Quality Assurance and Maintenance of the DLESE Community Collection

Report to Quality Workshop, 30 June 2003

The DLESE Catalog System (DCS) allows staff to view community contributed records in all stages of completion and passage through to accessioning. This allows us to help catalogers as they are working on records, as well as review and edit records prior to making them accessible in the library. This ensures that all records have required metadata, that the resources in the library are within the scope of DLESE, and that they are and remain functional. It is a manual, human mediated process.

The DCS provides a mechanism to set the status of records, thus identifying their location in the path towards accessioning. One status is called "holding". This bin contains records that have been submitted to the library, but are not accessioned at the present time. These records may be there for a variety of reasons. They are:

1. They did not meet initial selection criteria for DLESE
2. They are duplicate records (e.g. a mirror site)
3. They have gone offline permanently
4. They have gone offline temporarily
5. They have changed significantly such that they no longer meet DLESE selection criteria

Unfortunately, the current system does not allow for tracking the reason why a record ends up in "holding". We have periodically maintained an external document to attempt this, but clearly, with the rapid growth planned, we will need a more automated system. Investigation into this is currently underway. We do, however, periodically (e.g. when the incoming queue of new records is little to none) go back and check on records to see if any might be resurrected and can be re-accessioned. Currently there are 324 records in this bin. Most are off-line or broken in some way.

A. Item Quality Assurance

The current DLESE Broad selection criteria are:

> -Relevant to Earth system science education
> -Technically functioning electronic resource

On occasion we receive items that do not meet these criteria. In these cases, items are held back from accessioning. In the case of the few odd items that are clearly not science-related (see below) the contributor is contacted and informed that the item is not of acceptable content. In the other cases a discussion with the cataloger ensues and a decision is reached as to whether to accession the item or not. Contributions from the community at large are small (approx. 12%), and issues with those contributions have primarily been ones of proper cataloging protocol not resource selection. Because the majority of the DCC has been contributed by NSF-funded efforts at AGI and MSU (64% and 19% respectively) we have had the most discussion about issues of resource appropriateness with those catalogers and their collectors.

Here are a few examples of the types of things that have not been accessioned because they did not meet the DLESE selection criteria:

1. Not relevant to Earth system science education
   a. out of scope science
      http://superstringtheory.com/index.html
      The Official String Theory site
      http://www.genome.gov/
      Human Genome Project
      http://livefromcern.web.cern.ch/livefromcern/antimatter/index.html
      Antimatter: Mirror of the Universe

   b. not science
      http://www.rkamstra.demon.nl/index.html
      Dogs Worldwide Photo Album
      http://www.moscow-lady.com
      Moscow Brides
      http://www.ask.com/
      AskJeeves
      http://url.unavailable
      Science Fiction Writing

2. Not technically robust
   a. broken links in activities that prevent use
      http://www.oar.noaa.gov/k12/html/ocean_currents.html
      Ocean Currents
   b. fatal loop in "Return to" link
      http://seaborg.nmu.edu/earth/default.html
      Earth History resources
   c. site under construction
      Glaciation in the Sawtooth Mountains (never finished, site now gone)
      http://wapi.isu.edu/Field_Exercise/Glaciers/glacier_index.htm

3. Other issues include:
   a. granularity issues:
      Search results pages from larger site
      AskERIC: Ecology Lesson plans
      http://askeric.org/cgi-bin/lessons.cgi/Science/Ecology/index.html
   b. no direct access to material
      Mapquest Orthophotos
      http://www.mapquest.com/maps/main.adp

B. Linkchecking and On-going Maintenance

Over the past 2 years, we have been checking links in the accessioned collection on a regular basis. This turns out to be a challenging process, as many resources may be down just transiently and do not warrant immediate response. It is also difficult to find or build reliable software to perform this task. The following is our process, with occasional lapses as new software is transitioned to and refined.

　　　1. Pass over all URLs in the library twice a day, maintaining history of accessibility over time.
　　　2. Reports are generated on Mondays and Thursdays, listing all records with access or other issues.
　　　3. Resources that have been inaccessible consistently for 3 days or longer are flagged for follow-up.
　　　4. Follow-up includes:
　　　　　a. If report indicates a redirect, replace with current URL if content has not changed such that description no longer matches or resource is now outside of scope.
　　　　　b. Independent investigation to discover new URL
　　　　　c. Email to creator (if known) requesting new location or status of site (for DCC collection only)
　　　　　d. Make changes as indicated by creator, or if no response after 3-5 days, place record in the holding directory

　　　Examples:
- All 23 of the Newton's Apple sites were retired by PBS
- The Park Service and some other government sites had security breaches last year, such that approx. 100 records went off-line suddenly. When resurrected, many had new URL's assigned. All needed to be temporarily removed, and then updated.
- A set of 3 sites changed URLs. When investigated, we discovered it had changed to a fee-based system, such that resources were removed from public access.
  http://www.biology4kids.com/index.html

C. The Grey Area: What is in the library now that is of questionable quality, but current process could not clearly exclude?

A number of items have been contributed that have sparked discussion about other possible selection criteria for DLESE. These items are currently in the library, though in the case of large suites of objects, some remain in holding pending the outcome of this workshop. Examples of these include:

    1. Distracting or inappropriate advertising

        How Stuff Works, 30 records
        http://www.howstuffworks.com/flood.htm

        Solarviews
        http://www.solarviews.com/eng/homepage.htm
        at one point, had ads for surveillance equip, prescription drugs, last week, a pop-up ad for anti-pop-up software!

        WorldAtlas.com
        http://www.graphicmaps.com/aatlas/**world**.htm
            multiple popup ads accumulate as you click through
        Weather.com
        http://www.weather.com/education/index.html
            illegal drugs info, mortgage rates, match.com

    2. Biased science
        Global Warming
        http://www.**globalwarming.org**/index.htm
        Kentucky Coal Education
        http://www.coaleducation.org/

    3. Learning materials absent
        Fuel Economy for Cars (emissions data)
        http://www.fueleconomy.gov/index.htm

D. Current Source of Records in the DLESE Community Collection

Other Community 12%

DPC 5%

MSU 19%

AGI 64%

3693 total resources