

DLESE Developers Workshop, 2003

Report on Workshop Outcomes

Editors:

Michael J. Wright

DLESE Program Center

University Corporation for Atmospheric Research (UCAR)

Tamara R. Sumner

Dept. of Computer Science

University of Colorado at Boulder



DLESE Developers' Workshop 2003

Report on Workshop Outcomes

Editors:

Michael J. Wright

*DLESE Program Center (DPC), University Corporation for Atmospheric Research (UCAR),
Boulder, Colorado*

Tamara R. Sumner

Department of Computer Science, University of Colorado at Boulder

Contributing Authors:

Part 1: Michael J. Wright, Greg Janee, Tamara R. Sumner

Part 2:

Data Session:

Recorders: Ethan Davis, Tom Boyd

Write up: Rajul Pandya, Michael Wright

Collections Session:

Recorders: Martin Ruzek, Robert R. Downs

Write up: Tamara Sumner, Michael Wright

Services Sessions:

Recorders: Rajul Pandya, Peter Burkholder, Ethan Davis, Tamara Sumner

Write up: Tamara Sumner, Michael Wright

Appendix 3: John Weatherley

Appendix 4: Katy Ginger

Appendix 5: Project Summaries

- Anderson, Kenneth M. – METIS Workflow System – UNIVERSITY OF COLORADO
- Arko, Robert – DLESE Community Review System – COLUMBIA UNIVERSITY
- Davis, Ethan – THEMATIC Real-Time Environmental Distributed Data Services (THREDDS) – UCAR/UNIDATA
- Downs, Robert R. – Digital Collections Management – COLUMBIA UNIVERSITY
- Fox, Sean – Science Education Resource Center – CARLETON COLLEGE
- Klaus, Christopher – Atmospheric Visualization Collection (AVC) – ARGONNE NATIONAL LABORATORY
- Ludaescher, Bertram and Meertens, Charles - GEON: A Research Project to Create Cyberinfrastructure for the Geosciences – UNIVERSITY OF CALIFORNIA AT SAN DIEGO AND UNAVCO, INC.
- Olds, Shelley – Earth & Space Science Education Collections (ESSEC) (formerly Earth Science Education Digital Library, ESEDL) – NASA
- Pandya, Rajul – Virtual Geophysical Exploration Environment (VGEE) – UCAR/DPC
- Quintana, Christopher - IdeaKeeper—Digital Library Services for Information Analysis and Synthesis – UNIVERSITY OF MICHIGAN
- Ruzek, Martin – Journal of Earth System Science Education (JESSE) – UNIVERSITIES SPACE RESEARCH ASSOCIATION

- Shipman, Frank - Using Spatial Hypertext as a Workspace for Digital Library Providers and Patrons – TEXAS A&M UNIVERSITY
- Shipman, Frank - Design and Evaluation of Maintenance Tools for Distributed Digital Libraries – TEXAS A&M UNIVERSITY
- Sumner, Tamara - Strand Maps as an Interactive Interface to NSDL Resources – UNIVERSITY OF COLORADO
- Warnock, Andrew – Digital Water Education Library (DWEL) – COLORADO STATE UNIVERSITY

Copy editor & workshop Web site: Eileen McIlvain, DLESE Program Center (DPC)

The DLESE Developers' Workshop 2003 and this report were supported by the National Science Foundation under Grant EAR 0215640, to the University Corporation for Atmospheric Research (UCAR). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) or originators and do not necessarily reflect the views of the National Science Foundation.

©2003 University Corporation for Atmospheric Research (UCAR), PO Box 3000, Boulder, CO 80307, USA. Permission is granted to reproduce in part or in whole, for non-profit training, research, or educational purpose, and not for direct or indirect commercial purpose or advantage.

Table of Contents

Executive Summary	4
Introduction	7
Part 1: The Building Process for DLESE	
1.1 The Path to DLESE	8
1.2 DLESE and the Digital Library Definition	9
1.3 A Community-Based, Distributed Digital Library	10
1.4 DLESE Product Planning Process	11
1.5 DLESE Version 2.0 Architecture	12
1.6 Preparing for Version 3.0	15
1.7 Issues in Georeferenced/Geospatial Digital Libraries	15
Part 2: Thematic Sessions	
2.1 Bringing Data into the Library	
2.1.1 Presentations	19
2.1.2 Session Summary	20
2.1.3 Recommendations	22
2.2 Building Collections	
2.2.1 Presentations	23
2.2.2 Session Summary	24
2.2.3 Recommendations	26
2.3 Building Services and Service Integration	
2.3.1 Presentations	26
2.3.2 Session Summary	28
2.3.3 Recommendations	33
Part 3: References	34
Appendices	
<i>Appendix 1: Workshop Agenda</i>	36
<i>Appendix 2: Participants (by individual, by project)</i>	39
<i>Appendix 3: DLESE Core Services Architecture v.2.0</i>	45
<i>Appendix 4: DLESE Metadata and Collections</i>	49
<i>Appendix 5: Reports about Presented Projects</i>	52
<i>Appendix 6: Glossary</i>	73

Executive Summary

The DLESE Developers' Workshop 2003 was hosted by the DLESE Program Center (DPC) in Boulder, Colorado, on February 19-20. Sponsored by the Geosciences Directorate of the National Science Foundation, the workshop brought together technical representatives from a number of projects developing library infrastructure, services, or resources for DLESE, or that include collaboration with DLESE as a part of their project goals.

The workshop was organized in recognition that DLESE is now a broad-based effort with technical contributions coming from an array of projects and organizations. The goals of the workshop were to give participants a view of the breadth of development projects being undertaken under the DLESE umbrella, to share experiences from project developments, and to identify future needs and areas for development.

The workshop agenda consisted of two parts. First, a morning was devoted to the DLESE building process, which included an overview of DLESE, project introductions, a review of the current DLESE library versioning plan, DLESE version 2.0 architecture, the ADN metadata framework (a partnership between the Alexandria Digital Library, DLESE, and NASA), the National Science Digital Library (NSDL), and commentary on georeferenced services. Second, the remaining workshop time was devoted to three thematic areas:

1. Bringing data into the library
2. Building collections
3. Building services and service integration

Projects were invited to the workshop on the basis of being example efforts in one of the three thematic areas.

This executive summary of the workshop report condenses the workshop discussions and recommendations. The workshop and the report provide a current snapshot of DLESE development. A wide audience is envisaged for this report, including principle investigators and developers of DLESE-related projects, people considering developing services or collections for DLESE or NSDL, or people considering developing a proposal for funding with the idea of working with DLESE, or including a contribution to DLESE as part of the proposed work. What follows are the recommendations resulting from sessions devoted to each of the three thematic areas.

Theme 1: Bringing Data into the Library

There has been a strong desire over the past few years in the geosciences research and education communities to bring real data into teaching and learning. At present, there are major data archives and real-time data feeds in a number of Earth science areas (ocean sciences, atmospheric sciences, seismology etc.) that are used on a daily basis by the research community. Efforts are now underway to broaden access to data and scaffold that data into effective learning activities and resources; some of these efforts were highlighted in this session.

Based on discussions, the participants identified a number of primary recommendations with respect to data in the library:

1. Create a special library section aimed at educational material developers who will include data and data access in the materials developed. This library section could be a digital library in its own right, including the notions of not only cataloging and providing access to resources but serving as an *intellectual commons* for its audience. It could fulfill this broader mission by:

- a. Conducting workshops to bring providers and developers together
 - b. Hosting an ongoing dialog about data and data tool interoperability with the goal of agreeing on common frameworks and outlining interoperability tools
 - c. Archiving and preserving data
 - d. Connecting to key data-intensive research efforts, e.g. EarthScope, GEON
 - e. Migrating this resource toward the teacher/learner in subsequent versions of the library
2. Develop a series of data-use exemplars, which would populate a special collection in the broader library. Suggestions that could guide these exemplars include: that they be interdisciplinary, that they be developed by teams integrating the education, information technology, and science communities, and that they use proven educational strategies consistent with national recommendations (e.g. inquiry-based approaches), and that the exemplars target different grade levels or audiences (as opposed to focusing exclusively on the undergraduate audience). The development of these exemplars could be helped by development grants from programs such as the NSDL.
 3. Develop a translation tool to link the information technology, science, and education communities. This could include definitions of common acronyms and uses in each domain, as well as a glossary of common terms. This is particularly important for terms that have different meanings or nuances in different fields. A nice addition to this might be a bibliography of key references for each domain.
 4. Include educators in planning the use of data in the library (analysis of the *Using Data in the Classroom* NSDL-sponsored workshops, though primarily directed at undergraduate science courses, could be beneficial) to help better understand the needs and realities of working in K-12 classrooms.

Theme 2: Building Collections

The purpose of this thematic session was to take a total lifecycle perspective on collection building; i.e., to discuss issues related to collection building, collection management, and resource persistence. Presenters were asked to describe their projects and to comment on concrete lessons they have learned to date about collection building. During the discussion, the group identified and prioritized common issues and needs arising across collection building projects.

Based on discussions, the participants identified several priority areas for future work in the area of collections building:

1. The current approach towards cross-walking collections based on multiple metadata frameworks, and multiple versions of the same framework, is effortful and time-consuming. Improved mechanisms and tools are needed that help to alleviate this effort.
2. Methods for automatically tracking changes in resources are an extremely promising technology that is not yet being used in operational libraries. More research should be conducted in this area, specifically research involving significant and real operational test beds.
3. Building better collections is a prime concern. Collections assessment techniques that address utility, not just gap identification and coverage, and that can be easily embedded in a collection building process, are needed.
4. Distributed collection building and community reviewing both lie at the heart of DLESE's *users as contributors* philosophy. The library needs to devise ways to support, motivate, and acknowledge this challenging intellectual work. Better tools and processes to supported distributed collections builders are needed, as well as effective training for collection building participants. The library also needs to systematically and publicly acknowledge and reward these contributed efforts.

5. Outreach to major resource providers is needed. These groups should be generating and providing their own metadata, both to the library and as “advanced organizers” to their own direct users. That is, the summary information contained in the metadata is also useful to human users that find these resources outside of the library’s discovery service.

Theme 3: Building Services and Service Integration

The purpose of this thematic session was to think about the types of services a digital library such as DLESE needs, and how these may be integrated. Presenters were asked to discuss their approach and needs, and the lessons learned in developing their services. In the discussions that followed the presentations, the group identified common issues and needs arising across their projects.

Based on discussions, the participants identified several priority areas for consideration for service development and integration:

1. Training, documentation and project awareness. Technical workshops, technical documentation, descriptions of tools, services or frameworks available from projects should be available in the library as a “one-stop shop,” or collection. Projects need mechanisms that support staying abreast of what other projects do (both in DLESE, NSDL, and beyond), and mechanisms to support collaboration between projects.
2. More support of *user as contributor*. Tools to support collaborative content development are needed that are end-user friendly, i.e. to a teacher or educator, not a software developer. Workshops and professional development of teachers to induct them into library processes such as resource reviewing or collection building are needed.
3. Open access to, and help with, creating library content underpinning services. Many services are built on top of library content beyond metadata; i.e. annotations (e.g. reviews), vocabularies, crosswalks, knowledge spaces, and in a few cases, primary library content. Extending the *open access* philosophy of DLESE and NSDL around metadata to other types of library content would be beneficial to service developers.
4. The area of integration of compound, or embedded, services is important. For these to work, service interfaces, standards and protocols need to be available and understood. DLESE and NSDL need to establish standards for their community, and provide timely tools and support for using and implementing them. Developing (or funding) examples of compound services that others can use as models to understand the process would be useful. Example: authentication and registration systems (authentication, verification, user profiles, access control) were highlighted as one type of compound service needed by many projects.
5. Shared outreach and marketing services are needed to help projects be successful. Time and resources are not always available to each individual project, but through the consortium of DLESE and NSDL, much could be gained from shared marketing.
6. Different models of sustainability need to be articulated and investigated. What happens to a service when the funding comes to an end, or the project principal investigator(s) no longer want to or are unable to continue supporting the service? For some service projects, sustainability is conceived as long-term, stable funding to support continued operations. For other service projects, sustainability refers to devising a *hand-over* model, where the developer group hands the project over to another group for further operations and maintenance.

DLESE Developers' Workshop 2003 Report

Introduction

The Digital Library for Earth System Education (DLESE) Developers' Workshop was hosted by the DLESE Program Center in Boulder, Colorado, on February 19-20, 2003. Sponsored by the Geosciences Directorate of the National Science Foundation, the workshop brought together technical representatives from a number of projects developing library infrastructure, services or resources for DLESE, or that include collaboration with DLESE as a part of their project goals. The workshop agenda is available as *Appendix 1* and the list of participants as *Appendix 2*.

The workshop was organized in recognition that DLESE is now a broad-based effort with technical contributions coming from an array of projects and organizations. The goals of the workshop were to give participants a view of the breadth of the types of development projects being undertaken under the DLESE umbrella, to share experiences from project developments, and to identify future needs and areas for development.

The workshop agenda consisted of two parts. First, a morning was devoted to the DLESE building process, which included an overview of DLESE, project introductions, a review of the current DLESE library versioning plan, DLESE version 2.0 architecture, the ADN metadata framework (a partnership between the Alexandria Digital Library, DLESE, and NASA), the National Science Digital Library (NSDL), and commentary on georeferenced services. Second, the remaining workshop time was devoted to three thematic areas:

1. Bringing data into the library
2. Building collections
3. Building services and service integration

Projects were invited to the workshop on the basis of being example efforts in one of the three thematic areas.

This publication is a report of the workshop discussions. The workshop and report provide a current snapshot of DLESE development. A wide audience is envisaged for this report including principle investigators and developers of DLESE-related projects, people considering developing services or collections for DLESE or the National Science Digital Library (NSDL), or people considering developing a proposal for funding with the idea of working with DLESE, or including a contribution to DLESE as part of the proposed work.

The structure of this report follows the workshop agenda. Part 1 covers the DLESE building process including a review of previous DLESE development and discussion of future versions. Part 2 reviews the three thematic sessions held at the workshop. Each section in Part 2 is structured to provide an overview of the theme, a summary of the invited presentations and the group discussion, and recommendations arising from the session. The three sets of recommendations are included in the executive summary.

The appendices contain the workshop agenda, a list of workshop participants, overviews of DLESE core infrastructure and metadata, background information on each of the projects that were presented in the three thematic areas, and a glossary of terms.

Part 1: The Building Process for DLESE

1.1 The Path to DLESE

The need for a digital library was articulated at an NSF sponsored workshop in August, 1999, called *Portal to the Future*, which brought together researchers and educators in Earth and Space science to create a plan to develop a national digital library to support education in Earth and Space science. A history of the evolution of DLESE is available at <http://www.dlese.org/about/history.html>. The development of the library is also in line with a number of NSF reports, e.g. NSF's *Shaping the Future* and *Geosciences beyond 2000* (NSF 1996; NSF 1999). The first version of the library was released at the DLESE Annual Meeting in Flagstaff, Arizona, in August 2001.

It is envisioned that DLESE be a national resource of collections of peer-reviewed teaching and learning resources, interfaces and tools to allow exploration of Earth data sets, and services to help users effectively create and use materials. This vision is being enacted through a community-based, distributed building process overseen by a community-based governance structure of steering and standing committees (DLESE 2002a).

The Steering Committee oversaw the development of a Strategic Plan (DLESE 2001) for DLESE in October 2001. As part of the Strategic Plan, a DLESE mission statement was articulated:

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.

A central tenet of the distributed building process is the idea of *users as contributors*, where DLESE users can also develop resources, collections, services, and technology through independent contribution or as part of funded activities. The NSF Geosciences Directorate sponsors core DLESE services through funding of the DLESE Program Center (the DPC, which is responsible for coordination of the development of the technological infrastructure, and governance support), and in 2003, three new core service areas: Community, Data and Evaluation, and a partially-funded fourth service area for Collections activities. Together, these core services work in coordination and collaboration with the broader community of DLESE library builders, both funded and unfunded, to meet the mission for DLESE.

A key recommendation of the DLESE Strategic Plan is the advocacy of DLESE as the Earth system education node of the larger National Science Digital Library (NSDL) effort. A number of projects that are part of the DLESE building effort are funded through the NSDL program, and the NSDL provides a mechanism for leveraging digital library developments from the broader digital library and science education community to support the DLESE effort. In return, DLESE helps participating projects to be "NSDL-ready" by providing tools, technical support, and technical workshops, and DLESE community members and projects in turn provide active participation in the development of the NSDL, new technologies and test beds for the NSDL, and bring a focus on Earth system science education (ESSE) to the NSDL.

In addition to the NSDL, DLESE will also benefit from participation in other Geoscience-based efforts (e.g. [Earthscope](#), [Geoinformatics](#), GEON), and broader information technology initiatives such as the current NSF [Digital Library Initiative](#) (DLI), [Grid Computing](#) (and the Open Grid Services

Architecture – OGSA), and new programs such as the developing [Cyberinfrastructure](#) (CI) program at the NSF (Atkins, Droegemeier et al. 2003).

1.2 DLESE and the Digital Library Definition

The term *digital library* is used very broadly in the realm of information technology. For the basis of this workshop, four specific technical digital library models were noted as being relevant to the workshop agenda. A number of the projects represented at the workshop can be placed in these categories.

1. A digital library comprising digital resources housed and managed as a system with services minimally supporting discovery of items, browsing of items, and the use of items. This type of library has the close analogy to a “bricks and mortar” library. Examples include The Alexandria Digital Library (ADL), and NASA’s Earth and Space Science Education Collections (ESSEC) (formerly NASA Earth Science Education Digital Library).
2. A digital archive where the primary goal is the persistent archiving of digital items. General access to archive items is not necessarily a primary goal, where the analogy can be to a deposit library. The San Diego Supercomputer Center (SDSC) provides such archiving to the NSDL as an initial “backup” type of service.
3. *Metadata Repository*, *Metadata Registry*, or *Subject Gateway* are three terms used in different communities to describe a type of library that holds only metadata records about digital items that are held elsewhere (distributed over the Internet, for example). These libraries minimally provide discovery (search and browse) services over the metadata holdings. This is analogous to the union catalog of the traditional library. The NSDL community and Open Archives Initiative (OAI) community use the term Metadata Repository. Metadata Registry is commonly used in e-commerce and government agencies. The term Subject Gateway is commonly used in Europe. The holdings of such libraries are not usually under the control of the library, and so issues with persistence and versioning of digital items arise. Both the core NSDL and DLESE discovery services are based on this model.
4. *Data Archive* (or *Data Library*) is a more infrequent term used to describe an online source of scientific data. The library holds collections of data, and potentially provides browsing and visualization services. This model could easily be defined under 1 or 2 above, but is called out specifically here due to the importance of scientific data to the DLESE community. Examples at the workshop included the Thematic Realtime Environmental Distributed Data Services (THREDDS), the Visual Geophysical Exploration Environment (VGEE), and the Atmospheric Visualization Collection (AVC).

As noted, different projects within DLESE are examples of the models described above; as such, DLESE cannot easily be defined in terms of any single model. This is an important observation as DLESE considers how to accommodate the diversity of approaches and technologies being pursued and implemented by community-based projects. The library must consider the combination, or integration, of many models of collection and service provision, and yet still provide users with a meaningful library experience in line with the overall DLESE mission.

Collections

In developing collections, a number of matters need to be addressed: metadata and classification standards, scope (what is being collected and why), the process of locating, acquiring, collating and classifying content, and mechanisms to support these processes. For DLESE collections to grow in breadth and depth, technical infrastructure and tools must either directly support or enable all of these

activities. To make the library useful (beyond being an archive), collections need to support the development of associated services, such as discovery over collections, reference services, integrated discussion forums, etc. In many respects, these parallel services found in many of today's libraries (not just the digital variety).

Services

DLESE is an education digital library, i.e. it is aimed at supporting teaching and learning in the Earth sciences. This has impacts on the collections and services provided by DLESE, as the collections and services need to meet this requirement. To do this, DLESE broadens the scope of services provided by a library to include various review services as an integral part of collections development and to augment existing collections, and these services can build on a number of review models, e.g. scholarly peer review, community peer review, the Amazon model or the Expert Exchange model. Another suite of services to consider is *recommender services* such as teaching guides, teaching tips, and learning support services. In addition to these services, extended services to support innovative forms of library use based on concept maps, education standards, Earth system science thesauri, and Earth system event gazetteers are needed, along with services that can be embedded in other applications, e.g. discovery application programming interfaces (API). Some of these services are already under development by some of the workshop participating projects.

Resources

Another dimension to DLESE builds on the part of the mission "...by developing and providing access to high quality educational resources..." Building on the idea of *user as contributor*, the community is encouraged to create resources for the library. This brings a new role to the digital library concept, that of publisher and distributor of content. That content can apply to the added services such as review systems and recommender systems. Many of the resources found in DLESE have been "self-published" on the Web by individual creators or groups with little of the editorial and version management expertise that a publisher typically brings to the process. This has impacts on collection development, cataloging, and review processes—impacts we are just beginning to comprehend.

1.3 A Community-Based, Distributed Digital Library

So far, we've described DLESE with respect to collections and services, but one important feature of DLESE is the distributed nature of these collections and services. Resources, collections and services are being developed by various groups, some at single institutions, some as collaborative efforts between institutions, and from various funding sources. Development is proceeding in a heterogeneous environment of hardware, software, operating systems, and architectural approaches. Yet for DLESE to succeed in its mission, it is important to always stay focused on providing a coherent and compelling user experience.

Interoperability

Interoperability is the key to allowing these distributed functions to operate as an entity perceived as being "DLESE." Interoperability is needed at a number of levels: policy, social, semantic, and technical. Examples of each include the library's policies developed collectively by the DLESE community (e.g. collections, privacy); the social dimension of governance, workshops, annual meetings; the cataloging procedures and standards (e.g. metadata frameworks and crosswalks, cataloging best practices, and controlled vocabulary development); and technical, e.g. use of Web services frameworks, Open Archives Initiative (OAI) protocol for metadata sharing (interoperability with NSDL), Shibboleth for access control (interoperability with NSDL).

There are already a breadth of technical interoperability mechanisms being used in the DLESE community, in addition to the OAI approach advocated by NSDL, e.g. ADL as distributed nodes with query and result interoperability being done over JAVA RMI; THREDDS that builds upon the OPeNDAP (formally DODS) data exchange mechanisms; Unidata Internet Data Distribution (IDD) and Local Data Manager (LDM) technologies; and ESSEC using JINI for brokering and update. These groups have adopted these mechanisms to support the needs of their specific communities, and these communities are considered part of the larger DLESE community. Further interoperability mechanisms are likely to appear from the grid computing community (Carpenter 2003) and cyberinfrastructure initiatives to support specific services that would be useful to DLESE.

1.4 DLESE Product Planning Process

The development of DLESE has been guided by the DLESE Community Plan (Manduca and Mogk 2000) and by the DLESE Strategic Plan. From these, a product planning process has identified the major versions of DLESE to be available over the next five years. Within the context of DLESE, versioning fulfills two important goals. First, version descriptions provide a conceptual framework that helps individual projects to construct goals and work plans that contribute towards the overall library effort. Second, version roll-outs provide a useful mechanism for synchronizing distributed development efforts around periodic releases of library functionality. Two primary versions have been identified, version 2.0 to be released in the summer of 2003, and version 3.0 in 2006. The version timeline and functionality are shown in Figure 1. The functional goals the library will provide at these times are:

Version 2.0: Users will be able to search across multiple collections, including many that are peer reviewed, according to an Earth system perspective and a variety of benchmarks and standards. Community forums and library services supporting the effective use of resources and professional development are available.

Version 3.0: Users will be able to search across spatial, temporal, and event-based data, maps and images. Integrated tools and services to assist with age-appropriate exploration of data are available. Users will create and share a variety of personalized collections.

Between major versions, intermediate releases of library functionality will occur as development of features meeting parts of the version goals are attained. This progression is highlighted in the DLESE versioning document (DLESE 2002b). Much of this development necessary to achieve library advances involves collaboration between DLESE community projects, DLESE core service groups and the larger NSDL community in a multi-dimensional network of interaction. It is this facet of DLESE development that underpins the community-building philosophy of the library.

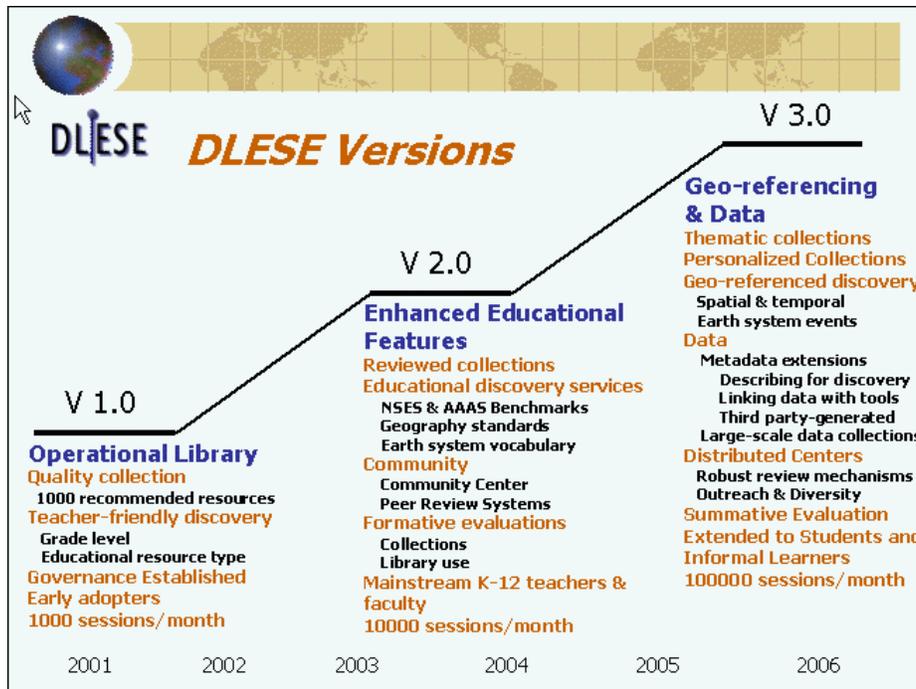


Figure 1: DLESE Versions

1.5 DLESE Version 2.0 Architecture

Version 2.0 of the library becomes a reality in the summer of 2003, at which point the library will contain a number of collections and review systems, a cataloging tool, and a discovery system to allow searching and browsing over the collections in addition to associating reviews from various review systems to resources as part of search results. The library has implemented a number of protocols and mechanisms to allow the library to meet the version 2.0 goals. The overall library will comprise distributed collections and services with a core set of services (illustrated in Figure 2) interoperating using agreed standards of metadata frameworks and inter-service, Internet-based protocols. More detail on the core services architecture of DLESE v2.0 can be found in *Appendix 3*.

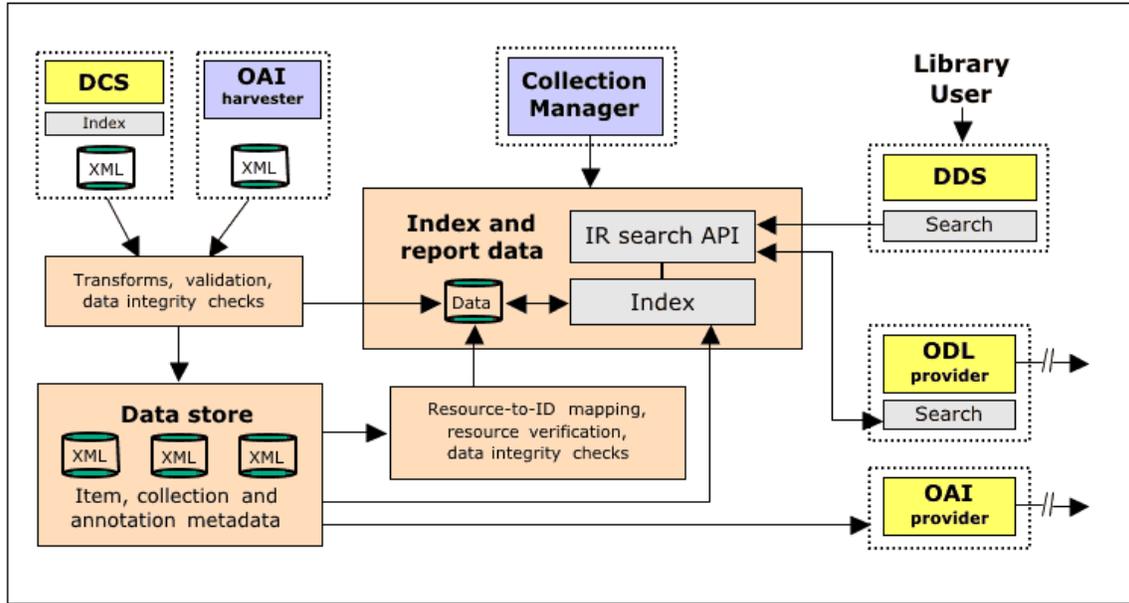


Figure 2: Core Collection Management and Discovery Architecture

OAI-PMH

A key component supporting the current interoperability for the library is the use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for passing metadata records between collections and services and the DLESE discovery management system (DDS). The OAI-PMH has been adopted by the NSDL for metadata exchange, and thus, DLESE is interoperable with the NSDL at this protocol level. In addition, the discovery management system provides access to the DLESE collections of metadata by this same method so others can harvest metadata records from DLESE. This is how the NSDL obtains records from DLESE.

ADN (Alexandria Digital Library-DLESE -NASA)

A second key component has been the use of a metadata standard (expressed in XML—eXtensible Markup Language), initially the DLESE-IMS framework (based on XML with *DTD*, or document type definition) and soon to be replaced by the ADN framework (based on XML with *schemas*, an XML document that describes the DLESE metadata framework in terms of structure, data types, number of field occurrences, and controlled vocabularies). The initials *ADN* denote the framework development collaborators Alexandria Digital Library, DLESE, and NASA. Crosswalks from these DLESE-based metadata frameworks to the Dublin Core framework have also been developed to allow DLESE collections to be harvested and made discoverable in the NSDL (the NSDL will take native metadata such as ADN, but the NSDL has normalized on qualified Dublin Core for its services such as discovery). In addition, DLESE has developed policies needed to define the controlled vocabularies used in the frameworks, cataloging best practices to promote consistent and complete metadata records, and articulated the documentation information a collection provider (or builder) needs to provide in order to be part of DLESE. This work entailed close cooperation with the NSDL. For more on DLESE metadata, see *Appendix 4*.

The OAI-PMH and DLESE metadata frameworks have allowed collections developers such as the Atmospheric Visualization Collection (AVC) to develop their collection metadata and share it with DLESE by having the records harvested into the DLESE discovery management system.

DLESE Cataloging System

The DLESE Cataloging System (DCS) is a Java-based Web application developed by the DPC to support the cataloging of resources to the DLESE metadata framework as part of a collection building effort. An instance of the DCS is used at the main library Web site to allow the cataloging of resources into the broad collection. Collections builders can download the DCS for use if they need a tool to support the creation and management of metadata (DLESE 2002c).

Exchanging Annotation Information

Just as there are collections of resources that would use ADN, there are collections that are annotations about resources. Annotation collections use the DLESE annotation metadata framework, which is based on and will provide feedback into the proposed NSDL annotation framework. The DLESE annotation framework not only provides annotation information, it is the gatekeeper to the DLESE Reviewed Collection. For example, the Community Review System (CRS) uses the OAI-PMH mechanism to send their collection of annotation records, thus indicating Reviewed Collection items.

Workflows

For the version 2.0 release, the discovery management system is being further developed to incorporate the needed workflows for managing and quality-assuring collections to be made available to users through the discovery interfaces. This is in addition to the cataloging workflow and best practices supported in the DCS.

Individual collections and services manage their own collection metadata and OAI-PMH mechanisms. An OAI-PMH implementation (in Java) developed by the DPC is available in addition to the DCS to support cataloging, but collections and services providers can implement their own mechanisms if they choose. However, they must make sure that the data shared to DLESE conforms to DLESE requirements to assure ingest into the discovery management system.

Web Services Framework

The basis of interoperability between distributed collections and services is a mediated mechanism, passing DLESE-acceptable metadata as XML encapsulated in a message container, the OAI-PMH, and transferred over an Internet transport mechanism, HTTP—the basic transport mechanism of the Web. This “stack” of data layer, message encapsulation layer, and transport layer mirrors the Web services framework that underpins much of today’s Web services development in both academia and industry. In the future, it is possible that DLESE could add to, or change, the implementations used at a particular layer, e.g. perhaps using SOAP—Simple Object Access Protocol, an XML-based lightweight protocol for information exchange in distributed environments.

While the exchange of XML-based metadata records over OAI-PMH is the primary mechanism for information exchange with the discovery management system, it is not the only mechanism found in the broader DLESE. THREDDS is a project working on the problems of cataloging and access to disparate data systems and works with data-specific access protocols such as OPeNDAP (developed from the Distributed Oceanographic Data System – DODS), ADDE (Abstract Data Distribution Environment), and with general protocols such as FTP (File Transfer Protocol). Query mechanisms are another possibility in addition to the harvest model’s exchange of metadata records. OPeNDAP and ADDE allow for clients to query the data servers for specific data, and the query protocol is part of those standards. Z39.50, SDLIP, and ODLP are examples of protocols used by, or proposed for, the digital library community to support distributed queries, or what is often termed *federated search*.

This indicates that many technology-based protocols for interoperability may be in use by sub-communities of DLESE and DLESE will need to accommodate this technical diversity.

1.6 Preparing for Version 3.0

Much of the library infrastructure that will be in place for version 2.0 in the summer of 2003 will be a base on which to start building for version 3.0. As noted under the earlier section *DLESE Product Planning Process*, version 3.0 will expand upon the collections and services available in line with the functional goal:

Version 3.0: Users will be able to search across spatial, temporal, and event-based data, maps and images. Integrated tools and services to assist with age-appropriate exploration of data are available. Users will create and share a variety of personalized collections.

Figure 3 illustrates a scenario for version 3.0. In this scenario, the mechanisms required to meet the scenario goal combine the work of distributed efforts in the larger DLESE. How well such distributed efforts integrate and interoperate will underpin the long-term success of the library.

DLESE Scenario: DLESE Version 3.0

Beth teaches an undergraduate introductory meteorology class in South Carolina. The class is studying hurricanes, and Beth is trying to explain the differences between Atlantic and Caribbean hurricanes and cyclones typical of the Philippines. She is interested in incorporating both real-time and archived data in her lab activities.

DPC:	Community:
<ul style="list-style-type: none">✦ Extends metadata framework to link data and tools, describe spatial and temporal coverage (geo-referencing)✦ Provide gazetteer services – place names and earth system events✦ Extend cataloging and collections tools for geo-referencing and events✦ Extend discovery system to support geo-referencing and events	<ul style="list-style-type: none">✦ Designs metadata for describing data and provides tools for creating third party metadata (THREDDS)✦ Contributes data sets, data collections, visualization and analysis tools, e.g., AVC, Geoinformatics✦ Provides information and services supporting effective use of data in the classroom

Figure 3: A Scenario for Version 3.0

In addition to the various strategic plans and versioning documents, instruments such as this report will also play a role in the development of version 3.0 along with the plans of the new DLESE core service areas being funded by the NSF in 2003.

1.7 Issues in Georeferenced/Geospatial Digital Libraries

In section 1.4, the DLESE versions illustrated in Figure 1 indicate that an important aspect of version 3.0 of the library is the inclusion of georeferencing. This is again highlighted in the scenario presented in section 1.6. While it may seem straightforward to include georeferenced discovery, there

are many issues to address that underlie the inclusion of georeferencing. A short background and discussion of the issues is provided in this section.

Georeferenced information is information that is relevant to a definable and explicitly stated subset of the Earth's surface; we call the subset the information's *spatial coverage*. Any kind of document that is about a particular geographic place (e.g. *A Tale of Two Cities*, *History of the Decline and Fall of the Roman Empire*, etc.) is an example of georeferenced information. There is a large subclass of georeferenced information—such as maps, remote-sensing imagery, etc.—that is distributed over the extent of its spatial coverage, and that is typically graphically visualized. We distinguish this subclass as *geospatial information* to emphasize its spatial characteristics.

The remainder of this section will provide a discussion of some issues in storing georeferenced and geospatial information in a digital library, and more importantly, making such information discoverable by and usable to a broad spectrum of library users.

Discovery

The first and most basic issue is: how can georeferenced information be discovered? The traditional approach to georeferenced discovery is to use text-based information retrieval techniques on the metadata associated with the information, and in particular, to base discovery on matching of textual place names. Using this technique, a user desiring a map of Boulder, Colorado, would use these two words as query, and the search service would return items whose metadata contain these two words.

This technique works well for certain classes of information and in certain contexts. Clearly, such a technique would return an item whose title is “Downtown street map of Boulder, Colorado.” But as a general technique it suffers from two serious drawbacks. First, there is a whole class of georeferenced information having *no* associated place-names, namely, data gathered from moving sensors. Examples include satellite imagery and aerial photography, which carry only technical metadata such as the coordinates and attitude of the camera at the time of exposure.

Second, discovery based on textual place-name matching is largely unreliable. Consider for a moment a user desiring a map of the Flatiron rock formations in the Front Range of the Rocky Mountains just outside of Boulder, Boulder County, Colorado. The underlined words in the previous sentence all describe the desired area to varying degrees of specificity, but it turns out that the most specific phrase, and realistically the only useful phrase, that will retrieve data from the U.S. Geological Survey for this area is Eldorado Springs, which happens to be the name of the quadrant that covers the Flatirons per the USGS's standardized grid system. This is but one instance of a very general problem—there are many names for any given place, and a document that is about a place is cataloged by, at best, a handful of those names.

To address the general problem of reliable georeferenced discovery, some kind of structured (non-textual) search technique must be employed. Defining a controlled vocabulary of places (so that, for example, “Boulder, Colorado” would be treated as one of a set of discrete terms) is one approach. The U.S. Geological Survey's use of named quadrants mentioned above is a controlled vocabulary. This approach resolves the place-name multiplicity problem, but it also places a considerable burden on both library catalogers and users to understand and agree on the vocabulary, which will inevitably be limited. A more general and powerful technique is to support range searching (i.e., searching that employs inequality testing) over, for example, numeric latitude/longitude coordinates. This technique allows a user to discover information without the user and library having to agree on anything except the coordinate system.

Gazetteer Integration

The above discussion should not be construed as implying that place-names are unimportant or not useful in a georeferenced discovery. On the contrary, place-names play a critical role in georeferenced discovery, and in georeferenced libraries in general, because human spatial cognition relies on relationships to and among known features, and to the extent that those features are named, to place-names. (Proof: think about how you would answer the question, “Where are you right now?”) As a result, it is necessary that a *gazetteer* (a kind of dictionary that supports translation between place-names and coordinates) be integrated into a georeferenced digital library.

Gazetteer data is typically too incomplete and imprecise to support automatic, sophisticated geospatial reasoning, and so it is not practical to integrate the gazetteer into a digital library at a low, hidden level. That is, because gazetteer interactions often require human mediation, gazetteers should be considered extensions to the library’s user interface.

Heterogeneity

In the area of georeferenced/geospatial information there are many different metadata formats and content standards in use, and there are multiple coordinate systems and georeferencing techniques. A digital library that intends to accommodate this heterogeneity will have to provide mapping mechanisms at some level.

Data Typing

Adding a structured discovery technique to a digital library (e.g., coordinate-based range searching) means adding a notion of data types to the system. This has many ramifications:

- Input validation is required; for example, geographic metadata cannot be treated as undifferentiated text
- Relatively complex internal structures and external representations are required to describe spatial coverages and spatial query regions
- The library must provide the means to express, and of course perform, different types of query constraints (spatial constraints, textual constraints, etc.) and Boolean combinations of different types of constraints
- Type-specific ranking methods are required, as are methods for combining different types of rankings
- Specialized user interface components are required to input and to view typed data; in the case of geographic coordinates, this means interactive map browsers and underlying map servers

Scalability

It is easy to accumulate geospatial data since it is often generated by automatic means (satellites, sensors, etc.), and thus scalability can become an issue in terms of accommodating large numbers of library items.

Georeferenced discovery techniques are nicely scalable in theory (the two-dimensional R-tree-based index structures in common use today offer logarithmic search time), but in practice, scalability is more limited. At the time of this writing, commonly available, commercial georeferenced search engines reach the limits of practical use (e.g., one day of compute time and several gigabytes of RAM) to index as few as 10^6 items. Larger numbers of items can be accommodated, but only with exponentially increasing amounts of custom engineering.

A more significant problem is the difficulty of *joining* (in the database sense) conjunctions of spatial constraints with textual and other types of constraints. A typical and reasonable gazetteer query such as “find a city named ‘Boulder’ near the Rocky Mountains” is a conjunction of a discrete constraint (find a city, where “city” is one of a set of discrete types), a textual constraint (named “Boulder”), and a spatial constraint (near the Rocky Mountains). When large numbers of items are involved, generating efficient query plans becomes paramount, but query plans are unavoidably sensitive to the queries and to the distribution of the data, which is inevitably far from uniformly distributed. Whether the library is working within the framework of a single relational database or across multiple, distributed search indexes, the heterogeneous join problem is a significant practical problem and appears to be a research problem as well.

Spatial Context

In any information space, users need context to understand where they “are” in that space and what the information is “about” in order to formulate queries and interpret query results.

In the case of a georeferenced digital library, the geographic context is especially critical because, as noted previously, humans reason about space symbolically and rely heavily on relationships to known features. Consequently, users typically want to express georeferenced queries either symbolically (“Boulder”) or by relationship to known features, such as by indicating a region of interest on a context-providing background map. The implication is that georeferenced digital libraries must provide interactive background maps with sufficient detail to allow users to relate their location to the landmarks they recognize. By contrast, libraries and other systems that require input of numeric coordinates are generally unusable without the aid of some external context-providing device.

Context is also critical in interpreting query results. And beyond result sets, context is required to evaluate individual items. Suppose for a moment that we have gotten past the first spatial query that everybody does (“let’s see if I can find my house”), an exercise that is highly misleading because we are intimately familiar with our own surroundings. Instead, suppose that the aerial photograph to the right has been returned as a response to the previous query about the Flatiron rock formations near Boulder, Colorado, an area we are unfamiliar with. Does this give us the information we were after? More to the point, where *are* the Flatirons in this image? Clearly we could answer this question by importing the photograph into a geographic information system (GIS), layering map and place-name data over it, but that assumes a high level of sophistication and resources on the part of the user. To be useful to the general user, a geospatial digital library must provide a certain amount of its own (lightweight) GIS functionality to be able to show this image positioned over a reference map, perhaps, or by labeling the features within the image.



Content Access and Integration

Geospatial information has complex structure and is often quite large. It can consist of multiple parts, require specialized viewing tools, and be accessible via multiple formats, protocols, and interfaces. For example, georeferenced imagery is often distributed as a pair of files: an image proper (e.g., a TIFF file) together with a file containing georeferencing information. A client may have to independently access one or both of these files to use the information successfully. To take another

example, the industry-standard ESRI “shapefile” format is in fact a set of 4–7 coordinated files of different types. And geospatial information is also often accessible via programmatic services such as the OpenGIS Web Map Server (WMS) protocol (Kolodziej 2003) or ESRI’s ArcIMS interface (ESRI 2003). Access via service becomes more critical as the size of the geospatial data increases because service access typically allows navigation to and interaction with just a desired subset of the item.

Geospatial information is also closely tied to geographic information systems and other types of data exploration environments. In some cases GIS systems are simply a common means of viewing and working the information; in other cases, the information is entirely unusable outside the appropriate analysis or visualization package. Either way, the ability to process geospatial information meaningfully strictly within the confines of a standard Web browser is limited.

In summary, these characteristics of geospatial information—its complex structure and its close ties to analysis environments—call out the need for libraries to be able to describe, in detail and for the benefit of both programmatic clients and human users, the components that make up a library item and the different modes by which it may be accessed. It is insufficient to describe the content of a geospatial library item by, for example, a single, unadorned URL (“click here”). But a more detailed description can allow the user to make an informed decision about the best way to access the item, and can allow the library to seamlessly hand the item off to a software visualization package.

Part 2: Thematic Sessions

Three important areas to DLESE were organized as breakout sessions for the workshop participants. In these sessions, short project presentations provided participants with a view of the developments being pursued, and time to discuss issues related to these areas. The three areas were *Bringing Data into the Library*, *Building Collections*, and *Building Services and Service Integration*.

2.1 Bringing Data into the Library

There has been a strong desire over the past few years in both the geosciences research and education communities to bring real data into teaching and learning. At present, there are major data archives and real-time data feeds in a number of Earth science areas (ocean sciences, atmospheric sciences, seismology etc.) that are used on a daily basis by the research community. Efforts are now underway to broaden access to data and scaffold that data into learning resources; some of these efforts were highlighted in this session.

Bringing access to data (both collected from observations including remote sensing, and from the output of large-scale models and simulations) to the broad teaching and learning community in DLESE has been highlighted as a major library goal. The scaffolding of access through tools to support visualization and manipulation in the context of learning resources has also been identified as being a core need for both DLESE and the larger NSDL. However, providing these access mechanisms, tools and resources is not a simple task. The breakout group for this theme was tasked to review how data is usable in educational digital libraries, to consider how data can be both a service and a collection, and to answer the questions of what special challenges or opportunities this holds for data in the library. What are the common problems/themes faced by the represented projects?

2.1.1. Presentations

An important aspect of this theme is to highlight how data can be incorporated into the digital library. By data, we mean large data sets accumulated from the output of various instrumentation mechanisms

in the geosciences. Incorporation of data into daily Earth science teaching and learning has already been identified as a need DLESE should address (DLESE 2001; Manduca and Mogk 2000). The four presentations give some views of how this may be accomplished, and provide a starting point for discussion. Short reports by presenters along with more information on their projects can be found in *Appendix 5*.

AVC – Chris Klaus

The Atmospheric Visualization Collection (AVC) was introduced as a collection based on Atmospheric Radiation Measurement (ARM) data that incorporates data products (primarily images) into a series of educational resources that also include visualization and modeling tools. Collaborative tools are used to allow a distributed community to develop the collection.

GEON – Bertram Ludaescher and Charles Meertens

The goal of GEON (GEOsciences Network) is to build an infrastructure for the geosciences such that scientists can work with disparate data sets from across disciplinary and organizational boundaries. The project is planning to build demonstrators that integrate disparate data sets in two test-bed areas: the Rocky Mountain region and the Mid-Atlantic. Another important aspect of this ITR funded project is to investigate the integration of knowledge structures (ontologies) to support data navigation and querying by domain scientists.

THREDDS – Ethan Davis

The goal of the Thematic Real-time Environmental Distributed Data Services is to build a framework for dealing with distributed scientific data sets, and to integrate this data, along with related tools for visualization and analysis, into digital libraries. Just as the Web can be used for publishing and accessing documents, THREDDS aims to provide an infrastructure for publishing and accessing scientific data.

VGEE – Rajul Pandya

The Visual Geophysical Exploration Environment (VGEE) is an educational package for undergraduate, entry-level Earth science students, that combines an inquiry-based curriculum with learner-centered interfaces to a scientific visualization tool (the Integrated Data Viewer) and authentic research data. This environment has been used in teaching undergraduate meteorology.

2.1.2 Session Summary

The four presentations gave an indication of the breadth of possibilities with respect to the access and use of what Pandya of VGEE termed *authentic data*. Two of the presentations looked at the tasks involved in providing access to distributed and disparate data sets: THREDDS as an access service to distributed data sets through cataloging of data sets with respect to thematic criteria and the attendant issues of dealing with different access protocols and metadata standards; and GEON, a recently funded ITR project to investigate the integration of disparate data sets (utilizing mediation strategies to enable data discovery) related to two geographic regions. Both these projects introduce the idea of thematic data, data in context to a place or event. In many cases, the context is both a place and event (e.g. an earthquake, hurricane, tsunami, etc.) Another important dimension these talks addressed is the distinction between syntactical (e.g. the way data is arranged in an XML document) and semantic (e.g. the conceptual model) integration of data. The GEON talk emphasized the need for semantic mediation to integrate disparate data sets and that semantic integration allows questions (discovery) to result in showing connections that didn't exist explicitly within the data structures.

The other two presentations on the AVC and VGEE gave insight into how data can be incorporated into learning environments—resources—to give context to data discovery and access in a digital library through a scaffolding framework that includes tools for data visualization and data manipulation. A common theme for these two projects is the blurring of the distinction between research-based tools and education-based tools, e.g. the VGEE incorporates the use of a scientific visualization tool initially developed for researchers. Both the items in the AVC collection and the VGEE are already available through DLESE. The AVC presentation discussed the NSDL K-12 Education Portal Workshop held in October 2002, where a data collection life cycle was developed. This generated discussion of where digital libraries lie in terms of that life cycle. There was general agreement that digital libraries have a role to play in supporting collection developers and educational material developers, a theme echoed in other discussions throughout the workshop.

Following the presentations, the group began a general discussion of the way in which data might be integrated into DLESE and into educational digital libraries. The discussion centered on common themes, obstacles, and/or issues that emerge from trying to use data in educational digital libraries. The group focused on how collaborative efforts that spanned many projects might develop services, protocols, or strategies that could be employed by a number of different projects in a variety of contexts.

One idea to quickly emerge from the discussion was the distinction between data providers—educational material developers who would include the use of data—and students and teachers who would use the educational content developed. The group felt that it would be best to target the educational material developers who want to use data. This approach was felt to be appropriate for a variety of reasons:

1. The expertise and background represented by the group was best suited to addressing the needs and goals of educational material developers rather than data providers or teachers
2. Direct, student/teacher use of the data would not be common. Instead most teachers/students would use the data as part of educational materials that included tools and curricula in addition to the data.
3. Library development in the area of data is a version behind its general development. While the library as a whole is transitioning away from early adopters as its intended users for version 2.0, data services are still developing and can't yet support a broad user base.
4. Although we agreed with the need to collaborate with data providers, we felt it would be unproductive to task them with the additional burden of modifying their data for educational users.

After agreement on targeting educational material developers using data, the discussion centered around what tools, resources, and support digital libraries might provide this audience, and how this audience, broadly speaking, could direct its efforts most efficiently. General points of discussion and recommendations included:

- Need for themed data collections
- Need for reviewed collections of data
- Need transparent mechanisms for users to access and download data streams
- Need to understand how to encode/translate the semantics of data sets
- Need to gather data providers to understand how we can bridge the variety of data formats and data semantics; this problem is not as much technical as social
- Need to exemplar data/curriculum integration like VGEE; a potentially powerful component of this would be an exemplar that pulls together data from several sources

- Need to encourage collaboration between educators, scientists, and data providers
- Need to develop service and/or tool registry
- Need a way to link data to tools that can work with the data
- Can agreement be reached (or at least guidance given) on what data access tools should be used? (For data providers and data users)
- Develop services for data content providers
- Data browsing services, e.g. thumbnail views, and data set search services
- Easy submission of data to the library
- Tracking multiple copies of a data set; unique ID service to register data repositories so that it is easy to track duplicate copies (example: the Grid Replication service)
- Tools are needed to help students build instrumentation and share data with each other, e.g. GLOBE project
- There are lots of metadata standards; how will we ensure that each of these contributes to a geospatially useful library? Need a way to change various reference systems into a standard way that allows the user to find things geospatially
- Footprint and gazetteer tools—a way for catalogers to attach spatial information to items (e.g. draw a box and get it converted to coordinates, or type in Boulder and get a set of coordinates)

2.1.3 Recommendations

Out of the discussion, the group identified a number of primary recommendations.

1. Create a special library section aimed at educational material developers who will include data and data access in the materials developed. This library section could be a digital library in its own right, including the notions of not only cataloging and providing access to resources but serving as an *intellectual commons* for its audience. It could fulfill this broader mission by:
 - a. Conducting workshops to bring providers and developers together
 - b. Hosting an ongoing dialog about data and data tool interoperability with the goal of agreeing on common frameworks and outlining interoperability tools
 - c. Archiving and preserving data
 - d. Connecting to key data-intensive research efforts e.g. EarthScope, GEON
 - e. Migrating this resource toward the teacher/learner in subsequent versions of the library
2. Develop a series of data-use exemplars, which would populate a special collection in the broader library. We had a number of suggestions that could guide these exemplars: that they be interdisciplinary; that they be developed by teams integrating the education, information technology, and science communities; that they use proven educational strategies consistent with national recommendations (e.g. inquiry-based approaches); and that the exemplars target different grade levels or audiences (as opposed to focusing exclusively on the undergraduate audience). The development of these exemplars could be helped by development grants from programs such as the NSDL.
3. Develop a translation tool to link the information technology, science, and education communities. This could include definitions of common acronyms and uses in each domain, as well as a glossary of common terms. This is particularly important for terms that have different meanings or nuances in different fields (e.g. semantics). A nice addition to this might be a bibliography of key references for each domain.
4. Include educators in planning the use of data in the library (analysis of the *Using Data in the Classroom* NSDL-sponsored workshops, though primarily directed at undergraduate science

courses, could be beneficial) to help better understand the needs and realities of working in K-12 classrooms.

2.2 Building Collections

The purpose of this thematic session was to take a total lifecycle perspective on collection building; i.e., to discuss issues related to collection building, collection management, and resource persistence. Presenters were asked to describe their projects and to comment on concrete lessons they have learned to date about collection building. During the discussion, the group identified and prioritized common issues and needs arising across collection building projects.

2.2.1. Presentations

Five presentations were given on five different collection-related projects and issues. Short reports by presenters along with more information on their projects can be found in *Appendix 5*.

NASA ESSEC (formerly ESEDL) – Shelley Olds

NASA has much educational content scattered among distributed partner organizations. The ESSEC, or Earth and Space Science Education Collections (formerly Earth Science Education Digital Library, ESEDL), aim to create a “one NASA” interface that provides a single access point to a distributed library network. ESSEC has created free interoperability software that runs on multiple systems (a JINI networking module), which when combined with OAI, enables metadata to be created and held at distributed library sites.

DWEL – Andrew Warnock

The Digital Water Education Library (DWEL) is building a collection of 500 exemplary resources on water in the Earth system targeted at K-12 users in both formal and informal learning contexts. Resources in DWEL are selected, reviewed and cataloged by four working groups of teachers and informal educators that are distributed across the nation. DWEL has articulated a multi-step collection-building workflow process, and has created the DWEL Work Hub interactive system to guide teachers step-by-step through all facets of this collection-building process.

METIS – Ken Anderson

METIS is a lightweight tool to support workflow in a digital library. It is a set of Java servlets and a Web-based interface that enables users to define workflow processes, in terms of events and actions that are triggered by events. A series of events, loops, forks etc. can be defined, and deadlines, outcomes etc. can be assigned. Action editors help facilitate activities that cannot be automated (assign role, send email, create user etc). The flexibility of METIS suggests that it might be a promising tool for managing a variety of digital library workflows, such as distributing cataloging efforts, the DWEL collection-building model, the Community Review process, etc.

Managing Change in Distributed Collections – Frank Shipman

A strength and weakness of Web-based educational resources is their capacity for change and innovation, as resource creators have new ideas, add new content, etc. Sometimes a resource can change significantly over time so that the metadata description is no longer accurate or the resource itself may no longer be suitable for inclusion in the library. Using a technique that collapses complex Web resources into a compact digital signature, this project is investigating how to automatically track content, structure, and navigational changes in Web-based resources.

Digital Collections and Management – Robert R. Downs

The issues and concerns associated with collection building take different forms when considering a range of collection types from centrally housed collections (focusing on archiving) to distributed collections (focusing on providing access). A four-part conceptual framework—acquisition, cataloging, dissemination, and maintenance—based on the experiences to date at the Center for International Earth Science Information Network (CIESIN) of Columbia University was presented.

2.2.2. Session Summary

Two of the presentations (DWEL and ESSEC) provided concrete examples of how diverse different approaches to collections building can be, as groups tailor goals and work processes to meet the needs of specific user constituencies. ESSEC collection-building efforts have focused on producing technical mechanisms for integrating distributed collections based on heterogeneous computing platforms and metadata frameworks. The DWEL collection-building effort has focused on producing a conceptual model and technical system for integrating distributed human efforts. The DWEL project offers a model for community-based collections development that emphasizes the collaborative and knowledge-construction aspects of collection building.

The CIESIN presentation provided a framework for comparing and contrasting different collection building projects, including those present at the workshop and many that were not. The framework advocates that decisions about how to approach the tasks of acquisition, cataloging, dissemination, and maintenance depend upon the intended use of the collection (archiving versus access) and the available resources (how to prioritize what gets archived, or what types of access are needed). The framework also distinguished between the primary activities of *building* and *operating*. Most of the DLESE collections are funded only for the building part and do not have funds for operations and ongoing collection management. Many session participants expressed concerns about the long-term sustainability of their projects.

Group discussion touched on the sheer effort involved in creating and managing a substantial digital collection. Issues discussed included training, item-level cataloging, and dealing with multiple metadata frameworks. The DWEL effort benefited from a series of workshops at the beginning and midway points of the project that focused on inducting participants into the collection building process and training participants on how to catalog resources using the DCS. In addition to these workshops, project participants received individualized follow-up advice on their first five cataloged records. This level of training and support is quite time and resource intensive. Even with training, generating and managing item-level metadata is very time and resource consuming. For many community members it is also not a particularly enthralling task. A particular challenge faced by the DWEL project has been to motivate participants to catalog resources and they report that participants appear to have particular trouble with writing brief descriptions.

Within ESSEC, considerable effort is being spent on managing collections that are in multiple versions of the same metadata framework (various versions of DLESE IMS). It will take considerable effort to transform all these variants to ADN. Two approaches for managing heterogeneous frameworks were discussed, including crosswalks (the mechanism currently being used in DLESE) and the “bucket” architecture used in the Alexandria Digital Library. In addition to managing multiple frameworks, managing changes to controlled vocabularies and taxonomies is also challenging. Participants expressed concern about the effort involved in updating legacy collections when a controlled vocabulary changes. One promising model for helping manage these types of changes is the Dublin core metadata elements registry.

Session participants agreed that quality was an important goal for DLESE collections and noted that many factors contributed to our overall perception of collection quality, including the quality of individual resources, the usefulness of a collection as a whole, the reliable maintenance of a collection, and the richness of services that add value to resources and collections, such as reviews. CIESIN relies on a User Working Group to review and approve all objects considered for accessioning into the archive. DWEL reviews every resource before accessioning it into their collection. These models emphasize filtering for quality prior to inclusion in the collection. The Community Review system, on the other hand, emphasizes identifying the best of an existing collection after resources have been accessioned. In this model, library users can then elect to search for only peer reviewed resources and can read reviews to learn about the strengths and weaknesses of individual resources.

The challenges behind all of these methods are scalability; each of these approaches requires significant human effort and it is not clear that these methods scale to large numbers of resources when relying mainly on volunteers. Session participants agreed that consideration must be given to motivating people to contribute time and effort to these activities and to acknowledging the work that people have contributed. Participants also noted that one way to improve resource quality was to inculcate resource creators to the review criteria and concerns, with the aim of effecting positive changes in the resources as they are being designed.

Participants also distinguished between quality and utility. For instance, even though the individual items in a collection may be scientifically sound or pedagogically innovative, how do you know that you are building a useful collection for teaching and learning? Making decisions about what the granularity of items in a collection should be is still challenging and little is known about how these decisions affect collection use. Cost effective formative evaluation techniques are needed that help guide collection developers during the building phase to create useful collections. The assessment techniques used in curriculum development projects were suggested as a useful model for educational collections assessment.

Discussions turned to the potential payoff of automatic techniques and tools to relieve some of the issues surrounding scalability and human effort, namely in the areas of metadata generation, workflow support, and tracking resource changes. Automatic metadata generation is a promising technology for reducing the effort to build collections, though several participants expressed reservations about the quality of metadata that is produced, particularly for important fields like the brief description. Perhaps rather than viewing automatic techniques as replacements for human activities, a more fruitful perspective would be one of cooperative problem solving (Fischer, Lemke et al. 1991) where the automatic techniques are used to augment human cataloging activities. Another approach, often called *scraping*, would be to encourage Web developers to incorporate basic metadata into their pages that can be easily harvested (e.g., large sites such as the Monterey Bay Aquarium).

Several participants expressed excitement over the potential of workflow tools, such as METIS, for helping to coordinate and automate many collection-building workflows. The distributed cataloging and ingest process being coordinated by the DLESE Program Center was volunteered as a candidate process to test METIS' viability for operational workflows. A third area of technology for helping with collection management was highlighted in the presentation by Shipman. Different approaches for identifying and tracking changes to resources based on link checking, caching and comparing, and digital signatures were discussed. While this research is very promising, many major issues are still unresolved, like identifying what changes are significant and warrant notifying collection managers. For instance, the major headlines in an online news service should change on a regular basis and collection managers would not want to be notified when this happened.

2.2.3. Recommendations

Based on these discussions, the group identified several priority areas for future work:

1. The current approach towards cross-walking collections based on multiple metadata frameworks, and multiple versions of the same framework, is effortful and time-consuming. Improved mechanisms and tools are needed that help to alleviate this effort.
2. Methods for automatically tracking changes in resources are an extremely promising technology that is not yet being used in operational libraries. More research should be conducted in this area, specifically research involving significant and real operational test beds.
3. Building better collections is a prime concern. Collections assessment techniques that address utility, not just gap identification and coverage, and that can be easily embedded in a collection building process, are needed.
4. Distributed collection building and community reviewing both lie at the heart of DLESE's *users as contributors* philosophy. The library needs to devise ways to support, motivate, and acknowledge this challenging intellectual work. Better tools and processes to supported distributed collections builders are needed, as well as effective training for collection building participants. The library also needs to systematically and publicly acknowledge and reward these contributed efforts.
5. Outreach to major resource providers is needed. These groups should be generating and providing their own metadata, both to the library and as "advanced organizers" to their own direct users. That is, the summary information contained in the metadata is also useful to human users that find these resources outside of the library's discovery service.

2.3 Building Services and Service Integration

The purpose of this thematic session was to think about the types of services a digital library such as DLESE needs, and how these may be integrated. Presenters were asked to discuss their approach and needs, and the lessons learned in developing their services. In the discussions that followed the presentations, the group identified common issues and needs arising across their projects.

2.3.1. Presentations

An important aspect of this theme is to highlight the rich variety of services a digital library can host. Some can be direct services that support the library user (review systems, bulletin boards etc.) Others provide background, or compound services (Fulker and Janée 2002), which are used by other library components but are not immediately apparent to a library end user. The library itself can also be treated as a compound service, and as such can be embedded into other applications. In these cases, the end user of the application may well have no knowledge of the particular library being used to provide a service that underpins the application.

Seven presentations were given to illustrate these different types of services. These seven presentations were grouped into three broad categories of service offerings:

- Services providing information enrichment
- Personal collection building services
- Services to support embedded interfaces

Short reports by presenters along with more information on their projects can be found in *Appendix 5*.

Services providing information enrichment

CRS – Robert Arko

The DLESE Community Review System (CRS) allows items in the DLESE broad collection to become part of the Reviewed Collection through reviews of the item by teachers and educators. The CRS technology is hosted at Columbia, and interacts with the core infrastructure hosted by the DPC so review status information can be made available to users of the discovery system. Reviews follow a rubric to reflect actual use of the resource in an educational setting.

JESSE – Martin Ruzek

The Journal for Earth System Science Education (JESSE) is an electronic journal that conducts peer-review of education resources aimed for use in Earth system science. The journal hosts resources in a system that allows reviewers to interact with the resource and enter their reviews online, and the review process is open, allowing the author and reviewers to interact. The journal hopes to provide professional recognition to authors and creators of resources through the peer-reviewed journal framework. In addition to the reviews, a peer-commentary area is also provided for others to give comments on the resource.

SERC – Sean Fox

The Science Education Resource Center (SERC) conducts workshops and develops materials that add value to resource collections by developing encompassing materials provided through themed Web areas. These activities are aimed at supporting faculty to improve their teaching. One project is the development of an early NSDL specialized portal to support using data in the classroom following a successful workshop on that subject.

Personal collection building services

Digital IdeaKeeper – Chris Quintana

This project is funded to design, build, and deploy into middle schools in the Detroit area the Digital IdeaKeeper. The Digital IdeaKeeper is an application individual students use to search libraries and collect resources in a portfolio. The application provides a framework in which to scaffold student inquiry and support them to analyze and synthesize the information they find. As an end-user application, an instance of the Digital Ideakeeper, will be used by a student to access libraries (through a search API) from within the application's interface environment.

Spatial Hypertext – Frank Shipman

The current Web is based on the idea of using explicit links between objects to provide relationships and meaning. The hypertext community has also been investigating how such links can be inferred from spatial arrangements of information. The Visual Knowledge Builder (VKB) is a tool supporting spatial hypertext. A user of the VKB application can create their own information space through dragging items from Web sites and libraries into a 2-D VKB workspace and arranging them as needed. In the context of a digital library, this is an example of a different paradigm for information to be organized.

Services to support embedded interfaces

Strand Map Service – Tamara Sumner

There is a strong premise that providing conceptual browsing interfaces to a library will help educators and learners more easily find useful information. This project is working to bring the AAAS strand maps for K-12 education into the digital library as an underlying service that may be accessed through both user interfaces and programmatic interfaces. The AAAS has published their strand maps in the *Atlas of Science Literacy*, but no structured digital form exists. The project is developing a data model for the maps, and a query language for accessing the service.

Digital Library Integration with Virtual Learning Environments (VLEs) – Mike Freeston

The presentation reviewed a new, international project bringing together US and UK partners to investigate how a digital library (in this case, the Alexandria Digital Library) can be used from within a Virtual Learning Environment (VLE). Some early examples of VLEs include WebCT and Blackboard, but there are others developed by various education enterprises and institutions. The value of a digital library will be in allowing VLE users access to library objects linked to the concepts being taught in the learning situation.

2.3.2. Session Summary

After the presentations, the participants broke into two parallel sessions for discussion devoted to the theme of building services and service integration. The results from these two sessions were combined and are presented below. Each session was asked to systematically examine issues and challenges arising from library services based on a five-point framework—service definition, building, integration, operation, and sustainability. Participants were advised to focus on the specific needs and lessons learned from their projects, rather than considering library services as an abstraction. Additionally, some discussion was devoted to the specific topic of data services; these comments were incorporated into section 2.2.2 on the data theme.

Service Definition

Participants agreed that *library services* defies definition, even when focusing on the projects represented at the workshop. This lack of definition makes it difficult to have concrete discussions with other projects, and to build on the results and experiences of other projects since the differences could be vast. As such, a taxonomy of service types would be extremely useful for furthering the state of knowledge in this area. What are the common characteristics? Within the projects represented, there were (at least) three major types of services present:

- Those that provided programmatic interfaces for constructing views of collections (e.g., library discovery services, or the strand map service)
- Those that provided tools for directly manipulating library content in specific ways (e.g., Visual Knowledge Builder, or Digital IdeaKeeper)
- Those that combined library content with community collaboration tools such as list servers or peer review systems (e.g., Community Review System, or SERC)

Service Building

Given that most of the projects present are currently in the building phase, it is not surprising that much discussion was devoted to this area. A number of key needs were articulated in the following areas:

- training and documentation
- collaboration and project awareness
- access to user populations and special constituencies
- open access to, and help with, creating library content underpinning services
- the establishment of standards in a few key areas

Training and documentation. Opportunities for more training were identified as a need. Workshops are needed on specific technologies and protocols, e.g. OAI, metadata frameworks and crosswalks. Better documentation is needed of library technologies and protocols across the full spectrum of development efforts, i.e., DLESE, NSDL, and specific services and projects. Minimally, *ReadMe* files should be included with each distribution describing the technical requirements, the skill levels recommended for installing, maintaining, and using the system or service, what the tool or service does, and who it is for. Ideally, this documentation could also be provided as a library collection; i.e., a one-stop shop for finding out about distributed projects and services under development. Participants specifically requested short one-page summaries of the tools and services that both DLESE and NSDL could provide collection or service developers, preferably with a schedule or timeline for provision. One interesting idea that emerged was the notion of mapping different development tools, services, and protocols onto a generalized lifecycle of a library collection or service, illustrating what comes into play when, during a project's life.

Collaboration and project awareness. Many participants expressed a desire for the Swiki (or similar tools) to be made available as a service to DLESE developers. Several participants had either used the Swiki at the 2002 DLESE Annual Meeting, or were using a variant (the *Wiki*) available through the NSDL Communications Portal. One project, the AVC (Klaus), is using the *Wiki* to support collaborations across teachers developing lesson plans around the Atmospheric Visualization Collection. Other tools needed by developers include tools for managing distributions and updates of the services they are making available. The need was also expressed for end-user-friendly version management tools; i.e., as teachers collaboratively develop educational resources they need to manage successive versions of the resource but do not want to engage with full-fledged Concurrent Versioning Systems (CVS).

Overall, it is regarded as extremely important to keep abreast of other projects and potential collaborators. Participants expressed the need for a wide spectrum of services and information in this area. Creating opportunities for people to talk to each other, and with DLESE and NSDL developers, was considered to be important. Meetings like the NSDL Annual Meeting and the DLESE Developers' Workshop help to enable this, but schedules are often too full and do not provide enough networking time. In terms of online information, short summaries of ongoing projects are needed, as well as concise listings of tools, services, collections etc. that distributed projects are making available to the community to build on. Other services, both online and human-mediated, could help projects to locate collaborators or people with skills and knowledge important to project goals, e.g., the *Collaboration Finder* developed in NSDL. Participants also discussed the possibility of having a recommender-style service (Resnick and Varian 1997) that helps to match tools and services with developers; such a service should enable users to recommend services that are provided outside of the digital library community but still potentially useful to digital library users and developers, such as *Wikipedia*, *PlaceWare*, and SRI's *Terravision*. Participants also expressed interest in learning more about international digital library efforts and services.

Access to user populations and special constituencies. Many projects are trying to follow best practices in user-centered design, or are trying to involve users as contributors as aligned with the broader DLESE philosophy. However, locating potential users and sustaining significant interactions

with users is often difficult. Projects need help finding focus group participants, beta testers, subjects for usability tests, resource reviewers, resource catalogers, etc. It would be fruitful to consider how events like the DLESE Annual Meeting could be leveraged to provide projects with time during the meeting for these types of activities, or at least to provide projects with a time to make the necessary contacts.

Open access to, and help with, creating library content underpinning services. Many services are built on top of library content; i.e., metadata, annotations (e.g. reviews), vocabularies, knowledge spaces, and in a few cases, primary library content. Oftentimes, individual services must do considerable amounts of work to create necessary content to illustrate their service or to broker partnerships to provide access to content. While both DLESE and NSDL provide access to metadata through harvesting of their metadata repositories, it is rare that projects make available other forms of library content, such as primary content or annotations. As such, extending the *open access* philosophy of DLESE and NSDL around metadata to other types of library content would be beneficial to service developers.

Given the importance of content to many services, discussions also considered computer and human-mediated services to ease metadata and annotation creation efforts. For instance, a “lite” version of the DCS could be made available to resource creators that helped them embed metadata into their content at the time of creation. Given the importance of involving users as contributors, workshops and professional development courses of teachers are needed that induct them into the processes of resource reviewing, collection building, and eventually using digital libraries. Organizations that engage in both digital library development and teacher professional development, such as CSMATE at Colorado State University, are well equipped to engage in these types of activities. Ideally, teachers and other participants could receive continuing education credits for participating.

Establishment of standards in a few keys areas. Many participants expressed the viewpoint that a key role for both DLESE and NSDL is to assist in the establishment of standards in a few key areas. Specifically cited areas were educational standards and geospatial metadata. For instance, there is no consistency across libraries in how educational resources are aligned with the National Science Education Standards (conceptually), and indexed as being aligned with a standard (technically). Participants also expressed frustration with the proliferation of content standards for georeferenced resources and data and expressed the need for leadership in picking a direction.

Knowing that your project correctly implements or complies with established standards is an ongoing concern. Several participants praised the OAI Validation Tool provided by the NSDL Core Integration team and expressed the desire for analogous automatic validation tools to ensure that they were complying with NSDL and DLESE standards. Similarly, standard reports should also be provided to collection developers about their harvested data, letting projects know about their schema validation, link integrity, etc.

Service Integration

Currently, the most widespread form of service integration involves building an interface or set of operations on top of metadata repositories. Most services use the OAI protocol to harvest metadata records, and optionally filter the metadata records to create a local repository, which they must subsequently manage. Another form of integration involves building services on top of protocols to support federated searching, such as the NSDL approach (SDLIP) (Paepcke 2000), the Alexandria Query Language (Janeé and Frew 2002), and the Open Digital Library Protocol (ODLP) (Suleman and Fox 2002). At the moment, few services make use of these protocols but several projects at the workshop are moving in this direction, primarily because they are reluctant to get into the “metadata

repository management” business. Finally, a third form of service integration builds on the NSDL’s incipient annotation framework. Projects such as the Community Review System (Kastens and Butler 2001) integrate with the DLESE discovery system via the exchange of annotation status information. Many DLESE projects are building types of annotation services that are pushing the envelope of the current definitions and capabilities of the NSDL annotation framework. It is imperative that work on the annotation framework move forward within NSDL.

There are very few cases of operational compound services, i.e., services built by combining other services through public programmatic interfaces. In part, this is because there are few standards or best practices within NSDL for defining and describing service interfaces. Perhaps a new NSDL track is needed that would explicitly fund the development of compound services. These pioneering services would serve as experimental models to help articulate the process of service combination and the interfaces needed. This also speaks to the need for service registries; is the business community already leading this effort with standards such as UDDI and WSDL?

The Alexandria Place Name Gazetteer protocol was cited as a positive example of an embedded service that provides a public programmatic interface to its place-name/geospatial footprint mapping service. Some new projects, such as the Strand Map Service, are also providing public protocols as explicit project goals. These types of services are relatively rare; the majority of services focus on providing custom interfaces or tools for end-users. Overall, participants expressed the need for more embedded services that developers can use over the wire without having to install software locally and without having to manage local metadata repositories.

Specifically called out were a variety of mapping services which translate between terms or information types; e.g., thesauri, knowledge spaces, etc. Another mapping service that would be useful to both developers and end-users is a service that maps national educational standards to state and local standards. To realize their full potential, future services of these types funded by NSDL must include the provision of public programmatic interfaces as explicit project goals.

Another type of mapping service critical to a distributed library network would be a metadata-mapping registry enabling developers to look up a way to transform from one metadata content standard to another. Such a registry would enable developers to submit their content standard and the transforms currently available, for different crosswalks. Tools are also needed to assist with the construction of crosswalks.

One class of embedded service was called out as being crucial infrastructure across a broad spectrum of library functions—namely authentication, user verification, and user profile or role management services. Many of the projects at the workshop anticipate needing these capabilities in the near future. NSDL is promoting Shibboleth (Internet2 2003), a tool that mediates between a registration service and an end-service, but does not itself provide a registration service. Participants would also like to see a reusable registration service that can be easily embedded with their libraries and services. Also, while Shibboleth supports authentication and authorization, it does not gather and store other information typically associated with user profiles. Participants discussed a variety of ways that LDAP capabilities could be merged with the Shibboleth service to meet this spectrum of needs.

Service Operation

Two key areas were identified as being important issues for operational services: marketing services and customer support infrastructure. Effective outreach and marketing are necessary for most projects to be successful. Most projects do not have the time, funds, or expertise to engage in significant marketing activities about their project or service. Both DLESE and NSDL could, and should, provide

shared marketing services to other projects. In fact, these two organizations could promote a culture of “helping us to help each other” get the word out. For instance, projects would like to provide DLESE with brochures, CDs and other give-aways for display in DLESE booths at conferences. Projects would like to be featured, when appropriate, in DLESE and NSDL advertising, product reviews, or press releases.

Projects that provide operational services also recognize the importance of good customer support; this includes topical *ask-a* services and more traditional forms of technical support or problem resolution services. NSDL Core Integration is taking a leadership role in providing an NSDL-wide Ask-A service (reference desk), as demonstrated in the December 2002 release of the NSDL. Projects also expressed the need for more traditional forms of technical support for bug reports, problem resolution, and other customer queries. A number of commercial and open source packages available for “trouble ticket” types of services; e.g., *Request Tracker* (RT). Other projects, such as *Windows to the Universe* (Windows to the Universe team 2003) and the NSDL (through building the Ask-NSDL on the services of the Virtual Reference Desk organization) have already created help desks that might be reusable. Projects need assistance identifying these types of open source or reusable tools and services. Alternatively, smaller projects might like to sign up for a trouble ticket brokering service; i.e., the support software is hosted by another organization that routes requests to the right service or project and tracks response time metrics for performance reports.

Service Sustainability

Discussions centered on the need to articulate different models of sustainability. For collections, sustainability involves collection maintenance and dealing with resource or content persistence. For some service projects, sustainability is conceived as long-term, stable funding to support continued operations. For other service projects, sustainability refers to devising a *hand-over* model, where the developer group hands the project over to another group for long-term operations and maintenance. An example of the latter is the Strand Map Service project, where the University of Colorado developer group will hand over the service to the DLESE Program Center for interim hosting (for up to a 5 year period), who will then hand the service over to Project 2061 at AAAS for long-term operations. This arrangement was negotiated during proposal development and was enumerated in the proposal as an integral part of the overall project plan.

Different types of support are needed for these different sustainability models. For projects interested in provided long-term operations, assistance with locating funding sources or workshops on developing new funding sources would be useful. For projects based on the *hand-over* model, assistance with locating and developing operational partners would be most useful.

Discussions often linked evaluation activities with project sustainability. It was felt that all projects would benefit from data showing that their service is useful to various user constituencies. For instance, it would be extremely helpful if DLESE and NSDL could communicate patterns and amounts of service and collection use to developer organizations. Also, basic information about overall library use could help developers to tell stories about how their service potentially benefits various libraries or populations. Such basic information could be disseminated through reports on library usage; what are people searching for, what tools and services get used, what are the demographics of library users, etc. Studies on the educational impact of digital libraries in general would be useful for projects seeking further funding. Projects also need standard tools and assessment instruments to help them conduct their own formative, summative and impact evaluations.

2.3.3. Recommendations

Based on these discussions (summarized above), the groups identified several priority areas:

1. Training, documentation and project awareness. Technical workshops, technical documentation, descriptions of tools, services or frameworks available from projects should be available in the library as a 'one-stop shop', or collection. Projects need mechanisms that support staying abreast of what other projects do (both in DLESE, NSDL, and beyond), and mechanisms to support collaboration between projects.
2. More support of *user as contributor*. Tools to support collaborative content development are needed that are end-user friendly, i.e. to a teacher or educator, not a software developer. Workshops and professional development of teachers to induct them into library processes such as resource reviewing or collection building are needed.
3. Open access to, and help with, creating library content underpinning services. Many services are built on top of library content beyond metadata; i.e. annotations (e.g. reviews), vocabularies, crosswalks, knowledge spaces. Extending the *open access* philosophy of DLESE and NSDL around metadata to other types of library content would be beneficial to service developers.
4. The area of integration of compound, or embedded, services is important. For these to work, service interfaces, standards and protocols need to be available and understood. DLESE and NSDL need to establish standards for their community, and provide timely tools and support for using and implementing them. Developing (or funding) examples of compound services that others can use as models to understand the process would be useful. Example: authentication and registration systems (authentication, verification, user profiles, access control) were highlighted as one type of compound service needed by many projects.
5. Shared outreach and marketing services are needed to help projects be successful. Time and resources are not always available to each individual project, but through the consortium of DLESE and NSDL, much could be gained from shared marketing.
6. Different models of sustainability need to be articulated and investigated. What happens to a service when the funding comes to an end, or the project principal investigator(s) no longer want to or are unable to continue supporting the service? For some service projects, sustainability is conceived as long-term, stable funding to support continued operations. For other service projects, sustainability refers to devising a *hand-over* model, where the developer group hands the project over to another group for further operations and maintenance.

Part 3: References

- Atkins, D. E., K. K. Droegemeier, S. I. Feldman, H. Garcia-Molina, M. L. Klein, D. G. Messerschmitt, P. Messina, J. P. Ostriker, and M. H. Wright, 2003. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. National Science Foundation, Arlington, VA, 84 p. http://www.communitytechnology.org/nsf_ci_report/report.pdf (15 April 2003).
- Carpenter, B., 2003. What is grid computing? *ISOC Member Briefings: (Grid Computing #11)*. <http://www.isoc.org/briefings/011/index.html> (15 April 2003).
- DLESE, 2002a. Governance plan for DLESE, V. 2.0. Digital Library for Earth System Education. http://www.dlese.org/documents/plans/GovernancePlan_ver2.0.html (5 November 2003).
- DLESE, 2002b. DLESE versioning document. Digital Library for Earth System Education (DLESE). <http://www.dlese.org/documents/plans/versioning.html> (5 November 2003).
- DLESE, 2002c. DCS and OAI Quick Start Overview. Digital Library for Earth System Education. <http://www.dlese.org/Metadata/tool/index.htm> (15 April 2003).
- DLESE, 2001. DLESE Strategic plan, Version 12.0. <http://www.dlese.org/documents/plans/stratplanver12.html> (5 November 2003).
- ESRI, 2003. ArcIMS geographic information systems (GIS) and mapping services. Environmental Systems Research Institute (ESRI). <http://www.esri.com/software/arcims/> (5 November 2003).
- Fischer, G., A. C. Lemke, T. Mastaglio, and A. Morch, 1991. The role of critiquing in cooperative problem solving. *ACM Transactions on Information Systems*; **9**(2), pp. 123-151.
- Fulker, D., and G. Janée, 2002. Components of an NSDL architecture: Technical scope and functional model. <http://arxiv.org/ftp/cs/papers/0201/0201027.pdf> (15 April 2003).
- Internet2, 2003. Shibboleth project. Internet2. <http://shibboleth.internet2.edu/> (15 April 2003).
- Janee, G., and J. Frew, 2002. The ADEPT Digital Library Architecture. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL '02)*, Portland, Oregon (July 14-18). ACM Press, pp. 342-349.
- Kastens, K. A., and J. C. Butler, 2001. How to identify the "best" resources for the reviewed collection of the Digital Library for Earth System Education. Column: "Another Node on the interNet," *Computers & Geosciences*; **27**(3), pp. 375-378.
- Kolodziej, K., Ed., 2003. *Open GIS Web Map Server Cookbook*. OGC Document Number: 03-050r1, Version 1.0.1; Open GIS Consortium, Inc. <http://www.ogcnetwork.org/docs/03-050r1.pdf> (5 November 2003).

- Manduca, C. A., and D. W. Mogk, 2000. *The Digital Library for Earth System Education: A Community Plan*. Final Report from the University of Oklahoma to the NSF, Grant 99-06648, 44 p. <http://www.dlese.org/documents/plans/CommPlanFinal.pdf> (25 March 2002).
- NSF, 1999. *NSF Geosciences beyond 2000: Understanding and predicting Earth's environment and habitability: Summary. Report of a long-range planning effort for geosciences in the 21st century*. NSF 00-28; National Science Foundation, Arlington, VA, 24 p. <http://www.geo.nsf.gov/adgeo/geo2000.htm> (20 November 2002).
- NSF, 1996. *Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology*. NSF 96-139; National Science Foundation, Arlington, VA, 90 p. http://www.nsf.gov/pubsys/ods/getpub.cfm?ods_key=nsf96139 (12 March 2003).
- Paepcke, A., 2000. Search middleware and the simple digital library interoperability protocol. *D-Lib Magazine*; **6**:(3, March). <http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=2000-53&format=pdf&compression=> (28 May 2003).
- Resnick, P., and H. Varian, 1997. Recommender systems. *Communications of the ACM*; **40**:(3), pp. 56-68.
- Suleman, H., and E. A. Fox, 2002. Designing protocols in support of digital library componentization. *European Conference on Digital Libraries 2002 (ECDL), Sept. 16-18, Rome*. http://oai.dlib.vt.edu/odl/pubs/ecdl_2002_odl_paper_revised.pdf (15 April 2003).
- Windows to the Universe team, 2003. *Windows to the Universe*. ©2000-01 University Corporation of Atmospheric Research (UCAR), ©1995-1999, 2000 The Regents of the University of Michigan. <http://www.windows.ucar.edu/> (5 November 2003).

Appendix 1: Workshop Agenda

Location

UCAR ATD Atrium, Bldg FL1
University Corporation for Atmospheric Research
Boulder, Colorado

Wednesday, February 19, 2003

8:30 AM

- Introduction to Workshop, logistics, group introductions – Wright, DPC
- DLESE Versions: Version 2.0 release in Summer 2003; Version 3.0 in 2006. This presentation will review plans for DLESE over the next few years as articulated through the Strategic Plan and versioning document – Sumner, DPC
- DLESE Governance, committees, new service areas. Short review of DLESE structures and relationships – Marlino, DPC
- Cyberinfrastructure and DLESE – Marlino, DPC

10:00 Break

10:30

- Overview of V2.0 discovery and Open Archives Initiative – Weatherley, DPC
- ADN metadata framework – Ginger, DPC
- A perspective on NSDL Core Integration – Phipps, Cornell University
- Utilizing geo-spatial services – Janee, UCSB

12:00 Lunch

1:30 Two parallel sessions:

Break-out Session A: Bringing Data into the Library

Four 10-minute presentations:

- Atmospheric Visualization Collection (AVC) – Klaus, Argonne National Lab
- GEOsciences Network (GEON) – Ludaescher, UCSD, and Meertens, UNAVCO
- Thematic Realtime Environmental Distributed Data Services (THREDDS) – Davis, Unidata Program Center
- Visual Geophysical Exploration Environment (VGEE) – Pandya, DPC

A long-term goal of DLESE is the incorporation of authentic data into the education process, and the linking of research (that utilizes data) with education. The presentations will highlight a number of projects giving access to data, both in its raw form, and scaffolded within education objects (or resources). This group will look to the future of how this goal can be achieved, e.g. through identifying good demonstrators, and what impediments there may be.

Break-out Session B: Building Collections

Four 10-minute presentations

- Earth and Space Science Education Collections (ESSEC) – Olds, NASA
- Digital Water Education Library (DWEL) – Warnock, CSU
- A Lightweight, Flexible, and Web-Based Approach to Supporting Workflow in Digital Libraries – METIS project – Anderson, CU

- Managing Change in Digital Collections – Shipman, Texas A&M
- Digital Collections Management – Downs, Columbia

Core components of a library are the collections provided to its users. The presentations will highlight two DLESE collections (that are digital libraries in their own right) and then two on tools and ideas for managing collections. What has become evident is that collections don't necessarily share all the same workflow or QA processes. This group will consider these past cases and look to what development areas are necessary to support distributed collections and aggregation.

3:30 Break

4:00 Group report out and larger discussion

5:00 Adjourn

Thursday, February 20, 2003

8:30 AM

Breakout Session C: Building Services and Service Integration

10-minute presentations:

- Services providing information enrichment:
 - DLESE Community Review System – Arko, Columbia University
 - Journal for Earth System Science Education (JESSE) – Ruzek, USRA
 - Science Education Resource Center – Fox, Carlton College
- Personal collection building services:
 - Digital IdeaKeeper – Quintana, University of Michigan
 - Spatial Hypertext for Digital Library Providers and Patrons (Visual Knowledge Builder) – Shipman, Texas A&M
- Services to support embedded interfaces:
 - Strand Maps as an Interactive Interface to NSDL Resources – Sumner, University of Colorado
 - Integration with Virtual Learning Environments (VLEs) – Freeston, University of California at Santa Barbara

Secondary core areas of the library are the various services provided. The short presentations here give a taste of the variety, from review and annotation systems, to other applications that require specific service API (Application Programming Interface) access (e.g. to discovery), to pure service mechanisms that support further services. As DLESE grows, the number of services and their distribution over the Internet will increase. This session will provide an opportunity to consider aspects of having distributed services: discovery, implementation, protocols, and interoperability standards.

10:00 Break

10:30 Break into discussion groups for Session C

12:00 Report out and larger discussion

12:30 Lunch

- 1:30 Workshop Wrap and User Interface Focus Group
- Conclude the workshop and any unfinished business
 - Timeline for white paper/workshop report production
 - Chance to participate in a focus group supporting the design of the next library User Interface (UI)
- 3:00 Adjourn

Appendix 2: Participants by Project and Alphabetical by Surname

Section 1: Participants – by Project

Projects represented	Participants	Organization	Additional affiliations
	Mike Freeston	University of California, Santa Barbara	
Alexandria Digital Library/ADEPT	Greg Janece	University of California, Santa Barbara	NSDL Core Integration technical team; Co-PI, Strand Maps as an Interactive Interface to NSDL Resources
Atmospheric Visualization Collection (AVC)	Chris Klaus, PI	Argonne National Laboratory	
	Mark McCaffrey	National Oceanic and Atmospheric Administration (NOAA)	
Center for International Earth Science Information Network (CIESIN)	Robert Downs R.	Columbia University	
A Digital IdeaKeeper for K-12 (NSDL services)	Chris Quintana, PI	University of Michigan	
	Srikanan Reddy	University of Michigan	
Digital Water Education Library (DWEL)	Ed Geary, PI	Colorado State University	
	Andrew Warnock	Colorado State University	
DLESE Community Review System	Robert Arko	Columbia University	
	Tom Boyd	Colorado School of Mines	
DLESE Technical Committee representatives	Frank Shipman	Texas A&M, Center for the Study of Digital Libraries	Visual Knowledge Builder Walden's Paths

Projects represented	Participants	Organization	Additional affiliations
	Peter Burkholder		
	Lynne Davis		
	Ryan Deardorff		
	Holly Devaul		
	Katy Ginger		
	Karon Kelly		
DLESE Program Center	Mary Marlino	PI, DLESE Program Center UCAR	co-PI, Strand Maps as an Interactive Interface to NSDL Resources ; Collaborator, GEON project
	Rajul Pandya		co-PI, Visual Geophysical Exploration Environment (VGEE) ; Collaborator, THREDDS
	Steve Sullivan		
	John Weatherley		
	Michael Wright		co-PI, Strand Maps as an Interactive Interface to NSDL Resources
GEOsciences Network (GEON)	Bertram Ludaescher, co-PI	San Diego Supercomputer Center	
	Charles Meertens	UNAVCO , UCAR	
Journal of Earth System Science Education (JESSE)	Martin Ruzek, PI	Universities Space Research Association (USRA)	co-PI, A Lightweight, Flexible, and Web-Based Approach to Supporting Workflow in Digital Libraries
A Lightweight, Flexible, and Web-Based Approach to Supporting Workflow in Digital Libraries	Ken Anderson, PI	University of Colorado	
NASA Earth and Space Science	Shelley Olds	NASA Earth Science Enterprise	

Projects represented	Participants	Organization	Additional affiliations
<u>Education Collections</u>			
<u>NASA Earth and Space Science Education Collections</u>	Robert Antonucci	<u>NASA Earth Science Enterprise</u>	
<u>NSDL Core Integration</u>	Jon Phipps	<u>Cornell University</u>	
<u>Science Education Resource Center (SERC)</u>	Sean Fox	<u>Carleton College</u>	
	Tamara Sumner, PI	<u>University of Colorado</u>	Co-PI, <u>DLESE Program Center</u> Co-PI, <u>DWEL</u>
	Faisal Ahmad	<u>University of Colorado</u>	
	Sonal Bhushan	<u>University of Colorado</u>	
<u>Strand Maps as an Interactive Interface to NSDL Resources</u>	Qianyi Gu	<u>University of Colorado</u>	
	Francis Molina	<u>American Association for the Advancement of Science (AAAS)</u>	<u>Project 2061</u>
	Brian Sweeney	<u>American Association for the Advancement of Science (AAAS)</u>	<u>Project 2061</u>
<u>Thematic Realtime Environmental Distributed Data Services (THREDDS)</u> (NSDL services)	John Caron	<u>Unidata Program Center, UCAR</u>	
	Ethan Davis	<u>Unidata Program Center, UCAR</u>	
	Ben Domenico, PI	<u>Unidata Program Center, UCAR</u>	

Section 2: Participants - Alphabetical

Participants	Organization	Project represented	Additional affiliations
Faisal Ahmad	University of Colorado	Strand Maps as an Interactive Interface to NSDL Resources	
Ken Anderson, PI	University of Colorado	A Lightweight, Flexible, and Web-Based Approach to Supporting Workflow in Digital Libraries	
Robert Antonucci	NASA Earth Science Enterprise	NASA Earth and Space Science Education Collections	
Robert Arko	Columbia University	DLESE Community Review System	
Tom Boyd	Colorado School of Mines	DLESE Technical Committee representative	
Sonal Bhushan	University of Colorado	Strand Maps as an Interactive Interface to NSDL Resources	
Peter Burkholder	UCAR	DLESE Program Center	
John Caron	Unidata Program Center, UCAR	Thematic Realtime Environmental Distributed Data Services (THREDDS)	
Ethan Davis	Unidata Program Center, UCAR	Thematic Realtime Environmental Distributed Data Services (THREDDS)	
Lynne Davis	UCAR	DLESE Program Center	
Ryan Deardorff	UCAR	DLESE Program Center	
Holly Devaul	UCAR	DLESE Program Center	
Ben Domenico, PI	Unidata Program Center, UCAR	Thematic Realtime Environmental Distributed Data Services (THREDDS)	
Robert R.	Columbia University	Center for International	

Participants	Organization	Project represented	Additional affiliations
Downs		Earth Science Information Network (CIESIN)	
Sean Fox	Carleton College	Science Education Resource Center (SERC)	
Mike Freeston	University of California, Santa Barbara	Alexandria Digital Library/ADEPT	
Ed Geary, PI	Colorado State University	Digital Water Education Library (DWEL)	
Katy Ginger	UCAR	DLESE Program Center	
Qianyi Gu	University of Colorado	Strand Maps as an Interactive Interface to NSDL Resources	
Greg Janee	University of California, Santa Barbara	Alexandria Digital Library/ADEPT	NSDL Core Integration technical team ; Co-PI, Strand Maps as an Interactive Interface to NSDL Resources
Karon Kelly	UCAR	DLESE Program Center	
Chris Klaus, PI	Argonne National Laboratory	Atmospheric Visualization Collection (AVC)	
Bertram Ludaescher, co-PI	San Diego Supercomputer Center	GEOsciences Network (GEON)	
Mary Marlino, PI	PI, DLESE Program Center ; UCAR	DLESE Program Center	co-PI, Strand Maps as an Interactive Interface to NSDL Resources ; Collaborator, GEON project
Mark McCaffrey	National Oceanic and Atmospheric Administration (NOAA)	Atmospheric Visualization Collection (AVC)	
Charles Meertens	UNAVCO , UCAR	GEOsciences Network (GEON)	
Francis Molina	American Association for the Advancement of Science (AAAS)	Strand Maps as an Interactive Interface to NSDL Resources	Project 2061
Shelley Olds	NASA Earth Science Enterprise	NASA Earth and Space Science Education	

Participants	Organization	Project represented	Additional affiliations
		Collections	
Rajul Pandya	UCAR	DLESE Program Center	co-PI, Visual Geophysical Exploration Environment (VGEE) ; Collaborator, THREDDS
Jon Phipps	Cornell University	NSDL Core Integration	
Chris Quintana, PI	University of Michigan	A Digital IdeaKeeper for K-12	
Srikaran Reddy	University of Michigan	A Digital IdeaKeeper for K-12	
Martin Ruzek, PI	Universities Space Research Association (USRA)	Journal of Earth System Science Education (JESSE)	co-PI, A Lightweight, Flexible, and Web-Based Approach to Supporting Workflow in Digital Libraries
Frank Shipman	Texas A&M, Center for the Study of Digital Libraries	DLESE Technical Committee representative	Visual Knowledge Builder Walden's Paths
Steve Sullivan	UCAR	DLESE Program Center	
Tamara Sumner, PI	University of Colorado	Strand Maps as an Interactive Interface to NSDL Resources	Co-PI, DLESE Program Center Co-PI, DWEL
Brian Sweeney	American Association for the Advancement of Science (AAAS)	Strand Maps as an Interactive Interface to NSDL Resources	Project 2061
Andrew Warnock	Colorado State University	Digital Water Education Library (DWEL)	
John Weatherley	UCAR	DLESE Program Center	
Michael Wright	UCAR	DLESE Program Center	co-PI, Strand Maps as an Interactive Interface to NSDL Resources

Appendix 3: DLESE Core Services Architecture v. 2.0

by John Weatherley, DLESE Program Center

In the summer of 2003 the Digital Library for Earth System Education (DLESE) will provide several new interactive and technical enhancements with release v2.0 of the library. Many of these enhancements will be visible to the public in the form of new search and discovery features while others will affect the behind-the-scenes operations of the library. Users will see new features such as the ability to search over multiple collections, the ability to view annotations associated with individual records, and the ability to search over educational standards. To support these features, the operational side of the library will see enhanced tools and services for collection management, data integrity, and quality assurance (QA), and the sharing of data and collections with external partners.

Core Collections Management and Discovery Architecture

Figure 1 gives an overview of the major components of the upcoming DLESE v2.0 core library architecture for collections management and discovery. Applications and services that will provide access to library users, collection builders, and external partners are shown in yellow (numbers 1a, 7, 8, 9), and applications that will provide access-controlled interfaces for library administration are shown in blue (1b, 6). The tan boxes (numbers 2, 3, 4, 5) indicate non-application software components and data stores.

The flow of data and library management in Figure 1 goes primarily from left to right following the

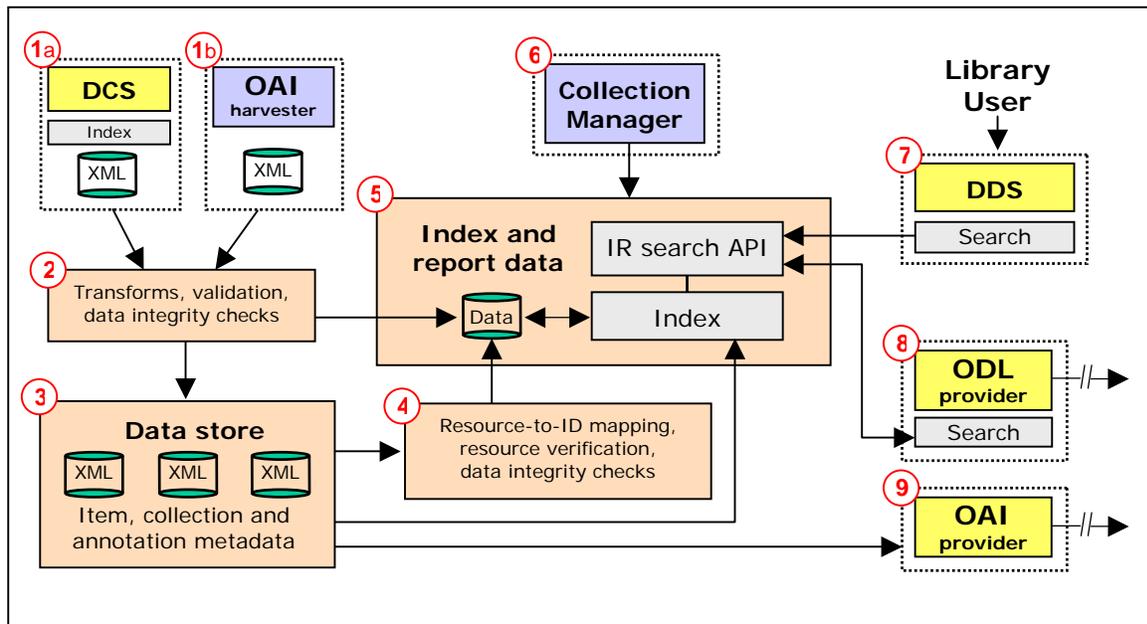


Figure A3-1 – v2.0 Core Services Architecture.

numbered items sequentially. On the left, items 1-3 involve the creation, editing and ingest of record metadata and collections. In the middle, items 4-6 are related to library and collection management and quality assurance. On the right, items 7-9 represent the external search and data-providing interfaces the library will provide. Together these components make up the DLESE core library architecture for collections management and discovery as it is envisioned for v2.0. Each of these items is detailed below, highlighting some of the core user and library administrative processes for which the supporting software is being designed.

Metadata Creation and OAI Harvesting

The creation and import of XML metadata records will continue to be accomplished using the DLESE Catalog System (DCS) and an Open Archive Initiative (OAI) metadata harvester, shown at numbers 1a and 1b. The DCS is used to create, edit and remove XML item-level metadata records that are part of the DLESE Community Collection (DCC), which is currently managed at the DLESE Program Center (DPC). Both internal and external users can contribute new resources to the DCC using this instance of the DCS, however administrative processes such as the QA and accession of the records is limited to DLESE library staff.

Metadata collections that are created and managed externally are imported to the library using an OAI harvester. External collection builders are responsible for maintaining data integrity and performing QA over their metadata. External collections that are harvested include both item-level and annotation metadata. For example, item-level metadata from the Atmospheric Visualization Collection (AVC) and annotation metadata from the Community Review System (CRS) are currently being harvested in this fashion.

Metadata Check and Storage

The DCC and other harvested collections will be transformed to the Alexandria-DLESE-NASA (ADN-I) metadata format (number 2) if not already in that format, checked for conformance to the ADN-I XML schema and then stored in XML form to the data store shown at number 3. Annotation and collection-level metadata will also be stored in the data store. Errors that appear during the harvest, transform or validation of individual records or collections will be saved as reports, and automated or manual error notification provided to DLESE staff and external collection builders.

Metadata Management, Administration, QA and Indexing

Once the data has been validated and stored, a number of automated and human processes provide management, administration and QA over the library collections. A resource-to-ID mapping service (step 4) will detect duplicate records that reference the same resource. Duplicate records within a single collection will be flagged in error per DLESE policy. Records across collections that refer to the same resource will be identified for presentation in the user interface (UI) as a single conglomerate to the user at search time. The mapping service is also used to check for resource URLs that are unavailable or that may have moved, providing reports used to notify DLESE library administrators and external collection builders.

An index that uses Information Retrieval (IR) technology is then built over all metadata records in the library (number 5). This index will be used for keyword and fielded searches performed by library end-users as well as administrative staff and will also be accessed by Web services as described below. The index supports the DLESE Query Language (DQL), which is derived from the fields and vocabularies defined in the ADN framework and also contains fields for searching by collection and annotation.

DLESE Discovery System

The DLESE Discovery System (DDS), shown at number 7, provides a Web-based search interface for library end-users that exposes the DQL by making the fields (such as grade level) and vocabularies (such as primary K-2) visible as selectable options in the UI, and provides free-text search and browsing

capabilities. The current DDS will be expanded in Summer 2003 to provide searching over multiple collections, the ability to discover and view records that have annotations associated with them, and the ability to search over standards. In the DDS, free-text searches return an ordered set of results to the user's Web browser. Relative term frequency and proximity is used to order the results. In addition, the order of results is determined in part by weighted fields. For example, if the user enters a search term that exists in the title field for a particular record, it is given more weight and thus appears higher in the result list than records that contain the term in the description field. In the new version of the DDS, records that are referenced by annotations will be assigned higher weight than those that do not.

In the new DDS, the user will be presented a list of results that show a brief description of the record including its associated collection, a link to the record's full description, a link to the associated collection level metadata and links to annotations that refer to the record. If a resource has more than one record associated with it from multiple collections, the record that matches the search criteria best will be presented to the user in the initial brief display and links will be provided to allow access to the other records.

Collection Manager

The Collection Manager (CM) application (number 6) will provide administrators with an interface from which they can control which items and collections are visible to end-users and Web services. The CM will provide the administrator full access to the index using a superset of the query options and data views that will be available to the end-user. The CM will also provide access to the error and URL reports, and allow configuration of which collections may be accessioned automatically upon ingest and which require manual inspection prior to ingest. The CM will provide options such as the ability to control the accessioning behaviors and notifications that occur when errors in new records are detected.

OAI Provider to support Harvesting and Searching

An OAI provider will be used to export the DCC to the NDSL and others in ADN, Dublin Core and other metadata formats (number 9). Externally managed collections that were harvested in step 1 may also be served using this OAI provider, allowing collection builders to use DLESE as a data provider proxy if requested.

A specialized OAI provider that implements the Open Digital Library (ODL) search protocol will provide metadata to external partners (number 8). Through the ODL search provider, external partners may query the DLESE metadata collections using the DQL and receive raw metadata back as an ordered set of results within the OAI protocol framework. The metadata may then be used to render custom interfaces or be embedded in remote clients on the fly. This ODL search provider will initially be tested as a prototype system with the Digital IdeaKeeper project and others to determine its viability as a formal library service in the future.

Component-based Architecture

Within the DLESE core library architecture, the software components are loosely coupled, allowing each to operate independently by providing data and functionality to one another internally using a service model. Most components are Java-based and use shared object-oriented DLESE code libraries. Components also leverage open source libraries and products including the Apache Lucene search API, which is used in the index, the Apache Struts Web application framework, which is used in the DDS, CM and OAI applications, and MySQL databases, which are used by several components to store and access

reports. Communication among the internal components is accomplished by sharing files and serialized Java objects stored on disk, accessing shared databases, and by sharing the Lucene index.

Conclusion

The DLESE library is undergoing a number of enhancements with the v2.0 release in the summer of 2003. User-centered enhancements include the ability to search over multiple collections, the ability to view annotations associated with individual records, and the ability to search over educational standards. To support these publicly visible enhancements, the library will see new technology and administrative tools such as the Collections Manager application, the resource-to-ID mapping service, automated resource URL verification, and a feature enhanced index and searching scheme.

Appendix 4: DLESE Metadata and Collections

by Katy Ginger, DLESE Program Center

Metadata and Collections Defined

In the Digital Library for Earth System Education (DLESE) metadata can be defined as descriptive information about a resource that is either returned to users upon a search or shared with other digital libraries. In this context, the word resource is very broad. Resources may include Web-based learning materials, DLESE policies and plans, or annotations about learning materials.

Metadata about these resources is the actual content of the DLESE core collection management and discovery system. This metadata provides:

- Pointers to Web-based learning resources
- Services to indicate when Web-based learning resources are part of the DLESE Reviewed Collection (an exemplar collection of resources within DLESE)
- Services to provide news and opportunities information relevant to the DLESE community
- Access to library documents regarding DLESE development, history, and policies

The metadata is in a variety of formats from specialized metadata frameworks developed in the eXtensible Markup Language (XML) to indexed documents using an open source search engine for organization. Metadata about individual resources is called item-level metadata. When individual item-level metadata about resources is grouped, a collection is born.

A DLESE collection is a grouping of individual item-level metadata records that are organized around a theme, organization, or other identified criteria. Once a group is organized, an overall description of the grouping is needed. This description is called the collection-level metadata as opposed to the item-level metadata describing the individual resources. Consequently, collection building is a process over time that shapes a collection into a balanced, cohesive and sought-after set of materials.

Metadata Specific to DLESE

With the brief introduction above, this section provides a closer inspection of the metadata formats within DLESE. Several *item-level* metadata formats are in use:

- **ADN-I:** A metadata framework developed as a joint effort between DLESE, the Alexandria Digital Library and the NASA Earth Enterprise Education Office; activated Feb. 2003
- **DLESE -IMS:** An old metadata format that serves the same purpose as ADN-I but will be decommissioned in Fall 2003
- **Annotation:** A metadata framework developed locally that will transition into the NSDL (National Science Digital Library) annotation framework
- **News and Opportunities:** A locally developed metadata framework for news and opportunities postings.

The *ADN-I*, *annotation*, and *news and opportunities* metadata frameworks are implemented using XML schemas and have a unique purpose within DLESE. The *ADN-I framework* is used to describe Web-based learning resources that can be used in a classroom setting, e.g. activities, data sets, animations, models, assessment tools, problems sets, instructor guides, Webcasts. The older, more limited, *DLESE-IMS framework* was used for the same purpose. The *annotation framework* is used to make statements, comments, or questions about a resource. Annotations are not meant to be stand-alone information. Inherently, annotations need to be connected to a resource to have meaning. Annotations are not resources

for use (if something is for use it should be directly cataloged as such). Rather, annotations are additional information that adds value to the resource being annotated. The *news and opps framework* is used to describe events of a time-sensitive nature. This includes workshops, fellowships, internships, announcements, etc.

Each of these item-level metadata formats has its own unique required metadata that must be completed before the metadata record can be accessioned. Within required metadata there is cataloger-provided required metadata and collection-builder required metadata. Cataloger-provided required metadata is information gathered about the resource at the time of cataloging (e.g. title and description). Collection-builder required metadata is administrative information (e.g. catalog number for the metadata record, catalog date). The cataloger-provided required metadata for ADN-I is: title, description, subject, grade range, resource type, technical requirements, URL to the resource, name of the resource creator and name of the cataloger, cost, and copyright information. The collection-builder required metadata for ADN-I is: a catalog number for the metadata record, catalog date, record status (e.g. is it complete), metadata terms of use, copyright of the metadata, language of the resource, and language of the metadata. Further information on ADN-I is available on the DLESE metadata Web site at <http://www.dlese.org/Metadata>.

For the annotation and news and opportunities frameworks, the required metadata is much less and it is specific to the type of information being provided. For example, if describing a workshop in the news and opportunities framework, start and end dates for the workshop are required. No matter which metadata framework is used, there is the ability to provide more than just required metadata. For example, in ADN-I, this would include providing geospatial, educational standards, and temporal information.

Collections Specific to DLESE

Once item-level metadata is grouped, it can become a collection. Within DLESE there are two primary overarching collections for educational resources. These two collections are:

- **The Broad Collection:** Provides access to resources that are relevant to Earth system education, basically bug-free, and available at little or no cost. Items (individual resources) are described using the ADN-I metadata framework. Annotation metadata records may be provided if appropriate and refer to one of the specific item records.
- **The Reviewed Collection:** Provides access to resources that meet the criteria for the Broad Collection and the additional Reviewed Collection criterion stated in the DLESE scope statement (http://www.dlese.org/documents/policy/CollectionsScope_final.html). One goal of the Reviewed Collection is to showcase exemplar resources and to create community-based review mechanisms that reward and recognize contributors. Collection builders provide both ADN-I and annotation metadata framework records or just annotation metadata framework records depending on how the collection is being built.

Within these two primary collections, DLESE contains individual collections, which are considered to be either formal or informal:

- **Formal collections:** Provide a high level of recognition for a collection. Individual descriptions and other information about each resource within a collection are available. Items within formal collections may reside in either the Broad or Reviewed Collection.
- **Informal collections:** Individual resources within a collection are not identifiable, only a description of the collection as a whole is available. Informal collections reside only in the Broad Collection as a single metadata record in the ADN-I metadata format. Some examples are large portal Web sites.

In order to be accessioned as a collection into DLESE, the requirements of the Interim Collections Accession policy (http://www.dlese.org/documents/policy/collections_accession.html) must be met at two levels; a documentation (general collection information) level and a technical level (correct metadata format completion of a collection-level metadata record and use of a metadata harvesting protocol). Once these are met, the collection can be harvested into the core collection management and discovery system for discovery through the library search interfaces.

Further Information

Further information is available on the metadata Web site at <http://www.dlese.org/Metadata> .

Appendix 5 – Reports about Presented Projects

1. Anderson, Kenneth M. – <i>METIS project</i> – UNIVERSITY OF COLORADO.....	53
2. Arko, Robert – <i>DLESE Community Review System</i> – COLUMBIA UNIVERSITY.....	54
3. Davis, Ethan – <i>THematic Real-Time Environmental Distributed Data (THREDDS)</i> – UCAR/UNIDATA.....	56
4. Downs, Robert R. – <i>Digital Collections Management</i> – COLUMBIA UNIVERSITY.....	57
5. Fox, Sean – <i>Science Education Resource Center</i> – CARLETON COLLEGE.....	59
6. Klaus, Christopher – <i>Atmospheric Visualization Collection (AVC)</i> – ARGONNE NATIONAL LABORATORY.....	60
7. Ludaescher, Bertram and Meertens, Chuck - <i>GEON: A Research Project to Create Cyberinfrastructure for the Geosciences</i> – UNIVERSITY OF CALIFORNIA AT SAN DIEGO AND UNAVCO, INC.	61
8. Olds, Shelley – <i>Earth & Space Science Education Collections (ESSEC)</i> – NASA.....	63
9. Pandya, Rajul – <i>Virtual Geophysical Exploration Environment (VGEE)</i> – UCAR/DPC.....	65
10. Quintana, Chris - <i>IdeaKeeper—Digital Library Services for Information Analysis and Synthesis</i> – UNIVERSITY OF MICHIGAN.....	66
11. Ruzek, Martin – <i>Journal of Earth System Science Education (JESSE)</i> – UNIVERSITIES SPACE RESEARCH ASSOCIATION.....	67
12. Shipman, Frank - <i>Using Spatial Hypertext as a Workspace for Digital Library Providers and Patrons</i> – TEXAS A&M UNIVERSITY.....	68
13. Shipman, Frank - <i>Design and Evaluation of Maintenance Tools for Distributed Digital Libraries</i> – TEXAS A&M UNIVERSITY.....	69
14. Sumner, Tamara - <i>Strand Maps as an Interactive Interface to NSDL Resources</i> – UNIVERSITY OF COLORADO.....	70
15. Warnock, Andrew – <i>Digital Water Education Library (DWEL)</i> – COLORADO STATE UNIVERSITY.....	71

Presenter Name: Kenneth M. Anderson
Institution: University of Colorado, Boulder
Email: kena@cs.colorado.edu

Project Name: A Lightweight, Flexible, and Web-Based Approach to Supporting Workflow in Digital Libraries

Project URL: <http://www-serl.cs.colorado.edu/metis/>

Overview of Project:

The METIS project is developing workflow technology designed for use in digital libraries by avoiding the assumptions made by traditional workflow systems. In particular, digital libraries have highly distributed sets of stakeholders who nevertheless must work together to perform shared activities. Hence, assumptions made by traditional workflow technology that all members of a workflow belong to the same organization, work in the same fashion, or have access to similar computing platforms are invalid. The METIS approach makes use of event-based workflows to support the distributed nature of digital library workflow and employs techniques to make the resulting technology lightweight, flexible, and integrated with the Web. The METIS conceptual framework defines an event as a named set of attribute value pairs. Constructs such as forks, loops, and sequences allow these events to be arranged in workflows. METIS also provides constructs that allow the participants of a workflow (such as users, groups, and roles) to be modeled. We have developed a proof-of-concept prototype that has been used to both model and execute a peer-review process from a real-world digital library. This prototype is now being deployed to a small set of digital libraries for additional evaluation.

Focus of the Presentation:

The purpose of the presentation was to introduce the workshop participants to the difficult issues surrounding the task of supporting workflow in digital libraries including why it is difficult to apply traditional workflow technology in these settings. The presentation included a demo of the METIS prototype to demonstrate its ability to create community-defined, shared events that can be arranged into workflows that support sequence, looping, and branching constructs. Workshop discussion and questions focused on the ability of METIS to handle a variety of workflow situations that occur in digital libraries, including whether METIS could support "exploratory behavior" of students searching/using the resources of a digital library; the idea being that a workflow could not be pre-defined for such a situation but that the ability to analyze the "event traces" of the student's actions at a later time would be very useful. As it stands, METIS can indeed support such a task although it would be pushing the limits of the current design. Based on this feedback, we intend to enhance the METIS architecture to more cleanly support such a task.

Key References:

- Kenneth M. Anderson, Aaron Andersen, Neet Wadhvani, and Laura M. Bartolo. (2003). METIS: Lightweight, Flexible, and Web-based Workflow Services for Digital Libraries. 2003 Joint Conference on Digital Libraries, Houston, TX.

Project Award Info: A Lightweight, Flexible, and Web-Based Approach to Supporting Workflow in Digital Libraries, NSF/DUE, NSDL Services Track, 0121460

Presenter Name: Robert Arko
Institution: Columbia University
Email: arko@ldeo.columbia.edu

Project Name: DLESE Community Review System (CRS)
Project URL: <http://crs.dlese.org>

Overview of Project:

A high quality, relevant, useful collection of Earth system science educational materials has been a highly desired characteristic of DLESE from its inception. For a collection which is community driven like DLESE, it is critical to have a variety of methods of evaluating and assessing the collection as a whole and the resources within it, to ensure the desired high quality. The DLESE Community Review System (CRS) is one such pathway, providing a mechanism for formal review of individual resources in the DLESE Broad Collection, which aims to select the “best” resources for inclusion in the DLESE Reviewed Collection. The criteria for admission to the Reviewed Collection are scientific accuracy, pedagogical effectiveness, ease of use for teacher and learner, quality of documentation, importance or significance of content, ability to motivate or inspire learners, and robustness as a digital resource. The CRS combines two types of reviews: reviews delivered via a Web-based recommendation engine from educators in the DLESE community who have used the resource in their classroom, and specialist reviews mediated by an Editorial Review Board. Development of the CRS is led by Kim Kastens (kastens@ldeo.columbia.edu) at Columbia University.

Focus of the Presentation:

The presentation was intended to provide an overview of the project, summary of progress to date, technical details of design and implementation, and future direction. The CRS Web-based review engine gathers feedback from three categories of commenters: non-educators; educators who examined a resource but did not use it in their own classroom; and educators who used a resource with real learners. Commenters may submit reviews via a *CRS* button on the resource itself; via a list of resources available for review from the CRS home page; or via the DLESE Discovery System (in DLESE v2.0). The results from educators who used a resource are cast as annotations which are linked to the resource and enrich it. These annotations (Scoring Rubrics, Teaching Tips, Challenging Situations, and Editor Summary) are stored in a relational database backend, and served to the community via a OAI provider. As of February 2003, 82 resources have entered the CRS and have a status of Nominated, Declined, Withdrawn, In Review By Community, In Review By Specialist, or Accepted. Our work goals for the next 6-9 months include replacing DLESE-IMS metadata (using XML DTDs) with the new ADN metadata (using XML schemas); replacing our custom *Review Status* XML records with fully specified and verified Annotation XML records; replacing our flat XML files with a live SQL-to-OAI bridge; installing an internal workflow system; developing rubrics to gather input from commenters who used the resource as learners rather than educators; and continuing to aggressively advertise and solicit reviewers.

Key References:

- Kastens, K.A., 2000, “A ‘Community-Review’ Mechanism to Build the Collection of the Digital Library for Earth System Education”, *EOS Transactions AGU*, v. 81, no. 19.
- Kastens, K.A. and J.C. Butler, 2001, “How To Identify the ‘Best’ Resources for the Reviewed Collection of the Digital Library for Earth System Education”, *Computers and Geosciences*, v. 27 (3), pp. 375-378.

Project Award Information:

Original: September 1, 2000 - August 31, 2003, EHR-DUE-NSDL-0085827 Collections Track, “Collaborative Project: To Gather, Document, Filter and Assess the Broad and Deep Collection of the Digital Library for Earth Systems Education (DLESE)”

Renewal: September 15, 2002 - August 31, 2004, EHR-DUE-NSDL-0226292 Collections Track,
“*Collaborative Project: To Enhance the Depth, Breadth, and Quality of the Collections of the Digital Library for Earth Systems Education (DLESE)*”

Presenter Name: Ethan Davis
Institution: UCAR/Unidata
Email: edavis@ucar.edu

Project Name: THematic Real-time Environmental Distributed Data Services (THREDDDS)
Project URL: <http://www.unidata.ucar.edu/projects/THREDDDS/>

Overview of Project:

The overarching goal of Unidata's Thematic Real-time Environmental Distributed Data Services (THREDDDS) is to integrate scientific data and tools into digital libraries and the Web, thus providing students, educators, and researchers with coherent access to a large collection of real-time and archived data sets from a variety of environmental data sources at a number of distributed server sites. Just as the World Wide Web and digital-library technologies have simplified the process of publishing and accessing multimedia documents, THREDDDS will provide needed infrastructure for publishing and accessing scientific data in a similarly convenient fashion.

Focus of the Presentation:

The focus of this presentation was a status report of the THREDDDS project and the framework being developed for dealing with scientific data sets. At present, we are focusing on data that is available on-line, e.g., via OPeNDAP, ADDE, or FTP. From that foundation, THREDDDS is adding a way for data providers (or educational materials developers) to communicate to a user an inventory of available data sets. The catalogs can then be used in various ways to allow users to search for the particular data they are interested in. There are two search and discovery issues on which THREDDDS is currently working. For large or quickly changing collections of data, THREDDDS is developing the Dataset Query Capabilities schema. The DQC allows data set providers to specify the set of queries (aimed at subsetting the collection) a user can make of a data set collection. For more general search and discovery, THREDDDS is focusing on enhancing the catalogs with metadata and developing software to allow digital libraries to harvest that metadata, e.g., using OAI-PMH. In particular, THREDDDS is working on tools to automatically extract metadata from the data sets referenced by catalogs.

Key References:

- Domenico, B., Caron, J., Davis, E., Kambic, R. and Nativi, S., 2002. Thematic Real-time Environmental Distributed Data Services (THREDDDS): Incorporating Real-time Environmental Data and Interactive Analysis Tools Into NSDL. *Journal of Digital Information*, <http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Domenico/>.

Project Award Information: NSF/DUE, NSDL (Collections Track), #0121623.

Presenters Name: Robert R. Downs
Institution: Columbia University
Email: rdowns@ciesin.columbia.edu

Project Name/Presentation Title: Digital Collections Management
Project URL: <http://www.ciesin.columbia.edu>

Overview of Projects:

The Center for International Earth Science Information Network (CIESIN) is a center within the Earth Institute of Columbia University and is located on the Lamont-Doherty Earth Observatory (LDEO) campus in Palisades, New York. In achieving its mission to advance understanding on human interactions in the environment, CIESIN actively practices digital collections management to archive and disseminate research-related information and online learning resources that are acquired and developed collaboratively for its programs and projects. CIESIN operates the Socioeconomic Data and Applications Center (SEDAC) for the National Aeronautics and Space Administration (NASA), and operates the US Global Change Research Information Office (GCRIO) for the US Climate Change Science Program and the US Global Change Research Program (USGCRP). CIESIN also has been designated the World Data Center for Human Interactions in the Environment by the International Council of Science (ICSU). Representing various disciplines of the natural, social, and information sciences, staff members of CIESIN cooperate with partners, within Columbia University, and with other organizations, to conduct research and improve interdisciplinary analysis and access to scientific and technical information by communities of users pursuing their scientific, decision-making, and educational objectives.

Focus of the Presentation:

The presentation prescribed a cooperative communities-centered approach to digital collections management. Collaborating with the communities of users informs the management of digital collections to continually serve their information needs (Downs & Chen, in press). Taking this perspective, digital collections management can facilitate both the development and operations of digital collections. Digital collections development activities create new collections to serve potential communities of users and continuously improve existing collections to improve services for current communities of users and meet their evolving needs. These activities involve identifying, acquiring, and creating new resources for existing and new online collections as well as cataloging these resources with descriptive metadata to facilitate their discovery and use. Collections development activities also include collections assessment and evaluation to improve practices for acquiring, creating, and describing digital resources.

Emphasizing digital collections operations efforts addresses the need to sustain digital collections for the long-term and to continually support access to collections by current and future communities of users. Digital collections operations activities include efforts to disseminate resources contained within the collections as well as efforts to maintain and preserve these resources for long-term access and use. Similarly, the services supporting access to and use of digital collections also need to be managed to foster improvement and provide long-term support to the communities using the collections. Preparation for the long-term management of digital resources also should be integrated with collections development activities (Downs, 2003).

Both digital collections development and operations activities are influenced by the nature of the digital collections services provided, which can be described as falling within the continuum that ranges from those providing archival services to those disseminating resources stored at distributed archives. Differences between managing archival and distributed collections were described for attaining acquisition, cataloging, dissemination, and maintenance goals. While archival collections manage the individual resources within their collections, quality assurance presents challenges for distributed

collections that do not archive these digital resources and, as a result, have less control over the resources disseminated from within their catalogs.

Key References:

- Downs, Robert R., and Robert S. Chen. (in press) Cooperative Design, Development, and Management of Interdisciplinary Data to Support the Global Environmental Change Research Community. In *Science and Technology Libraries* (Binghamton, NY: Haworth Press).
- Downs, R. R. (2003). Planning for Long-Term Management of Online Learning Resources: A Digital Library Approach. *Collected Presentations of the Eighteenth Computers in Libraries Conference*, March 12-14, 2003. (pp.84-89). Washington, DC: Information Today.

Project Award Information:

SEDAC is operated for NASA under contract NAS5-98162.

Presenter Name: Sean Fox

Institution: Science Education Resource Center: Carleton College

Email: sfox@carleton.edu

Project Name: NAGT On the Cutting Edge Professional Development Program, NSDL Using Data in the Classroom Portal, DLESE Community Issues and Groups: Teaching Quantitative Skills in the Geosciences

Project URL: <http://serc.carleton.edu>

Overview of Project:

The Science Education Resource Center works to enhance science education through workshops and Web sites that support faculty in improving their teaching. Projects include:

The NAGT/DLESE *On the Cutting Edge* project helps geoscience faculty stay up-to-date with both geoscience research and teaching methods. The workshop series and Web site combine to provide professional development opportunities, resources, and opportunities for faculty to interact on-line and in person with colleagues around the world who are focused on improving their teaching.

The *Using Data in the Classroom* project includes an NSDL specialized portal which supports using data in the classroom including learning resources, data access tools, and pedagogical information. Web site content derives from a series of workshops addressing how educational leaders are currently using data in the classroom, their vision for the future, and the implications of this information on the Core Integration team of NSDL, and NSDL overall.

The *Teaching Quantitative Skills in the Geosciences* Web site is an example of how DLESE can support geoscience educators in addressing an issue related to classroom instruction. To date this site explores the development of intellectual entry points into an education issue, linking pedagogic and teaching resources, integration of discussion opportunities with resources, and development of thematic collections for faculty and students.

Focus of the Presentation:

The focus of the presentation was an overview of SERC projects of interest to the DLESE community. This included the 3 projects mentioned above, the DLESE Community Issues and Groups site and two upcoming NSDL funded projects: the *Earth Exploration Toolbook*, and *Starting Point-Teaching Entry Level Geosciences*. A brief demonstration was given showing how the tools developed can be used in conjunction with the existing digital library infrastructure to manage and display cataloged resources within the pedagogic context provided by the Web sites.

Project Award Information: NSF-DUE: CCLI-ND 0127141, NSF-GEO: 0085600, NSF-DUE: CCLI 0235007, NSF-DUE: CCLI 0083251, NSF-DUE: NSDL-CI 0127298, NSF-DUE: NSDL 0226243, NSF-DUE: NSDL 0226199

Presenter Name: Chris Klaus
Institution: Argonne National Laboratory
Email: klaus@anl.gov

Project Name: Atmospheric Visualization Collection
Project URL: <http://www.nsdل.ارم.gov>

Overview of Project:

The intent of the Atmospheric Visualization Collection (AVC) is to enhance physical science education and research through visualization of atmospheric data. This collection includes an archive of atmospheric data images and educational material based on these images. By utilizing collaborative digital library tools, a growing user community assists in the development of this collection.

The data image archive focuses on the Atmospheric Radiation Measurement (ARM) program's Southern Great Plains (SGP) site, which has the largest collection of ground-based, remote-sensing atmospheric instruments in the world. Automated visualization tools create these data images for archival and real-time uses. ARM instrument mentors and scientists, as well as other researchers involved with the SGP site, review the visualization work to ensure scientific integrity.

A growing educational community develops the AVC educational material using the Lesson Plan Sandbox. The Lesson Plan Sandbox allows teachers to submit their lesson plans to share with others, to review other submitted lesson plans, and to improve existing lesson plans, while keeping a copy of the previous version.

The AVC currently averages 450 unique users per week, with visitors from over 64 countries. Workshops and conferences continue to disseminate the AVC to the educational and research communities.

Focus of the Presentation:

The presentation covered why producing digital library data collections is important, a view of the stakeholders involved in data collection development, and an overview of the Atmospheric Visualization Collection as a data collection.

Key References:

- Klaus, Christopher M. , K. Andrew, and T. McCollum, 2003. Teacher involvement in the NSDL. In print. *Knowledge Quest*.
- Klaus, Christopher M., 2002. The National Science Digital Library: Focusing on ARM Data Images for Education. *ARM Climate Education Update*. June. <http://www.arm.gov/docs/outreach/news/0602newsltr.pdf> (2 April 2003).
- Klaus, Christopher M., K. Andrew, and G.G. Mace, 2002. Atmospheric Visualization Collection: Developments in the NSDL. *Journal of Digital Information (JoDI)*. 2:(4,May). <http://jodi.ecs.soton.ac.uk/Articles/v02/i04/Klaus/> (2 April 2003).

Project Award Information: National Science Foundation, NSDL Program, Grant 0086225; Department of Energy, ARM Program, Funded under ARM Engineering

Presenter Name: Bertram Ludaescher

Institution: San Diego Supercomputer Center, U.C. San Diego

Email: ludaesch@sdsc.edu

Presenter Name: Charles Merteens

Institution: UNAVCO, University Corporation for Atmospheric Research

Email: chuckm@unavco.ucar.edu

Project Name: GEON: A Research Project to Create Cyberinfrastructure for the Geosciences

Project URL: <http://www.geongrid.org>

Overview of Project:

The Geosciences Network (GEON) is a large collaborative project building a cyber-infrastructure for the geosciences. GEON has its roots in a broad-based community effort to make numerous and very heterogeneous geoscience data sets available and interoperable. While many disciplinary geoscience database projects are already underway, GEON provides an overarching and standards-oriented Data Grid infrastructure, guaranteeing the interoperability of such databases across disciplinary and organizational boundaries. For example, the emerging Open Grid Services Architecture (OGSA) is based on a fabric of high-end networks, computing and storage systems, and includes protocols for authentication and authorization, data access, transformation, replication, and layers for data federation, discovery, and pipelined processing. In addition to employing such a generic Data Grid infrastructure, GEON also addressed a number of specific geoscience IT issues such as situating data sets in a 4D spatio-temporal context, as well as thematic and process contexts. Navigation and querying between seemingly unrelated data sets will be accomplished through formal representations of *glue knowledge*, i.e., ontologies provided by specialized teams of data mediation engineers having both domain expertise in the geosciences and in databases and knowledge representation (KRDB). Thus, GEON brings together computational grid infrastructure including cluster computing and scientific visualization, and data grid services for data handling and mediation of very large and/or very complex geoscience data sets.

Focus of the Presentation:

The presentation covered the following issues: A broad overview of the GEON project, an introduction to the data integration challenges faced by the Geoinformatics community, and a brief discussion of the overall GEON system architecture (Ludaescher). This was followed by a short introduction of GEON data sets, in particular from the test bed regions (Rocky Mountains and Mid-Atlantic) and a discussion of possible links to Earthscope data management concluded the presentation (Meertens).

The importance of scientific *concept spaces* and ontologies for making scientific data more interoperable was highlighted throughout the first part of the presentation. For example, ontologies can be seen as *smart metadata* which facilitate querying and integration of collections of data sets using various concept spaces. In addition to linking data via a 4D (spatiotemporal) reference, different ontologies can be used to provide various linkages to mediate and query across heterogeneous geoscience data: e.g., terminological or process ontologies for rock classification (composition, texture, fabric, genesis), geologic time, theme (e.g. plutons), etc. Parts of the GEON infrastructure have a functionality similar to that of digital libraries, and conversely, digital libraries such as DLESE can make scientific data and associated analysis functions available via the emerging GEON infrastructure. In this sense, GEON and DLESE can work together to provide digital libraries with scientific data and mediation technology.

A GEON concept space workshop focusing on Igneous Rocks was held at SDSC (March 26th-28th). The GEON All-Hands Meeting will be held from April 16th-18th at SDSC.

Key References:

- GEON: A Research Project to Create Cyberinfrastructure for the Geosciences, project description of NSF/ITR proposal, 2002. http://www.geon.org/docs/GEON_description.pdf
- GEON: Developing the Cyberinfrastructure for the Earth Sciences, A Workshop Report on Intrusive Igneous Rocks, Wilson Cycle and Concept Spaces, A.K. Sinha et al., GSA Today, to appear, 2003.
- Data Integration and Model-Based Mediation for Geoscience Data, B. Ludäscher, B. Brodaric, K. Lin, in Introduction to Geoinformatics and the Digital Earth, A.K. Sinha and C. Baru, editors, forthcoming, 2003.

Project Award Information: NSF/EAR: ITR Collaborative Research: GEON: A Research Project to Create Cyberinfrastructure for the Geosciences, #0225673, Oct. 2002.

Presenter Name: Shelley Olds

Institution: NASA Education Office/ NASA Earth Science Education Office/ SSAI

Email: shelley.olds@gsfc.nasa.gov

Project Name: NASA Earth and Space Science Education Collections (ESSEC)
(formerly Earth Science Education Digital Library, ESEDL)

Project URL: <http://jdl.gsfc.nasa.gov>

Overview of Project:

The education community desires easy access to the multitude of NASA education resources currently spread across thousands of Web site. To serve our community's needs, NASA has needed a digital library to link its resources together to a single Web-interface and to provide sufficient information about each resource for community members to make decisions about its appropriateness to their needs (descriptive metadata).

To solve these basic challenges, the NASA Earth Science Digital Library created a pilot system which links together distributed collection metadata and provides a multiple-field search interface to our community. This Web-based, decentralized Library is a multi-collection, multi-Center, and multi-Enterprise effort with multi-agency collaboration.

To minimize operational and maintenance costs, the backbone of the library utilizes JINI™ technology to join individual collections together in a peer-to-peer system with a shared communication network and supports OAI resource harvesting. Owner built collections are hosted in MySQL databases (a free, Open Source software) or in individual XML files. Individual collections are maintained by their owners and retain control of their collection holdings, further minimizing costs to the Library system while allowing for collection branding. The NASA ESSEC project is a collaborating partner with the DLESE Program Center.

Focus of the Presentation:

The ESSEC is a distributed library system using Java's peer-to-peer toolkit, JINI™ technology. This JINI-based technology uses two modules for brokering exchanges, a module on the library side (any library or Web interface) and a module on the collections and services side. JINI provides a stable mechanism for registration and notification across the Internet all through a Java API.

The modules broker the connection between the Library (and other library front ends) with services and OAI-enabled collections. Libraries approve the connection of the collections and services for use. Using the companion module, collections and services register their location and configuration information with the library-side module.

The brokerage between the library module with the collection/services module provides a number of automated tasks:

- If a collection or service changes its configuration information (its location), the library-module automatically sees the change and uses it.
- Instead of manually maintaining a list with the OAI collections/services and configuration information, this information is held by the library module and is automatically updated from collection/service module updates.
- The broker knows when collections or services are available or unavailable. When a service is not available, the broker issues a notification to the library when the service goes offline and again when it comes online.

- The library is notified when the collection has changed its metadata, allowing a library to be provided with up-to-date information without frequent polling.
- For security, ESSEC's module uses public-private keys to sign services (and collections), thereby ensuring the library is using a service/collection that has been approved. One future enhancement will be the ability to revoke a collection or service's access to the library if necessary.

Key References:

- http://see.gsfc.nasa.gov/ppt/15Edu_DigitalLibrary/AMS2003Final_Web/AMS2003_ESSECfinal.html
- http://see.gsfc.nasa.gov/ppt/15Edu_DigitalLibrary/JDLdemoHQ/JDLdemo.html

Project Award Info:

Funding Agency: NASA Learning Technology Program & NASA Earth Science Enterprise.

Presenter Name: Rajul Pandya
Institution: DLESE Program Center
Email: pandya@ucar.edu

Project Name: The Visual Geophysical Exploration Environment (VGEE)
Project URL: <http://dpc.ucar.edu/vgee/>

Overview of Project:

The VGEE is an online learning environment that helps undergraduate students, especially non-science majors, learn geoscience. The VGEE includes data tailored for student use, an inquiry-based curriculum to guide students, and a specially enhanced scientific visualization tool. Modifications to the visualization tool include a learner-centered interface and embedded concept models. Concept models are simple, interactive models that help students learn basic physics. In the visualization tool, concept models can be used to probe data. By helping students ‘see’ basic physics in real data, concept models can help students apply theoretical understandings to real geophysical phenomena. To assess the effectiveness of the VGEE in student learning, we compared two classes, one that used the VGEE and another that did not. The group that used the VGEE showed a greater increase in test scores between pre- and post- essay tests. The VGEE group showed the most improvement in the questions closely related to the VGEE curriculum, but they showed improvement across the curriculum.

Focus of the Presentation:

The presentation focused on the VGEE as a way to allow students and learners to discover usable data in the context of digital libraries. To do this, the VGEE builds on the THREDDS notions of a *compound document* that connects text, data and data tools. The power of compound documents lies in their ability to leverage text-based searching and cataloging tools for discovery of embedded data and tools. In the context of DLESE, this means learners and educators can discover the data and tools in the VGEE by discovering the curriculum that describes their use. Compound documents also provide a way of addressing one of the key barriers to the use of research data in educational contexts. Most environmental data are in highly specialized files and require complex tools for access and analysis. Further, these data are created by researchers and aren't well connected to specific learning objectives or educational contexts. By using data as part of integrated packages that include tools to use data as well as curricula to connect data to appropriate learning goals, compound documents are a vehicle for providing data, in usable forms, within digital libraries.

Key References:

- Pandya, Rajul E., Mary R. Marlino, and Ben Domenico, 2003. Finding and using data in educational digital libraries. *Joint Conference on Digital Libraries 2003 (JCDL 2003)*, Houston, TX, 23-28 May 2003. Institute of Electrical and Electronics Engineers, Inc. (IEEE) and the Association for Computing Machinery (ACM).
- Pandya, Rajul E., and D. Murray, 2002. A tool for exploring geophysical data: The VGEE-IDV. *Eos Trans.* 83:(47, Fall Meet. Suppl. Abstract ED71C-09).
- Pandya, Rajul E., C. Contrisciane, C. Seider, and J. Yoder, 2002. Using inquiry- and visualization-based curricula in real classrooms. *GSA 2002, Fall Meeting of the Geological Society of America*, Denver, CO. Geological Society of America (GSA).

Project Award Information: NSF/EHR, CCLI (Educational Material Development), #9972491

Presenter Name: Chris Quintana
Institution: University of Michigan
Email: quintana@umich.edu

Project Name: IdeaKeeper—Digital Library Services for Information Analysis and Synthesis
Project URL: <http://www.eecs.umich.edu/~quintana/ideakeeper.htm>

Overview of Project:

K-12 students engaging in substantive science inquiry need supportive tools for such challenging activity. The NSDL can give students access to a range of scientific information. However, learners also need support to help them *analyze* and *synthesize* information they find in libraries with respect to their scientific questions. Using our learner-centered design process, IdeaKeeper addresses these additional learner needs through:

1. Development of specialized scaffolded analysis and synthesis services for K-12 science learners. IdeaKeeper will have extensible scaffolds to support students in analyzing NSDL resources and synthesizing the information into scientific arguments. We will contribute IdeaKeeper to the NSDL for other learners to use and other researchers to extend with new functionalities and scaffolds.
2. Deployment of IdeaKeeper in middle school classrooms to assess its impact on learners and articulate possible new scaffolds and capabilities. We hope to not only describe the utility of IdeaKeeper, but also to extend scaffolding research for other related projects.

Thus IdeaKeeper can act as a guidepost to services that can more effectively support science inquiry. Not only will we try to identify what works with IdeaKeeper, we hope to illuminate new kinds of support or functionalities to consider for learners.

Focus of the Presentation:

The focus of the presentation was to introduce the goals and initial work for the IdeaKeeper project. The presentation was centered on the information needs of learners engaging in science inquiry, and some initial design directions for IdeaKeeper as we begin our software design. Looking at information needs for science learners, we outlined the fact that learners need to not only find different sources of information (i.e., the traditional digital library focus), but they also need to judge, read, comprehend, and synthesize the information they find to address the science questions they are investigating. The key areas of IdeaKeeper support are aimed at helping learners engage in more thoughtful, intentional searches. More importantly, IdeaKeeper hopes to include aspects from common reading strategies (e.g., Directed Reading) to help learners tie their current knowledge as they read library resources and give direction to their reading activities. Finally, IdeaKeeper will include tools to help learners organize their notes and thoughts so they can synthesize their different pieces of information into a final scientific argument. We are completing our initial software design and implementation as we aim for focus group software testing later in 2003.

Key References:

- Quintana, C., Krajcik, J., & Soloway, E. (2001). Exploring a Description and Methodology for Learner-Centered Design. In W. Heineke & L. Blasi (Eds.), *Methods of Evaluating Educational Technology* (Vol. 1). Greenwich, CT: Information Age Publishing.

Project Award Information: NSF/DUE, NSDL (Services Track), #0226241

Presenter Name: Martin Ruzek
Institution: Universities Space Research Association
Email: ruzek@usra.edu

Project Name: Journal of Earth System Science Education (JESSE)
Project URL: <http://jesse.usra.edu>

Overview of Project:

The Journal of Earth System Science Education is an interdisciplinary electronic journal aiming to foster the study of the Earth as a system and promote the development and exchange of interdisciplinary learning resources for formal and informal education. JESSE serves educators and students by publishing and providing ready electronic access to Earth system and global change science learning resources for the classroom and provides authors and creators with professional recognition through publication in a peer reviewed journal. The Journal will publish a wide ranging variety of electronic content, with minimal constraints on format, targeting undergraduate educators and students as the principal readership, expanding to a middle and high school audience as the journal matures. JESSE employs an open peer review process in which authors and reviewers discuss directly the acceptability of a resource for publication using a software tool called the Digital Document Discourse Environment. Reviewer comments and attribution are available with the resource upon acceptance for publication. JESSE also implements a moderated peer commentary capability where readers can comment on the use of a resource or make suggestions. Copyright of materials submitted remains with the author, granting JESSE the non-exclusive right to maintain a copy of the resource published on the JESSE Web server, ensuring long term access to the resource as reviewed.

Focus of the Presentation:

The presentation walked through the motivation for JESSE and its contribution to the Digital Library in general, as one of the pathways to the DLESE Reviewed Collection, and as a bridge between traditional and electronic publishing. JESSE also fills a niche for authors seeking to share their teaching and learning resources while receiving professional recognition in a more traditional journal publication. JESSE has yet to meet its initial goals of a fast turn around and review time, and needs continued focused editorial guidance to achieve this. The infrastructure of the Journal is in place with several resources that have gone through the review process. The presentation walked through the Web site. The Digital Document Discourse Environment (D3E) can be a bit daunting to new reviewers and may be part of the reason for less than full participation on reviews. The basic JESSE concept seems to have general support from authors—more time and energy are needed to exercise the system fully and revise procedures as necessary to gather a diverse and active community of contributors and reviewers. There will also be some benefit to an association with a professional society as JESSE proves itself of value to the educator community.

Key References:

- Johnson, D.R , Ruzek, M., Weatherley, J., “The Journal of Earth System Science Education: Peer Review for Digital Earth and Digital Library Content" *Eos Trans. AGU*, 82(20) Spring Meet. Suppl., S107, 2001
- Johnson, D.R , Ruzek, M., Ford, R., “Implementing an Electronic Peer Reviewed Journal of Earth System Science Education Resources (JESSE): Pathfinder for SMETE Resource Peer Review" *Eos Trans. AGU*, 81(48) Fall Meet. Suppl., F295, 2000
- Ruzek, M., D. R Johnson, M. Kalb “An Electronic Peer-reviewed Journal of Earth System Science Education" abstract and paper IGARSS 2000, Honolulu, HI (<http://www.usra.edu/esse/jesse/jesse2000/igarss2000.html>)

Project Award Information: NSF Grant EAR-9907764, DUE- 0085793

Presenter Name: Frank Shipman
Institution: Texas A&M University
Email: shipman@cs.tamu.edu

Project Name: Using Spatial Hypertext as a Workspace for Digital Library Providers and Patrons
Project URL: <http://www.csd.tamu.edu/VKB/>

Overview of Project:

This Targeted Research project is investigating the use of spatial hypertext by digital library patrons to build personal and shared annotated digital information spaces, and by digital library providers to organize, annotate, and maintain collections of digital information. Spatial hypertext is a class of information workspace in which users collect source materials as information objects in a set of two-dimensional spaces and imply attributes of and relationships between the materials via visual and spatial cues. The ease of expressing evolving interpretations makes spatial hypertext well suited for tasks where the task and materials (or the user's understanding of these) change over time. The PI and colleagues are extending an existing spatial hypertext system, the Visual Knowledge Builder, in the following areas: (1) suggestion-based methods supporting the incremental specification of metadata; (2) dialog generation for converting the visual interpretation that occurs in the workspace into metadata; and (3) history annotation, filtering, and editing for viewing how collections change over time. By expanding the means to create personal digital information spaces beyond textual modes into visual ones, this investigation promises to broaden the impact of the NSDL on users.

Focus of the Presentation:

The focus of the presentation was a brief history of how spatial hypertext emerged from navigational hypertext and a demonstration of the Visual Knowledge Builder (VKB). Navigational hypertexts of the 80's (e.g. Xerox's NoteCards) started including maps (graph visualizations) to show the hypertext reader where they were in the network of material. But, when people began authoring new hypertexts in these map-based hypertexts, often they would not include explicit links. They used spatial position and visual similarity to represent relationships between nodes. This practice is what spatial hypertext aims to support. VKB, a second-generation spatial hypertext, actively supports the visual organization of materials via a spatial parser and a series of suggestion agents. Materials from NSDL, the Web, and the desktop may be dragged into VKB and organized alongside notes and visual annotation. A version of VKB with embedded support for searching and organizing NSDL materials is planned for summer 2003. Questions should be sent to Frank Shipman at shipman@cs.tamu.edu.

Key References:

- F. Shipman, R. Airhart, H. Hsieh, P. Maloor, J.M. Moore, and D. Shah, "Visual and Spatial Communication and Task Organization in the Visual Knowledge Builder", Proceedings of the ACM 2001 GROUP Conference, 2001, pp. 260-269.
- F. Shipman, H. Hsieh, R. Airhart, P. Maloor, and J.M. Moore, "The Visual Knowledge Builder: A Second Generation Spatial Hypertext", Proceedings of the ACM Conference on Hypertext, 2001, pp. 113-122.
- F. Shipman and C. Marshall, "Spatial Hypertext: An Alternative to Navigational and Semantic Links", *ACM Computing Surveys Electronic Symposium*, Vol. 31, No. 4es (Dec. 1999).

Project Award Information: NSF/DUE, NSDL (Targeted Research Track), #0226321

Presenter Name: Frank Shipman
Institution: Texas A&M University
Email: shipman@cs.tamu.edu

Project Name: Design and Evaluation of Maintenance Tools for Distributed Digital Libraries
Project URL: <http://www.csdl.tamu.edu/walden/>

Overview of Project:

This targeted research project is investigating and developing tools and social protocols to make more feasible the management and maintenance of distributed digital library collections in which authors put material into the library and librarians (collection managers) organize and annotate it for the library patrons. While such "author-based" approaches enable a digital library to grow rapidly, they can create confusion as resources are added, deleted, or changed without warning. In the case of Web sites that are pointed to, noticing when those Web sites go away, are rehosted, or change their underlying structure is very time consuming. Additionally, when changes are made to the content of the resources, the collection manager must decide if the new version is still suitable for the collection and, if so, whether the document needs to be recategorized. Towards the goal of improving the ability of collection managers to maintain distributed digital libraries this project is developing: i) algorithms and heuristics for identifying resources no longer available; ii) methods for identifying the relocation of resources; iii) methods for categorizing and evaluating the significance of changes to resources; and iv) tools supporting social mechanisms (between resource authors, library managers, and library patrons) to contend with document changes. The project plan includes the evaluation of identification and categorization algorithms based on technical and social metrics. These evaluations answer whether the algorithms correctly identify network and server problems, whether resources that have been moved are successfully located, and whether ratings of significance of change match those of human evaluators.

Focus of the Presentation:

The focus of the presentation was an overview and status report on the Walden's Paths Path Manager project. The problem being addressed is identifying materials that are changing in a distributed digital library. The Path Manager uses a set of heuristics based on a study of how people perceive and evaluate changes to Web pages to flag materials that should be reviewed by a collection manager. The tool is available for download on the projects Web site and will work on any list of URLs, including the bookmark list for your browser or URL lists of materials in any other form of personal or public collections. Contact Frank Shipman at shipman@cs.tamu.edu with any questions on how to get or use this tool.

Key References:

- L. Francisco-Revilla, F. Shipman, U. Karadkar, R. Furuta, and A. Arora, "Perception of Content, Structure, and Presentation Changes in Web-based Hypertext", Proceedings of the ACM Conference on Hypertext, 2001, pp. 205-214.
- L. Francisco-Revilla, F. Shipman, R. Furuta, U. Karadkar, and A. Arora, "Managing Change on the Web", Proceedings of the ACM and IEEE Joint Conference on Digital Libraries, 2001, pp. 67-76.

Project Award Information: NSF/DUE, NSDL (Targeted Research Track), #0121527

Presenter Name: Tamara Sumner
Institution: University of Colorado at Boulder
Email: sumner@colorado.edu

Project Name: Strand Maps as an Interactive Interface to NSDL Resources
Project URL: <http://swiki.dpc.ucar.edu/StrandMapService>

Overview of Project:

Conceptual browsing interfaces can help educators and learners to locate and use learning resources in educational digital libraries; in particular, resources that are aligned with nationally-recognized learning goals. Towards this end, we are developing a Strand Map Library Service, based on the maps published in the Atlas of Science Literacy by the American Association for the Advancement of Science (AAAS). Strand maps are visual representations of learning goals that provide compact functional groupings around topics important to science literacy (e.g., weather and climate). These groupings provide an overview of K-12 learning goals, or benchmarks, for a particular topic organized into ‘strands’ reflecting key ideas within that grouping (e.g., heat, water cycle, atmosphere). Rather than building static presentations of existing maps, we are creating a service middleware capable of generating visualizations of strand maps and map components from a benchmark database. When complete, this service middleware will interoperate with NSDL discovery systems to locate resources aligned with specific benchmarks or learning goals. Thus, the Strand Map Service will include two public interfaces: (1) a graphical user interface for use by teachers and learners and (2) a programmatic interface that enables developers to construct conceptual browsing interfaces based on the strand maps using dynamically generated components.

Focus of the Presentation:

The focus of the presentation was a status report on the Strand Map service project, which began in January 2003. Our development efforts have focused on two key activities: (1) developing the content standard for modeling the benchmark and strand map knowledge space, and (2) identifying, designing, and evaluating promising interface components for constructing useful and usable conceptual browsing interfaces. To date, we have developed and evaluated four interfaces comprised of a suite of interconnected components. In doing so, we have identified a number of promising components, such as the Cluster Navigator, the Map Tree Navigator, and the Benchmark Extended Information Viewer. We have used cognitive walkthrough analyses and think-aloud tests with teachers and library designers to evaluate these interfaces and components. A participatory design workshop for library developers interested in using the strand map service is being planned for fall 2003. Projects interested in participating in this workshop should contact Tamara Sumner.

Key References:

- Tamara Sumner, Mike Wright, Mary Marlino, Greg Janee, Ted Willard, and Francis Molina (2002). Strand Maps as an Interactive Interface to NSDL Resources, Distribution version of NSDL Services Proposal.
- Tamara Sumner, Sonal Bhushan, Faisal Ahmad, and Qianyi Gu (2003). Designing a Language for Creating Conceptual Browsing Interfaces for Digital Libraries, To appear in Proceedings of JCDL 2003, Houston TX (May 23-28).
- Sonal Bhushan, Tamara Sumner, Susan Hendrix, Gabe Johnson, Sreeraman Vaidyanathan (2002). Elen: Conceptual Browsing Interfaces for Educational Digital Libraries, Poster presented at ECDL 2002, Rome IT (Sept 16-18).

Project Award Information: NSF/DUE, NSDL (Services Track), #0226286

Presenter Name: Andrew Warnock
Institution: Colorado State University
Email: warnock@csmate.colostate.edu

Project Name: Digital Water Education Library (DWEL)
Project URL: dwel.dlese.org

Overview of Project:

The Center for Science, Mathematics, and Technology Education (CSMATE) at Colorado State University in partnership with the National Science Teachers Association (NSTA), and the Center for Lifelong Learning and Design at the University of Colorado at Boulder are developing an exemplary collection of digital K-12 materials and resources built around the theme of "Water in the Earth System". Our purpose in creating this collection is to enhance the ability of K-12 teachers, students, and parents to easily find, access, and use high-quality, standards-based water resources in their classrooms, at home, and in informal learning environments. The primary goals of the DWEL project are to (a) create a collection of approximately 500 "exemplary" K-12 water resources (scientific, economic, and policy) that can be used to investigate and learn about important water concepts, processes, and issues, (b) develop and test a model for the adaptation, enhancement, inclusion, and support of existing curriculum resources in the Digital Library for Earth Science Education (DLESE), and (c) conduct research on collaborative collection processes critical to building high-quality, user-friendly K-12 collections. To achieve these goals, exemplary water resources, and the associated tools and services needed to easily use these resources are being identified, reviewed, and cataloged by a variety of scientists and educators. K-12 teachers are leading this development effort. In creating the DWEL K-12 Collection, we are working closely with a number of organizations including: federal and state government agencies, professional scientific and education societies, academic institutions and centers, and other National SMETE Digital Library projects such as the one at Cornell.

Focus of the Presentation:

The challenge of managing the workflow of teachers and scientists distributed across the United States has been met with the rapid in-house development of a Web-based tool called the DWEL Work Hub (<http://129.82.204.180/dwel/workhub.html>). This presentation focused on the features of the Work Hub that were developed to address the specific needs of working group members as they began building the DWEL collection. The work hub has evolved into a well-tuned system for collecting and reviewing digital resources that can be easily adapted to other digital library collection efforts.

Project Award Information: This project was funded by the National Science Foundation's National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) program (#DUE-0121724).

Appendix 6: Glossary

Items in *italic* are terms that are defined elsewhere within the Glossary.

ADL

[Alexandria Digital Library \(ADL\)](#), a project of the [University of California at Santa Barbara](#).

ADN

[ADL - DLESE - NASA](#), a partnership between the *ADL* project, *DLESE*, and NASA's ESSEC - the NASA Earth and Space Science Education Collections, for the development of a common and extensible *metadata framework* known as ADN, an *XML schema-based framework* with strong data typing and controlled vocabulary support. The partnership leverages the ideas, technology, architecture, and communities of all three organizations in order to strengthen and unite digital library efforts in the geosciences.

API

Application programming interface, a set of routines, protocols, and tools for building software applications.

Attribute

Describes additional information about an *XML* element. An example is `<price currency="Euro">`. Currency is the attribute.

Broad Collection

All resources in DLESE meet a basic set of standards, outlined in the [Collections Policy](#). By definition, these resources are: Web-based; have relevance to Earth system education; are basically bug-free; and are available at little or no cost.

Cataloging

The process of providing *metadata* information for a resource that enables discovery.

Categories

The first order division of *metadata fields* for the *DLESE* metadata framework. Categories are made up of *fields*.

Collection

Generically, this means a group of *metadata* records that are organized around a theme, or some other criteria. For specific types of collections within DLESE, please see the individual entries for *Broad Collection*, *Reviewed Collection*, *formal collection*, and *informal collection*.

Collection-Level Metadata

Metadata that describes the overall characteristics of a grouping of item-level *metadata* records. This information includes who maintains the collection, how many resources are in the collection, what type of resources are in the collection and what is the scope of the collection.

Collections Development

A process conducted over time that builds and shapes a collection of materials into a balanced, cohesive, and sought-after set of user resources. This process includes assessing the information needs of users, analyzing usage statistics and demographic projections, formulating and articulating selection criteria, planning for resource sharing, creating a well-defined cataloging plan, and creating a selection and deselection mechanism for library items.

Controlled Vocabularies

Words or phrases that catalogers use to complete *metadata* information. Use of common words and phrases ensures better searching capabilities for library users.

Crosswalk

A semantic or technical mapping (sometimes both) of one *metadata framework* to another *metadata framework*.

CRS

DLESE [Community Review System](#), the *Reviewed collection of DLESE* (an NSDL Collections project).

DC

Dublin Core - the full name and acronym are [Dublin Core Metadata Initiative](#) (DCMI). The Dublin Core *metadata* element set is a commonly used standard for cross-domain information resource description.

DCS

An abbreviation for the [DLESE Catalog System](#).

DDS

An abbreviation for the DLESE Discovery System.

DLESE

[Digital Library for Earth System Education](#)

DLESE Catalog System

A Web-based tool for generating, editing, managing, discovering, and sharing *metadata* records.

DLESE IMS

The current *DLESE metadata framework* that uses an *XML DTD* and has a foundation in *IMS*.

DLESE Resource Cataloger

The Web-based tool for entering *required metadata*. It is one component of the *DLESE Catalog System*.

DPC

[DLESE Program Center](#), charged with building and maintaining the core library infrastructure and with supporting community development and integration of core library services.

DTD

Document Type Definition file that specifies how elements inside an *XML* document should relate to each other. It provides "grammar" rules for an *XML* document and each of its elements. DLESE's *metadata* records are *XML* documents.

Earth and Space Science Education Collections (ESSEC)

A NASA collection; also a DLESE partner in the development of the *ADN* metadata framework.

Element

The smallest division within an *XML* document that is defined within a *DTD* or *schema*. An example is `<body>formatted text</body>`. Body is the element.

Fields

The smallest division within a *metadata framework*. *Fields* become *elements* in a *DTD* or *schema*. *Fields* become *tag sets* within *XML* documents when the *field* is surrounded by a "<" and a ">" sign.

Formal collection

A recognized collection within the *DLESE* discovery system. This means that information records output in search results will indicate a relationship to a particular collection. Please see [Contributing to DLESE: A Reference Guide](#) for further information.

Framework

The systematic format and technical structure that supports *metadata* concepts, contents, and *controlled vocabularies*. For *DLESE*, the systematic format is a variation of the *IMS framework* and the technical structure is *XML*.

FTP

File Transfer Protocol. The protocol used on the Internet for exchanging files. FTP uses the Internet's TCP/IP protocols to enable data transfer. FTP is most commonly used to download a file from a server using the Internet or to upload a file to a server (e.g., uploading a Web page file to a server).

Gazetteer

A kind of dictionary that supports translation between place-names and coordinates, enabling or promoting georeferenced discovery of resources.

IDD

Internet Data Distribution, a system for disseminating near real-time earth observations via the Internet, developed by the Unidata Program Center of the University Corporation for Atmospheric Research (UCAR).

IMS

One of many different *metadata* schemes that exist. The abbreviation no longer stands for anything. The IMS Project is part of the non-profit EDUCAUSE consortium of U.S. institutions of higher education and their vendor partners that work to develop open market-based standards for online learning, including specifications for learning content metadata.

Informal collections

Collections discoverable in the *DLESE* discovery system through a single overview *metadata* record. This *metadata* record is part of the *Broad Collection*. Please see the [Guidelines for Contributing Resources](#) for further information.

Interoperability

The ability to share (provide and harvest) *metadata* records via standard protocols.

Java Remote Method Invocation (RMI)

A protocol enabling a programmer to create “distributed Java technology-based applications to remote Java technology-based applications in which the methods of remote Java objects can be invoked from other Java virtual machines, possibly on different hosts.”

JINI

Jini Network Technology is an open software architecture developed by Sun Microsystems enabling the creation of adaptable and flexible networking solutions.

LDM

The Unidata Local Data Manager, a collection of cooperating programs that select, capture, manage, and distribute meteorological data products, developed by the Unidata Program Center of the University Corporation for Atmospheric Research (UCAR).

Metadata

Descriptive information (e.g. title, description, audience, geospatial coverage, keywords) that can be used to describe, index, and discover learning resources for particular user needs.

NASA

National Aeronautics and Space Administration

NSDL

National Science Digital Library for science, technology, engineering, and mathematics (STEM) disciplines.

OAI

An abbreviation for the [Open Archives Initiative](#), which develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content.

OAI-PMH

Open Archives Initiative-Protocol for Metadata Harvesting - an application-independent *interoperability framework*.

Object-Level Metadata

Metadata that generally describes a single item or object.

Open Digital Library (ODL) and Open Digital Library Protocol (ODLP)

Open Digital Libraries (ODLs) are systems built as networks of extended [Open Archives](#). Adopting the notions of simplicity and reusability from the Open Archives Initiative, ODL adds extensibility and componentization to the effort.

OPeNDAP

Open-source Project for a Network Data Access Protocol, the name of an organization, as well as the protocol the organization developed, providing “a discipline-neutral means of requesting and providing data across the World Wide Web.”

ReadMe

A small text file that comes with many software packages and contains information not included in the official documentation. Typically, readme files may contain late-breaking information that could not be included in the printed documentation.

Required Metadata

The minimum level of *metadata* information that enables searching and indexing of resources within *DLESE*. Some examples of required metadata include: title, description, audience, technical information, copyright etc. Every resource suggested to *DLESE* would include this information.

Reviewed Collection

Materials that pass the filters to be in the *Broad Collection* and have been reviewed to ensure that the following criteria stated in the scope statement are met: well documented; pedagogically effective; easy to use; scientifically accurate; bug-free; foster mastery of significant understandings or skill

motivational for learners; are classroom tested.

Robust Metadata

Includes all *metadata* above and beyond required *metadata*. Allows for *DLESE* community-specific searching. Examples include the National Science Education Standards, related resources, and geospatial coverage.

Schema

An *XML* document that describes the *DLESE metadata framework* in terms of structure, data types, number of *field* occurrences, and *controlled vocabularies*.

SDLIP

[Simple Digital Library Interoperability Protocol](#). A protocol to facilitate query and response between clients and servers. Clients use SDLIP to request searches to be performed over information sources. The resulting documents are returned synchronously, or they are streamed from service to client as they become available.

Shibboleth

A project of [Internet2/MACE](#), developing “architectures, policy structures, practical technologies, and an open source implementation to support inter-institutional sharing of Web resources subject to access controls.”

SOAP

Simple Object-Access Protocol, an *XML*-based lightweight protocol for information exchange in distributed environments.

STEM

Science, technology, engineering, and mathematics education; a common abbreviation for these combined disciplines.

Swiki

Swiki is a popular implementation of Ward Cunningham's [WikiWikiWeb](#) (Squeak + Wiki = Swiki; *wiki-wiki* is Hawaiian for "quick"). Swiki may also be referred to as CoWeb, short for Collaborative Website. Swiki is implemented by [Mark Guzdial](#)'s [Collaborative Software Laboratory](#) at [Georgia Tech](#). It is a collaborative work area and Web page modifiable by its users.

Tag Set

An element defined in a *DTD* or a field enclosed in a greater-than and a less-than sign. An example is <body>. The element, body, is defined in the *DTD*. Body may also be the field name within the metadata scheme (but not necessarily). The tag set is then "<body></body>".

Valid or Validity

An *XML* document that adheres to the specifications outlined in the *DTD*. This generally refers to how an *element* can occur, the name of the *element* and the number of times the *element* can occur.

Well-formed

An *XML* document that adheres to the following *XML* syntax rules:

Uses a *DTD* file or contains an *XML* declaration with the "standalone" attribute set to "no." *Element* attributes are in quotes. *Elements* have an opening and closing tag unless it is an empty *tag set*. *Tag sets* nest correctly.

XML - The abbreviation for the eXtensible Markup Language. The XML language is a document processing standard that allows one to create and format document markups.