Dave Mogk's Comments on the DLESE Collections
Letter to the Committee

I'm sorry that I can't join you, but Kim has accepted my offer to provide some written comments about the DLESE collections, having served as the DLESE Collections Coordinator for the first two years of DLESE.

1. An Historical Note: The DLESE "broad" collection was originally created as a developmental collection to test the cataloging system, metadata, and search systems. The original charge was to populate the collection with resources that covered a) the range of topics that encompass the Earth system, the range of resources that could be included (e.g. visualizations, activities, lab exercises, images…), and the intended audiences (e.g. (elementary, secondary, undergraduate, informal…)

Similarly, the topics that are currently used in the "browse" system were installed to help facilitate the review of the holdings of the collection by providing histograms that demonstrated number of resources by topic, audience and types of resources. There was considerable discussion among interested colleagues regarding the 30 or so topics that were identified—these were somewhat controversial, but consensus was achieved to settle on the current terms so that cataloging and testing of the discovery system could proceed.

The point here is that the original broad collection and supporting browse structures were created to facilitate the early DEVELOPMENT of DLESE systems. Once established, they have *de facto* become the operational systems. That is, the original systems are not necessarily designed to be optimal for general use as a functional, operational digital library. To the best of my knowledge, there has never been a comprehensive review of these systems, either in usability studies, or in comparison with other digital libraries, to determine whether or not the initial design of these systems can scale up to more universal use. Further, the browse structure was developed to allow direct comparison of the library holdings to an idealized profile of library holdings (as explained to me by Barb de Felice). I am not aware that the profile of the model library, defined by community needs and expectations, has ever been fully described. So, it is difficult to determine how close the current collections track to this idealized standard.

2. In defense of a "broad collection". There can be no doubt that the science education community has repeatedly called for access to high-quality, reviewed instructional resources. This has been true as articulated in a series of reports, starting with the visionary Shaping the Future of SMET Education (NSF 96-139), Shaping the Future of Undergraduate Earth Science Education (AGU, 1997), the DLESE Community Plan (Manduca and Mogk, 2000). The broad collection was originally envisioned as

- A general collection of resources that were to be contributed by the community for a wide variety of unspecified uses
- A pool of resources from which the "reviewed " collection could emerge

- A place where colleagues could post works in progress (e.g. "homegrown" resources), hopefully to receive some formative assessment to improve the product.

Arguments for a relatively open broad collection include the following

- The composition of the broad collection does indeed meet standards: these are the initial filters of relevance to the Earth system and some assurance that the resource works.
- The initial ~3500 resources in the broad collection have almost entirely come from credentialed sources. Most of the initial cataloging has been done at Montana State University and by the AGI staff. Selection of these resources has come from a) a number of lists/collections from our early collections compatriots John Butler, Suzanne O'Connell, Mark Francek, all of whom have NSF funding to pursue their independent projects. Other resources were discovered by Chris di Leonardo's group. At MSU, I tasked graduate RAs to systematically search for materials in a sequence of topic areas.
- All learning is contextual, and there is no way to predict whether or not a given resource will be useful in a particular learning environment, for a given student population, in a specific geographic setting, etc. So, I strongly believe that it is essential to have a broad base of resources available to support the creative abilities of our teaching colleagues—why put them in a mental straight jacket and dictate that only certain materials are worthy of being used?
- A libertarian view: I think I agree with most librarians that an open and free flow of information is essential. I am very troubled by the view that there should only be a very small collection of "quality" items in DLESE—quality according to whom? If you take the sum of all interests across the Earth sciences, and to meet needs of all audiences, I think you'll find that you're right back to a large broad collection. To prescribe a small set of resources as the "right" stuff, will necessarily be exclusive of large segments of our community. And philosophically, I find this point of view to be very Orwellian. We do not have a central "canon" of core content across the Earth Sciences, as perhaps physics or chemistry might have.
- What if an "offensive" resource is found in the broad collections? This issue came up with respect to a single instance of a URL from the oil/gas industry that attacked global change. Similarly, what about items from the Creation Science Institute? Well, I use their propaganda directly in my classes to expose them for the frauds they are—this approach isn't for everyone, but if you decide to take on the evolution/creationist debate in your class (even in Kansas), shouldn't you have access to the appropriate information. It is the nature of Science to engage controversies. DLESE should support this.
- However, I would not let this controversial material to be available in DLESE out of context. So, I would advocate rapid deployment of a) metadata records and b) annotation services that clearly define the caveat emptor: warning the user that his material is from a source with a particular point of view, and to capture the comments from the community of users of how this material

worked or didn't work in different learning settings. Further, if something is truly offensive (e.g. of a sexual or defamatory nature), as per any library, we have in place (I think, or should have) a deaccession policy.

- Finally, it must be realized that there are a great number of resources "out there" that have tremendous value to education, that do not have a specific educational application identified. For example, most of the federal sites such as USGS, EPA, NOAA, etc. have a lot of great materials—hydrograph and weather data, images, charts, etc. As above, we have no idea what a teacher might do with these resources when s/he finds them, but isn't the point of DLESE to help connect people and ideas? These resources would not meet the criteria required of the reviewed collections—but they do indeed have value. So, I would argue that the broad collection is the place for records of these URLs to reside. In reality, DLESE has a very small user base. I would recommend that we let the experiment play out into the future, and take a good hard look at what different groups of users are really using in DLESE. That should help formulate the next steps in collection development.

Some other comments about collections:

- Community-based cataloging is a failed experiment as far as I'm concerned. Even community-based recommendations of resources has only had limited success. To grow the collections, DLESE will really need to have a number of dedicated collection groups to populate the "holdings". Despite almost three years of outreach and advocacy, I could not find a lever big enough to get folks to take a few minutes to even recommend a resource, let alone try to catalog. So, plan to budget for an army of catalogers.
- Cataloging of individual resources into the broad collection is neither sustainable nor scalable. I firmly believe that brute force cataloging of individual resources must be replaced with strategic development of targeted thematic collections (more below).
- Overall, the DLESE search and discovery system does not provide much value to users beyond other web search engines, and there is no way that DLESE can compete with Google. This was the point of my last AGU talk (Mogk, AGU, 2002 Fall meeting). DLESE should be less about collections and much more about services. We can find plenty of information on the open web. What DLESE needs to deliver are the Community Review system, annotation systems, linkages across science, pedagogy and related services (e.g. assessment and data toolkits…)
- Thematic collections—the original design laid out in the Community Plan was to enable interest groups across the E. Sciences to nucleate, define their needs, aggregate their resources, etc. So far, most of the funding and effort has been focused on the centralized DPC. There simply has not been enough nurturing or encouragement of community groups from paleontology to extreme weather to the deepest oceans to form and develop their own thematic collections. Yes, some are just now being considered for inclusion as a DLESE collection. But this is at least a year too late. The deadline for

inclusion this year was set by the DPC for mid June—this should be a regular, ongoing process whenever a group is ready. The process that determines which collections are "in" or "out" is very vague. My understanding is that applications would be sent to the Collections Committee and that upon further advice from DPC technical staff that a recommendation would be made to the Steering Committee for their approval. I'm further concerned that earlier efforts to include other collections have not been pursued: Dex Perkins had GeoEd funding to develop the Geoscience Digital Image Library—this was never integrated into DLESE due to technical difficulties. But for a project that was funded with the intent to integrate with DLESE should not have been dismissed because of technical difficulties (i.e. the question was how to ingest 3000 + image records on one hand vs. how to categorize an entire collection as a single record on the other hand, I think was the issue). Similarly, I worked to get the National Academy Press collections from the Earth and Life Sciences Board integrated into DLESE. NAP was motivated to make their database compatible with DLESE and the NSDL, but I have not heard that progress has been made in the past year since these discussions were initiated.

- My point in the above is that there will be many types of collections that emerge from the community—GeoDIL and NAP are just two examples. The DPC must be responsive to a wide range of community-based collections that represent a wide range of interests and consequently, technical challenges. I would suggest that decisions about whether or not to devote technical resources to help develop these collections should be made at a high level—starting with the Management Team and perhaps ultimately seeking advice from the Steering Committee. Whether or not a collection is included in DLESE is an important decision, particularly when groups have devoted significant time and energy in their development, and in recognition of the technical support that might be needed. These decisions should not be made at the DPC staff level.
- Finally, if it is decided to pursue a collection development policy that will seek out thematic or disciplinary collections, there is one other "quality" issue to be dealt with: what will you do if a given collection has some resources that are deemed to be "objectionable". For example, I'm working on life in extreme environments—there is a lot of E. Sci in the materials I'm developing, but there is also genomic information. What will the policy be if I apply for inclusion in the DLESE collections, and it's decided that microbial genomics is outside of the interests of DLESE? Am I in or out? I think you run into real problems if you presume to pick and choose among records that are part of an otherwise integrated and coherent collection.

I'm sorry I couldn't attend this meeting, and I ran out of time to write a more coherent letter—this is basically a dump off the top of my head. I hope that some of the points resonate and will be useful to the group.

Best to all,
Dave Mogk