

## **Networking Issues For**

### **A Digital Library of the Middle East**

**Alexandria, Egypt, January 15-17, 2006**

by Dale Smith, University of Oregon

Digital libraries are an enticing way of extending the reach and impact of libraries. An eloquent description of this vision can be found in the United States President's Technology Advisory Committee February 2001 report "Digital Libraries: Universal Access to Human Knowledge" <http://www.nitrd.gov/pubs/pitac/pitac-dl-9feb01.pdf>:

All citizens anywhere anytime can use any Internet-connected digital device to search all of human knowledge. Via the Internet, they can access knowledge in digital collections created by traditional libraries, museums, archives, universities, government agencies, specialized organizations, and even individuals around the world. These new libraries offer digital versions of traditional library, museum, and archive holdings, including text, documents, video, sound, and images. But they also provide powerful new technological capabilities that enable users to refine their inquiries, analyze the results, and change the form of the information to interact with it....

Very-high-speed networks enable groups of digital library users to work collaboratively, communicate with each other about their findings, and use simulation environments, remote scientific instruments, and streaming audio and video. No matter where the digital information resides physically, sophisticated search software can find it and present it to the user. In this vision, no classroom, group, or person is ever isolated from the world's greatest knowledge resources.

As we work towards implementing such a vision, there are very important questions that must be addressed. I do not have library background, but the core challenges in implementing a digital library seem to include:

1. Digital Preservation. The issues here include agreement on formats of digital data, conversion from outdated formats to new ones, and maintaining technology systems. This issue of preservation is hard enough with simple digital data such as JPEG and PDF, but as the digital data gets to be more complex such as databases and custom software systems, this problem will grow to be very hard to solve.
2. Meta Data. This seems like a difficult issue. The vision of having a collaborative effort for cataloging really emphasizes that there must be agreement ahead of time about how meta data is going to be used, what the source vocabulary and language for the meta data will be

3. Cataloging. How do we prioritize and select the order of items to catalog? As we digitize holdings, do we provide research-quality images/data/measurements, or do we provide lower quality holdings?
4. Standards. Standards for digital library archiving as well as the digital formats are a moving target and continue to evolve and change at a rapid rate. This makes the preservation issue more difficult.

If we don't address these issues, particularly the meta data one, the digital collections will be disjoint, making it very difficult to provide a unified view and seamless experience to the library user.

My area of expertise is network engineering, so my contribution to this effort will be focused on the technical networking aspects of building such a distributed digital repository. To address network design and engineering, we need to answer the following types of questions:

1. Collaborators. Who are the collaborators that will be contributing to the development of the unified digital holdings?

The collaborators will be working closely on many of the issues I raised in the first part of my discussion (meta data, preservation, etc.). As this work is done in a distributed computing environment, it is very important that we provide systems for authentication (is this collaborator who they claim they are) and authorization (what rights does this collaborator have to change and modify the holdings). There is significant work in this area of distributed authentication and authorization systems, including Shibboleth which has emerged from the United States Internet2 project. We must have agreement on how we are going to provide such a distributed authentication and authorization system.

To provide a seamless unified digital repository, all collaborators will need to have access to a common high speed network. Much work will need to be done to determine appropriate and cost effective ways to accomplish this. We may find that we can get all collaborating entities attached to the research and education networks in each country, then connect those networks together with high speed cross border links.

2. User base. Are we targeting researchers or the general public for the digital holdings? How are these users connected to the Internet?

The identification of the target audience will drive issues such as the quality (and size) of the digital holdings, but we need to understand who these users are so we can provide good connectivity to the networks that provide service to those users. For example, if the primary target audience is the world wide general public, then the networks that service the digital repositories need to have connections to a variety of global Internet service providers. On the other hand, if the target audience is the global research community, then we should be developing high

speed links to the various research and education networks scattered around the globe (GEANT in Europe, Internet2 in the US, CANARIE in Canada, etc.).

3. Bandwidth. What kind of digital information will be in the repositories? How big is it? Is it presented in ways that require special networking characteristics (delay, jitter, bursty bandwidth)? Does it refer to other digital collections elsewhere? What are the characteristics of those digital holdings?

My guess is that much of the digital holdings will be oriented toward the research community, which implies to me that the data will be quite large (high resolution digital photographs, 3 dimension images, high resolution x-rays and cat scans, etc.). This means that the bandwidth requirements for both the library collaborators and the research user will be large. It also implies that the connectivity discussion in the user base item above will be critical to provide sufficient and affordable bandwidth for both the libraries and the user base.

The core networking issue that needs to be addressed really can be simplified to identifying where the network traffic will be anticipated, providing links to handle that capacity, and configuring the network equipment to utilize those links once they are in place.