

Context and background information on Data Access issues in DLESE

In February 2001, 27 self-identified members of the DLESE Community met as the Data Access Working Group (DAWG) to discuss how they could facilitate access to data for education. The goals of the meeting were to

1. determine the issues surrounding incorporation of datasets as a learning tool,
2. to develop a vision for how the DLESE and Earth system science communities can work together to reach this goal, and
3. to develop a six-month action plan.

Overall goals of the DAWG included

- Facilitate discovery across distributed data archives
- Provide tools to help instructors and learners, parse, process, and visualize datasets
- Facilitate the integration of seemingly disparate datasets, and
- Facilitate the development and dissemination of educational content that utilizes datasets and datastreams

Recommendations produced by the 2001 DAWG Meeting

Approaches for short-term implementation

1. Identify collections that we would initially like to include in the DLESE collection and then address issues specific to these collections.
2. Identify a few use cases to frame the educational content and the datasets that should first be exploited.
3. Identify the breadth of efforts currently underway and enable communication and collaboration between those groups.
4. Demonstration project linking a few distributed Earth data representing datasets from ocean, atmosphere, solid Earth, social and life sciences with a basic tool kit. Constructed and evaluated within one or more educational contexts.
 - a. (possible projects are locally based - how has the climate changed in my city? What is the seismicity/geology of my state?)
 - b. (possible projects are of a more general nature. What are the best examples of a particular example? Small projects done by individuals)
5. Linking with related groups (non-Earth scientists, mathematicians, physicists) and commonly used tool developers who would like access to Earth data in their presentations.
6. Demonstration focused on a small region showing in-depth development of tools and data access.
7. Develop automated package to construct datasets and learning resources around timely Earth events.
8. Developing underlying support structure for end-to-end delivery and use of datasets.
9. Enhance metadata framework for datasets and catalog existing datasets and learning resources using datasets.
10. Research on effective use of datasets in the classroom.

11. Encouraging the development of resources and services supporting end-to-end delivery.
12. Plan for sustainability and maintenance.
13. Explore dataspace and develop demonstration project that exercises a variety of data types.

To gain early success:

- Pick a few common data-set types and tools
- Offer a few representative data sets
- Augment discovery with a gazetteer
- Ensure end-to-end effectiveness
- Place/feature specific data discovery
- Tool specific data discovery
- Data specific tool discovery
- Application-based viewing/analysis (thick client)
- Browser-based data viewing/analysis (thin client)

To aim for long-term success:

- Address key challenges
- Persistent identifiers
- Personalized data collections
- Middleware matching tools and data types
- Scalability limits on human-generated metadata

Follow emerging standards/patterns

- Support embedding DLESE data discovery/access in applications and into educational materials
- Conceptualize and support chains of data derivations/transformations
- Link strategically to research needs (remote, real-time, etc.)
- Improved data access (remote, real-time, etc.)
- Facilitated data use (i.e., enhance usage metadata)
- Improved data discovery

Subsequent Developments

To address some of these recommendations, the THREDDS (Thematic Realtime Environmental Data Distributed Services) project was funded by NSDL in 2001.

<http://my.unidata.ucar.edu/content/software/thredds/index.html>

The central role of THREDDS is to provide middleware that enables interactions between data tools and services.

The Earth Exploration Toolbook, a collection of step-by-step guides for using data in education was funded by NSDL in 2002. Five chapters are currently available at

<http://serc.carleton.edu/eet>

In April 2002, NSDL sponsored an interdisciplinary workshop on Using Data in Undergraduate Science Classrooms. The workshop Web site and report appears at <http://serc.carleton.edu/usingdata>

The report includes examples, activities, and scenarios for using data.

In February 2003, a group of educational resource developers participated in the first DLESE Developer's Workshop. The "**Bringing Data into the Library**" subgroup at that meeting offered the following observations and recommendations in the DLESE Developer's Workshop 2003 Report. The full, finalized report is available at http://www.dlese.org/libdev/workshops/2003_Dev/Dev_Wkshop_Report2003.pdf

Observations and Recommendations from the DLESE Developer's Workshop 2003

...[The group held] 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, we felt that 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 dataset 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 of them (example: the Grid Replication service).
- Tools are needed to help students build instrumentation and share data with each other. Example: 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 the standard way that allows you 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).

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. The vision for this section of the library includes 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, 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 (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 (or at least looking at the results of the *Using Data in the Classroom* NSDL-sponsored workshops) to help better understand the needs and realities of working in K-12 classrooms.