate ranking by relevance. This ranking is determined by two basic values:

- Words – the location and frequency of a search term within a result account for one half of the algorithm. This is known as static ranking.
- Links – the number of links to a page account for the second half of the algorithm – the more often a page is referred to by other pages, the higher it is ranked. This is known as dynamic ranking. Overall ranking is the weighted sum of the static and dynamic rank values. Scirus does not use metatags, as these are subject to ranking-tweaking by users.

2.6.2. Google Scholar

(http://scholar.google.com/)

Google Scholar allows us to search specifically for scholarly literature, including peer-reviewed papers, theses, books, preprints, abstracts and technical reports from all broad areas of research. Google Scholar can be used to find articles from a wide variety of academic publishers, professional societies, preprint repositories and universities, as well as scholarly articles available across the web.

How Google Scholar works

Google Scholar orders the search results by relevance, so the most useful references should appear at the top of the page. This relevance ranking takes into account the full text of each article, as well as the article's, author, the publication in which the article appeared and how often it has been cited in scholarly literature. Google Scholar also automatically analyses and extracts citations and presents them as separate results, even if the documents they refer to are not online. This means that search results may include citations of older works and seminal articles that appear only in books or other offline publications.

Google Scholar provides the support for libraries by supporting Institutional Access and Library Search facility.

Institutional Access: It offers the option to include a link of library resources available via a link resolver as a part of the Google Scholar search results. On-campus users at participating schools will see additional links in Google Scholar search results, which facilitate access to their library's resources. These links lead to the library's servers, which, in turn, direct them to the full-text of the article.

Library Search: This is available to libraries that have their holdings listed in OCLC's Open WorldCat. It gives a link for each Google Scholar book result that leads to the Open World database where users can find the book in a local library. All users of Google Scholar will see a "Library Search" link for book results. Clicking on this link will direct them to the WorldCat system, which will allow them to find a list of nearby libraries that have the desired book.

Google Scholar also supports publishers by boosting the worldwide visibility and accessibility of their content, which increases the traffic to their website. One condition for the publisher is that they must offer the complete abstracts of items.

3. Digital Libraries in Science: Trends, Needs and Challenges

3.1. Trends and Challenges Identified from DL Examples

National level Internet Resources Gateway

The "Resource Discovery Network" and the "Virtual Libraries" (ViFas) in German are gateways to quality resources on the web. The vast and unrestricted growth of the Internet has made it difficult to identify authentic and quality information on the web. Experts help identify and categorize information available on the web and make it available through these portals and gateways.

Content Creation

Other than gateways, NSDL has sponsored several research projects that look to generate content for digital libraries employing new methods. For example, by using competitive intelligence, capturing lectures from professional society meetings, reviewing materials collaboratively for quality, etc.

DL Services

A number of traditional library services are also provided by DLs. For example, the virtual reference desk is based on the reference service provided by libraries. Additionally new forms of service, such as personalized content delivery, recommendation systems, visualization of data, etc., are being developed.

Aggregation of services and interoperable DLs

NSDL, eBank UK and vascoda are examples of how diversified DL systems are integrated to form a national level DL. While the content and services are developed in a distributed manner, some features, such as uniform metadata, search and, retrieval services, are provided by the integrated portal. Other than national level integrated portals, digital libraries (for example, institutional repositories) also provide their own interoperable services, such as aggregation based on harvesting mechanism or based on common user profiles.

Open Archives Initiative

OAI-PMH is fast being adopted as a standard for interoperability of content. It is being used for integrating distributed digital libraries or repositories in many of the examples studied.

Institutional Repositories

A large number of academic institutions across the world now have their own institutional repositories. FAIR (Focus on Access to Institutional Resources), ePrints UK and Australia's ARROW and RUBRIC projects show the development of repositories in institutes of higher education.

Metadata Schemas and application development

While it is generally agreed now that different metadata schemas may be required to describe different sets of resources, there is a need to develop standards for each one. When designing integrated or distributed DLs there is also a need to set up minimum standards that all the DLs have to meet. NSDLs Metadata Repository is central to its portal and NSDL has sponsored several research projects for metadata development. For example, MetaTest is for evaluating the utility of metadata, Quality Analysis of Metadata Records in the NSDL Metadata Repository, Metadata Generation for automating classification and metadata generation; Dissertations Online (Germany) has developed MetaDiss standard; JISC IE Metadata Schema Registry is used for describing resources of relevance to teaching, learning and research.

Digital Preservation and access

From persistent URLs to MD5 hash, storage in multiple copies (LOCKSS) etc. are used for reservation of digital documents for long term. The JISC has established a Digital Curation Centre to look into all aspects of long-term preservation of digital materials. A report on e-science preservation http://dev.dcc.rl.ac.uk/twiki/bin/view/Main/LordMacdnaldReportRecommendations at the Digital Curation Centre makes several strategic and tactical recommendations for preservation of digital data.

Grid computing

In the UK, The JISC (http://www.jisc.ac.uk/) and the Engineering and Physical Science Research Council (http://www.epsrc.ac.uk/) have both undertaken to fund projects to progress the development of a Semantic Grid. Australia's DART project is also aimed at using datasets in research institutions.

Geographical information systems

Geographical and geospatial information systems are being developed. Integrating different types of information related to a particular geographic location as layers is a concept that has gained significance.

Security and Access

Traditional libraries were accessible to everyone, but the growth of DLs has raised new issues of intellectual property rights. While the social and legal implications are studied and addressed, it is also necessary for DLs to develop technological solutions to allow only authorized people to view certain content. In NSDL access to some materials is restricted. Hence an access management service based on the Shibboleth protocol enforces an access mechanism that allows only authorized user to view such items in the library. There is a similar project sponsored by the ARIIC in Australia.

Information Retrieval

Searching, extracting and accessing textual information from large-scale resources is being researched at the National Centre for Text Mining in the UK. Content Based Retrieval for images, audio and video collections are also challenges that are being addressed by various DL systems. BioMed Image Archive and Dental Images in the UK and DIRECT in the USA are some of examples of projects researching image retrieval systems.

3.2. Trends and Challenges Identified from International DL Conferences

Conferences provide a bottom-up approach for identifying trends in DL research. The session headings of three international conference series from 2001 to 2004 were noted and similar topics grouped together. The conferences are:

JCDL: Joint (IEEE/ACM) Conference on Digital Libraries
ECDL: European Conference on Digital Libraries
ICADL: International Conference of Asian Digital Libraries

This methodology helps identify what problems are being tackled by researchers today. It also helps to see if a certain topic is gaining or losing importance. For example "Digital Preservation" was grouped with other topics in 2002 ECDL conference but has a separate session in 2003. This shows the increasing attention this area of research work is getting. The following are the major topics identified. The conference sessions (and if available their numbers) are also given under each heading.

Academic Publishing

Scholarly publishing is going through tremendous change. Several factors like easy accessibility of technological tools and the serials crisis have made academic institutions take their publishing seriously. Libraries, as storehouses of scholarly literature, often become involved in publishing. Academic digital libraries, in addition to looking at technological solutions for production, storage and access of scholarly material, also tackle intellectual property problems, measuring usability and other issues involved in digital scholarly publishing.

2003 ECDL Session 4 – Panel 1: The Future of Academic Publishing
Plenary 2002 ECDL Session 10 – Panel 2: International Perspectives on Creating a Usability Methodology for Academic DLs

Architecture and design

Just as traditional libraries, each digital library has its own unique needs depending on the type and volume of content, users, usage, etc. Most libraries have huge amounts of text-based information and need effective text organization, indexing and retrieving mechanisms. An art library may have deal with pictures, an archaeological library may have to store three-dimensional images of objects, and a geographical library may have to deal with a host of variables regarding a location on a map.

The software that form the digital libraries therefore need to be designed with different architectures based on the types of libraries they cater to, the amount and type of content they have to store, and the mechanisms they need to locate the right information from their store. User-centred design needs to focus on allowing the users to create and customize the digital library interface, etc.

2004 JCDL Session: Repository architectures
2004 ECDL Session 2A: Digital library architectures
2004 ECDL Session 12A: User-centred design
2002 ECDL Session 8A: Architecture I
2002 ECDL Session 9A: Architecture II
2002 ECDL Session 13A: Architecture III
2003 ECDL Session 7A: Architectures and systems
Panel 2004 Session: Cross-cultural usability of digital libraries: user-centered design for the global society
2003 ICADL Session: Machine architecture and organization

Books

2004 JCDL Session: Books and reading
2002 ECDL Session 2B: e-Book

How e-Books can be incorporated in a digital library and used effectively by a user in his or her work environment is an issue that needs much discussion.

Collection Development and Management

Libraries are about collections. How to build and mange these are discussed in these sessions. Obtaining digital content or converting content to a digital format, that can be easily accessible and read by the end users, is itself a huge task. Tagging them with the appropriate metadata also needs to be done while ingesting the content in the DL. These tasks are manpower intensive. The answer is to automate the processes. The varied nature of content in a digital library and hence the difficulty in finding universal “patterns” for machine processing, makes the task very challenging.

2004 JCDL Session: Collaboration and group work
2004 JCDL Session: Surrogates for physical artifacts
2004 JCDL Session: Automated techniques for managing collections
2003 JCDL Session: Managing resources and services
2003 ECDL Session 8A: Collection building and management
2002 ECDL Session 3A: Collection building
2003 ICADL Session: Development of contents

Collaborations

There are numbers research organizations, agencies and academic institutes working in the area of digital libraries. There is a tendency to develop a specialization in a particular domain or handling a specific problem posed by digital libraries. Hence the need for collaboration between different organizations. Also a generic model, architec-

ture or piece of software may need to be modified to meet local requirements (e.g. a generic DL search engine to search Chinese language text). A synergy can thus be created by sharing the knowledge and working in collaboration.

2004 ICADL Session: Collaboration and localization
2004 ICADL Session Poster: Collaboration and localization

DL Initiatives and Usage

Federal governments and international agencies are funding several initiatives in DLs. The experiences and specific project developments are shared in the DL conferences, either as reports of progress made or as case studies. In fact, entire sessions are devoted to dealing with large initiatives like NSDL.

2002 JCDL Session: NSDL
2002 ECDL Session 4B: Case studies
2003 ICADL Session: Building and using digital libraries
2002 ICADL Session: Building and using digital libraries
2002 ICADL Session: Posters: Digital library initiatives and services

DL Infrastructure

Digital libraries are infrastructure intensive. The complexity increases as volume of content, usage traffic and enhances tools and interfaces become available. Therefore some sessions in the DL conferences deal with infrastructure issues.

2003 ICADL Session: Digital library infrastructure
2002 ICADL Session: Digital library infrastructure

DL Technology

Digital library is one of the outcomes of information technology. As new innovations arise, for example in cryptography, compression techniques or protocols, they need to be explored to see how they can benefit, enhance or make more secure the digital library experience to the end user.

2004 ICADL Session Poster: Digital library technology
2002 ICADL Session Posters: Technology
2003 ICADL Session: Cryptography and compression
2002 ICADL Session: Algorithms and protocols

Digital Preservation

The volatile nature of digital content and the tools to render and retrieve the content makes preservation a very important challenge. Preserving for posterity is one of the major concerns of digital libraries. This research problem is gaining in importance with several initiatives in the US, UK and other parts of Europe specifically devoted to finding solutions.

2004 JCDL Session: Digital preservation
2002 JCDL Session: Preserving, securing and assessing digital libraries
2003 ECDL Session 11B: Digital preservation
2003 ECDL Session 12B – Panel 3: Digital preservation: are metadata really crucial?
2002 ECDL Session 12B: Preservation, classification, user studies

Education and DLs

How digital libraries are used in a virtual learning environment has gained considerable importance in recent years. The opportunities and challenges created for digital libraries by distance, mobile and e-learning both in traditional mainstream education and also for continuing education are being explored.

2004 JCDL Session: Educational aspects of digital libraries
Panel 2004 JCDL Session: Demonstrating education impact: challenges in the years ahead
2003 JCDL Session: Digital libraries in the classroom
2002 JCDL Session 5A: Digital libraries for education
2002 ECDL Session Workshop 1: E-books + e-readers + e-journals = e-education?
2003 ICADL Session: E-learning and mobile learning

Evaluation

For a long time the emphasis of digital library research was on creation of digital libraries. In the recent years, there has also been a stress on its actual usage by the end users. Evaluation on either the usability of DLs or the actual usage of DLs are being studied.

2004 JCDL Session: Evaluation
2003 JCDL Session: Correction and analysis
2004 ECDL Session 2B: Evaluation & usability
2002 ECDL Session 9B: Evaluation
2003 ECDL Session Workshop: CLEF 2003 – Workshop of the cross-language evaluation forum

Future of DLs

At many conferences there are panel discussions or sessions by eminent persons from the area on where digital libraries are going. A vision or a visualization of the future digital library often sets the research topics for today. It defines the path we need to follow. Emerging standards and technologies are used to see how they will affect DLs of the future.

Panel 2004 JCDL Session: Library leaders on digital libraries and the future of the research library: a panel discussion
Panel 2004 JCDL Session: Digital libraries settling the score: 10 years hence and 10 before
2003 ECDL Session Workshop: Digital library evaluation – metrics, testbeds and processes
Panel 2004 Session: Semantics, syndication and social networks: mechanisms for future structured information spaces

Geographical/Spatial DLs

Geographical and spatial DLs are unique in their content and ways of access and presentation of data. Much work has been done in this area.

2004 JCDL Session: Geographic aspects of digital libraries
2002 JCDL Session: Digital libraries for spatial data

Human Resource/Training

Not much is being discussed about the training of personnel to manage DLs. Only one session in 2003 at the ICADL concentrated on this aspect.

2003 ICADL Session: Human resource and training

IPR

Similarly one session in ICADL 2003 was allotted for intellectual property rights in the context of digital libraries.

2003 ICADL Session: Intellectual property rights and copyright

Information Retrieval (IR)

The tools and techniques for retrieval of the right content for the user is probably the most important feature of any digital library. It is a very challenging area because of the large variety and volume of content being stored in digital libraries. Add to this indexing and searching across many languages and translating queries or text. Summarization of

large texts, datamining and question-answering requiring machine understanding, searching for images based on form or colour, etc. are major challenges.

IR – Indexing, Searching and Querying

2004 JCDL Session: Search and query strategies
2004 JCDL Session: Automatically structured and translated queries
2002 JCDL Session: Novel search environments
2002 JCDL Session: Searching across language, time and space
2004 ECDL Session 3B: New approaches to information retrieval
2004 ECDL Session 5B: Enhanced indexing & searching methods
2002 ECDL Session 7B: Audio/video retrieval
2002 ECDL Session 8B: IR
2003 ECDL Session 11A: Information retrieval in different application areas
2003 ECDL Session 12A: Indexing and searching of special document and collection information
2003 ICADL Session: Information retrieval in asian languages
2002 ICADL Session: Information retrieval in asian languages
2004 ICADL Session: Information retrieval techniques
2003 ICADL Session: Information retrieval techniques
2002 ICADL Session: Information retrieval techniques

IR – Data Mining

2004 JCDL Session: Mining and disambiguating names
2003 JCDL Session: Information retrieval and data mining
2003 ICADL Session: Data storage and retrieval
2003 ICADL Session: Data mining in digital libraries
2002 ICADL Session: Data mining in digital libraries

IR – Other

2002 JCDL Session: Summarization and question answering
2002 ECDL Session 7A: Navigation/query language
2003 ECDL Session 6B: Topical crawling. Subject gateways

Interoperability

DLs and the protocols for interoperability are discussed in some conferences in separate sessions.

2004 JCDL Session: Interchange and interoperability
2002 JCDL Session: Federating and harvesting metadata
2004 ECDL Session 5a: Interoperability

Knowledge Representation

Good organization of knowledge is a essential requirement for effective retrieval. Concepts of knowledge representation, classification, indexing and discovery have taken up many sessions in DL conferences.

2003 JCDL Session: Knowledge and representation
2002 JCDL Session: Classification and browsing
2003 ECDL Session 3B: Automatic classification and indexing
2003 ECDL Session 7B: Knowledge organization: concepts
2003 ECDL Session 8B: Knowledge organization: authorities and works
2004 ICADL Session: Multimedia processing and knowledge discovery in digital library
2004 ICADL Session: Knowledge organization and representation
2003 ICADL Session: Knowledge management

Metadata and Standards Issues

Automatic metadata creation, metadata standards and applications are much discussed topics in most conferences.

2003 JCDL Session: Automatic metadata creation
2003 JCDL Session: Standards, mark-up and metadata
2003 ECDL Session 2B: Metadata applications
2003 ICADL Session: Metadata issues
2002 ICADL Session: Metadata issues
2003 ECDL Session Workshop: Networked Knowledge Organization Systems/Services (NKOS): evolving standards

Models/Tools for DLs

The tools and models for creating digital libraries have had separate sessions in JCDL (2002 and 2003).

2003 JCDL Session: Tools for building digital libraries
2002 JCDL Session: Models and tools for generating digital libraries

Multimedia DLs

Separate sessions in all DL conferences have been devoted to digital libraries containing multimedia content and the problems and challenges in this area.

2002 JCDL Session: Image and cultural digital libraries
2004 JCDL Session: Image and video digital libraries
2003 JCDL Session: Multimedia issues in digital libraries
2002 JCDL Session: Video and multimedia digital libraries
2002 ECDL Session 12A: Multimedia/mixed media
2003 ICADL Session: Multimedia digital libraries
2002 ICADL Session: Multimedia digital libraries

Music DLs

Similarly, Music DLs also have deserved special attention because of the nature of their content and the tools for storage and retrieval of content. In JCDL 2004 it has been combined with the ”Chinese text” issue.

2004 JCDL Session: Indexing music and Chinese text
2003 JCDL Session: Music and digital libraries: from users to algorithms
2002 JCDL Session Music digital libraries
2004 ECDL Session 7B: Music digital libraries

OAI

The Open Archives Initiative, together with one of the most popular protocols for interoperability – the OAI-PMH, has been discussed extensively.

2003 JCDL Session: OAI in action
2002 JCDL Session: OAI application
2004 ECDL Session 10A: Open Archives Initiative (OAI)
2002 ECDL Session 4A: OAI Applications

Types of DLs

DLs in specific domains, such as science, culture, humanities or catering to special needs, have been discussed in a number of sessions.

2003 JCDL Session: Designing and accessing scientific digital libraries
2002 JCDL Session: Building and using cultural digital libraries
2003 ECDL Session: Workshop: digital libraries in healthcare
2002 ECDL Session: 13B: Humanities
2003 ICADL Session: Special purpose digital libraries
2002 ICADL Session: Special purpose digital libraries
2004 ECDL Session 9A: Personal digital libraries

Personalization

Personalization is one of the advanced features of DLs which allows the user to access a DL in a way that is unique to the user. In the recent years it has gained importance.

2004 JCDL Session: Supporting personalization
2004 ECDL Session 7A: Personalization & annotation
2003 ECDL Session 3A: Annotation and recommendation
2004 ICADL Session: Personalized issues in digital library

Services

As DLs advance in features, they no longer are just sites to access content. They can push services to the users. Some of these are similar to those provided by traditional libraries (like SDI-Selective Dissemination of Information) or new to the digital environment (like allowing users to rank content).

2004 ICADL Session: Service and management
2004 ICADL Session Poster: service and management
2003 ICADL Session: Digital library services
2002 ICADL Session: Digital library services

User Interface design

Humans and the machines (DLs) interact through interfaces. Good design and presentation of interfaces help effective communication between the two. A user is able to get more from a DL if the interface is intuitive and pleasant to use.

2002 JCDL Session: Novel user interfaces
2004 ECDL Session 3A: User interfaces & presentation
2003 ICADL Session: Human computer interface
2002 ICADL Session: Human computer interfaces

User Studies/Interaction

Studying users and communities of users of DLs before and after the creation of DLs is considered essential today. We check their requirements for designing a DL that will be most useful to the user/community. After the DL is built we see how effectively they use it. User studies are therefore very important components of DLs.

2003 JCDL Session: User interaction
2002 JCDL Session: Studying users
2002 JCDL Session: Digital library communities and change
Panel 2004 JCDL Session: The virtual and the real: panel on current research on museum audiences and library users
2003 ECDL Session 2A: Uses, users and user interaction
Plenary 2002 ECDL Session 6: Uniting communities: the digital library as a practical tool
2002 ICADL Session: Digital libraries for community building
2004 JCDL Session: Interacting with collections

Web Archives & Web Technologies

The web itself is a huge but highly disorganized library of material. One of the methods employed to create a DL is to collect "good and relevant" content from the web and organize it "well". Hence web technologies and archiving techniques on the web have been given special attention at DL conferences.

2004 JCDL Session: Crawling the web
2003 ECDL Session 6A: Web technologies
Plenary 2002 ECDL Session 1: WebBase: Web capture and distribution
2002 ECDL Session 3B: Web technologies
2003 ECDL Session Workshop: Web archives
2002 ECDL Session 2A: Web archiving
2002 ECDL Session Workshop 2: Web archiving

3.3. Trends and Challenges Identified from DELOS

DELOS is a Network of Excellence on Digital Libraries; it is a framework for international cooperation on research activities and research agendas in the digital library domain. Funded by the EU's Sixth Framework Programme, its activities started in January 2004.

DELOS has a number of Research clusters, working on topics of current relevance in DLs. The activities of these research clusters are reported through publications or through the newsletter of DELOS.

Cluster 1: Digital Library architecture
(http://www.delos.info/WP1.html)

This cluster is working on the following topics:

- Architectures for digital libraries;
- Standards and protocols;
- Mobile information and information dynamics.

Such studies are done by organising workshops on DL architecture, conducting surveys of significant contributions in DL architectures, comparison and feasibility studies, evaluations of approaches to information management, development of standards, protocols and benchmarks.

The first workshop looked at the following DL Architectures:

- Peer-to-Peer Data Management;
- Grid Computing Middleware;
- Service-Oriented Architecture.

Cluster 2: Information access and personalization
(http://www.delos.info/WP2.html)

In DELOS, Information Access in Digital Libraries is studied from three different aspects:

- Information Access: interaction with a single information provider;
- Information Integration: interaction with multiple information providers;
- Personalization: customization of information and interaction to the user.

This cluster's objectives are to promote knowledge and research about the topics and create a common understanding of problems and methodologies involved. It also aims at the construction of a common, comprehensive framework for information access and personalization approaches. Current activities are organized into the following tasks:

Task 1: Creation of Common Foundation for Information Access
This task seeks to create a common conceptual and infrastructural foundation in respect of Information Access, and is subdivided into three sub-tasks:

T1.1: Information and Interaction Models and Processing Schemes;
T1.2: Metadata Models;
T1.3: Accessing time/locality information.

Task 2: Creation of Common Foundation for Information Integration
Here the aim is to achieve a common conceptual and infrastructural foundation in respect of Information Integration which deals with multiple, heterogeneous DLs that need to be treated in a cohesive fashion. It is subdivided into three sub-tasks:

T2.1: Integrated Interaction Management Schemes;
T2.2: Integrated Metadata;
T2.3: Data Provenance.

Task 3: Creation of Common Foundation for Personalization
This task aims to create a common conceptual and infrastructural foundation with regard to Personalization and Customization of the behaviour of a DL system. It is subdivided into two sub-tasks:

T3.1: User Modelling for Personalization;
T3.2: Content, Service and Interaction Personalization.

Cluster 3: Audio-visual and non-traditional objects
(http://www.delos.info/WP3.html)

This cluster focuses on metadata capture for audio-visual content, universal access and interaction with audio-visual libraries, together with the management of audio-visual content in digital libraries.

The objectives of this cluster are to establish a common ground of knowledge for European researchers about the state of the art, to identify the direction that research will take, and to realize important new applications for digital libraries with audio-visual and non-traditional objects.

Other than the management of the cluster and establishment of an audio-visual forum, the cluster aims to assemble demonstrators and test beds to demonstrate innovative solutions and applications for audio-visual libraries; investigate automatic and semi-automatic metadata extraction; explore information access and interaction with audio-visual content for different media and devices, and explore users preferences about the content; and also study management aspects of audio-visual data by experimenting with prototype systems.

Cluster 4: User interfaces and visualization
(http://www.delos.info/WP4.html)

The goal of the cluster is to develop methodologies, techniques and tools to enable future digital library (DL) designers and developers to meet the user-oriented requirements in a balanced way. This is done by studying two different aspects:

User Requirement-Related Activities: The end-users and other DL stakeholders, such as librarians, content providers and maintainers, are studied. The DL life cycle will be related to both functional and non-functional requirements. They plan to develop a set of character profiles of DL users and also explore how users can make use of a multi-modal DL-user interface which will meet their particular needs.

User Interface and Visualization (UIV) Design Activities: The cluster will carry out studies towards developing taxonomy of relevant context models and other related topics.

The UIV cluster will investigate the exploitation of existing visualizations, as well as consider novel visualizations and how they present DL results/views; it also expects to examine certain aspects of the DL life cycle. The cluster will build a theoretical framework from which user interface designers and developers can design interfaces for DL-users.

The goal is to develop a generic user interface, which will enable a design methodology, and associated guidelines, which will help define appropriate technical solutions. The aim is to create an integrated DL architecture which combines both user- and application-oriented functions, for example, query and navigation features, and which will adapt to the needs of its users.

The final integrated results of this work should be the theoretical framework for user interface developers together with the design methodology and supporting documentation which will allow designers to create specific technical solutions. Additionally, they also aim to provide profiles of the different classes of users, together with analyses of the DL life cycle.

Cluster 5: Knowledge extraction and semantic interoperability
(http://www.delos.info/WP5.html)

Semantic interoperability applies to the application of different vocabularies and terminology used in descriptions of digital objects for both learning and research, collections of those objects, collections of datasets and resources used in the wider cultural heritage sector and in e-research.

In addition, the application of algorithms for the mining and analysis of digital resources (text, data, complex objects) offers exciting opportunities for the extraction of new knowledge and the reuse of data and information in new ways.

The Knowledge Extraction and Semantic Interoperability research cluster has two key strategic goals: to coordinate activities, that facilitates the sharing of expertise in the practitioners form both DL and Grid computing science backgrounds; and to explore new models, algorithms, methodologies, etc. in a variety of technical applications that will lead to creation of guidelines and recommendations of best practice for dissemination to the communities of interest.

This is achieved through a series of activities, such as creating a forum for the exchange of experience and research in the areas/themes of the cluster. The meeting of the forum would end in an evaluative report and workshops to disseminate the findings. In addition, the cluster aims to conduct studies to determine the requirements for and usage of extracted knowledge for biblio-metrics, domain analysis, issue tracking and community modelling.

Semantic interoperability is being addressed initially by producing a state-of-the-art overview of DL semantic issues including the application of standards, thesauri, anthologies, Knowledge Organization Systems and the implementation of metadata schema registries.

Cluster 6: Preservation
(http://www.delos.info/WP6.html)

The DELOS Preservation cluster aims to interconnect people, organizations and projects working in the areas of digital preservation, coordinate research and projects, identify, collect and disseminate expertise, develop test beds and metrics and tools for evaluating digital preservation strategies, relate research with development of exploitable product, develop links with the industrial sectors, eliminate duplication of effort, ensure that work has a direct impact on DL architecture and working of other clusters.

Cluster 7: Evaluation
(http://www.delos.info/WP7.html)

The Evaluation cluster works on evaluation methodologies in general, as well as on providing the infrastructure for specific evaluations. Thus, the following objectives are addressed: development of theoretical frameworks for DL evaluation; research on new methodologies; and development of toolkits and test beds.

These are achieved by conducting workshops, creating an evaluation forum, and the development of new approaches, methods, toolkits and test beds.

CLEF (Cross-Language Evaluation Forum, http://www.clef-campaign.org/) provides test beds for the evaluation of cross-language information retrieval.

INEX (Initiative for the Evaluation of XML Retrieval, http://inex.is.informatik.uni-duisburg.de:2004/) deals with the evaluation of information retrieval methods for XML documents.

Most of the current activities in this cluster are targeted towards providing the necessary infrastructure for the INEX and CLEF campaigns. As a general infrastructure for DL evaluation, an evaluation forum is under development, which will support communication between DL researchers and evaluation specialists. In the future, research on evaluation models and methods will be enforced in DELOS, along with the development of appropriate evaluation toolkits and test beds.

3.4. Summary

Preservation for long-term sustainability of digital documents is of particular concern to researchers in the area. Use of standards is no longer considered as the only feasible option, nor is the involvement of the industry which provides applications to create documents. The solution may not be a single approach but a multi-pronged one, requiring the involvement of standards, persistent identifiers, etc.

While standards for interoperability, such as OAI and Z39.50, are in use today, new methods of interoperability are being studied. New interoperable, modular architectures for handling massive amounts of data and content in multiple languages and formats are being built. DLs that can provide traditional and new forms of services based on the content are also being built.

Audio, video and question-answering information retrieval systems pose a challenge to digital library developments. Building effective user interfaces with personalization and how the DLs can be interwoven with the work system used by the end user (for example, in education or medical records systems) are also a matter of study. This all leads to a better usage of DLs. Mechanisms to evaluate the effectiveness of DLs are also gaining importance.

There is a profusion of metadata standards in place today. It is generally agreed now that “one size does not fit all”. What is needed instead are systems that can crosswalk between the standards. Automatic metadata creation and extraction are also important areas of study.

The creation and management of digital libraries is a multidisciplinary activity, requiring cooperation among a whole range of specialists from computer engineers, psychologists, librarians, etc. All of them have their roles to play. What is listed above is a list of topics that seem to have captured the interests of researchers in the recent past. By the large the technological issues faced by DLs is the same for science, humanities or any other domains.

4. Digital Libraries in Science: National and International Policies and Initiatives

4.1. National Policies, Initiatives and Programmes

4.1.1. United States of America

Figure 22. NSDL homepage

NSDL: Governance, Policies and Projects

Governance: The NSDL governance model is built around committees, subcommittees, taskforces and interest groups that provide several domains of guidance. NSDL is governed through a leadership structure that is shared among NSDL Core Integration (CI), the National Science Foundation, the NSF-NSDL National Visiting Committee, Assembly and the NSDL Policy Committee. The Core Integration (CI) project, funded in 2002, engages the community to cooperatively develop technical standards, infrastructure and organizational processes for providing resource collections and services in NSDL. CI also operates the NSDL infrastructure and supports projects in community and governance activities. NSF-NSDL National Visiting Committee is appointed by the NSF to provide high-level strategic advice to CI and continuing evaluation of the project to NSF. Assembly is composed of one representative from each project and elects the Policy Committee by majority vote. The Policy Committee advises the CI team, other NSF grantees, and NSF on operational strategies, policies and implementation issues.

NSDL Projects: To realize its vision, the NSDL programme is supporting projects focused on the development or enhancement of resource collections, implementation of digital library services, and a small set of targeted research investigations. The Core Integration (CI) activity mentioned above is also a project-based activity that coordinates distributed resource collection and service providers to ensure reliable and extensible access to and usability of the resulting network of learning environments and resources. NSDL supports projects in the three tracks: ***Pathways***, ***Services*** and ***Targeted Research***.

- **Pathways** projects assume a stewardship role for the educational content and/or the services needed by a broad community of learners.
- Within the **Services** track, two particular types of projects are emphasized: (1) selection services and (2) usage development workshops. Selection services projects focus on increasing the amount of high-quality STEM educational content known to NSDL. Usage development workshops promote the use of NSDL and its resources by various communities of learners.
- **Targeted Research** projects explore specific topics that have immediate applicability to collections, services and other aspects of the development of education digital libraries.

Collection Development Policies and Processes

Several policies guide the development of collections. Two such polices are the "NSDL Collection Development Policies" and "Metadata Policies". The Collection Development Policy document includes the following:

- Definition and objectives of the NSDL collection;
- Clientele;
- Scope in terms of breadth, subject coverage, geographic and language coverage, cultural scope and formats of resources;

- Components of the NSDL collection in terms of resources, resource organization and persistent naming;
- Resource identification and selection responsibility;
- Selection criteria and resource quality in terms of open or closed access, scientific and/or educational quality of the resource, selection criteria, accessibility and resources considered unsuitable for NSDL collection;
- Collection maintenance, evaluation and archiving;
- Initial release strategy;
- Technical options for resource ingest into the NSDL;
- Recommended practices for metadata and metadata harvesting.

The "Metadata Policy for NSDL" document defines the categories of metadata used in the NSDL to describe collections and individual educational resources within collections for provision of digital library services and the administration of those resources. The policy defines the rights of third parties (including collection holders) and the NSDL with regard to the display and distribution of metadata held by NSDL.

NSF International Digital Library programme: One of the programmes of NSF is to fund joint research programmes on International DLs.

This activity is supported by the Division of Information and Intelligent Systems of the Directorate for Computer and Information Science and Engineering and by the Division of International Programs of the Directorate for Social, Behavioral and Economic Sciences. It builds on and extends prior Foundation efforts in digital library research. It focuses on multi-country, multi-team projects involving at least one research team in the United States and one in another country. The NSF would support the US part of a joint project while the non-US parts needed to gain their support from other sources.

NSF Partnerships

NSF-JISC (http://www.jisc.ac.uk/) (US-UK)
NSF-DFG (http://www.dfg.de/foerder/biblio/neues/0812diglib.html) (US-Germany)
NSF-EU (http://www.dli2.nsf.gov/internationalprojects/nsfec.html) (US-European Union)

NDLTD: A current major US initiative is the development of a Networked Digital Library of Theses and Dissertations (NDLTD). The project which is led by the Virginia Polytechnic Institute and State University involves a growing number of US universities, University Microfilms International and industry partners. As a result of this work, standards for creation, deposit, conversion, storage, archiving and printing have already been established. The focus of the NDLTD has shifted from a US-based one to a growing international network of universities. Its aim is to improve graduate education, increase sharing of knowledge, help universities build their information infrastructure, and extend the value of digital libraries.

4.1.2. Germany

A series of joint projects, especially in the areas of research and provision of information and electronic documents in full-text format is contributing to the creation of a national Digital Library.

The BMBF, the Ministry of Education and Research, supports the setting up of subject-related information networks in order to improve access to full-text electronic documents. Virtual specialist libraries are being developed with the financial support of the DFG, the German Research Foundation. The individual virtual specialist libraries offer subject-related access to academic information and documents. Such libraries make it possible for research to be done centrally and for the desired material to be delivered directly to the place of work.

The DFG (Deutsche Forschungsgemeinschaft – German Research Foundation)

The Deutsche Forschungsgemeinschaft is the central, self-governing research organization that promotes research at universities and other publicly financed research institutions in Germany. The DFG serves all branches of science and the humanities by funding research projects.

Part of the DFG's mission is to support German information centres, linked to the global network of information systems through a wide array of cooperative programmes and activities, in developing infrastructures and services.

The Committee on Scientific Library Services and Information Systems

The Committee on Scientific Library Services and Information Systems *(Bibliotheksausschuss)* is responsible for advising the DFG on all projects and activities relating to the development and support of scientific library services and information systems in Germany. In order to be able to plan, review and monitor the success of projects, the committee has established four standing subcommittees: "Nationwide Library Services", "Electronic Publications", "Information Management" and "Cultural Tradition", as well as two temporary working groups: "Information Systems for Researching Mediaeval and Early Modern Times" and "Information Management of Archives".

The committee is currently made up of seven library directors, four researchers, and one representative each from an information and documentation centre, an archive, and a computer centre. The committee is appointed by the Joint Committee for a two-year term and convenes twice a year, in spring and autumn. The Scientific Library Services and Information Systems Division is responsible for managing the committee.

International Cooperation

The DFG and the National Science Foundation (NSF) in Washington have set up joint funding initiatives to support German-American cooperative efforts in the area of digital libraries. The aim is to promote high-quality cooperative projects and thus the participation of German partners in innovative international developments. Following the establishment of the first joint funding initiative in 1999, the DFG and the NSF established a second joint funding initiative in the field of digital libraries in 2002.

Under the first funding initiative "International Digital Libraries Research Programme. A Joint NSF/DFG Initiative," the DFG and the NSF have jointly funded four cooperative projects since early 2001, and three further projects were approved in 2003.

DINI (Deutsche Initiative für NetzwerkInformation)

(http://www.dini.de/)

DINI aims to support the development of information infrastructures and the cooperation between relevant German institutions. This is done by developing recommendations, standards, requirements for information services and communication networks and by creating net-based digital publication opportunities.

4.2. International Policies, Initiatives and Programmes

4.2.1. DELOS

(http://www.delos.info/)

The DELOS Network of Excellence on Digital Libraries is a framework for international cooperation on research activities and research agendas in the digital library domain. Funded by the EU's Sixth Framework Programme, its activities started in January 2004.

The DELOS network conducts a joint programme of activities aimed at integrating and coordinating the ongoing research activities of the major European teams working in DL.

The Network will also disseminate knowledge of DL technologies to many diverse application domains by means of workshops, conferences, exchange programmes, etc. To this end a Virtual Digital Library Competence Centre has been established which provides specific user communities with access to advanced DL technologies, services, test beds and the necessary expertise and knowledge to facilitate their take-up.

Membership: DELOS has 63 members (as of December 2003) from 16 EU countries. They include major European research centres in informatics, 30 University departments, representatives of interested application domains in ICT industry like broadcasting (BBC), electronic publishing (ELSEVIER, Springer Verlag), libraries, UK Office for Library and Information Networking (UKOLN), etc.

Working: DELOS has two boards: the Scientific Board and the Advisory Board. The Scientific Board comprising of members form various EU nations, administer the working of the network. The Advisory Board consists of five external members. They evaluate the working of the network and provide guidance at the strategic level.

DELOS has a number of Research clusters (see Part 3.3), working on topics of current relevance in DLs. The activities of these research clusters are reported through publications or through the newsletter of DELOS.

4.2.2. ICSU

(http://www.icsu.org/)

Founded in 1931 as the International Council of Scientific Unions (ICSU), it changed its name in April 1998 to International Council for Science (retaining the acronym ICSU). It is a non-governmental organization representing a global membership that includes both national scientific bodies (103 members) and international scientific unions (27 members).

ICSU's mission is to strengthen international science for the benefit of society. ICSU mobilizes the knowledge and resources of the international science community to:

- Identify and address major issues of importance to science and society;
- Facilitate interaction amongst scientists across all disciplines and from all countries;
- Promote the participation of all scientists – regardless of race, citizenship, language, political stance, or gender – in the international scientific endeavour;
- Provide independent, authoritative advice to stimulate constructive dialogue between the scientific community and governments, civil society and the private sector.

ICSU recognizes the importance of data and information to science and has addressed issues related to it in its various initiatives over the past half-century. It has identified scientific data and information as a priority area in developing its strategic plan for the coming years.

ICSU has established three bodies: the Committee on Dissemination of Scientific Information (CDSI), the International Network on the Availability of Scientific Publications (INASP), and the Committee on Data for Science and Technology (CODATA), that specialize in scientific data and information issues.

There are several interdisciplinary ICSU bodies whose principal focus is the management and use of large scientific data sets: the Panel of the World Data Centres (WDC) and the Federation of Astronomical and Geophysical Data Analysis Services (FAGS). ICSU is also a co-sponsor of the Global Observing Systems (GOS). Many members of ICSU are very actively involved in issues related to scientific data. For example, an important policy initiative was the creation of an Inter-Union Bioinformatics Group, which produced an authoritative report and recommendations regarding biological databases in 2002.

ICSU had also been playing a very active role in the negotiations leading up to World Summit on the Information Society (WSIS) in Geneva in December 2003. In particular, an agenda for action – Science in the Information Society – had been developed as the result of an online discussion forum and international workshop in March 2003, which focused on the central role of scientific data and information in society.

In summary, the ICSU family has an array of disciplinary or topic focused activities in which scientific data and information use are a major component. It also has several overarching committees with special responsibility for data and information.

An international panel of independent experts was appointed in 2003 by the Committee on Scientific Planning and Review (CSPR) to perform a Priority Area Assessment (PAA) on Scientific Data and Information. This panel was charged with both assessing the strategic issues in this arena and reviewing ICSU's current activities. The assessment focuses primarily on the needs of science.

At the end of 2004, the report of the Assessment Panel on Scientific Data and Information was published. The report has a large number of recommendations. These recommendations are addressed to multiple audiences: the ICSU Executive (Secretariat, Board and CSPR), specific ICSU interdisciplinary bodies and committees, the ICSU membership, science funding bodies, governmental organizations, commercial publishers and data producers, and individual scientists.

Most importantly, this concluded that the research community should assume responsibility for building a robust data and information infrastructure for the future and produced a number of recommendations to support this objective.

Recommendations to ICSU with respect to digital libraries include: ICSU should work with its members and relevant bodies in encouraging the coordinated development of digital libraries and their integration with journal publishing and data systems.

Other recommendations include: Free and open access to research data by scientists, with financial support for data. Ensuring the long-term accessibility of increasing quantities of scientific data and information will necessitate increased public (and private) investment in data management and long-term institutional support. ICSU and its members should explore various solutions to meet the financial challenge of providing full and open access to scientific data and universal and equitable access to publications.

A workshop "Science in the Information Society" was organized by ICSU, CODATA and UNESCO in Paris on 12th March 2003. The workshop report mentions these among the action points:

- Promote interoperability principles and metadata standards to facilitate cooperation and effective use of collected information and data;
- Provide long-term support for the systematic collection, preservation and provision of essential digital data in all countries;
- Promote electronic publishing, differential pricing schemes and appropriate open source initiatives to make scientific information accessible on an equitable basis.

4.2.3. INASP

(http://www.inasp.info/)

The International Network for the Availability of Scientific Publications (INASP) was established in 1992 by the International Council for Science (ICSU) to improve access to information and knowledge through a commitment to capacity building in emerging and developing countries. Initially based within the International African Institute, INASP moved to Oxford in June 1996. ICSU provided an umbrella legal status under its Committee for the Dissemination of Scientific Information until 2003 when INASP was registered in the United Kingdom as a company limited by guarantee. In October 2004, it was registered as a charity.

The vision is that all people are able to access and contribute information, ideas and knowledge necessary to drive sustainable and equitable development. The mission of INASP is to enable worldwide access to information and knowledge with particular emphasis on the needs of developing and transitional countries. They work with partners and networks around the world to encourage the creation and production of information, to promote sustainable and equitable access to information, to foster collaboration and networking, and to strengthen local capacities to manage and use information and knowledge. They achieve these through activities and programmes that:

- Improve access to scientific and scholarly information;
- Catalyse and support local publication and information exchange;
- Strengthen local capacities to manage and use information and knowledge;
- Foster in-country, regional and international cooperation and networking;
- Advise local organizations and funding agencies on ways to utilize information and publishing to achieve development goals.

The major roles of INASP are:

- Networking – catalyzing people-to-people exchanges;
- Accessing – facilitating sustainable access to information;
- Sharing – supporting local exchange of information and knowledge;
- Publishing – strengthening local knowledge creation and dissemination;
- Training and capacity development – enhancing skills of information and communication professionals and local trainers.

These roles are achieved through:

- Country- and sector-focused support for networking, accessing, sharing and publishing of information and knowledge;

- Strengthening the efforts of information and communication "enablers" – individuals and organizations supporting research, science, education and development;
- Skills and capacity development in areas of information networking, accessing, sharing and publishing;
- Providing advisory and liaison services and acting as an information exchange point and clearing house on information and communication issues in development;
- Partnerships and joint ventures with other organizations.

Programme Areas

INASP activities are grouped in five programme areas:

- Education and Research – supporting networking, accessing, sharing and publishing of information for researchers, scholars, scientists and academics;
- Health – supporting networking, accessing, sharing and publishing of information for health professionals;
- Library Support – strengthening the work of libraries and information management professionals, especially those serving research and educational communities;
- Publishing Support – strengthening the work of editors and communication professionals, especially those in research and science;
- Rural and Agricultural – supporting networking, accessing, sharing and publishing of information for rural development practitioners.

A key focus of all programmes is capacity strengthening, which is normally addressed through training and workshops.

Project and Partners

INASP implements its activities through partnerships and joint ventures with many organizations. Here we list some of them:

African Journals Online: AJOL was incubated by INASP as part of its efforts to support the wider dissemination and visibility of research results published through African journals.

Bandwidth optimization and management: INASP is coordinating capacity strengthening project that will provide information and resources and develop skills and knowledge in selected universities and research institutions on strategies and techniques for managing and optimising bandwidth usage.

CARINDEX: With the University of the West Indies, INASP hosts an index to journals and other research outputs published in the Caribbean region.

Dgroups: INASP is a member of this collaboration that aims to foster effective electronic dialogue and information exchange – mainly by email.

ItrainOnline: INASP is a founder partner in this initiative to bring together training materials on information and ICTs in development.

Programme for the Enhancement of Research Information: The PERI project strengthens research capacities in developing and transitional countries by reinforcing local efforts to produce, disseminate and gain access to scholarly information and knowledge.

INASP is an interdisciplinary body of the International Council for Science (ICSU) and it works closely with members of the ICSU "family" on information access and dissemination issues.

4.2.4. WSIS (World Summit on the Information Society)

(http://www.itu.int/wsis/)

The United Nations General Assembly Resolution in 2001 endorsed the International Telecommunication Union (ITU) proposal to hold a World Summit on the Information Society (WSIS) at the highest possible level and under the patronage of the United Nations Secretary-General. The main objective of WSIS is to develop a common vision and understanding of

the information society and the adoption of a declaration of principles and a plan of action for implementation by governments, international institutions and all sectors of civil society. The first phase of WSIS took place in Geneva hosted by the Government of Switzerland from 10 to 12 December 2003, where 175 countries adopted a Declaration of Principles and Plan of Action. The second phase took place in Tunis hosted by the Government of Tunisia, from 16 to 18 November 2005.

WSIS offers a unique opportunity for the world community to discuss and give shape to the information society by bringing together key players: governments, international organizations, private sector, civil society/NGOs. It addresses the central issues raised by the information society and related to policy formulation, socio-economic development, as well as cultural, ethical and technological concerns, which include:

- Building the ICT infrastructure;
- Universal and equitable access to the information society;
- Services and applications;
- User needs;
- Developing a framework for building the information society;
- ICTs in education.

It addressed the broad range of themes concerning the Information Society and adopted a Declaration of Principles and Plan of Action. UNESCO encourages the broadest possible participation by decision-makers, professional communities, representatives of civil society, bilateral and multilateral partners and the private sector in a debate on the conditions for the development of an information society for all.

With respect to Digital libraries and science, the WSIS Action Plan specifically gives the following calls for national governments to implement in the declaration of Plan of Action (PoA, http://www.itu.int/wsis/documents/doc_multi.asp?lang=en&id=1160%7C0):

C3. Access to information and knowledge

10 h) Support the creation and development of a digital public library and archive services, adapted to the Information Society, including reviewing national library strategies and legislation, developing a global understanding of the need for "hybrid libraries", and fostering worldwide cooperation between libraries.

i) Encourage initiatives to facilitate access, including free and affordable access to open access journals and books, and open archives for scientific information.

http://www.itu.int/wsis/docs/geneva/official/poa.html

C4 Capacity Building

11 k) Design specific training programmes in the use of ICTs in order to meet the educational needs of information professionals, such as archivists, librarians, museum professionals, scientists, teachers, journalists, postal workers and other relevant professional groups. Training of information professionals should focus not only on new methods and techniques for the development and provision of information and communication services, but also on relevant management skills to ensure the best use of technologies. Training of teachers should focus on the technical aspects of ICTs, on development of content, and on the potential possibilities and challenges of ICTs.

ICT Applications: benefits in all aspects of life

22. E–science

a. Promote affordable and reliable high–speed Internet connection for all universities and research institutions to support their critical role in information and knowledge production, education and training, and to support the establishment of partnerships, cooperation and networking between these institutions.

b. Promote electronic publishing, differential pricing and open access initiatives to make scientific information affordable and accessible in all countries on an equitable basis.
c. Promote the use of peer–to–peer technology to share scientific knowledge and pre–prints and reprints written by scientific authors who have waived their right to payment.
d. Promote the long–term systematic and efficient collection, dissemination and preservation of essential scientific digital data, for example, population and meteorological data in all countries.
e. Promote principles and metadata standards to facilitate cooperation and effective use of collected scientific information and data as appropriate to conduct scientific research.

Second WSIS Summit: In preparation for the second summit, several regional and thematic meetings were held. On 10 November 2004 a CODATA (ICSU) – WSIS Session was held in Berlin (Germany). The WSIS session of CODATA 2004 discussed, among other things, the role of the international science community in implementing its agenda for action prior to the second phase of the Summit in Tunis in November 2005.

Prior to phase II of the World Summit, there was a three days event in November 2005 focused specifically on the role of research in the Information Society, that is, the production of knowledge. The symposium involved policy makers and programme managers, as well as working scientists and scholars. Sessions were organized by a variety of individuals and organizations, focusing on: (1) the role of science in developing the Information Society; (2) current uses of information and communication technologies for the production of knowledge; (3) future involvement of developing areas in the global scientific community.

4.2.5. Open Archives Initiative (OAI)

(http://www.openarchives.org/)

Discussing access to scholarly information, one cannot overlook the Open Access Movement. Stevan Harnad's "subversive proposal" posted to a Virginia Tech mailing list in 1994 is a landmark in the history of open access. It advocated the continuance of peer-reviewed scholarly publishing in print form but it also advocated scholarly articles be digitally self-archived by authors and made freely available through the Internet, much like the online network of preprint archives (arXiv) of the physics community lead by Paul Ginsparg. Though the movement is more than a decade old, the phrase "open access" was coined in the Budapest Open Access Initiative (BOAI) in 2002. This Initiative is an attempt to see how much the Open Society Institute (OSI) could help resolve the scholarly communication.

The development of technological standards like OAI-PMH and LOCKSS and software (e.g. EPrints, DSpace, etc.) that implement them have made the philosophy a possibility.

Advocacy campaigns, such as SPARC (Scholarly Publishing and Academic Resources Coalition – http://www.arl.org/sparc/); the Public Library of Science (http://www.publiclibraryofscience.org/), the Digital Preservation Coalition (http://www.jisc.ac.uk/dner/preservation/prescoalition.html), the Budapest Open Access Initiative (http://www.soros.org/openaccess/), and the Open Archive Initiative (http://www.openarchives.org/) are all seeking to create and coordinate radical change in the scholarly communication process.

To achieve open access to scholarly literature, BOAI recommends two strategies:

- Self-archiving (also known as the "green route"): Scholars deposit their refereed journal articles in open electronic archives, a practice commonly called, self-archiving.
- Open-access Journals (also the "gold route"): Scholars need to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access. These new journals will no longer invoke copyright to restrict access to and use of the material they publish. Instead they will use copyright and other tools to ensure permanent open access to all the articles they publish.

New economic models are emerging to support the open access journals. Alternative sources of funds include the foundations and governments that fund research, the universities and laboratories that employ researchers, endowments set up by discipline or institution, friends of the cause of open access, profits from the sale of add-ons to

the basic texts, funds freed up by the demise or cancellation of journals charging traditional subscription or access fees, or even contributions from the researchers themselves. Organizations like BioMed Central (http://www.biomedcentral.com/) and Public Library of Science (http://www.plos.org/) already have such publishing models in place.

There are also growing "independence statements" from journal article editors/staff, as well as formal statements from institutions of higher education to move away from the traditional modes of publishing. Most recently, governments (like the UK and US) are advocating open access publishing at least in the publicly funded research.

A recent study done by Kristin Antelman at North Carolina State University indicates that freely available articles do have a greater research impact. Though Antelman only looked at four specific disciplines, it was clear that the more often something is downloaded, the more likely it is to be cited. In an article entitled **Online or Invisible** (http://citeseer.ist.psu.edu/online-nature01/), Steve Lawrence of the NEC Research Institute also opines that free online availability can substantially increase a paper's impact. Scholars in various disciplines are being rewarded for choosing open access publishing.

Because OA journals are a relatively new phenomenon, there are few impact studies thus far. But despite the paucity of data, certain subject areas rank OA journals among the top 9%.

4.3. Summary

The US and Europe are making substantial progress in DL. This is due to the large amount of funding given to research in the area. Government organizations like the NSF and JISC are set up at the national level to channel science R&D and supporting activities in science. A lot of impetus is given to DLs by these bodies.

Policy Issues: Role of National-level organizations:

- **Funders for building and Research work in DLs:** Large centralized organizations at national levels like NSDL, JISC act as organizations that fund the development of DLs. They also fund R&D activities based on challenges and problems faced by DLs.
- **Evaluation and quality control:** Such organizations should have mechanisms in place to ensure that quality of the DL is maintained. If the organization involved in implementation of DL like NSDL also fund Research, this ensures that current and practical problems are identified and addressed or researched to find a solution.
- **Aggregators and service providers:** Their functioning is not limited to funding, they also act as aggregators of the DLs. They ensure uniform interfaces and provide services at the top of all the DLs, such as portals and tools for information/knowledge discovery.
- **Policy makers:** They also frame policies based on evaluation and feedback they obtain from the experiences of the various parties involved including those who actually build DLs. Regular review of past experiences, evaluation of usage of DLs and monitoring of trends enables them to frame effective polices.
- **Long-term sustainability and preservation:** Once a project is finished, organizations concerned with preservation and archiving can take the outcome and ensure its upkeep and availability for long term.
- **Exchange of Knowledge:** This takes place in the form of workshops and meetings conducted by the organization.
- **Usage of DL:** Such large organizations are highly visible to the public and there is wide publicity that ensures that the DLs built individually get wider usage.
- **Standardization:** They ensure a minimum standardization in metadata and quality of content.

D. Waters describes the nature of such an organization: "New organizational models are needed that are sensitive to the dynamics of particular scientific communities, driven by academic mission, and able to sustain themselves over time as integral parts of the broader cyberinfrastructure." (Waters, 2003)

There is a need for creating awareness at the national government levels of the importance of setting up digital library systems in place. But there is equally a need to create enough local content in the languages of the region, both by digitising existing documents and creating new content. International bodies like UNESCO and ICSU, through summits like WSIS, should induce national bodies to implement digital libraries and encourage research and development work in this area. Setting up national level organizations like NSDL to coordinate the activities in digital library developments should be encouraged. Mechanisms for quality control and evaluation should also be in place. What is also required is the need to share the knowledge generated by such activities by conducting conferences and workshops not just at the national levels but also at the international.

Multinational organizations: While there have been some activities in DL at the international level, it is quite small compared to the national level activities. There have been some explorative activities of the NSF International DL Initiative in cooperation with institutes in Europe and China (CMNet). DELOS is funded by the "Sixth Framework" programme of the EU and mainly funds research projects in DL. Some inter-university programmes also have taken place but mainly in the form of R&D.

What does take place is a lot of DL based conferences and seminars. While most countries might have awareness of the advancements done in DL, they may lack a clear direction as to how to go about in terms of implementing DLs. This may also be due to the huge amount of funding and trained manpower required for DL activities.

Bibliography

1. Ananiadou, Sophia, et al. "The National Centre for Text Mining: Aims and Objectives", *Ariadne*, 42, 30 January, 2005. Available: http://www.ariadne.ac.uk/issue42/ananiadou/intro.html

2. Antelman, Kristin "Do Open Access Articles Have a Greater Research Impact?" *College & Research Libraries News*, 65(5), 2004, pp. 372-382. Available: http://eprints.rclis.org/archive/00002309/

3. Awre, Chris "The JISC's FAIR Programme: disclosing and sharing institutional assets", *Learned Publishing*, 17, 2004, pp. 151-156. Available: http://eprints.rclis.org/archive/00001331/01/Learned_Publishing_FAIR_article.pdf

4. Fox, Edward A., et al. "Networked Digital Library of Theses and Dissertations", *D-Lib Magazine*, September 1997. Available: http://www.dlib.org/dlib/september97/theses/09fox.html

5. Laura M. Bartolo, et al. "NSDL MatDL: Exploring Digital Library Roles", *D-Lib Magazine*, 11, no. 3, March 2005. Digital Object Identifier (DOI):10.1045/march2005-bartolo

6. Lawrence, Steve "Online or Invisible?" *Nature*, 411, no. 6837, 2001, p. 521. Available: http://citeseer.ist.psu.edu/online-nature01/

7. Lyon, Liz "eBank UK: Building the links between research data, scholarly communication and learning", *Ariadne*, 36, 30 July 2003. Available: http://www.ariadne.ac.uk/issue36/lyon/intro.html

8. Nelson, Michael L. "A Digital Library for the National Advisory Committee for Aeronautics", NASA/TM-1999-209127, April 1999, pp. 15. Available: http://library-dspace.larc.nasa.gov/dspace/jsp/handle/2002/13202. Digital Object Identifier (DOI): http://hdl.handle.net/2002/13202

9. Nelson, Michael L., Rocker, JoAnne, and Harrison, Terry L. "OAI and NASA Scientific and Technical Information", *Library Hi-Tech*, 21(2), 2003, pp. 140-150. Available: http://library-dspace.larc.nasa.gov/dspace/jsp/handle/2002/13223. Digital Object Identifier (DOI): http://hdl.handle.net/2002/13223

10. Morgan, Eric Lease "Open Access publishing". Available: http://www.infomotions.com/musings/open-access/open-access.pdf

11. Morgan, Peter and Wilson, Michael "The future of the scientific journal: Workshop report", 16 April 2002. Available: http://www.lib.cam.ac.uk/Future/Report.doc

12. Paepcke, Andreas, et al. "Search Middleware and the Simple Digital Library Interoperability Protocol", *D-Lib Magazine*, 6, no. 3, March 2000. Available: http://www.dlib.org/dlib/march00/paepcke/03paepcke.html

13. Pinfield, Stephen "Managing electronic library services: current issues in UK higher education institutions", *Ariadne*, 29, 2 October 2001. Available: http://www.ariadne.ac.uk/issue29/pinfield/intro.html

14. Powell, Andy and Lyon, Liz "The DNER Technical Architecture: scoping the information environment", in UKOLN, University of Bath, 18 May 2001. Available: http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/dner-arch.html

15. Powell, Andy "JISC Information Environment Architecture. Standards Framework Version 1.1", in UKOLN, 14 May 2004. Available: http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/

16. Rusbridge, Chris "Towards the Hybrid Library", *D-Lib Magazine*, 4, July/August 1998. Available: http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html

17. Shrum, Wesley "Past, Present and Future of Science in the Information Society", in *The Information Society: New horizons for Science Proceedings for the 19th International CODATA Conference*, Berlin, Germany November 7-10,

2004. Available: http://www.codata.org/04conf/abstracts/WSIS/Shrum-PastPresentandFutureofScienceintheInformationSociety.htm

18. Warner, Simeon "Open Archives Initiative protocol development and implementation at arXiv, Los Alamos National Laboratory, USA", 23 January 2001. Available: http://www.cs.cornell.edu/people/simeon/talks/OAI_DC_23Jan2001/talk.pdf

19. Whitelaw, Alan and Joy, Gill. ESYS Limited "Summative Evaluation of Phases 1 and 2 of the eLib Initiative: Overview". Available: http://www.ukoln.ac.uk/services/elib/info-projects/phase-1-and-2-evaluation/overview.pdf

20. Zia, Lee L. "The NSF National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) Program: A Progress Report", *D-Lib Magazine*, 6, no. 10, October 2000. Digital Object Identifier (DOI):10.1045/october2000-zia

21. Zia, Lee L. "The NSF National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) Program: New Projects and a Progress Report", *D-Lib Magazine*, 7, no. 11, November 2001. Digital Object Identifier (DOI):10.1045/november2001-zia

22. Zia, Lee L. "The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program: New Projects in Fiscal Year 2002", *D-Lib Magazine*, 8, no. 11, November 2002. Digital Object Identifier (DOI):10.1045/november2002-zia

23. Zia, Lee L. "The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program: New Projects in Fiscal Year 2003", *D-Lib Magazine*, 10, no. 3, March 2004. Digital Object Identifier (DOI):10.1045/march2004-zia

24. Zia, Lee L. "The NSF National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program: New Projects from Fiscal Year 2004", *D-Lib Magazine*, 11, no. 3, March 2005. Digital Object Identifier (DOI):10.1045/march2005-zia

Websites

1. ARL, “Scholarly Publishing & Academic Resources Coalition”, The Association of Research Libraries, http://www.arl.org/sparc/
2. ARROW, “Australian Research Repositories Online to the World”, http://arrow.edu.au/
3. Australian Government, Department of Education, Science and Training, “Australian Research Information Infrastructure Committee (ARIIC)”, http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/australian_research_information_infrastructure_committee/default.htm
4. Australian Government, Department of Education, Science and Training, “ARIIC Projects”, http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/australian_research_information_infrastructure_committee/ariic_projects.htm
5. BioMed Central, “BioMed Central – The Open Access Publisher”, http://www.biomedcentral.com/
6. Council of Australian University Librarians, “ADT”, Australasian Digital Theses Programme, http://adt.caul.edu.au/
7. David Newman, “CombeChem”, CombeChem, http://www.combechem.org/
8. DELOS, “DELOS Network of Excellence on Digital Libraries”, http://www.delos.info/
9. DFG, “DFG – Scientific Library Services and Information Systems”, http://www.dfg.de/lis
10. Digital Curation Centre, “Digital Curation Centre”, http://www.dcc.ac.uk/
11. Digital Library Initiatives 2, “International Projects”, http://www.dli2.nsf.gov/intl.html
12. Digital Library Production Service, “Oaister”, University of Michigan, http://www.oaister.org/o/oaister/
13. Dissertation Online, http://www.dissonline.de
14. Edinburgh University, “Theses Alive! At Edinburgh University Library”, http://www.thesesalive.ac.uk
15. EPSRC, “EPSRC Website”, EPSRC Engineering and Physical Sciences Research Council, http://www.epsrc.ac.uk/
16. Federal Ministry of Education and Research, “BMBF: Federal Ministry of Education and Research”, http://www.bmbf.de/en/index.php
17. Federal Ministry of Education and Research, Germany, “EconDoc”, Economic Online Documents, http://www.econdoc.de/
18. GetInfo, “Specialist Information on science and Technology”, http://www.getinfo-doc.info/en/
19. HaIRST, “Harvesting Institutional Resources in Scotland Testbed”, http://hairst.cdlr.strath.ac.uk
20. ICSU, “ICSU – International Council for Science”, http://www.icsu.org/
21. INASP, “International Network for the Availability of Scientific Publications”, http://www.inasp.info/
22. Internet 2, “Shibboleth project”, http://middleware.internet2.edu/shibboleth/
23. JISC Comms, “JISC: The Joint Information Systems Committee (JISC)”, JISC, http://www.jisc.ac.uk/
24. John Hopkins University, “Scholarly Communication Group”, https://openaccess.jhmi.edu/publishing.cfm
25. KnowLib, “Digital library projects, initiatives and research centers”, http://www.it.lth.se/knowlib/resources.htm
26. Los Alamos National Laboratory, “ArXiv”, http://lanl.arxiv.org/
27. Loughborough University, “The RoMEO Project (Rights MEtadata for Open archiving)”, Project RoMEO, http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html
28. MedPilot, http://www.medpilot.de/
29. Monica Duke and Traugott Koch, “UKOLN eBank UK Project”, UKOLN, http://www.ukoln.ac.uk/projects/ebank-uk/
30. NASA, “National Technical Report Server”, http://ntrs.nasa.gov/
31. National Library of Australia, “Pandora Archive – Preserving and Accessing Networked DOcumentary Resources of Australia”, PANDORA, http://pandora.nla.gov.au/
32. National Science Digital Library, http://www.nsdl.org/
33. NDLTD, “National Digital Library of Theses and Dissertations”, http://www.ndltd.org/
34. Networked Computer Science Technical Reference Library, “NCSTRL”, http://www.ncstrl.org/
35. Penn State, “Citeseer.IST”, http://citeseer.ist.psu.edu/
36. PloS, “Public Library of Science”, http://www.plos.org/
37. Rachel Bruce, “Focus on Access to Institutional Resources (FAIR) Programme”, JISC, http://www.jisc.ac.uk/programme_fair.html
38. RDN, “RDN – The ePrints UK Project”, http://www.rdn.ac.uk/projects/eprints-uk/
39. RDN, Resource Discovery Network, http://www.rdn.ac.in/
40. Resource Discovery Network, “EEVL: The Internet Guide to Engineering, Mathematics, and Computing”, http://www.eevl.ac.uk/

41. Resource Discovery Network, “GEsource: geography and environment gateway for UK HE and FE”, http://www.gesource.ac.uk/
42. Resource Discovery Network, “Physical Sciences Information Gateway”, PSIgate, http://www.psigate.ac.uk/
43. Scirus, “About Us”, http://www.scirus.com/srsapp/aboutus/
44. SOROS, “Budapest Open Access Initiative”, http://www.soros.org/openaccess/
45. SOSIG, “ Social Science Information Gateway”, http://www.sosig.ac.uk/
46. Steven Harnad, “CogPrints”, CogPrints, http://cogprints.org/
47. The Apache Software Foundation, “The Apache Jakarta Project”, http://jakarta.apache.org/
48. The Robert Gordon University Library, “RGU ETD Electronic Theses”, http://www.rgu.ac.uk/library/e-theses.htm
49. The Universities of Manchester and Liverpool, “The National Centre for Text Mining – NaCTeM”, NaCTeM, http://www.nactem.ac.uk/
50. The University of Hull, “PORTAL: Presenting national Resources To Audiences Locally”, http://www.fair-portal.hull.ac.uk
51. The University of Nottingham, BIOME, “The Hub for Internet Resources in the Health and Life Sciences”, http://biome.ac.uk/
52. UKOLN Project IEMSR “JISC IE Metadata Schema Registry”, UKOLN, http://www.ukoln.ac.uk/projects/iemsr/
53. University of Bath, “UKOLN”, http://www.ukoln.ac.uk/
54. University of Bristol, “Bristol Biomedical Image Archive”, BIOMED Image Archive, http://www.brisbio.ac.uk/
55. University of Cambridge, “DSpace at Cambridge”, http://www.dspace.cam.ac.uk/
56. University of Cambridge, “Dspace@Cambridge Project”, http://www.lib.cam.ac.uk/dspace/index.htm
57. University of Glasgow, “DAEDALUS – A JISC FAIR Project”, http://www.lib.gla.ac.uk/daedalus/
58. University of Nottingham, “SHERPA”, http://www.sherpa.ac.uk
59. University of Southampton, UK, “TARDis University of Southampton e-Prints”, TARDis: Targeting Academic Research for Deposit and Disclosure, http://tardis.eprints.org
60. University of Waikato, “Greenstone Digital Library Software”, http://www.greenstone.org/
61. Vascoda, “Vascoda – Entdecke Information”, http://www.vascoda.de/
62. Western Colleges Consortium, “FAIR Enough”, Focus on Access to Institutional Resources Programme (FAIR), http://www.fairenough.ac.uk
63. WSIS, “World Summit on Information Society”, http://www.itu.int/wsis/

DIGITAL LIBRARIES IN CULTURE

1. Introducton

This report will identify the way that digital libraries interact with culture and how this interaction may develop and deepen in the future. It will discuss current uses of digital libraries for cultural purposes, particularly for increasing and democratising access to the Arts. This report shows the important role that digital libraries play in preserving culture and in connecting people with their national and regional identities. Digital libraries are also a means by which displaced peoples may remain in contact with their indigenous cultures and where family history (genealogy) may be discovered and cherished. The author considers digital libraries used for cultural purposes as being a key factor that will help to bridge the growing digital divide in the world. The report finishes with a view to the digital future through "cultural commons" and emerging trends to consider the resource, policy and strategic implications for future action.

The author would like to thank all those who helped with this report, including:

Abby Goodrum, Alma Rivera Aguilera, Amalyah Keshet, Andrea de Polo, Andrew Lewis, Andy Holt, Anna Maria Tammaro, Austin David, Bertrand Couasnon, Cees Klapwijk, Chip Bruce, Christa Müller, Claire Warwick, Cushla Kapitzke, David Dawson, Diana Rosenberg, Doug Moncur, H.M. Gladney, Jody Perkins, John Pereira, Julia Bolton Holloway, Julie Harvey, Katrien de Kerpel, Laszlo Drotos, Laurie Gemmill, Linda Royles, Lois McLean, Lourdes Feria, Manfred Thaller, Manuela Speiser, Maria E. Gonzalez, Michael Gösele, Michael Hart, Michal Krsek, Michele Barbera, Minna Kaukonen, Nancy Allen, Peter Keelan, Peter Suber, Ria Smith, Richard Wright, Rob Davies, Robina Clayphan, Sabine Stadler, Sam Gustman, Sara Jane Pearman, Sarah Giersch, Shirley Williams, WG Kilbride and Ximena Cruzat.

Special thanks are due to Harold Short, Marilyn Deegan, Willard McCarty and Rebecca Finkel from the Centre for Computing in the Humanities, King's College London.

2. Digital Libraries and Culture: Definitions and Contexts

To understand the relationship between culture and digital libraries, it is important to delineate the parameters that define each concept. This is especially important within the contexts of cultural institutions and exponents, such as museums, galleries, festivals, arts organizations or broadcasters, whose information infrastructure has not been traditionally dominated by the physical library concept.

Whilst the "conversion of all sorts of cultural contents into bits and bytes opens up a completely new dimension of reaching traditional and new audiences by providing access to cultural heritage resources in ways unimaginable a decade ago" (Mulrenin and Geser, 2001), the roles and responsibilities of those providing cultural context within the digital library framework are not instantly clear and must necessarily be further defined.

2.1. Digital Libraries

Digital Libraries have been variously defined and many such definitions seek to find the role of the library in relation to the digital content. From its original etymological meaning of a "collection of books", a library could be redefined in modern terms to mean a collection of almost anything: software routines or digital images, for instance. Many libraries are different with unique features, every digital library is equally different, and different players are advancing many definitions for the digital library. Arms, for instance, defines a digital library as:

"A managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network. A crucial part of this definition is that the information is managed." (Arms, 2000)

For the Digital Library Federation in the US,

"Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities." (Greenstein, 2000)

Digital libraries are a development of a whole range of underlying theories and technologies which have come together to create a paradigm shift. This report will use as its definition of a digital library a set of principles devised by Deegan and Tanner that characterize something more clearly as a digital library rather than any other kind of set of digital objects. It is important to have a definition that characterizes the digital library as this report will record digital resources from a very wide range of cultural organizations not necessarily based in library infrastructures.

The Deegan and Tanner characteristics of a digital library are:

- A digital library is a managed collection of digital objects;
- The digital objects are created or collected according to principles of collection development;
- The digital objects are made available in a cohesive manner, supported by services necessary to allow users to retrieve and exploit the resources just as they would any other library materials;
- The digital objects are treated as long-term stable resources and appropriate processes are applied to them to ensure their quality and survivability.
  (Deegan and Tanner, 2002)

2.2. Culture

The term "culture" is far more than just a synonym for "art" as it can refer to heritage, arts and creativity, leisure activities and the ways and beliefs of a society. The Cambridge Dictionary defines culture as:

"The way of life, especially the general customs and beliefs, of a particular group of people at a particular time" and "the continuing traditions of art, music, literature, etc. of a particular society or group within a society" (Cambridge Advanced Learner's Dictionary, 2003).

UNESCO views culture according to heritage, arts and creativity, museums, creative industries, tourism, and activities involving societal customs and diversity. This report will include all of the definitions of culture included here, spanning from the arts to leisure activities to heritage to traditions of a society.

Clearly cultural resources can thus have an audience of a few persons or of a whole community or population. Digital cultural resources assume the potential for a worldwide audience but frequently are focussed upon a specific community or audience as will be evidenced by the resources listed in this report. Digital cultural resources thus should address or contain, to some extent, one of the following elements:

- Heritage;
- Arts and creativity;
- Museums;
- Creative industries and tourism;
- Societal customs and diversity;
- Ways of life.

2.3. Digital Libraries and Culture

Digital libraries have a major function to enhance our appreciation or engagement with culture. The underlying information and communication technologies are still generally referred to as "new" or "high" technologies – they remain highly visible, and have not yet, despite their pervasiveness, become part of the natural infrastructure of society. "Technology", as the computer scientist Bran Ferren memorably defined it, is "stuff that doesn't work yet" (Adams, 1999).

As Castells points out, "technology does not determine society... technology is society and society cannot be understood or represented without its technological tools" (Castells, 1996). Society also has to be ready, both technically and psychologically, for major technological change to happen. Progress seems unlikely to follow a smooth linear rate of change, but more likely "a series of stable states, punctuated at rare intervals by major events that occur with great rapidity and help to establish the next stable era" (Gould, Ebdon, 1999), a process characterized by Kuhn as "paradigm shifts" (Kuhn, 1970). Gleick suggests that these changes are happening more regularly than before (Gleick, 2000), which is probably why change appears as a process of constant acceleration.

The need to deliver cultural resources, especially from major cultural organizations, such as museums or national libraries, has become an imperative closely associated with the core mission of these organizations to educate and elucidate, to promote and disseminate and to preserve culture. These attempts to reach out to new audiences and to refresh current audiences are major driving factors behind many digitization programmes and the shift towards digital repositories. As has been shown, the justifications for delivering cultural resources digitally are rarely made on economic or mercantile grounds as the returns on investment are relatively small (Tanner, 2004), but the cultural, educational and prestige returns are quite high.

With the digital revolution, data and information can now be transmitted to all corners of the world. Some predict that we are reaching a period of cheap access for all, but the reality is that there are still political, cultural and financial issues which prevent low cost access in certain strata of society and many parts of the world. The digital divide exists and could further disadvantage the poor, the under-educated and those in developing countries as the better-off, the better-educated and the economically developed forge ahead into the digital domain.

Views on the democratising nature of electronic networks vary wildly and we need to be cautious in our evaluation of these: for some we are on the verge of global utopia, an "age of optimism" (Negroponte, 1995), for others the Internet (our conduit to the digital library) "continues to remain an expensive western toy" in a world where less than 2% of the population is connected to the Internet and where 80% of the population has never even made a telephone call (Taylor, 2001).

3. Digital Library Technologies as Applied to Culture

The cultural sector often needs to utilize innovative technology to deliver artefacts in a way that encompasses their intrinsically audio-visual nature. Documentary sources remain vital but it is in auditory and visual materials that culture identifies unique challenges.

There are a number of core areas common to most digital library systems as described by open source models, such as referenced in the Open Archival Information System[I], and as implemented by DSpace[II] or FEDORA[III], for instance.

- Archival storage – especially digital preservation;
- Data management – especially descriptive metadata;
- Access mechanisms – especially search/browse functions and user interfaces;
- Administration.

The key areas for culture's interaction with digital libraries are descriptive metadata and digital preservation.

3.1. Describing Culture Through Metadata

"Doing research on the web is like using a library assembled piecemeal by packrats and vandalized nightly" (Ebert, 1998). Time and technology have moved on but the underlying principle of this quip remains valid. Metadata and tools for resource discovery are needed to allow users to locate the items they seek, whether they know of their existence or not. Because the audience for cultural resources is likely to be extremely wide and diverse then this is a particular challenge for digital libraries.

Finding a known object is always going to be easier than finding a range of previously unknown pertinent objects and if the starting perspective of the searcher is unknown because of diversity (age, education, language, etc.) then making a resource findable when it might be text, audio, video, 3D, geographic, database or image based is a challenge to any digital library. In a known case (e.g. Monet's paintings of Giverny), searches can be constructed by inexperienced users that will almost certainly result in satisfactory retrieval. It is when the user knows only the field of enquiry, and not the precise resource, that search engines are very much less useful. In the context sensitive world of the **ambient intelligence (AmI) environment**[IV], widely divergent user needs would be supported when asking for information, such as "tell me some historical information about this village and anyone famous who lived here"; "places to buy Monet souvenirs near where I am now"; "French gardens and painting"; "where did Monet live and what colour is his house", or "why did Monet paint flowers and gardens". These questions are not unusually difficult in a human mediated environment, such as the traditional library, but in a non-mediated digital environment they are not answerable unless very rigorous description and context has been provided.

An ambient intelligence environment is not yet with us but the developments towards the **Semantic Web**[V] are developing. The Semantic Web (a term coined by Tim Berners-Lee) views a future in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The infrastructure of the Semantic Web would allow machines as well as humans to make deductions and organize information. The architectural components include semantics (meaning of the elements), structure (organization of the elements), and syntax (communication). The use of RDF (Resource Description Framework) and XML (eXtensible Markup Language) are essential elements of this approach. The International Council of Museum's common extensible semantic framework, CIDOC Conceptual Reference Model (CRM)[VI] provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. It is an excellent example of the power of semantic approaches to express common concepts in basic categories that will have real longevity in use.

Other new and open standards have been developed that have particular relevance to the cultural domain. In particular, the **Metadata Encoding and Transmission Standard (METS)**[VII] is a standard for encoding descriptive, administrative and structural metadata regarding objects within a digital library, expressed using XML. One such application is that within the Greenstone Digital Library[VIII], open source software from the New Zealand Digital Library Project at the University of Waikato.

The **DIG35** standard for digital imaging is particularly useful with the explosion in born digital photography. Its aim is to "provide a standardized mechanism which allows end-users to see digital image use as being equally as easy, as

convenient and as flexible as the traditional photographic methods while enabling additional benefits that are possible only with a digital format."[IX]

The **Material eXchange Format (MXF)**[X] is targeted at the interchange of audio-visual material with associated data and metadata. MXF was designed by the broadcast industry with input from the user community to meet their demands for improved file based interoperability, improved workflows, more efficiency with the mixed and pro proprietary file formats experienced in the broadcast environment. It is being put forward as an open SMPTE standard[XI], is not compression-scheme-specific and simplifies the integration of systems using MPEG and DV, thus working with video and audio between different applications. Its major contribution is the transport of metadata, enabling media management and improved content-creation workflows by eliminating repetitive metadata re-entry. The MXF development community is also working on a specific dialect for Descriptive Metadata, called MXF DMS-1, which is being designed to describe people, places, times, production billing, etc.

The aptly named **"Deep Web"**, those massive resources missed by search engines due to being in a database or other non-harvested format, remains an often unnoticed problem to be resolved. Ironically, these may be very content rich and from a reliable publicly funded source, such as a library, museum or archive. Bright Planet have issued a White Paper[XII] that estimates:

- Public information on the Deep Web is currently 400 to 550 times larger than the commonly defined World Wide Web.
- The Deep Web contains 7,500 terabytes of information compared to nineteen terabytes of information in the surface web.
- The Deep Web contains nearly 550 billion individual documents compared to the one billion of the surface web.

The end user is certainly not seeing the complete picture when using search engines, but the user behaviour and attitude towards the search engine as the answer to everything worth asking means that content providers must engage with this problem to bring their resources from the Deep Web to the surface.

An example of how a provider might resolve this issue is provided by the OCLC Open WorldCat programme. OCLC have engaged with Google and Yahoo! to "make records of library-owned materials in OCLC's WorldCat database available to web users on popular Internet search, bibliographic and bookselling sites. The result: OCLC member libraries are more visible on the web, and their catalogues are more accessible from the sites where many people start their search for information"[XIII].

One further enhancement to cultural resource discovery has been the increasing implementation and use of the **Open Archives Initiative Protocol for Metadata Harvesting (OIA-PMH)**[XIV]. With 286 registered providers as of March 2005, it is an efficient mode for metadata exchange and relies upon a minimum requirement for unqualified Dublin Core to enable effective cross-walking of resources. Of course, Dublin Core has such a basic structure that many complex resources are simplified so much that hierarchical richness in their encoding is lost at the point of cross-walking. Google is now using OAI-PMH to harvest information from the National Library of Australia (NLA) Digital Object Repository[XV]. Other major users include the Library of Congress[XVI] and the UK Resource Discovery Network (RDN)[XVII].

3.2. Digital Preservation

"Culture, any culture... depends on the quality of its record of knowledge." (Waters, 1997)

The documentary heritage has always been at risk of damage or destruction through natural or human forces: fire, flood, warfare or neglect (Deegan and Tanner, 2002). We now face a new threat in the form of digital obsolescence.

"Not brittle papyrus and crumbling mortar" is the most severe threat to our cultural heritage today, but, as Mary Feeney expressed it, "the death of the digit". This "death of the digit" is related primarily to two factors that put at jeopardy current efforts in archiving and preserving our digital cultural heritage:

- First, technology develops ever more rapidly, reducing the time before a particular technology becomes obsolete.
- And secondly, unlike their analogue counterparts, digital resources are much more "unstable" with the effect that the integrity and authenticity of digital cultural resources is corrupted. (Geser and Mulrenin, 2002)

Digital data is being produced on a massive scale by individuals and institutions both large and small. Some of this data is created through digitization programmes but increasingly much of it is born, lives and dies in only digital form. Initially, digital files were part of a process of analogue publication but soon the digital became the publication and process became product. Digital preservation of cultural resources faces a number of challenges:

- Preserving the data streams integrity;
- Preserving the means to interpret the data stream;
- Preserving the means by which the resource is experienced.

Precautions can be taken which will help significantly to reduce the danger of digital loss and the Digital Preservation Coalition[XVIII] list a number of options:

- Storing in a stable, controlled environment;
- Implementing regular refreshment cycles to copy onto newer media;
- Making preservation copies (assuming licensing/copyright permission);
- Implementing appropriate handling procedures;
- Transferring to "standard" storage media.

The Digital Preservation Coalition provides an essential starting point for information and guidance for digital preservation methods, strategies and models.

The DigiCULT Report (Geser and Mulrenin, 2002) provides an excellent overview of digital preservation and includes a schematic from Salzburg Research that clearly identifies the complexity of long-term preservation when dealing with cultural objects.

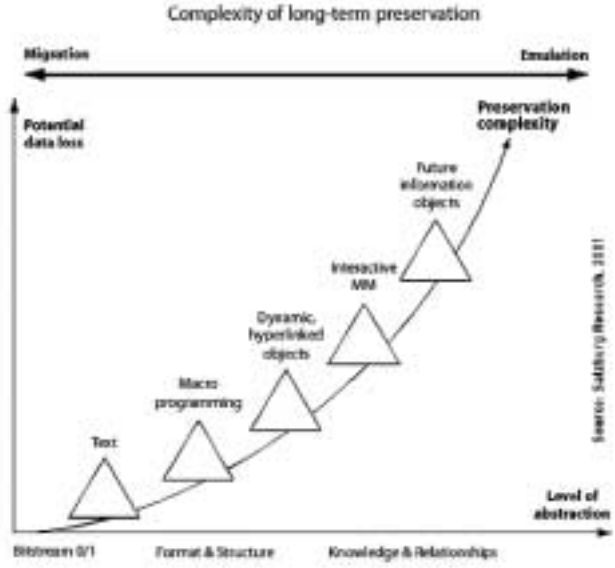

Figure 23. Complexity of long-term preservation

Some examples of good practice:

- Arts and Humanities Data Service (AHDS)[XIX] is a distributed service whose aim is to collect, preserve and promote reuse of electronic resources resulting from arts and humanities research.
- CURL Exemplars in Digital Archives (Cedars)[XX] addressed strategic, methodological and practical issues and provide guidance in best practice for digital preservation.
- Preserving and Accessing Networked Documentary Resources of Australia (PANDORA)[XXI] is Australia's Web Archive. The purpose of the PANDORA Archive is to collect and provide long-term access to selected online publications and web sites that are about Australia, are by an Australian author on a subject of social, political, cultural, religious, scientific or economic significance and relevance to Australia, or are by an Australian author of recognized authority and make a contribution to international knowledge.

4. Art for Everyone: Democratising Access to the Arts

The arts are the most obvious and visible of the cultural resources made available via digital library technologies and cultural organizations. The arts is generally assumed to include at least the visual arts, literature (written and oral), music and theatre arts. The democratising effect of the digital domain is due to its role in providing a platform for equality of access and engagement with the arts.

The arts are increasingly met by an incredibly diverse audience in an environment that delivers many benefits. Not least of these are:

- Benefits of aggregation, by making available a critical mass of works;
- Benefits of virtual unification, bringing together all the works of a single artist within one source rather than distributed across many institutions (possibly both physically and logically);
- Benefits of more open access, by making available resources that would otherwise be unavailable for public consumption;
- Benefits of enhanced access, by enabling new and refreshing ways to view or utilize the resource, including modifying the resource to benefit different audiences (e.g. multilingual);
- Benefits of integration, by enhanced finding aids to enable resources to be used in multiple environments and for multiple purposes but particularly for integration into teaching and learning materials;
- Benefits to mission, by democratising access cultural organizations can reach out to new audiences and demonstrate value to their community and stakeholders.

These benefits, whilst manifest, are offset by economic and copyright issues. Arts resources are not cheap to digitize, deliver or preserve access to, and thus the key questions for all digital library providers are who will pay for the resource, how will they pay, and what is worth the price of purchase? Whether resources are created for financial, mission or intellectual reasons, questions of sustainability will remain.

One significant barrier to delivering arts cultural content in particular is copyright. Arts resources, particularly in the performing arts, are legally complex as objects and this may present barriers for organizations wishing to digitize or deliver them. For instance, the choreographer who throughout her career takes video tape of dance rehearsals using her choreography does not necessarily own the rights to digitize those tapes – there may be performance rights, music copyright, broadcast rights, not to mention any contractual/licence agreements in force at the time of the performance. Contrary to the preconceptions held by many, owning art, artefacts or objects does not convey any automatic rights to reproduce and distribute – many museums of modern art will not own the reproduction rights for paintings in their own collections for instance (Tanner, 2004). This will block, delay or make more expensive attempts to digitize arts collections.

4.1. ARTstor

ARTstor[XXII] is a prime example of an arts resource created to address issues of aggregating access to content whilst establishing a sustainable business model to balance the needs of users with the interests of content providers regarding IPR and copyright[XXIII]. ARTstor, a non-profit entity initiated by The Andrew W. Mellon Foundation, is currently only available in the USA and Canada but is planned to become more widely available. The ARTstor Charter Collection currently contains approximately 300,000 images; by 2006 it is expected to contain 500,000. The Charter Collection documents artistic traditions across many times and cultures and embraces architecture, painting, sculpture, photography, decorative arts and design, as well as many other forms of visual culture.

ARTstor is a form of super-aggregation as the collections aggregate a number of source collections which themselves are the end product of collaborations with libraries, museums, photographic archives, publishers and scholars. ARTstor contains digital images and associated catalogue data, browser based JAVA and Flash tools to enable use of images, and a password based restricted user environment to protect the business model. This restricted environment has led to a number of technological issues for ARTstor in that there is a natural tension between wishing to provide flexible access and interoperability of a resource as desired by the user base with their need to protect the rights and licences that underpin the whole business model of the resource. In this, ARTstor expresses some fundamental technical issues in delivering arts via digital libraries of how to enable:

- Metadata in federated searches;
- Image views by searchers outside the restricted environment, so searchers can evaluate the usefulness of the object/collection;
- Stable URLs to enable linking from learning management systems or Image Groups;
- Downloadable images to authenticated users within a licensed environment;
- Further aggregation by incorporating institutions own collections within the same user environment as ARTstor uses to enable seamless art access (ARTstor offers to place images on ARTstor's servers and to map the institutionally provided data to ARTstor's collections).

This is done whilst protecting IPR and developing true economic sustainability.

ARTstor's digital library is orientated towards education institutions and the business model requires two fees to participate in the ARTstor Charter Collection. The Archive Capital Fee is one of fees with costs ranging from £40,000 to £1,000 depending on institution size. There is a smaller Annual Access Fee graduated in the same way from $20,000 down to $1,200. ARTstor notes that "the goal of ARTstor's fee structure is to make participation as broad as possible across a range of institutions while generating sufficient revenue to enable ARTstor to be a sustainable, long-term resource for the community"[xxiv].

ARTstor is an ambitious venture but one with sound backing and a track record from the JSTOR journals repository that suggest it may become just as ubiquitous within its user base. What ARTstor does address is a local North American need for low cost access to large aggregations of art for teaching, education and scholarship. The challenge for the long term is whether this resource could reach out to democratize access in a worldwide context and provide art for all.

4.2. RLG's Cultural Materials and Trove

The Research Libraries Group Cultural Materials[xxv] is a similar attempt to aggregate cultural content, with a "mission of providing greater access to primary sources and cultural materials – the rare and often unique works collected for their value to history, education and research"[xxvi]. RLG Cultural Materials contains a vast array of materials of paintings, photographs, posters, woodcuts, artist's books, maps, pamphlets and gravestones. The digital object standard METS (Metadata Encoding & Transmission Standard) provides a framework for representing works which have multiple digital media files associated with them[xxvii]. As of March 2, 2005 the service contains 231,819 works in 109 collections and is accessible through a subscription based business model. There is much useful advice within the resource as well digitization guidance[xxviii] and descriptive metadata guidelines[xxix].

RLG have uniquely made every effort possible within the constraints of the diverse collections and IPR interests to make as much content freely available through its public access site Trove.net[xxx]. Trove.net includes 26 unrestricted collections from RLG Cultural Materials, freely available in low resolution, watermarked form. RLG Cultural Materials and Trove.net combine to form a prime example of a mixed business model that allows more freedom of access to, and thus the democratising of, the arts and cultural heritage.

4.3. The State Hermitage Museum

The State Hermitage Museum[xxxi] resides in six magnificent buildings in the heart of St Petersburg. It now has a digital collection, which provides a virtual gallery of high-resolution artwork images from the State Hermitage Museum, including:

- Painting, Prints and Drawing;
- Sculpture;
- Machinery and Mechanisms;
- Arms and Armour;
- Furniture and Carriages;
- Ceramics and Porcelain;
- Applied Arts;
- Jewellery;
- Textiles;
- Numismatics and Glyptic;

- Costume;
- Archaeological Artefacts.

What is of particular interest, over and above the sheer quality of the artwork available in this digital collection, is the innovative use of technology for searching. They utilize IBM's experimental Query By Image Content (QBIC) search technology which offers the ability to find "a Gauguin masterpiece simply by recalling the organization of his subjects or locating a da Vinci painting by searching for its predominant colours"[xxxii]. This operates within the IBM DB2 Content Manager software platform, which although by no means unique in offering this functionality, has delivered a excellent exemplar at the Hermitage Museum. QBIC allows users to search a database of images by similarity, such that an image may be found through its degree of similarity within set parameters, such as colour percentages or colour layout. Searches may also be customized by specifying exact percentages of one or more colours within the image being sought.

Another example of the search-by-similarity technology is provided by the Virage VS Archive[xxxiii] video and rich content management system, but with expanded capabilities enabling comparison by colour, texture, colour layout and structure (general shape characteristics). Virage SmartEncode is also available for transforming video and audio, with automated indexing, analysis and encoding.

The Hermitage Museum Digital Collection was facilitated by a $2 million grant from IBM and again this demonstrates the high start up costs associated with creating cultural assets for the digital domain. It remains a useful exemplar, receives on average over 200,000 hits a day (Mintzer, et al., 2001) and has won awards including the Best Overall Website award at the Museums and the Web Conference 2000.

4.4. British Library

The British Library Treasures provides online access to their treasures and also a gallery function through "Turning the Pages" technology. The treasures are provided as complete works online with basic functionality, such as page views, zoom in function and expert commentary. The British Library Treasures[xxxiv] include:

- William Shakespeare in Quarto – 93 copies of the 21 plays by Shakespeare printed in quarto;
- Caxton's Chaucer – William Caxton's two editions of Chaucer's Canterbury Tales;
- Gutenberg Bible – two copies of Europe's earliest printed book;
- Magna Carta;
- Renaissance Festival Books – festivals and ceremonies in Europe from 1475 to 1700.

The British Library also offers an online gallery of great works. The key difference to the treasures is that these may not all be complete works online but rather snapshots to cast illumination upon their best features. The other key difference is the underlying technology for display – the Turning the Pages feature[xxxv]. This is originally intended as an interactive exhibition installation in the Library allowing visitors to turn the pages of virtual manuscripts, using touch-screen technology and animation. Other features include being able to zoom in on the digital images and read or listen to notes explaining features of each page. The technology allows for other innovative features such as turning around the text in a Leonardo da Vinci notebook, so users can read his "mirror" handwriting. This technology has been adapted for online viewing using Macromedia Flash technology to deliver the resources in an online environment. Whilst not as satisfying as the Turning the Pages exhibition pieces physically located in the British Library, the online versions are rich and provide a sense of performance that brings the remote viewer closer to the experience of being in the presence of the original artefact. The Turning the Pages resources include:

- Lindisfarne Gospels;
- Sultan Baybars' Qur'an;
- Sforza Hours;
- Golden Haggadah;
- Vesalius' Anatomy;
- Sherborne Missal;
- The Diamond Sutra.

Turning the Pages is also in use at the National Library of Ireland for their "James Joyce and Ulysses" Exhibition to mark the centenary of Bloomsday.

5. Preserving Culture

One of the key responsibilities held by libraries, museums and other memory organizations is that of preserving culture. This may be preserving the record of a way of life or the very roots of a culture, its language, literature, music and traditions. Digital libraries play an important role in preserving culture and in connecting people with their national and regional identities.

Cultural resources are at risk from natural disaster and human mediated destruction. The loss of archives includes examples, such as:

- The Catholic University of Louvain, Belgium which burned down in both 1914 and 1940 due to wartime bombing, destroying 230,000 books, 800 of them incunabula printed before the year 1500 and 900 manuscripts;
- The Anna Amalia Library in Weimar, Germany destroyed through fire in September 2004. 30,000 unique books plus paintings destroyed, many more damaged and "the literary memory of Germany has suffered severe damage", states German Culture Minister Christina Weiss;
- The Asian Tsunami had an impact on a number of museums, libraries and archives, as well as the people of many countries affecting not just lives but cultural memories. For example, the Museum Negeri Provinsi D.I. Aceh (Banda Aceh, Indonesia) lost six colleagues and had the ceramic collection heavily damaged[XXXVI]. The State Library of Aceh (Badan Perpustakaan Daerah Propinsi Aceh) lost all of its children's, youth, adult and reference collections, and 23 of its staff were killed. The Documentation and Information Centre of Aceh Library (Perpustakaan Pusat Dokumentasi dan Informasi Aceh) seems to have lost all of its collections[XXXVII].

Culture is at constant risk and the digital domain, to a certain extent, may exacerbate this risk by storing resources in formats that have problems for long term storage and retrieval. But the digital library itself is a powerful tool for holding back the tide of cultural diffusion and loss. Digital libraries are often the key tool by which citizens can engage with their own history, culture and language whilst also being of enormous value in restoring cultural artefacts to public view where the original is lost or too fragile for normal display.

5.1. Rebuilding Libraries in Kosovo

For people hit by war and conflict the damage and destruction of documents and objects of cultural heritage may have consequences that far outlast the period of conflict itself. Kosovo has been characterized by war and conflict for a long period of time. The UN Interim Administration Mission in Kosovo (UNMIK) and the library professionals in Kosovo thus face a major task to reconstruct library buildings and re-establish library services at all levels. It is urgent to secure the cultural heritage represented by books, documents and a large range of other objects and ensure they are both kept safely for the future and made widely available. A programme aiming to preserve and secure Kosovar cultural heritage has been proposed and includes suggestions to:

- Compile inventories of cultural property at institutional, regional and national level;
- Identify short term projects for emergency protection of property at risk;
- Establish links and partnerships with relevant institutions abroad holding collections relating to Kosovo;
- Begin digitization of Kosovar printed or written cultural heritage;
- Reassume the production of a Kosovo national bibliography;
- Establish emergency plans and training of staff in risk preparedness and recovery for relevant institutions;
- A digitization project should be initiated in order to ensure a systematic preservation and documentation of the cultural heritage of Kosovo and open access dissemination through the Internet. Special focus should be on a broad definition of cultural heritage including cultural, religious, historical and ethnic documents, items, images, sounds, etc.;
- Albanian and Serb languages should both be represented and the project should aim at a complete coverage in all fields possible;
- The far-reaching distribution of material on Internet will to some extent compensate for the damage done to important texts and documentation during repression and war. Material lacking in Kosovo should be digitized outside Kosovo in close cooperation with relevant museums, libraries and institutions in foreign countries.

The project should be done in close cooperation with the National Archive of Kosovo and the National Museum of Kosovo and a commission of Albanian and Serb experts should be established with the task of securing quality and

authenticity of material digitized and distributed on Internet. The project should be run in close cooperation with international experts (Frederiksen and Bakken, 2001).

The initiative here shows the important role that digital resources have on preserving and delivering cultural resources. Even in post-conflict circumstances digital libraries are important to the rebuilding of the cultural and national identity of a country. It also highlights the central role that cataloguing, and from this metadata, has in enabling effective digitization and preservation of resources – without the knowledge of what exists (or was lost) then planning and implementations may be significantly hindered.

5.2. Reconstruction of the Great Buddha of Bamiyan, Afghanistan

In the valley of Bamiyan, Afghanistan, two large standing Buddha were carved out of the sedimentary rock circa 2,000 years ago. In 2001 the Taliban militia demolished these statues (each 53 and 35 meters high) as they considered them an insult to Islam. Later, after the fall of the Taliban, a consortium was established to rebuild the Great Buddha of Bamiyan in its original shape, size and location. The Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland,[XXXVIII] have made a 3D computer reconstruction of the statue, which serves as a basis for the physical reconstruction. The 3D modelling is based upon point cloud generation and manual measurements from a number of existent digital images of the original statue. Currently the fundraising efforts are under way to support the physical reconstruction of the Great Buddha of Bamiyan.

5.3. Preserving Language and Literature

The preservation of indigenous language and literature is something that has increasing importance in European digital library resources. As English, French and German are the dominant languages of Europe then other traditions could become marginalized or ignored and indigenous traditions lost. The two examples given here are both large textual resources that thus preserve and promote the culture and literary traditions of those nations.

The **Digital Library of Dutch Literature**[XXXIX] is a growing collection of primary and secondary information on Dutch language and literature and its historical, societal and cultural context. Researchers and others who are interested, from the Dutch-speaking areas and beyond, can get direct or controlled access to this information via the Internet. The site is an initiative by the Digital Library of Dutch Literature Foundation (Stichting dbnl), founded by the Society of Dutch Literature (Maatschappij der Nederlandse Letterkunde), with financial support from the Dutch Language Union (Nederlandse Taalunie) and the Netherlands Organisation for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek).

The **Austrian National Library**[XL] has a mass digitization programme. In 2003 they began by scanning historical newspapers and journals and now there are 2.7 million pages online, there are also historical bills and parliamentary papers with first editions to follow in the future. At the moment there are only the images of the texts online and this is a significant issue to be resolved in the future to allow fully searchable text. Presently the resource resembles a traditional library in that the resources are easily available and well indexed at the level of the work but the content is not yet searchable. Conversely, the **Austrian Literature Online**[XLI] resource provides full searchability to 5,619 Austrian books and is a model for delivering texts online.

5.4. Keeping Memory Alive: the Memory Projects

There is a number of digital library projects which conform to the memory model as established by American Memory from the Library of Congress or the Making of America[XLII] digital resources. These North American resources remain the greatest examples that are followed and emulated worldwide.

American Memory[XLIII] provides free online access to written and spoken words, sound recordings, still and moving images, prints, maps and sheet music "that document the American experience... a digital record of American history and creativity". There are over 5 million items in the digital library which serves an audience of education and lifelong learning. The Library of Congress has expended over $45 million on its development since 1994 and views it "as an integral component of the Library of Congress's commitment to harnessing new technology as it fulfils its mission to sustain and preserve a universal collection of knowledge and creativity for future generations"[XLIV].

The Library of Congress has adopted the OAI Protocol for Metadata Harvesting to make records for some of its digitized historical collections available for integration into other services. It has also adopted the METS metadata

standard[XLV]. Several of the American Memory collections have associated archival finding aids and these have been marked up following the Encoded Archival Description (EAD) standard. As of late 2004, the finding aids are being converted from the SGML (Standard Generalized Markup Language) version of EAD to XML using a TEI-conformant DTD (see American Memory DTD for Historical Documents[XLVI]).

UNESCO's Memory of the World Programme[XLVII] also encourages digital library developments as it seeks "to guard against collective amnesia calling upon the preservation of the valuable archive holdings and library collections all over the world ensuring their wide dissemination". A key example offered to the Programme is provided by the Australian National Library in regard to web archiving – the PANDORA Archive[XLVIII]. Another example is the Czech National Memory project.

The **PANDORA Archive** is Australia's Web Archive. The purpose of the PANDORA Archive is to collect and provide long-term access to selected online publications and web sites that are about Australia and by an Australian author. Whilst leading the world on issues of digital preservation the National Library of Australia has also developed an integrated, web-based, web archiving management system, known as PANDAS[XLIX] (PANDORA Digital Archiving System).

PANDAS is designed to support the workflows defined by the staff of the Australian National Library's Digital Archiving Section, and also adopted by the other PANDORA partners. These workflows include:

- Identifying, selecting and registering candidate titles;
- Seeking and recording permission to archive;
- Setting harvest regimes;
- Gathering (harvesting) files;
- Undertaking quality assurance checking;
- Initiating archiving processes; and
- Organising access, display and discovery routes to, and metadata for, the archived resources.

PANDAS supports these work flows by means of the following functions:

- The management of administrative metadata about titles that have been either selected for archiving, rejected, or are being monitored pending a decision;
- The management of access restrictions;
- The scheduling and initiation of the harvesting of titles selected for archiving;
- The management of the quality checking and assurance process and associated problem fixing;
- The preparation and organization of archived instances for public display through title entry pages, and title and subject listings; and
- The provision of defined management reports.

The **Czech National Memory Project**, launched June 2003, aims to "create an archive to preserve the history of the Czech Republic from 1945-1989 in all its forms – written, spoken and material – by using young people to interview their elders about their experiences"[L]. The total project would become a collective memory of the Czech experience of Communism – a virtual "distributed" archive of oral history, eventually backed by a collection of real objects.

The **Uruguay Archivo Fotográfico de Montevideo**[LI] has the objectives to conserve and preserve the collections of glass negatives (35,000) showing the cultural life of Montevideo since 1865. The archive is also digitising images and seeks to promote the use of the images as widely as possible to "convert them into a common cultural property... to safeguard the knowledge of Montevideo in its actual state to the future generations through the photography" (Sambaíno, 2004). "Recovering a city it is not only improving its streets, its public squares or its drainage: recovering a city is also recovering its memory. For this reason the Municipality of Montevideo, through its Photographic Archive, is preserving the most important photographic collection of our city."[LII]

The **Memoria Chilena**[LIII] digital library provides "wide access and services to support educational process, to support our identity through the heritage, and to give more opportunities to the population to access the culture" (Amunategui, 2005). In particular, future plans include a new resource within this site especially designed for children between 4 and 8 years old to encourage "love and respect for the cultural heritage and identity". It is partnered by the **Biblioteca Virtual del Bicentenario**[LIV] which represents Chilean culture, history and literature through the representative literature of Chilean culture.

6. Connecting with Home and Family: Issues of Diasoera, Displacement, Family History and Cultural Identity

Cultural cohesion and social inclusion are issues that many heritage and cultural organizations increasingly seek to address. In this modern era, with worldwide forced and economic migration, there is a need for digital resources to help maintain a sense of home and family as peoples become displaced and their cultural identity potentially becomes diffused. Equally, indigenous peoples have the right to expect digital resources will help solidify their sense of home and family through family histories and resources that reflect their cultural identity.

The following resources are all examples of how digital library resources are addressing these issues. The advantages conveyed by the digital domain include the ability to speak to an audience not necessarily limited by borders and the capacity to maintain access to cultural resources that are meaningful to that audience at a distance. The digital domain also provides many opportunities for virtual communities to be created and maintained and some of these revolve around a cultural resource, whether a religious codex, a family history resource or oral histories.

6.1. Digital Shikshapatri

The Digital Shikshapatri[LV] is provided by the Indian Institute Library and Oxford University, in partnership with Forced Migration Online at the Refugee Studies Centre, Oxford University. The Shikshapatri Manuscript is one of the great treasures of British Hinduism and is an object of pilgrimage for Hindus. 900 devotees per year visit Oxford University to darshan, or venerate their holy manuscript. The digitization and online provision of this manuscript allow wider access, plus more extensive study and also contextualization with supporting materials. The wider access to the manuscript allows for more cultural cohesion, education, identity re-affirmation and cultural pride for this diaspora community.

6.2. Shoah Archive

The Shoah Foundation[LVI] has 120,000 hours of Holocaust survivor and witness video testimonies which have been digitized into a 200 Terabyte database and distributed over Internet2 to overcome prejudice, intolerance and bigotry – and the suffering they cause – through the educational use of the Foundation's visual history testimonies. In 1994, the Shoah Foundation was established by Steven Spielberg to document the experiences of survivors and other witnesses of the Holocaust. More than 90 percent of the testimonies are from Jewish Holocaust survivors; however, the archive also contains interviews with other survivors, including Jehovah's Witnesses, Sinti and Roma, homosexuals, political prisoners and survivors of eugenics policies. In addition, the Shoah Foundation has interviewed rescuers and aid providers, liberators and war crimes trials participants (Gustman, 2005).

Under an $8 million grant from the National Science Foundation the Shoah Foundation investigated the use of speech recognition for information retrieval from our oral archives. They have partnered with IBM Thomas J. Watson Research Center, Johns Hopkins University, University of Maryland, Charles University and University of West Bohemia in the Czech Republic. Digital archiving of the spoken word is an important emergent method for capturing the human experience. The Shoah Foundation feels that in the future a great deal of our cultural heritage will be archived in this form. "If we are to learn from our past, teachers, students, historians and others will need effective access to these resources. The enormous scale of these collections and the tremendous expense of manually cataloguing multilingual audiovisual materials will make it impractical to rely on manual techniques alone."[LVII]

The research project proposes to dramatically improve access to large multilingual collections of recorded speech by advancing the state of the art in technologies that work together to achieve this objective:

- Automatic speech recognition;
- Computer-assisted translation of domain-specific multilingual thesauri;
- Natural language processing techniques for automated creation of metadata;

- Support for efficient professional cataloguing;
- Support for search and exploration.

Advancing the state of the art in this technology will produce significantly improved access to many artefacts of cultural heritage.

6.3. Digital Imaging South Africa

Digital Imaging South Africa (DISA)[LVIII] based at the University of KwaZulu-Natal aims to make Southern African material of high socio-political interest, such as anti-apartheid material, accessible to scholars and researchers worldwide. This is a non-profit making initiative for cooperation among research libraries and archives in Southern Africa, sponsored by the Andrew W. Mellon Foundation. DISA have done digital imaging projects across Southern Africa, which results in important historical material becoming more widely accessible.

This model is important in that it encourages self determination for developing important historical resources and building capacity, knowledge and expertise in digital imaging technology in the library and archival community in the region.

6.4. Te Ara (The Pathway) New Zealand Encyclopaedia

The New Zealand government has funded an initiative to develop a new encyclopaedia of New Zealand thematically online. The Te Ara Encyclopaedia[LIX] combines short essays with a variety of multimedia additions to create a informative resource of wide appeal. Extensive links provide pathways to the digital collections of libraries, archives and museums around the country. Te Ara will allow the inclusion of community contributions, such as photographs, oral histories, or updates of descriptive information. The resource will have a very long gestation (published progressively over 9 years) which may possibly create long term management, editorial and preservation problems. In particular, the development of a completely new platform rather than utilising commercially supported technology may also prove problematic for the long term. Alternately, the slow burn of progressive development over a long term may convey the capacity to flexibly respond to technology changes. This resource will provide a useful barometer for this type of resource implementation. The first theme about the people of New Zealand has been recently launched and they have also digitized the only other complete encyclopaedia of New Zealand produced by the New Zealand government in 1966.

6.5. Aboriginal Resources in Australia

AIATSIS[LX] is gradually digitising their collection to be made available online via a digital asset management system. One of the key driving factors behind this initiative is to start providing digital cultural repatriation – i.e. AIATSIS will continue to hold and preserve the culturally significant items but provide open and free access to the originating communities. This is considered important as it is beyond the financial abilities of many individuals in remote aboriginal communities to afford the cost of travelling to Canberra to view the collections but with the development of indigenous knowledge centres increasingly these communities have access to the Internet.

The digital asset management project manager for AIATSIS thinks that digital repatriation may work well in this case. Because many traditional aboriginal communities had relatively low levels of material possessions as "nomadic hunter gatherer peoples [are] limited to what they can carry, but are rich in stories and music. Many of our holdings are of people speaking or performing chants – i.e. the cultural artefacts we hold are in the main intangible." (Moncur, 2005) There are some remaining issues relating to issues of cultural sensitivity and access control to sensitive recordings but it is hoped that in the near future this resource will become available.

Awaba[LXI] is another aboriginal resource. It is a small electronic database and guide to the history, culture and language of the indigenous peoples of the Newcastle and Lake Macquarie region of New South Wales, Australia. Awaba is the product of collaboration between the University's School of Liberal Arts and Wollotuka School of Aboriginal Studies, in partnership with representatives of the Awabakal Aboriginal community. The project accords with the University's proposed Charter of Aboriginal Reconciliation, which will formalize a commitment to bring Aboriginal and non-Aboriginal Australians together through consultation and cooperation amongst university staff and members of local Aboriginal communities. The resource provides access to documents and materials about the Awabakal people and language plus a thematic guide to these materials, with contributions prepared by various locally-based scholars, writers and community leaders.

6.6. Native American Heritage

The **Labriola National American Indian Data Center**[LXII], part of the ASU Libraries, is a research collection that brings together in one location current and historic information on government, culture, religion and world view, social life and customs, tribal history and information on individuals from the United States, Canada, Sonora, and Chihuahua, Mexico.

The **Cultural Heritage Preservation Institute (CHPI)** and its research component, the Digital Collective[LXIII], has the goals to:

- Create educational materials about Diné culture in Navajo voice;
- Create primary materials by and about contemporary Native American people and their cultural heritage for future use and preservation;
- Raise awareness of the role of museums and archives in preserving cultural materials;
- Build into the CHPI a process to capture, describe and preserve the digital images, drawings, texts and new knowledge created during the institutes.
  (Smith, 2002)

The Cultural Heritage Preservation Institute have developed the Digital Collective, "a model for storing and accessing shared information and knowledge, as well as for creating new knowledge and recreating global memory, and a place where people share personal and professional information and where they seek connections and build community" (Smith, 2002). The Digital Collective is a space where members of the community can add their own digital objects alongside stories, information and experiences about those objects or already existent objects in the digital library.

6.7. African Online Digital Library

The African Online Digital Library (AODL)[LXIV] has the goal of being an online digital repository that adopts "the emerging best practices of the American digital library community and apply them in an African context". Focussed mainly as an educational and scholarly resource it produces multilingual, multimedia materials. For example, the Passport to Paradise gallery highlights images found all over Dakar and focuses upon the urban visual culture of the Mourides, a Senegalese Sufi movement centred upon the life and teachings of a local saint named Sheikh Amadou Bamba.

The AODL consider that many African languages are seriously under-represented in digital libraries. They are carrying out research to find ways to digitize and deliver linguistic content on the web. The outcomes of this research may result in fully developed tools and methodologies that will help to represent and preserve African language content (particularly Pulaar) in a robust, open source, electronic environment.

7. Digital Futures and Culture

7.1. Metadata

Heritage and cultural resources lend themselves to multiple interpretations, uses and views and thus tend to require sophisticated systems and rich metadata to realize their full value and potential in the digital domain. The digital library community has a good range of controlled vocabularies and descriptive metadata schemas, but with this comes concomitant issues of crosswalking resources and enabling interoperability. The means that to effectively access a wide range of resources in varied environments in order to meet disparate audience expectations remains the key technical challenge for digital cultural resources.

As elegantly expressed in the DigiCULT Expedition Report, "We found many islands, with very different islanders and views of the future digital heritage space... There is very little likelihood of a future digital heritage space being created unless ways can be found to bring the different islands closer together." (Geser and Pereira, 2004)

7.2. Digital Preservation

Alongside access and usability there is a significant issue of digital preservation that digital libraries must address. Digital data is being produced on a massive scale by individuals and institutions, both large and small. Some of this data is created through digitization programmes but increasingly much of it is born, lives and dies in only digital form. Digital preservation of cultural resources faces a number of challenges:

- Preserving the data stream's integrity;
- Preserving the means to interpret the data stream;
- Preserving the means by which the resource is experienced.

Storage media continue to reduce in cost per terabyte and the technical issues underlying data stream storage and preservation thus become simplified to a certain extent. The technical issues surrounding interpretation and usability are not getting easier, but increased use of open standards for data representation is helping. What remain as key challenges are:

- The will and capacity to put in place a strategic preservation plan;
- Development of institutional policy that will continue for the long term;
- Building a sustainable economic model to enable the plans and policy to be carried forward.

7.3. Digital Divide

Readiness to exploit the available technologies is a key factor in digital library take-up for cultural resources. The capacity of a population to access digital resources will affect the investment and development opportunities for providers of cultural resources. Looking to the future it is clear that the top 12 countries in readiness terms will be better placed to develop resources than those lower down the scale. Readiness is not always defined by national wealth but it is clear that those at the lower end of readiness are defined by poor infrastructure, lower wealth per capita and sometimes ongoing conflict.

The top 12 networked readiness countries:

1. Singapore
2. Iceland
3. Finland
4. Denmark
5. United States of America
6. Sweden
7. Hong Kong
8. Japan
9. Switzerland

10. Canada
11. Australia
12. United Kingdom

The bottom 12 networked readiness countries:

93. Malawi
94. Zimbabwe
95. Ecuador
96. Mozambique
97. Honduras
98. Paraguay
99. Bolivia
100. Bangladesh
101. Angola
102. Ethiopia
103. Nicaragua
104. Chad
(World Economic Forum[LXV])

These network readiness estimates will provide evidence to assist those looking to address the digital divide in relation to culture worldwide. The digital divide in this context is defined by the inequitable access to local, national and world heritage and culture due to differences in access to computers and the Internet. The roots of the digital divide are social and economic inequality and these are difficult to overcome. A technology divide is also not a new phenomenon as new technology – for example, radio, television and computers – has historically opened a wide gap that has taken a long time to be narrowed and even then failed to close the information or education gaps adequately. Whilst we can hope for the digital divide to be eradicated, it might be more reasonable to expect information technology to address some of the worst information, education and cultural resource inequalities rather than solve all of them.

Rao defines success in the Internet economy through 8 factors and bridging the digital divide is achieved in his opinion by addressing the inequalities across these factors. Rao's factors are:

- Connectivity
- Content
- Community
- Commerce
- Capacity
- Culture
- Cooperation
- Capital

(Rao, 2000)

Indonesia has looked to address the digital divide through the national digital library network IndonesiaDLN[LXVI]. As Indonesia is an island nation, there are obvious infrastructural problems in developing and integrating information in an equitable way. However, the digital domain and telecommunication technologies can enable distributed information centres and environments for cooperative working that will make a contribution to narrowing the digital divide. The IndonesiaDLN will "work at the following goals:

- Content: The amount of local language content will increase, and many sectors will put content online that will have local relevance and usefulness.
- Community: The DL systems are not stand-alone systems; there are connections between them. Their collections are unified, managed and presented in an integrated DL network system. These collections will promote discussion within communities, and many forums will be established, whether using web-based application or extended by e-mail.
- Capacity: DL systems, especially at research and educational institutions, will contain useful information and knowledge available locally. For example, DLs of such materials as electronic theses and dissertations and research reports will be very useful in increasing the nation's information literate and narrowing the skill gaps. The DL network also can be utilized to deliver e-learning and distance-learning materials.

- Cooperation: Cooperation is a logical impact of communication among the communities. By publishing local information on DL networks, every institution will have chances to open collaboration with other people or institutions. This process can happen at the interpersonal, state, national or regional level."

(Fahmi, 2002)

In Vietnam, the national information infrastructure has been "rapidly developed enabling it to play an important role in socio-economic development. The digital content however, should be strengthened and developed as a major component of the IT master plan and the software industry... Participation in the InterCity Digital Library Demo Project is an important opportunity... for Vietnamese libraries, in general, to share professional experiences in the construction of digital libraries. It would be a really good chance for bridging the digital divide between countries in the region." (Hung, 2001)

In Africa, the digital divide is being addressed by many projects and in terms of infrastructure there have been great advances recently with every African country now having at least a web presence. Developments in wireless communication networking and low orbit satellites will enable access to even the most remote areas of Africa to potentially allow poorer communities to take an active part in the digital and information domains.

In the digital library realm there has also been significant digital appropriation of African cultural and historical resources by the developed world through many projects of interest to Africanists. A typical example is Africa Focus[LXVII], a project of the University of Wisconsin-Madison Libraries, which provides the general public with access to a digitized selection of images and sounds from the collection of the University's African Studies Program. However, there are also indigenous projects of real significance and one example is DISA: Digital Imaging Project of South Africa. As its first phase, DISA's participating institutions digitized some 40 antiapartheid periodicals of the three decades 1960-1990. Skills are actively shared and disseminated from DISA and foundations have been laid for future cooperative projects within the region.

Britz and Lor asked recently, "if African documentary heritage is digitized, how many Africans will be able to benefit?" – and came to the following conclusions. "Regardless of who coordinates and controls, there are some serious moral and legal concerns. Taking the case of African heritage, these are:

- Who has access to this information? For example, will African scholars be able to access this information free of charge?
- What control will the originating community have over their information once others have digitized it?
- Will originating communities be identified as the original creators of their cultural heritage and will they have the right to control access and non-disclosure of certain categories of their cultural heritage, for example sacred knowledge artefacts?
- To what extent will the global rules on intellectual property be able to protect this common heritage of Africa and prevent it from becoming exclusive, private property? Will the international intellectual property regimes be able to maintain the balance between private ownership and common heritage of the people of Africa?
- Will the people of Africa be fairly compensated for the use of their knowledge by others and what incentives will there be for them to make their body of knowledge available to the rest of the world?"

(Britz and Lor, 2004)

These issues are essential moral and ethical questions for all digital library developments and would be equally applicable for resource development across the entire developing world.

UNESCO itself is developing an international convention on cultural diversity (CCD)[LXVIII]. The CCD would be an international legal agreement to implement the principle that culture cannot be reduced to a commodity. It would potentially allow countries to exclude cultural policies and even media from "free trade" deals like the WTO. The Declaration emphasizes the principles of pluralism, respect for human rights, promotion of creativity and international solidarity. These principles are of importance to digital libraries and their adoption would be of great benefit to our digital futures.

In conclusion, looking to a more distant digital future, the goal for digital libraries will be to bring cultural resources equitably to every home, citizen and business. This will surely lead, eventually, to the vision of "ambient intelligence", a vision where digital cultural resources will provide "anywhere, anytime, natural and enjoyable access... for all" (Geser and Pereira, 2004).

Bibliography

1. Adams, D. (1999) *Predicting the Future*, H2G2 Entry ID: A216433. Available at: www.bbc.co.uk/h2g2/guide/A216433

2. Amunategui, X. C. (2005) Director, Biblioteca Nacional de Chile, in personal communication by email, 21 February.

3. Arms, W. (2000) *Digital libraries*. The MIT Press, Boston.

4. Britz, J. and Lor, P. (2004) *A Moral Reflection on the Digitization of Africa's Documentary Heritage*, IFLA JOURNAL 30, 3. Available at: http://www.ifla.org/V/iflaj/ij-3-2004.pdf

5. *Cambridge Advanced Learner's Dictionary* (2003) Cambridge University Press, Cambridge.

6. Castells, M. (1996) *The rise of the network society*, Blackwells.

7. Deegan, M. and Tanner, S. (2002) *Digital Futures: Strategies for the Information Age*, Library Association Publishing, London.

8. Ebert, R. (1998) Yahoo! Internet Life, September. Available at: http://marylaine.com/exlibris/xlib27.html

9. Fahmi, I. (2002) *The Indonesian Digital Library Network Is Born to Struggle with the Digital Divide*, Bulletin of the American Society for Information Science and Technology, 28 (4), April/May 2002. Available at: http://www.asis.org/Bulletin/May-02/fahmi.html

10. Frederiksen, C. and Bakken, F. (2001) *Libraries in Kosova/Kosovo: A general assessment and a short and medium-term development plan*, Joint UNESCO, CoE and IFLA/FAIFE Kosova Library Mission. Available at: http://www.unesco.org/webworld/kosovo_library/

11. Geser G. and Mulrenin A. (2002) *The DigiCULT Report: Technological landscapes for tomorrow's cultural economy Unlocking the value of cultural heritage*, European Commission, Directorate-General for the Information Society. Available at: http://www.digicult.info/pages/report.php

12. Geser, G. and Pereira, J., eds. (2004) *The Future Digital Heritage Space: An Expedition Report*, December 2004, European Commission, Directorate-General for the Information Society. Available at: http://www.digicult.info/downloads/dc_thematic_issue7.pdf

13. Gleick, J. (2000) *Faster*, Abacus.

14. Gould, S. and Ebdon, R. (1999) IFLA/UNESCO Survey on digitisation and preservation, IFLA Offices for UAP and International Lending. Available at: www.unesco.org/webworld/mdm/survey_index_en.html

15. Greenstein D. (2000) *Digital Library Federation draft strategy and business plan*, Digital Library Federation. Available at: www.diglib.org/about/strategic.htm#archs/

16. Greenstein, D. and Thorin, S. (2002) *The Digital Library: A Biography*, Second edition, ISBN 1-887334-95-5. Available at: http://www.clir.org/pubs/reports/pub109/contents.html

17. Gustman, S. (2005) Chief Technology Officer, Survivors of the Shoah Visual History Foundation, in personal communication by email, 15 March.

18. Hung, T. B. (2001) *Bridging the digital divide in Vietnam*, ASTINF Newsletter, July-September 2001, 16(3). Available at: http://www.stii.dost.gov.ph/astinfo2/frame/july_to_sep_2k1/pg_10_to_14.htm

19. Kuhn, T. (1970) *The structure of scientific revolutions*, University of Chicago Press.

20. Mintzer, F., Braudaway, G., Giordano, F., Lee, J., Magerlein, K., D'Auria, S., Ribak, A., Shapir, G., Schiattarella, F., Tolva, J., and Zelenkov, A. (2001) Populating the Hermitage Museum's new web site. *Communications of the ACM*, 44 (8), 52-60.

21. Moncur, D. (2005) Digital Asset Management System Project Manager, AIATSIS, Australia, in personal communication by email, 14 February.

22. Mulrenin, A. and Geser, G. (2001) *The DigiCult Report: Technological Landscapes for Tomorrow's Cultural Economy: Unlocking the value of cultural heritage*, Salzburg Research, on behalf of the European Commission. Available at: http://www.salzburgresearch.at/research/publications_detail_e.php?pub_id=27

23. Negroponte, N. (1995) *Being digital*, Coronet.

24. Rao, M. (2000) *Struggling with the Digital Divide: Internet Infrastructure, Content, and Culture*, On the Internet, September/October 2000. Available at: http://www.isoc.org/oti/articles/1000/rao.html

25. Sambaíno, S. (2004) *Images and sounds in Uruguay*, World Library and Information Congress: 70th IFLA General Conference and Council, Buenos Aires, Argentina. Available at: http://www.ifla.org/IV/ifla70/papers/134e_trans-Sambaino.pdf

26. Smith, K. R. (2002) *From Oral Tradition to Digital Collectives: Information Access and Technology in Contemporary Native American Culture*, RLG Digi News, 6 (6). Available at: http://www.rlg.org/legacy/preserv/diginews/v6_n6_feature1.html

27. Taylor, D. J. (2001) Curse of the teenage cybergeeks: review of Michael Lewis, The Future Just Happened, *Sunday Times Culture* (15 July), 34-35.

28. Tanner, S. (2004) *Reproduction charging models & rights policy for digital images in American art museums: A Mellon Foundation funded study*, King's College London. Available at: http://www.kcl.ac.uk/kdcs/USart.htm

29. Waters, D. J. (ed) (1997) *Digital archiving: the report of the CPA/RLG Task Force*, Preservation and digitisation: principles, practices and policies, National Preservation Office.

30. Waters, D. J. (2003) *Beyond Digital Libraries. The Organizational Design of a New Cyberinfrastructure*, NSF Post Digital Library Futures Workshop, Chatham, Massachusetts,USA. Available at: http://www2.sis.pitt.edu/~dlwkshop/paper_waters.html

References

I http://www.rlg.org/en/page.php?Page_ID=377
II http://www.dspace.org/
III http://www.fedora.info/
IV http://www.ehu.es/ami-life/
V http://www.w3.org/DesignIssues/Semantic.html
VI http://cidoc.ics.forth.gr/
VII http://www.loc.gov/standards/mets/
VIII http://www.greenstone.org
IX http://www.i3a.org/i_dig35.html
X http://www.mxfig.org/index.php
XI http://www.smpte-mxf.org/
XII http://www.brightplanet.com/technology/deepweb.asp
XIII http://www.oclc.org/worldcat/open/default.htm
XIV http://www.openarchives.org/OAI/openarchivesprotocol.html
XV http://www.nla.gov.au/digicoll/oai/
XVI http://memory.loc.gov/cgi-bin/oai2_0
XVII http://www.rdn.ac.uk/
XVIII http://www.dpconline.org/
XIX http://ahds.ac.uk
XX http://www.leeds.ac.uk/cedars
XXI http://pandora.nla.gov.au
XXII http://www.artstor.org
XXIII http://www.artstor.org/info/about/statement.jsp
XXIV http://www.artstor.org/info/about/history_mission.jsp
XXV http://culturalmaterials.rlg.org/
XXVI http://www.rlg.org/en/page.php?Page_ID=217
XXVII http://www.rlg.org/legacy/r-focus/i60.html#implements
XXVIII http://www.rlg.org/en/page.php?Page_ID=220
XXIX http://www.rlg.org/en/page.php?Page_ID=214
XXX http://trove.net/
XXXI http://www.hermitagemuseum.org/
XXXII http://www.hermitagemuseum.org/fcgi-bin/db2www/qbicSearch.mac/qbic?selLang=English
XXXIII http://www.virage.com/
XXXIV http://www.bl.uk/treasures/treasuresinfull.html
XXXV http://www.armadillosystems.com/ttp_commercial/home.htm
XXXVI http://icom.museum/disaster_relief/
XXXVII http://www.ifla.org/V/press/tsunami-Indonesia1.htm
XXXVIII http://www.photogrammetry.ethz.ch/research/bamiyan/
XXXIX http://www.dbnl.org
XL http://anno.onb.ac.at/anno.htm
XLI http://www.literature.at/webinterface/library
XLII http://www.hti.umich.edu/m/moagrp/about.html
XLIII http://memory.loc.gov/ammem/index.html
XLIV http://memory.loc.gov/ammem/about/index.html
XLV http://www.loc.gov/standards/mets/
XLVI http://memory.loc.gov/ammem/techdocs/amdtd.html
XLVII http://www.unesco.org/webworld/mdm/en/index_mdm.html
XLVIII http://pandora.nla.gov.au/
XLVIX http://pandora.nla.gov.au/pandas.html
L http://www.ngf.org.uk/html/cnmp.htm
LI http://www.montevideo.gub.uy/fotografia/
LII http://www.montevideo.gub.uy/fotografia/inicio/inicio.htm
LIII http://www.memoriachilena.cl

LIV http://www.dibam.cl/BVB/default.html
LV http://www.shikshapatri.org.uk/
LVI http://www.vhf.org
LVII http://www.clsp.jhu.edu/research/malach/
LVIII http://disa.nu.ac.za
LIX http://www.teara.govt.nz
LX http://www.aiatsis.gov.au
LXI http://www.newcastle.edu.au/group/amrhd/awaba/index.html
LXII http://www.asu.edu/lib/archives/labriola.htm
LXIII http://www.archimuse.com/mw99/abstracts/prg_1010.html
LXIV http://africandl.org/
LXV http://www.weforum.org/pdf/Global_Competitiveness_Reports/Reports/GITR_2004_2005/Networked_Readiness_Index_Rankings.pdf
LXVI http://hub.indonesiadln.org/
LXVII http://digicoll.library.wisc.edu/AfricaFocus/
LXVIII http://unesdoc.unesco.org/images/0012/001271/127160m.pdf