Final Report for

US – Korea Joint Workshop on Digital Libraries

August 10-11, 2000
San Diego Supercomputer Center
San Diego, California
http://fox.cs.vt.edu/UKJWDL

Edward A. Fox
Reagan Moore
Ronald Larsen
Myaeng Sung Hyon

April 11, 2003

1. Executive Summary

There are many barriers to the worldwide development of digital libraries, and the involvement of the US in such activities. These are of particular concern in the context of digital library support of collaboration on research and education between pairs of nations with very different languages and cultures, such as the US and Korea. We report on recommendations to remove such barriers. The recommendations were developed originally through a workshop involving digital library researchers from the US and Korea who met August 10-11, 2000 at the San Diego Supercomputer Center. These recommendations have been refined through a variety of subsequent meetings and discussions. They are presented early in 2003, when peace, economic development, and cross-cultural understanding are of particular importance with regard to US relations with the Korean Peninsula – and since various developments have occurred that finally make feasible both further productive collaboration and possible funding by both governments.

As can been seen from the above mentioned web site, extensive documentation of the workshop has been available online since August 2000, including:

- proposal to US National Science Foundation (NSF) for the workshop;
- agenda and other general information about the meeting;
- presentations given at the meeting (in PowerPoint and other formats);
- position statements of attendees (in PDF and other formats);
- reports of discussions: by working groups, breakout groups, and meeting organizers;
- draft of outlines and reports;
- pointer to Oct. 2002 D-Lib Magazine publication derived from this project;
- reports from visits to Korea in late 2002 by workshop participants Robert B. Allen and Ching-chih Chen

This final report was further edited after presentations and discussions in Seoul July 11-13, 2002 during the visit of PI Fox hosted by Professor Myaeng Sung Hyon. It was finalized after receipt of the last of the documentation items listed above.

Key recommendations, which are supported by the discussion below, include:

1. Since researchers in both US and Korea are involved in R&D related to digital libraries (DLs), there are real and potentially significant opportunities for collaboration in this area that should be nurtured.
2. Since education has high priority in Korea, and since in the last years there has been successful US-Korea collaboration regarding such educational resources as electronic theses and dissertations, an important opportunity to be explored is involvement of Koreans in the initiative led by the US National Science Foundation to develop the National STEM (Science, Technology, Engineering, and Mathematics) education Digital Library, NSDL (see NSF 02-054, http://www.nsf.gov/pubsys/ods/getpub.cfm?nsf02054, and NSF 03-530, http://www.nsf.gov/pubsys/ods/getpub.cfm?nsf03530, as well as www.nsdl.org).
3. Since NSF has supported work by US researchers through NSF 02-085 in connection with its Digital Libraries Initiative (DLI, see http://www.dli2.nsf.gov/), and in Spring 2002 launched a new International Digital Libraries Collaborative Research and Applications Testbeds Program Solicitation (NSF 02-085, http://www.nsf.gov/cgi-bin/getpub?nsf02085), US teams are encouraged to apply to such programs in concert with partners that may obtain support from similar programs in Korea.
4. Further focused programs for US-Korea collaboration on DLs should be developed and explored, especially if related to the following opportunities:
   a) Work on Korea Culture and Heritage Digital Libraries: has support in Korea; could provide an opportunity for US citizens (including students) to access Korean culture and heritage; can benefit from technologies and experience developed by US researchers; and could lead to important solutions to problems such as cross cultural, cross lingual, cross religion, and multidisciplinary interoperability.
   b) Advances in work on ontologies and on the various areas dealing with human languages (e.g., machine translation, cross-language information retrieval, text summarization) should allow DLs to support users in both nations.
   c) Application areas of particular importance, such as health and medical care, seem particularly amenable to US-Korea collaboration, especially with regard to special services such as annotation.
   d) A number of important DL problems, such as digital library architecture (including for managing data, information, and knowledge), interoperability, and user-centered design and evaluation, are of particular interest to US-Korea teams that may need new insights regarding solutions that cross national, linguistic, and cultural barriers.

2. Introduction

US-Korea collaboration on digital libraries (DLs) must be considered in context. The following subsections provide some background and motivation for this. They also may help readers to identify likely points to focus on, possible implications, and promising

supporting activities. Table 1 further establishes a context for this discussion by identifying application domains, institutions engaged in those applications, example activities, technical challenges, and possible benefits. The application domains column represents communities that are providing requirements that drive the development of digital library applications. The related institutions column lists representative groups that need the associated software systems. The examples column lists either types of software systems, or collections that are used by the institutions. The technical challenges column lists specific capabilities that are needed for a successful digital library. The benefit/impact column lists the broader results that accrue from the implementation of software to meet the technical challenges.

Table 1. Overview of Digital Library Applications, Benefits, and Challenges (where bold entries may have particular application in Korea)

Application Domain	Related Institutions	Examples	Technical Challenges	Benefit / Impact
Publishing	Publishers, Eprint archives	OAI	Quality control, openness	Aggregation, organization
Education	**Schools, colleges, universities**	**NSDL**, NCSTRL	Knowledge management, reusability	Access to data
Art, ***Culture***	Museum	AMICO, **PRDLA**	Digitization, describing, cataloging	Global understanding
Science	Government, Academia, Commerce	NVO, PDG, SwissProt, *UK eScience, European Union Commission*	Data models	reproducibility, faster reuse, faster advance
(e) Government	*Government Agencies* (all levels)	Census	Intellectual property rights, privacy, **multi-national**	*Accountability*, homeland security
(e) Commerce, (e) Industry	Legal institutions	Court cases, patents	Developing standards	Standardization, ***economic development***
History, **Heritage**	Foundations	American Memory	Content, context, interpretation	Long term view, perspective, documentation, recording, facilitating, interpretation, understanding
Cross-cutting	Library, Archive	Web, personal collections	***Multi-language***, preservation, scalability, interoperability, dynamic behavior, workflow, sustainability, ontologies, distributed data, infrastructure	Reduced cost, increased access, preservation, democratization, leveling, peace, competitiveness

Many of the technical challenges are cross-cutting, in that they impact almost all of the application domains, and will result in basic technology that will be important across all related institutions. In this sense, the cross-cutting issues can be considered the highest priority challenges for exploration within the context of a US-Korea joint development project. Also noteworthy are the bold entries, which may be of particular relevance to the growth of digital libraries in Korea.

2.1 Background

The US-Korean workshop was organized to identify joint research and development activities that would enable the development of multi-cultural digital libraries. The emphasis was on software, systems, applications, content collections, community needs, testbeds, and other matters. The challenge was to identify where technology and applications that were unique to Korea could be used to drive the development of new digital library technology. From the perspective of the US National Science Foundation, the workshop presented an opportunity to promote international collaboration, and to increase the general applicability of digital library technology.

The workshop was organized collaboratively between Korea and the United States through a joint program committee. Dr. Edward A. Fox led the US participation on the program committee, and proposed researchers within the US digital library community that were working on relevant technology. A representative sample of researchers who were dealing with multi-cultural issues, including multi-lingual digital library access, were selected, as were a number already engaged in or considering collaboration. Each researcher who was invited was asked to submit a white paper describing their research and associated technology. The selected researchers then were asked to give a short presentation at the workshop.

The workshop was hosted at the San Diego Supercomputer Center by Dr. Reagan W. Moore on August 10-11, 2000. The agenda for the workshop is posted at the workshop web site.

2.2 Motivation and context

The motivation for the workshop was to identify mutually beneficial applications that would justify either digital library research or the application of digital libraries to new communities. The presentations that were given at the workshop were evenly split between US researchers and Korean researchers, providing ample opportunity to present relevant technologies from both communities. In addition, the presentations were grouped into common themes to promote interactions between groups doing similar research. One of the outcomes from the workshop was a better understanding of the driving motivations for the development of digital library technology. From the Korean community, the importance of cultural heritage was strongly emphasized. From the US community, the use of digital libraries to further education and promote publication of data was clearly shown. These synergistic motivations led to the identification of cultural heritage education as a driver for new digital library applications. Work in this area

should not only lead to good science, but should have broad and positive impact on US and Korean societies, leading to improved cross-cultural understanding.

2.2.1 Strong Strategic Ties between US and Korea

The opportunity to build interoperable systems was recognized as an important mechanism for promoting strategic ties between the US and Korea, which have a long history of military and economic cooperation. In particular, the ability of persons of Korean descent within the US to access Korean cultural material resident in digital libraries within Korea was a strong motivation.

The development of interoperable systems requires the exchange of technology between the US and Korea. Technology to support a cultural education digital library will need to come from both communities, and must needs arise from, as well as lead to, extensive discussion (e.g., regarding protocols and data interchange standards) and cooperation. There was a common desire to not divide the problem into an application specification by researchers in Korea, and a technology specification by researchers in the US. Application drivers and viable software technology need to be provided jointly by researchers in both the US and Korea if interoperability is to be achieved as a result of what thus must be a successful collaboration.

2.2.2 Strong Academic Relationships

The optimism for success in this endeavor was driven by the strong academic relationships between researchers in the US and Korea. Many faculty at Korean universities received training in the US. Further, there are many joint conferences, researcher exchange visits, and presentations at both academic institutions and the San Diego Supercomputer Center. What is needed to strengthen these relationships is a set of well-defined workshops that alternate between the two countries. Alternatively, the exchange of meeting sites for appropriate conferences will facilitate academic collaborations. Finally, a sabbatical program could be established to promote strong academic relationships. (Note, for example, that after the workshop, two Korean researchers interested in digital libraries came to work with Professor Fox during their sabbatical. Further, in Spring 2002 a delegation from Chungnam National University, led by its President, came to Virginia Tech to sign an MOU for academic collaboration.).

2.2.3 Strong Economic Relationships

The development of strong academic relationships also can be fostered by improving economic relationships through jointly sponsored research. Joint projects can be used to facilitate information exchange, leading to interoperable systems. Since the relevant communities in Table 1 include both commercial as well as federal entities, it is clear that the development of interoperable technology can facilitate commerce.

Indeed the listed benefits from joint digital library technology development can lead to improved information access within each culture, as well as between cultures.

Information management and access is driving innovation within business communities, and revolutionizing the services that can be provided by governments. Facilitating information exchange on the technologies needed to manage information is a fundamental step in promoting strong economic relationships.

2.3 Regional Focus / Global Implications

The regional focus of a joint US-Korea development effort will have global implications. Extending systems developed in the US to handle Korean texts can help lead to their being able to work with Chinese and Japanese documents too. The technology needed to promote multi-lingual and multi-cultural holdings can be applied across all countries.

A technology development model that is used successfully to coordinate US-Korean joint research also can be applied to other countries, facilitating the spread of information management technology. While this discussion focuses on the explicit opportunities for joint projects between the US and Korea, we expect the process to be applicable to other countries and application areas. Thus the results of these collaborations can have a global impact.

2.4 Emerging Collaborative Activities: The Pacific Rim Digital Library Alliance

Digital libraries provide a mechanism to assemble collections that span multiple countries, enabling access to material that would otherwise be inaccessible. An example of such collaboration is the Pacific Rim Digital Library Alliance (PRDLA). The members of the alliance constitute institutions from the nations that surround the Pacific Rim, including a Korean university library. Each institution contributes collections, software infrastructure, or hardware support systems – to build a distributed digital library. One of the goals is to assemble the largest collection of Chinese text in the world by integrating access to multiple existing collections. Note that intellectual property rights restrictions limit access to some collections (but digital library management of intellectual property rights is an important precursor for establishing an information industry that engages the corporate sector). The following collections under PRDLA are operational:

- Bibliography of East Asian Studies provides an index to Western language journal articles on East Asian studies. The index development is led by Stanford University. To access this database, point your browser to http://libraries.ucsd.edu/prl/ and follow the "Bibliography of East Asian Studies" link.

- Beijing SuperStar Digital Library provides scanned images of some 80,000 titles of Chinese digital books on many subjects from China. Some of the books are mirrored in a repository at the San Diego Supercomputer Center, with the entire collection located in China. The collaboration includes Beijing SuperStar - a major vendor of Chinese digital books in China. To access this database, point your browser to prl.sdsc.edu and follow the "Beijing SuperStar Digital Library" link.

- The electronic version of Wenyuange Sikuquanshu (the Wenyuan Library collection from the Qing Palace) contains full-text and scanned images of some 3,700 titles. The collaboration includes a vendor of digital books in Hong Kong – Digital Heritage. Access to this database requires vendor supplied client software that is licensed to each university

The US-Korean joint workshop on digital libraries pointed to the need to extend these collaborations to include cultural information from Korea. The PRDLA shows that the policy and cooperative agreements needed to build a multi-national library can be implemented to further communication between nations.

Yet, PRDLA is not the only context for collaboration. Thus, the Networked Digital Library of Theses and Dissertations, which emerged as a result of US government funding beginning in 1996, has stimulated interest in electronic theses and dissertations around the globe, including in Australia, China, Germany, and India. Two government agencies in Korea, KISTI and KERIS, are engaged in projects that have led to collections already containing tens of thousands of such works.

There are opportunities for many additional collaborations between the US and Korea. Some will emerge naturally, while others, such as those that build on broad-based university volunteer efforts, are likely only to develop if initially encouraged by government support. High-level discussions, such as those undertaken in the US by the President's Information Technology Advisory Committee (PITAC), may be needed to identify the most important opportunities to explore.

2.5 Digital Library Overview Diagram

The opportunities that are provided by digital library technology are pervasive, affecting all organizations that need to manage data and information. As was explained briefly above at the start of Section 2, the wide range of opportunities that are enabled by digital libraries are characterized in Table 1, a digital library overview diagram. Seven different application domains are considered, with an eighth domain that represents cross-cutting issues. For each application domain, the principal related institutions and current example implementations are identified. Then an important technical challenge is listed, and important benefits and impacts from the use of digital library technology are shown. The three columns on the left of the matrix represent the types of opportunities. The two columns on the right of the matrix provide a cross section of the challenges and impacts. We note that the technical challenges may be encountered for any of the application domains, while most of the benefits apply to all of the application domains.

There are many ways to interpret the matrix. One can pick an application area and identify the important research areas (technical challenges). Or one can pick an institution that funds the development or application of digital library technology, and identify the important motivating benefits. Finally, one can pick an example and decide if comparable technology would benefit similar organizations within your country.

The challenges and benefits listed here demonstrate the advantages that arise from the organization of data within digital libraries. In the following sections, we discuss some of the technical challenges in more detail.

3. Technical Directions and Challenges for Global Digital Libraries

The majority of the discussions at the workshop focused on the availability of technology to overcome current challenges in digital library applications. These discussions were focused by selection of appropriate themes for organizing the multiple presentations. Applications were used to define new requirements for digital libraries, and scalability issues were selected to choose between software technology implementations. The goal is to identify an approach that can lead to a global method for the sharing and distribution of information. This requires the development of architectures that sustain access by residents of multiple countries, mechanisms to manage distributed repositories, and mechanisms to operate systems that span multiple countries.

3.1 Scalable Content and Collections

A dominant requirement for a digital library that supports access by residents of multiple countries is the ability to scale in capacity to manage arbitrarily large access rates. We note that events that cross cultural boundaries such as World Cup Soccer, or the Summer Olympics, or the landing of the Rover on Mars, can generate web accesses from all countries. The ability to give everyone equal access requires a degree of scalability that has not previously been attempted. Note that equal access implies use of the relevant language within the associated cultural context, and access rates that are not adversely impacted by network topologies (that arise because of geographic and economic concerns) or transcontinental latencies (that arise due to the finite speed of light). The scalability issues relate to many of the technical challenges listed in Table 1.

3.1.1 Globalization

Digital libraries extend the trend that 'going digital' encourages, with regard to application convergence, including across space and time. Thus, digital libraries must address the needs not only of libraries, but also of archives and museums, and on a global scale. Many will refuse to invest in the construction of a digital library if the concept of preservation, strongly associated with archives, is not fully addressed. Few users of digital libraries will understand their collections unless context is made clear, which occurs as a matter of course in museum exhibits. People searching in a digital library focused on a particular genre typically expect that all instances of that genre, from around the world, will be included: Why would a researcher in a scholarly discipline willingly ignore the relevant work, from any country?

Further, why would a researcher ignore relevant content, just because it was in the form of a sensor data stream? So, too, what serious scholar would complain if provided with relevant primary data, irregardless of its form – whether a satellite image or a photo of

ancient art, whether a CAD model or a 3D digitization of sculpture, whether GIS data or a digital representation of an ancient map? Capturing key data has always been fundamental to scholarship; in the context of digital libraries we refer to this as digitization, gathering metadata (including provenance and cataloging information), and recording of relationships. The latter, whether in the form of concept maps, or built into hyperbases that directly support navigation and linking, are crucial to the development of theories and models that capture the generalities that move us from data to information and knowledge – regarding art, culture, heritage, and history, as well as science and technology.

Modern scientific research in particular is dependent on amassing and integrating diverse related sources of data, information, and knowledge. In order to be collected, processed, and analyzed they must be represented and unified. Geographically connected data must be suitably registered to a consistent frame of reference. Measurements must have units reconciled, and terminology must be interpreted consistently. Only then can data be suitably modeled. Only then can complex dynamic systems (e.g., ecosystems) be described and understood. For example, the Intermedia digital library effort has brought together video data from around the Pacific Rim regarding El Niño, making clear that it had diverse effects in different regions: drought in parts of the US and China, but floods in regions of South America.

Primary data leads to datasets and formation of databases. Through analysis these lead to publication, at an increasingly rapid pace, that engages collaborating scholars from around the globe. All of the data, information, and knowledge that arise must be made accessible; replication often is involved to make that reliable and efficient. Long-term preservation is essential if the much more costly processes of scholarship are to bear fruit into the future; alas, this is often ignored by those providing funding. Yet digital libraries that have built-in mechanisms for preservation will assuredly facilitate re-use, as well as support education of the next generation of scholars.

Access to primary data, as well as the availability of powerful tools for analysis and visualization, and the encouragement of international collaboration, are beginning to transform education. Interactive learning is more effective than traditional approaches, and more fun – helping ensure that there will be adequate numbers of R&D workers who will continue to help solve the many crucial problems faced by humanity. Sharing and reuse of primary data is being followed by sharing of curriculum modules and educational resources. While a culture of sharing must be nurtured, so too must there be support for privacy and proprietary processes, as well as management of intellectual property, especially in highly competitive areas (e.g., bioinformatics and chemistry). Modern education is a global system that must be supported suitably by modern (digital) libraries, but they must be significantly extended to ensure scalability and the provision of a broad range of services.

Such need for global collections leads to particular challenges for digital library research and development. Distributed collections in today's internet environment lead to distributed systems, which must interoperate in order to be utilized in an integrated

fashion. The World Wide Web Consortium (W3C) has played a key role in encouraging standards for metadata annotation (e.g., XML). Activities like the Open Archives Initiative have established protocols that support data interchange. Yet without multi-lingual systems, and multi-national cooperation, global collections cannot be developed.

3.1.2 Lessons from Applications

When we look at the technical challenges that arise across each of the application domains, they can be divided into **technical**, **legal**, and **policy** challenges.

The **technical** challenges include quality control over the contents within the collections, re-purposing of collections to take advantage of their information content through knowledge management, the generation of digital images from paper or art holdings to promote network-based access, the cataloging of collections to facilitate discovery, and the specification of standard data models to ensure that everyone has an equal ability to display and manipulate the digital entities.

The **legal** challenges are formidable, and are centered upon the specification of intellectual property rights. How can material be shared within the interactive environment of the web, while control over the ownership of the material is properly maintained? A related issue is the management of privacy within web-accessible material. How can sensitive material be anonymized such that statistical analysis and data mining can proceed, while the rights of individuals are respected? The traditional example is the release of census data, which must take place without violating the rights of any individual.

When collections become multi-national, the legal challenges become more complex. The implication is that a joint collection will require a joint understanding of what constitutes intellectual property rights, and what constitutes private material. Such an understanding, however, may only arise if government support and project deadlines force interested parties to work diligently on hard problems, realizing that opportunities will be lost if challenges are not surmounted.

The **policy** challenges are related to the processes used to develop joint agreements on legal issues. Examples are the development of policies on the openness of the material, the conditions under which material may be re-used, and the mechanisms that will be used to provide context for a collection. Policy challenges are particularly important when specifying use conditions for restricted material. Equally challenging is the specification of the mechanisms that will be supported for presenting the material. With current technology, it is possible to expose relationships within collections that might be considered proprietary or private. Digital library technology is very powerful, and can provide mechanisms for query and discovery across multiple sites that allow relationships to be discovered that otherwise would remain hidden.

3.1.3 Overarching Challenges / Cross-Cutting Issues: Data, Information, and Knowledge

The overarching challenges listed in the table are identified as cross-cutting issues. These challenges include some of the major research issues related to digital libraries. We understand how digital libraries can be used to organize information. What is not yet apparent is how digital libraries can be used to organize knowledge across multiple cultures. We apply semantic tags to data entities to create information, and then look for relationships between the tagged elements to create knowledge. We manage the tagged data elements as attributes in databases. But we have no standard mechanisms for managing the relationships between the tagged elements. The purpose of creating collections of information is to facilitate the generation of knowledge. In cross cultural settings, the problem becomes even more difficult, as we differentiate between different cultural perspectives. The temporal or procedural relationships that are valid within one culture might be replaced with a different set of intermediate states in a second culture.

The management of knowledge is the overarching challenge for all digital library research. The relevant issues for cross cultural collaborations are the specification of ontologies for characterizing relationships between semantic terms, and the semantic interoperability mechanisms for deciding on common semantic terms.

The traditional cross cutting issues are related to economic sustainability of collections, technological sustainability of collections in the sense of preservation, and access sustainability in the sense of scalability across distributed data repositories. One would assume that a distributed digital library would require the use of common mechanisms for each of these sustainability issues. What is not apparent is whether it is possible for the global system to support multiple mechanisms for each of these issues, while providing interoperability mechanisms to unify access across the multiple choices. The emergence of grid technologies suggests that it may be possible to integrate all independent sustainability models within the global digital library. Data grid technology is an example of the federation of completely independent digital libraries into a common name space with common access mechanisms. This strongly suggests that it will be possible to create global digital libraries, in which differing technical, legal, and policy issues are overcome.

The development of a global digital library has unique requirements that are not present in current distributed digital library projects. While the architecture is substantially similar, the need to support multi-lingual interfaces and multi-lingual knowledge bases poses significant computer science research issues.

The simplest characterization of a digital library environment differentiates between the management environments required for digital objects (data), attributes about the digital objects (information), and relationships between the attributes (knowledge). Each environment requires explicit infrastructure support. Fortunately, storage of data has been facilitated by the development of data handling systems (e.g., SDSC Storage Resource Broker), and information management has been enabled by the acceptance of a

standard information markup language, XML. Management of knowledge has been a major research area for decades, but the recent integration of XML markup languages with Topic Maps for describing relationships promises to provide a suitable infrastructure. Ontologies now can be expressed as XML tagged relationships that associate topics. It is possible to create representations of ontologies that can be implemented on a variety of logic systems.

The challenge in managing knowledge is compounded by multi-lingual access. The concepts used in one language may not map directly into concepts used in another language. This means that even though the same domain knowledge is being described related to the same collection, a multi-lingual digital library will have to maintain multiple ontologies to describe the relationships used to organize the collection. Relevant research questions include whether it is possible to build a joint ontology, whether the mapping between the ontologies can be automated, and whether cross language information retrieval can facilitate the ontology mapping.

3.1.4 Transforming Education

Among the areas with greatest need for scalable support in terms of content and collection is education. Learning is a basic human need, so that we may understand our world and culture, and contribute to its improvement. Thus, education must be available to everyone, irregardless of language, culture, age, and prior experience. Advancing education proceeds only if we nurture communities of learners and teachers, facilitate their engagement in sharing activities, and have suitable educational resources as well as environments that facilitate meaningful learning. Digital libraries thus must provide a broad range of services related to learning. They also must support the collecting, organizing, and making available of suitable educational resources. Further, because of digital libraries we may improve those resources both in terms of quantity and quality.

Learners may utilize not only secondary materials like textbooks but also primary resources that heretofore were only available to scholars. Thus the entire set of resources that support research and scholarship, in addition to a broad diversity of novel resources including interactive multimedia developed by interested educators (or produced by peers as part of developing their own portfolios of constructive learning materials), can become available through digital libraries. With such a broad range of materials, automatic classification is essential to support discovery and retrieval, as well as routing to instructors eager to find the best materials for a particular situation or subject. We might say that every content item may help in education, presenting a significant challenge, especially when management of intellectual property rights is involved in conjunction with support of authentication and authorization. But this leads to a crucial additional challenge, which is to help learners gain access (directly, during self or distance learning activities, or indirectly, during teacher-directed activities) to the most appropriate items for their personal learning needs and objectives. When diverse languages and cultures are involved, and mappings must be made across those barriers so that resources may be properly understood, we may have what amounts to the grandest challenge for scalable digital libraries! An important instance of that is for learners in US to better understand

Korea, and for learners in Korea to better understand the US – real understanding, not just a boring section in a course on world history.

Transformation in educational practice is essential if the potential benefits of digital libraries are to achieve fruition. Support for sharing among innovators is essential, including handling of submission, testing, review, editing, annotation, and composition of new materials from a collection of selected resources. When production of the best materials is aided, and when they can be identified easily by busy educators or learners involved in an assignment or engaged in discovery, a radical improvement in meaningful learning may result. Such positive impact, however, requires an architecture supporting powerful, scalable, distributed, interoperable systems with effective operational management, and effective services for users.

3.2 Architectural Considerations

The architecture for a global digital library is simplified by considering the following infrastructure levels:

- Application/Configuration Layer: This level provides tools to help select digital objects for assembling a new digital library, as well as tools for defining context as concept spaces. Concept spaces are needed for both defining the level of expertise of the user, as well as the level of expertise of the collection.
- Multi-lingual Digital Library Services: Multiple services are provided by a global digital library, including presentation, translation, knowledge mining, information mining, and query support. Presentation services include tools for dynamically defining the user context and modifying the presentation context to be compatible. Views are needed to control user specified presentation versus collection specified presentation (such as choice of character set). Translation services provide an opportunity to develop correspondences between concept spaces that function at the level of a thesaurus. Knowledge mining provides tools for the identification of relationships. Information mining provides tools to seek out the existence of explicit relationships, such as content based information retrieval.
- Domain Knowledge Management: Ontologies manage concepts for the collection, and require explicit tools to facilitate their maintenance. Tools are needed to extend an ontology to include new concepts that describe common sets of information about the collection. Tools also are needed to support the inverse task, creation of metadata for concepts that are contained within the ontology, but not expressed within the collection metadata. Further, tools are needed to integrate existing ontology representations.
- Collection Management: The ability to build an infrastructure independent description of a metadata catalog is needed to support migration of collections between different types of database systems. Effectively, this consists of tools to support creation of schema, modification of schema, and publishing of schema for use by other applications. The collection then can be distributed across multiple information repositories, with local control over the digital objects, and access control lists used to protect intellectual property. A distributed collection can be

implemented across multiple administration domains, including internationally distributed sites.

- Data Handling System: When digital objects are distributed across multiple storage systems, data handling environments are needed to manage access. This is essential for integrating collections that are distributed across file systems, archives, and databases provided by multiple vendors. Distributed digital object management provides persistent identifiers, replicas of data sets, containers for aggregating digital objects before storage in archives, and archival storage interfaces.
- Storage System: The fundamental level of a global digital archive is the set of storage systems used to hold the data objects. The storage systems will be distributed internationally, with each site maintaining local control. A global digital library only becomes possible when the storage systems are integrated by a data handling system.

3.3 Distributed Repositories

Fundamental to modern scholarship is the notion of distributed collections. Each nation, state/province, town, university, department, center/group, and individual may run its own server, with collections of data, documents, and other resources. Building upon this base, digital libraries add unique types of support.

First, they allow sharing not only of various digital objects but also of information describing those objects. Such descriptive information is often called metadata. In many cases metadata is adequate for discovery of useful resources, and generally is more freely shared than are digital objects themselves. Through the Open Archives Initiative (OAI) Protocol for Metadata Harvesting, it is very simple for metadata to be shared, from personal collections (e.g., built using tools that run on laptops, such as Kepler, http://kepler.cs.odu.edu:8080/kepler/), to group or department or focused collections (e.g., using the eprints system – see www.eprints.org), on up to global repositories (e.g., those registered with OAI, see www.openarchives.org). The OAI concept is powerful. Scholars can easily share metadata, without also having to provide services (e.g., web sites with fancy interfaces, searching, browsing). Others who desire can use specialized tools to provide tailored services, after harvesting just the information really needed.

Second, distributed collections allow virtual documents to be constructed. A collage can be assembled from a group of images of art works. A homework assignment can be assembled by an instructor from a small set of problems, each made available as an educational resource. A report can be prepared by a student, drawing upon diverse materials, to argue a particular viewpoint, supported by data, images, and even video.

Further, virtual collections can be prepared. A portal might provide access to a virtual collection that is built up in turn from smaller subsidiary collections. Thus, the city of Berlin and the surrounding German province of Brandenburg worked together to establish a portal to all the hundreds of libraries in the region, which would act as if an integrated union collection had been established. Smaller virtual collections can be

assembled in a particular specialty area, or larger virtual collections may be built as national resources (e.g., see discussion below about NSDL).

3.4 Integration and Interoperability.

One way to assemble an global digital library is to integrate multiple existing libraries. As shown in the PRDLA, this can be achieved through multiple mechanisms. It is possible to integrate access at the:

- Data level through replication of resources. Each site then continues to provide their original services, but on a much broader set of primary source materials
- Access level through implementation of a data grid. As before, each site continues to provide their original services, but they are now able to access material that is managed and archived at another site.
- Service level through implementation of common web services. This makes it possible to apply a service developed at another site to your digital holdings on your site.

For these integration mechanisms to work across cultural and language domains, interoperability mechanisms are needed for both semantic and procedural knowledge. One can think of semantic interoperability as providing mechanisms to describe relationships between terms to define equivalent meaning. Procedural knowledge is the expression of the manipulation context that one expects to associate with a collection. This can be as simple as the organization of material for presentation for societies that read right-to-left, or bottom-to-top, or back-to-front. Procedural knowledge defines the manipulation context for the presentation of the material within a digital library.

One of the interesting challenges is that procedural knowledge is typically assumed by the style of display used within a portal for accessing the digital library, to conform to a cultural standard. It is interesting to note that the emerging support for handicapped persons is challenging the ideas of cultural standards for presentation of data and information. A fruitful area of research will be to understand how support for handicapped access addresses the same procedural knowledge issues that are encountered by multi-cultural access.

3.5 User Access

User access to digital libraries brings an additional layer of complexity to the definition of procedural knowledge. The challenge is to preserve not just the cultural context, but also the individual user perspective when exploring information. Users can define their preferred access context as a combination of their historical access patterns, their current interests, and their personal level of knowledge. In a multi-cultural setting, each access to a digital library should result in the presentation of a context that meets not only cultural expectations, but also personal preference expectations.

The area of digital library infrastructure where legal, policy, and technical issues most strongly interact is in the management of user access. The specification of what an

individual is allowed to access, the rights they have to reuse material, and the rights they have to remain anonymous represent possibly conflicting requirements. The resolution of these conflicts requires policy decisions for use of the digital library infrastructure. For multi-cultural situations this can become a "catch-22" situation. To correctly identify the rights of a person to anonymous access to material, the digital library may first have to identify the culture or organization through which the person is obtaining access. By representing individuals as members of a combination of cultural/organizational groups, it may be possible to decide which policies to invoke without having to explicitly identify an individual. These imply the ability to authenticate membership within an organization or cultural group. This type of identification is being pursued as part of the Shibboleth authentication environment, and is being promoted for use within digital libraries.

3.6 Operational Management

Digital libraries that span cultural systems also have substantial operational management concerns. Many of these issues can be cast in terms of policy management decisions, for the criteria under which actions will be taken. Operational management also includes management of user load, by redistribution of accesses to sites where resources are available to meet demand. This in turn implies the replication of data resources between sites to enable access load to be shifted. One possible value of US-Korea collaboration on digital libraries is to support such replication, where US resources that are of common interest are replicated in Korea, and vice versa. The mechanisms to manage replicas across sites are provided by data grid technology. Thus operational management for global digital libraries will be linked to the emerging global grid infrastructure.

An international effort has been underway for several years to develop consensus on the services and protocols that will be provided by global grids. The grid community meets at Global Grid Forums that are held three times per year, traditionally in the US or Europe. An effort has been made to plan for grid forum meetings in Asia, to promote a truly global grid environment.

The creation of global digital library infrastructure should carefully track the emerging grid technologies. The typical services include grid authentication mechanisms, support for remote job execution, support for characterizing grid resources, and data grids for building distributed data collections. These services will form the core operational framework for global digital libraries.

The second community that is providing relevant technology involves the standards efforts led by the World Wide Web consortium, ISO, and industrial consortia such as the Object Management Group. These communities are developing technologies to support semantic webs, network protocols, and model driven architectures for infrastructure independent representations of services. The hope is that the technologies developed by the digital library community, the grid community, and the standards groups will be able to interoperate, and create an infrastructure that minimizes the operational management requirements.

The development of a global digital library can be thought of as the creation of interoperability mechanisms for access to data, information, and knowledge. The competing approach is to rely upon standards for storage, schema definition, and ontology definition. The standards-based approach has been tried for large multi-national data collections, but requires a tremendous effort to reach consensus. The resulting system usually requires use of a single storage system (such as object-oriented databases), a common schema (with pre-specified lists of allowable attributes), and a common ontology (with explicitly defined relationships between the domain concepts and the sets of attributes used in the collections). The challenge is that all aspects of the digital library will evolve in time, with new concepts added to represent new findings, new attributes used to describe new data sets, and new types of storage systems created that are larger and faster. If a standards-based digital library is to persist, all of the components will have to be upgraded simultaneously. This will not be feasible in most global digital library efforts.

3.7 Sustainability of Infrastructure

By building an infrastructure that focuses on interoperability, new interfaces are built whenever technology evolves, and the system can continue to operate with both the original and evolved components. Standards are still important, but at the information markup language and knowledge representation level. The storage systems used, the contents of each collection within the digital library, the schema used to organize the collection, and the ontology used to represent relationships about the collection can be different. The interoperability mechanisms provide the ability to map between the different instantiations.

This approach is also needed to ensure persistence of the global digital library. The mechanisms used to support interoperability in space between heterogeneous systems are the same as those needed to support migration of collections in time onto new technology. The challenge that global digital libraries present is the need to agree on standards for information languages and knowledge representations. But this challenge is precisely what must be faced for information to become available and accessible globally.

Since there are so many challenges related to the digital library field, research efforts must focus. In the next section we highlight some challenges that are of high priority.

4. High Priority Challenges

As is discussed above, there are many challenges to be addressed by the digital library community. Here we suggest priority challenges that relate to possibilities of US-Korea collaboration.

First, there is the goal of ubiquitous accessibility. Particular challenges relate to crossing boundaries of culture, age, social level, and educational level.

Second, there is the related goal of interoperability. While technical approaches are crucial, they are not sufficient. Further, it is not enough just to have a research agenda, or to have a common project, or to have a common goal for societal impact. What is essential is to have joint funding to support work on all of these.

Third, there are language issues. Multi-lingual user interfaces are essential. Machine translation is of great value. Cross-lingual information retrieval is highly beneficial. Also from a cross-language perspective, it is valuable to support extraction and summarization. Ultimately, adaptive interpretation, with tailored presentations for users, is desired.

In the next subsection we focus on the window of opportunity that is a result of NSF-funding of projects in the Digital Library Initiative (DLI). Then, in subsequent subsections we briefly describe two special challenges. One relates to multimedia and metadata, and is of particular relevance since many media types are of common interest (without barriers of language) to users in both US and Korea. The other relates to the fact that collaborative efforts require funds, which makes it essential to have clear policies regarding intellectual property.

4.1 Leveraging NSF-Funded Digital Library Projects

One of the goals of the NSF Digital Library Initiative Phase II projects has been the development of interoperable services. A particular display or query service developed within one digital library should be usable with the data collections managed by a second digital library. Three of the DLI-II projects are collaborating on the development of interoperable services, with the technology demonstrated under production load at the San Diego Supercomputer Center, and then transferred into production use for the California Digital Library (CDL). This model for technology transition can be usefully followed for the development of global digital libraries. A prototype service can be developed and tested locally, then applied within a single production digital library setting, before transfer to the global digital library.

The NSF DLI-II efforts also are using production systems to identify the primary capabilities needed to create a viable global library. For the CDL effort, four critical capabilities have been identified:

- Support for persistent identifiers
- Support for replicas of data sets stored at multiple sites
- Archival storage interfaces
- Support for persistent collections

Research is conducted within the DLI-II projects to develop infrastructure to provide each of these capabilities. It is worth noting that the same critical capabilities are needed for a global digital library. Persistent identifiers are needed that are independent of the storage location and local storage protocol. For a distributed collection, this means that the name of the local file and the local access protocol need to be attributes that are stored with each data object. Support for replicas implies the ability to identify the official

original version, as well as maintain consistency between the replicas. Integration with archives requires the ability to maintain a copy in an archive as well as on disk for immediate access. Finally, persistent collections must handle technology evolution. The infrastructure components needed to support these capabilities are all present within the proposed architecture for the US-Korean digital library.

The development of a multi-national digital library will identify additional key research areas, as outlined in the previous sections. The practical experience gained by building early prototypes for a multi-national digital library can be used to focus future research activities, and provide the impetus for future NSF solicitations.

4.2 Multimedia Indexing and Access

Compared to text, much less is known about multimedia indexing and access. MPEG-7, the emerging multimedia description standard, should be the foundation for much of this effort. This program should provide opportunities to integrate MPEG-7 descriptions into the digital library services and end-user applications. For instance, a multimedia collection server might use MPEG-7 for indexing its collection, and groups of these multimedia servers could be organized as multimedia open archives.

Advanced services will require developing new multimedia content and actively re-purposing existing content. The current MPEG-7 proposals need to be greatly expanded to allow specification of multimedia services such as descriptions of problem sets, classroom lesson plans, and interactive hypermedia presentations of lectures.

4.3 Intellectual Property Rights

Experience in other digital library projects (e.g., Perseus) has established the need to clarify intellectual property arrangements early, and to avoid compromises and assumptions. Particularly in international projects, where intellectual property laws and mores may differ among the participants, it is essential to reach a full and complete understanding early, to make the obvious explicit, and to establish the principal of monotonic progress (i.e., once something is in the collection, it is never out of the collection).

Addressing the above mentioned challenges is best arranged in the case of particular application areas. In the next section we focus on particular areas of priority.

5. High Priority Application Areas

In Table 1 is a list of application domains; those shown in bold are of particular interest with regard to US-Korea collaboration. In this section we select three applications where the priority is particularly high. We discuss the first in-depth, as an illustration of how such efforts might be investigated, but due to limitations of space only explain the remaining two briefly.

5.1 Leveraging Korean Investment in Cultural Heritage

Korean cultural heritage provides a rich opportunity for digital library collaboration between the US and Korea.
Korea has large collections of historical resources that are only available *in Korea*, and, hence, not available to international researchers without those researchers traveling to Korea. Currently there are efforts within Korea to scan ancient documents for preservation purposes. Precious ancient documents and artifacts are at risk and must be preserved before they deteriorate beyond usability, are lost, or are destroyed. These efforts are capturing static images, rather than generating truly digital documents. Even those materials that are available digitally rarely conform to international standards that would make them usable by the international community, and researchers that are successful at finding digitized resources still find the language barrier a serious impediment to usage.

Cultural exchange between the US and Korea is out of balance. Koreans tend to be substantially more familiar with US culture and language than US citizens are about Korea, and many more Koreans come to the US for education and employment than US citizens go to Korea. Given the strength of the economic and diplomatic relationships between the two countries, substantial benefit can accrue to both countries through a more balanced exchange. This is clearly inhibited by a range of technical issues, such as standards conformant metadata support, as well as character and document encoding – and further by a significant (and lop-sided) language barrier. Seemingly straightforward and simple solutions such as the use of Optical Character Recognition (OCR) to convert page images into digital documents have not matured sufficiently to offer viable solutions for Korean materials.

Collaboration between the US and Korea on preserving and providing access to cultural resources offers a mutually rewarding opportunity for advancing the technologies underlying global digital libraries. Such a collaborative endeavor can be composed of three phases: proof of concept, focused development and marketing, and production operations.

5.1.1 It's new & novel!

The first phase (proof of concept) captures the interest of the technology community and the stewards of cultural resources, and strives toward concrete, well-defined objectives to set the stage for longer term expansion. This prototyping period requires the construction of a balanced core corpus of materials in which the breadth and depth of resources is represented, including content such as text, images (of art objects, people, places, rare books, …), maps, dictionaries, and semantic aids such as ontologies. This phase also presents the opportunity to construct 3-D representations of interesting objects and spaces, and to build bilingual resources such as lexicons and parallel corpora.

Phase 1 yields a "best of class" prototype based on the current state of the art, to prepare the way for long-term development and confront fundamental questions in areas such as preservation and access. Its operational goal is to demonstrate feasibility, functionality, and utility. Its pragmatic goal is to establish a critical mass of people, infrastructure, and information resources. Its technical goal is to evaluate the sufficiency of current approaches, techniques, and standards.

5.1.2 It's useful & desirable!

The second phase of development moves from proof of concept to deployment and expansion, bridging multiple disciplines and domains. An explosion of applications, users, and uses is accompanied by a growing recognition of the need for interoperability, metadata standards, ontologies, machine translation tools, and scholarly resources such as commentaries and handcrafted translations. This phase confronts issues of scalability and sustainability directly, validates the architecture and approach, and accommodates new advances from research and development. This is also the phase where user evaluation studies are most relevant and rewarding, as a statistically significant set of resources, applications, and users become available within a coherent domain for experimentation and study.

5.1.3 It's assumed & unnoticed!

Phase 3 may be the invisible phase, characterized by digital library resources and facilities being so common and fundamental to educational and scholarly pursuits that few would even consider working without them. This sets the stage for transformation of education and research. Phase 3 requires cross-lingual transparency, or the ability to operate in one's preferred language on a set of materials in other languages with proficiency. Cross-lingual transparency requires effective cross-lingual information retrieval (CLIR), content-based multimedia support, and information extraction and summarization tools. It enables seamless multilingual cross-disciplinary research and analysis, and cross-cultural learning and collaboration, leading to a fundamental transformation of manual scholarly practice.

Phase 3 also leads to documents that are *designed* for a digital library, which is to say that they would not be feasible without the infrastructure of a DL in place. One could easily imagine routinely time-tagging and geo-referencing new digital objects in the collection. Images, for example, would always have a geographic reference and time associated with them, enabling easy correlation of images across time and space. One also could attend to a vast array of details that confound scholarly use of information, such as disambiguating proper names, resolving co-references within and among documents, and establishing authority control over information resources.

5.1.4 Capturing the Moment

The time is right for US-Korean collaboration on a Cultural Heritage Digital Library. While digital library research and development will continue for a long time, the decade

or so of development to date has yielded substantial returns, resulting in a mature set of basic digital library technologies. Coupled with the dramatic growth in networking worldwide, and particularly the development of broadband capacity in the US and the extensive wired and wireless infrastructure of Korea, the basic technological infrastructure is in place.

Korea is committed to a massive digitization effort that will yield a wealth of Korean cultural heritage information resources, available for global access and utilization. This effort recognizes the need for conformance with global information standards and best practices, so is sure to provide an invaluable source of materials.

The US, through the National Science Foundation (NSF), is encouraging and facilitating the development of international digital library projects. Cooperative efforts have been established with several European countries, but few such relationships have been established with Asian countries. Korea provides a timely and propitious opportunity, with benefits accruing to the global community in the large, but with a special opportunity for the US to gain access to high quality information on Korean culture that is essentially unavailable anywhere in the US today. A joint US-Korean cultural heritage digital library project could also expand the accessibility of American cultural resources to Korea, and enable greatly enhanced multicultural education and collaboration opportunities.

A joint initiative between the US and Korea would necessarily engage researchers in industry and universities in both countries and their government sponsors. Its reach would extend beyond the traditional computer science researchers engaged in digital library work to include cognitive scientists, librarians, and other disciplines.

A successful cultural heritage initiative will require massive sets of digitized information resources produced through government resources (e.g., MIC in Korea), coupled with comparable corpora from existing DLs in the West. The initiative also will require an array of language technologies, both existing and under development. But, perhaps most important, the initiative will need a strong and coherent voice of leadership.
It will need to consider a broad range of performance measures and criteria by which success can be evaluated, including both quantitative and qualitative attributes. Quantitative measures must clearly include size and usage statistics, including the numbers of users, the number of digital objects, the relative proportion of US vs. Korean users, etc. Qualitative measures also must be developed to assess international usability, value, and return on investment.

5.2 Global Data

Multiple international science-based projects are being driven by the need to globally generate, access, and analyze data. The projects span the "hard" sciences, including

- High-energy physics – the CERN large hadron collider will generate petabytes of data per year starting in 2005, which will be analyzed by researchers in Europe and the US.

- Astronomy – a Virtual Observatory is being proposed that integrates digital image data from all-sky surveys generated from telescopes in both hemispheres.
- Biology – the Protein Data Bank distributes protein structures through international mirror sites
- Earth systems science – remote satellite sensor data is generated by multiple nations, and integrated through data assimilation to provide both weather forecasts and to serve as primary data for investigating global warming.
- Neuroscience – the Human Brain Project builds upon research at multiple institutions to build a representation of the "normal" human brain.
- Fusion – experiments to control fusion are underway in multiple nations, with data shared to facilitate progress.

Each discipline is characterized by the need to access data sources (sensors, experiments, simulations) that are nationally distributed, aggregate the data into either central or distributed catalogs, and then support analysis by replicating data at multiple sites. The collections contain billions of digital objects, organized by domain specific attributes, and are accessed over the Web. To maximize utilization of the data, mirror sites and replicas are created where the research is conducted. Support for multi-lingual access is minimized by using domain specific terms and by picking a single language for describing the scientific content. Of great interest is the ability to use alternate representations for both accessing the collections (multi-lingual support) and for organizing the data (multiple ontology support). Despite having a central repository, the high-energy physics community is already faced with over 200,000 different organizations of pertinent data subsets. This constitutes a massive ontology mapping problem for deciding which organization is appropriate for a particular research question. The research issues identified in the US-Korean joint workshop will greatly benefit the STEM (science, technology, engineering, and mathematics) global digital libraries.

The need for global digital libraries for social sciences is just as critical. Global digital libraries can serve as the primary data resources for understanding cultural heritage, developing educational curricula on world cultures, and providing multi-lingual corpora to support translation research. The scale of data is also huge, when images are taken of every page of every book or report. A single text collection can be terabytes in size, with millions of images. The technologies developed to handle the scale of scientific digital libraries, can be usefully applied to social science digital libraries.

5.3 Education

While the U.S. is actively pursuing the use of information technology in the NSDL, science education is an international concern. The bi-lateral Korean-American effort can be a model for international collaboration in STEM education and a step toward an International Science Digital Library (ISDL). UNESCO, through an analytical survey regarding digital libraries for education, has helped promote this goal. As with the U.S. NSDL, a joint Korean-American effort to extend the STEM educational digital libraries could be aimed at a broad range of users by providing rich resources and services into

primary and secondary education through universities, and to broader needs for scientific literacy and education by the public.

Educational content creation and access were discussed during the workshop as key dimensions for this work. Technical issues such as translation and cross-language information retrieval, ubiquitous access by wireless devices, personalization of educational resources, and the re-purposing of multimedia content would be central to meeting these two dimensions. Beyond the bilateral STEM digital library, education also provides a framework for additional research directions. Network technologies would support the collaboration of students in the two counties. In addition, education will be a major use for the Cultural Heritage resources.

6. Conclusions

This project has been concerned with further developing US-Korea collaboration in the digital library area. Given the current world situation, it appears particularly important to promote US-Korea understanding and cooperation. It is our hope that this process can be further advanced as a result of our 2000 workshop, and subsequent efforts related to preparation of this report.

A strategic approach to this process is needed. A number of initial collaborative efforts have been identified and should be nurtured. Some of the sections above highlight characteristics of desired future programs, and directly or indirectly identify those who might be involved.

However, real success in research collaboration on a large scale requires funding by suitable sponsors. In the USA, the most likely sponsor appears to be NSF. Clearly, NSF support should leverage large prior and ongoing investments in programs like DLI and NSDL. Cross-divisional cooperation between CISE, DUE, and the international programs division would seem appropriate, especially when application areas like education are considered. A list of other US agencies to involve might include (D)ARPA, NIH (especially NLM), IMLS, and NEH. Connections between Library of Congress and national library efforts in Korea also seem promising. Certainly, bilateral programs are preferred, so that all interested parties in Korea may contribute. A broad consortial approach might be appropriate since agencies and active parties in the governments of US and Korea do not have a simple 1-1 correspondence.

Regarding support, we recommend funding levels of at least $1M per year for 5 years. Initially there might be planning grants for workshops or small teams to develop into consortia, with the goal of 10 such efforts over a 1-2 year period. Later, there could be a balance of small, medium, and large projects. For example there might be 6 small, 3 medium, and 1 large effort. Since there is interest in other similar collaborations with China and other nations in the region, as well as in other areas like Europe, and since funding from other sources could help expand the program, the level and amount of funding might increase if the scope expands. Similarly, though it may lead to

complications and might disenfranchise certain talented groups, calling for cost sharing could lead to larger and more sustainable efforts.

Looking toward the long term, as well as because of concern about maximizing impact, we recommend dissemination grants that might follow after the completion of standard projects. In particular, when results are particularly promising, it is desirable to invest further in order to transfer results to the open source community or to work toward commercialization. With regard to content collections, ongoing community building activities are helpful so as to ensure sustainability.

We believe that one of the most exciting possibilities is that US-Korea collaboration on digital library research may have a bootstrapping effect. If there is study of cultural issues over the long term, the groups involved should advance cultural understanding inside their projects, as well as in the two nations. A rich cultural context should emerge in which future interactions will be facilitated. It is our earnest hope that this will occur, and that as digital libraries emerge as the result of US-Korea collaborations, that there can be peace and prosperity surrounding future interactions between the USA and the Korean Peninsula.