

DIGITAL LIBRARY: AN OVERVIEW

*M. Krishnamurthy**

This paper describes the objectives of digital libraries and the project in some detail. The role of the digital libraries, standards, collection management, cataloguing, stages of digital libraries, and evaluative of various aspects. A number of lessons emerged from this paper which are described, as are model for digital library.

KEYWORDS/DESCRIPTORS: Digital Libraries; Research; Standards; User issues

1 INTRODUCTION

The term 'digital library' 'electronic library', and 'virtual library' have appeared in the professional literature of library and information science for some years already, but rarely with explicit definitions. Borgman(1999), in an article "what are digital libraries? Competing visions", explores digital libraries as organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to interpret, distribute, preserve the integrity of and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities [1]. The digital library is making the library undergo a change in the paradigm of its role to create, organize, and distribution of information resources. A digital library is a distributed electronic collection that covers virtually all fields of human endeavor to serve a defined community.

2 ROLE OF DIGITAL LIBRARY

The term "digital library" connotes different things to different people, spanning many different types of information technology, on every conceivable subject. With the many different types of information available on any conceivable subject, the Internet can appear to serve some of the same purposes

* *Librarian*, Indian Statistical Institute, 8th Mile Mysore Road, R.V.College Post, Bangalore 560059, INDIA. Email: mk13murthy@hotmail.com; krish@isibang.ac.in

as a library [2]. One feature that all digital libraries share is a system of organization or management. Arms defines “digital library” to mean “a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network [3]. To some people it simply means carryout the functions of library in a new way. A digital library is an integrated set of services for capturing, cataloguing, storing, searching protecting and retrieving information, which provide coherent organization and convenient access to typically a large amount of digital information.

One of the main goals in developing digital library is to provide users with opportunities for accessing and using information in highly flexible and user-oriented ways not available in current information repositories [4]. The highest priority of a library, digital or otherwise, is to serve the research needs of its constituents [5]. The development, maintenance, and extension of its collection and technology must be supportive of this primary objective. Yet, this may at times be lost in the midst of more immediate and apparently more pressing tasks.

3 OVER VIEW OF CURRENT DIGITAL LIBRARIES

Digital library research projects have a common theme of bringing search to the net. The initiatives focus is to dramatically advance the means to collect, and organize materials in digital formats and make them available for searching, retrieval and processing via communication networks. Six major projects were involved in DLI-I,(1994-98) sponsored in the U.S by the National Science Foundation, the Advanced Research Project Agency, and the National Aeronautics and Space Administration. The following web pages contain current project information.

- **University of Illinois at Urbana–Champaign**
Building the Inter space: Digital Library Infrastructure for a University Engineering Community”
Federating repositories of scientific literature
<http://dli.grainger.uiuc.edu/>
- **Carnegie Mellon University**
“Intermedia: Integrated Speech, Image and Language Understanding for Creation and Exploration of Digital Video Libraries”
Full-content search and retrieval of video
<http://informedia.cs.cmu.edu/>

- **Stanford University**
“The Stanford Integrated Digital Library project”
Interoperation mechanism among heterogeneous services
<http://www.diglib.stanford.edu/diglib/pub>
- **University of California at Berkeley**
“The Environmental Electronic Library: A prototype of Scalable, Intelligent, Distributed Electronic Library”
Work-centered digital information services
<http://elib.cs.berkeley.edu/>
- **University of California at Santa Barbara**
“The Alexandria Project: Towards a Distributed Digital library with Comprehensive Services for Images and Spatially Referenced Information: Spatially-referenced map information
<http://alexandria.sdc.ucsb.edu/>
- **University of Michigan**
“The University of Michigan Digital Libraries Research Proposal”
Intelligent agents for information location
<http://www.sils.umich.edu/umdl/homepage>

These DLI-1 projects use many contrasting approaches. For example, University of Illinois and University of California, Berkeley projects both full plan systems with many users, with the Illinois project focusing on manually structured text documents, and the Berkeley project on automatically recognized image documents. These projects use complementary approaches, receiving materials in electronic format directly from publishers to take advantage of the embedded Standard Generalized Markup Language (SGML) structure, and receiving them in paper format in large volumes and automatically transforming the articles into digital format.

The Carnegie Mellon and Santa Barbara projects plan to provide the ability to manipulate new media that were previously impossible to index and search. Carnegie Mellon is investigating, segmenting and indexing video, using automatic speech recognition and knowledge about structure. Santa Barbara is indexing maps, using automatic image processing and knowledge about region metadata.

The Stanford and Michigan projects plan to investigate the intermediaries necessary to perform operations on large-scale digital libraries. These projects are trying to find the representations needed, on the one hand, to interoperate between the formats for different search services and, on the other hand, to identify the appropriate sources to be searched for a given query. All projects are building test beds with large collections to address their corresponding fundamental research questions in building large-scale digital libraries.

The second phase "Digital Libraries Initiative-phase (DLI-2' 1998-2004) is sponsored by the National Science Foundation, the Defense Advanced Research Project Agency, the National Library of Medicine, the Library of Congress, the National Aeronautics and Space Administration, and the National Endowment for the Humanities. DLI-2 awards involve eight agencies, to make for a full range of activities related to different digital libraries: These include interoperability, portability, data exchange, scalability, federation, extensibility and open network architectures in the following universities:

- **Carnegie Mellon University**
Million Books Project
<http://zeeb.library.cmu.edu/>
- **Columbia University**
"A Patient Care Digital Library: Personalized Search and Summarization over Multimedia Information,
<http://persival.cs.columbia.edu>
- **University of Arizona**
"High Performance Digital Library Classification System: From Information Retrieval to Knowledge
<http://ai.bpa.arizona.edu/go/dl/>
- **University of California Berkeley**
"Re-inventing Scholarly Information Dissemination and Use"
<http://elib.cs.berkeley.edu>
- **University of California Santa Barbara**
"Alexandria Digital Earth Prototype"
<http://www.alexandria.ucsb.edu/>

- **Harvard university**
“Operational Social Science Digital Data Library”
<http://thedata.org/>
- **University of South Carolina**
“A Software and Data Library for Experiments, Simulations and Archiving”
<http://weblab.badm.sc.edu>
- **Stanford University**
“Stanford Interlib Technologies “
<http://www-diglib.stanford.edu>
- **Gutenberg**
<http://promo.net./pg/>
- **Open Archive Initiatives**
“Coalition for Networked Information, NSF”
<http://cni.org>
- **Tufts University**
“A Digital Library for the Humanities “
<http://www.perseus.tufts.edu>
- **NCSTRL (Networked Computer Science Reference Library)**
<http://www.ncstrl.org>

The Electronic Libraries Program (e-lib) of the United Kingdom’s Joint Information Systems Committee (JISC), Australia’s Distributed Systems Technology Center (DSTC), the Canadian Initiative in digital Libraries (CIDL) and the DELOS Working Group of the European Research Consortium for Information and Mathematics (ERCIM) and many other digital projects are under way in Europe, Asia and elsewhere.

4 RESEARCH ISSUES

Digital library research builds upon a long history of related work in information retrieval, databases, user interfaces, networks, information seeking,

classification and organization, library automation, publishing and other areas [1]. Users increasingly have access to various types of digital collections and information systems. Defining the boundaries and characteristics of these information spaces and exploring ways in which they can be fused into a coherent whole is a central problem that cuts across all aspects of the research agenda. Since early 1996, OCLC has been engaged in research and demonstration project to prove the viability of centralized archive [6]. The most important component of Internet is information. Information retrieval in the Internet can draw upon years of results and practical experience in online information access as well as from traditional physical libraries. A repository is just an organized collection where information documents are indexed for effective search. A digital library is a group of these distributed repositories that users see as a single repository. How to incorporate technology that provides semantic federation of distributed repositories for scientific literature should be another research issue.

4.1 Interoperability

The examination of current digital library projects reveals a low level of interoperability between various servers. They tend to be specific to the discipline they serve, with no knowledge of the status or existence of other digital library projects in their own or other fields. Many projects are simply providing electronic access to journal articles. Defining interoperability is difficult. It is not simply a matter of providing coherence among passive object repositories. There is a spectrum of interoperability objectives, ranging from those that can be achieved in the near term to longer term challenge objectives.

4.2 Collection Development and Management

Collection development and management research is the area where traditional library missions and practices are reinterpreted for the digital library environment. Librarians have considerable experience in Digitization although the profession has tended to call it something else. The retrospective conversion of printed library cards in to machine readable catalogue records represents one of the earliest widespread, digitization efforts [7]. Some types of materials are more suited to be digitized, while other materials such as maps and books may not.

Considerable research in to what users need, how they use information, and whether digital formats serve their needs effectively are still needed.

4.3 Preservation

If we assume that libraries are able to build some types of digital collections, there remains a significant challenge inherent in preserving these collections. Preservation has been clearly defined and understood for the last twenty years by a set of widely accepted treatments. To preserve digital information, digital hardware and software configuration will require regular “refreshing” or “migration” to more current systems. The Report of the task force on Archiving of Digital Information suggests that rapid changes in the means of recording information, in the formats for storage, and in the technologies for those threaten to render the life of information in the digital age as “nasty, brutish and short” [8]. The library and archive community is currently engaged in defining standards and or best practices for converting research materials to digital format. Questions of quality, authenticity, validation and metadata requirements are being addressed.

4.4 Cataloguing and Indexing

We can expect to see influence and new communities due to the arrival of digital technology. To be productive and efficient in the new information world individuals now performing information services will have to rethink their actions and services. It is time to rethink about the likely locates of futures services and who will provide them. As library collections grew, a need evolved to locate books, and librarians began classifying collections. Online access over the Internet is now available for local librarians to catalog new acquisitions and make them available to users. New standardization of cataloguing tools as well as knowledge of new domains and formats are needed. Another key issue is indexing. One of the roles of the digital library is being defined as making vast array of information resources available to everybody through the global Internet. The creation of indexes to bibliographic material is a service that adds much value for greatly increases the precision and completeness of subsequent retrievals [11].

4.5 Reference Service

Much has been said about technological dimensions largely of the digital library of the future, but little attention has been given to the nature and quality service that will be required in a digital library. Buckland describes three phases of modern and future libraries; paper library, automated library, and electronic library. In the paper library, material collections and technical operations are based largely on paper. The automated library sees the computerization of most operations while collections remain largely paper. In the electronic library, both collections and operations are stored and used in electronic formats. The shift from the largely paper to the largely digital library accompanies a shift away from the model of library as locus for information [10]. Reference service has been one of the focal points for library responses to changes. The defining characteristic of traditional reference service is answering questions posed by users. Other activities carried on in the reference department or service are supportive of this central function. As libraries continue in an era of constant change under pressure to deliver value added services while continuously improving the quality of these services, they would do well periodically to rethink their core values and to bring into awareness new values to the evolving needs and expectations of the users. Direct user access to information in digital format and essential services through computer network environment are two powerful emerging phenomena for which librarians necessarily only must evolve a set of values that will shape services in academic libraries for the next several years.

5 CONCLUSION

The concept of the digital library will continue to evolve. Digital libraries are evolving in to digital communities uniting people with common interests in new ways. A digital environment enables cross-community interactivity and collaboration, regardless of physical location. Digitization presents opportunities for long term preservation of bodies of information. The standards and issues of the digital library require thinking about funding, common standards, collaborative management, intellectual property rights, interface and preservation design, and integration into existing structure. It should include all the processes and services of traditional library such as collection development, cataloging and indexing.

6 ACKNOWLEDGEMENT

The author would like to acknowledge Prof. Linda C. Smith, Graduate School of Library and Information Science, Urbana- Champaign for her guidance and valuable suggestion to prepare this paper as part of the Research Program in Fulbright Fellowship.

7 REFERENCES

1. Borgman (C L). What are digital libraries? *Competing visions, Information Processing Management*, Vol. 35(3); 1999.
2. Association for Computing Machinery. *Communication of the ACM*. Vol. 38 (4); 1995.
3. Arms (W Y). *Digital Libraries*. Cambridge, M.A: MIT press, 2000
4. Huser (C), et al., Knowledge based editing and visualisation for hyper media encyclopedia. *Communication of the ACM*. Vol. 38(4); 1995.
5. Levy (D M); Marshall (C). Going digital: A look at assumptions underlying digital libraries. *Communications of the ACM*. Vol.38 (4); 1995.
6. Bellinger (M). The Transformation from microfilm to digital storage and access. *Journal of Library Administratio*. Vol. 25(4); 1998.
7. Kuny (T); Cleveland (G). The Digital library: myths and challenges. *IFLA Journal*. Vol. 24 (2); 1998.
8. Task Force on Archiving of Digital Information, "Preservation Digital Information: Final Report and Recommendations". <http://www.rlg.org/Arch TF/>
9. Buckland (M). Redesigning library services. "Proceedings of the American Library" Association Annual Conference, Chicago, IL, 1992.
10. Ferguson, (C D); Bunge, (C A). The Shape of service to Come: Values-Based reference services for the largely digital library. *College and Research Library*. Vol.58, 1997.
11. Wiederhold (G). Digital libraries, value, and productivity. *Communication of the ACM*. Vol. 38(4); 1995.
12. Dorman (D). Exploring the success and failures of UIUC digital libraries. *Computers in Libraries*. Vol. 18 (6); 1998.
13. Schatz (B); Chen (H). Building large scale digital libraries. *IEEE Computer*. Vol. 29(5) ; 1996.

14. Fox (E A). et al. Digital Libraries. *Communications of the ACM*. Vol.48(4); 1995
15. Lesk (M E). *Practical digital libraries: books, bytes, and bucks*. San Francisco: Morgan Kaufman,1997
16. DLI (Digital Library Initiatives) 2. 2002. Digital Libraries Initiatives (DLI) phase 2. <http://www.dli1.nsf.gov>
17. DLF (Digital Library Federation). *A working definition of digital library*. <http://www.clir.org/diglib/dldefintion.htm>
18. DLF. *Digital library federation*. <http://www.diglib.org/>
19. Ann (P); Bishop. Working towards an understanding of digital library use. *D-Lib magazine*. Vol. 1, Oct;1995. <http://ww.dlib.org/dlib.oct95/10bishop.html/>

Genetic Diversity and Relationships Among the Tribes of Meghalaya Compared to Other Indian and Continental Populations

B.T. LANGSTIEH,^{1,2} B. MOHAN REDDY,² K. THANGARAJ,³ V. KUMAR,² AND LALJI SINGH³

Abstract The autosomal AmpFLSTR markers validated and widely used for forensic applications are used in this study to examine the extent of diversity and genetic relationships among nine Meghalaya populations. Altogether, 932 chromosomes from 9 populations were analyzed using 9 tetrameric AmpFLSTR loci. The included populations were all seven subtribes of the Austro-Asiatic Mon-Khmer-speaking Khasi and the neighboring Tibeto-Burman Garo. The Lyngngam, which are linguistically closer to the Khasi but are culturally intermediate between the Khasi and the Garo, are also included in the study. Although most of the microsatellite loci are highly polymorphic in each of these populations, the allele distributions are fairly uniform across the Meghalaya populations, suggesting relative homogeneity among them. Concurrent with this, the coefficient of gene differentiation (G_{ST}) is observed to be low (0.026 ± 0.002). This is naturally reflected in the lack of clear differentiation and clustering pattern of the Meghalaya tribes based on either geographic proximity or the historical or current affiliations of these tribes. Analysis of molecular variance (AMOVA) suggests no significant population structure. The structure analysis further suggests that, barring War-Khasi and Pnar, no other population shows any semblance of genetic identity. Even the position of the linguistically distinct Garo is not portrayed as separate from the Khasi. However, when comparable data from other Indian, Southeast Asian, and other continental populations were analyzed, the Meghalaya populations formed a compact cluster clearly separated from other populations, suggesting genetic identity of the Meghalaya populations as a whole. These results are concurrent with the hypothesis of a common and recent origin of these Meghalaya populations, whose genetic differentiation is overwhelmed by the homogenizing effect of continuous gene flow.

¹Department of Anthropology, North Eastern Hill University, Shillong, Meghalaya, India.

²Anthropology and Human Genetics Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata, India.

³Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad, India.

Human Biology, August 2004, v. 76, no. 4, pp. 569–590.

Copyright © 2004 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: KHASI SUBTRIBES, GARO POPULATION, LYNGNGAM POPULATION, MEGHALAYA POPULATIONS, INDIA, AMPFLSTR LOCI, D3S1358, D8S1179, D5S181, VWA, D21S11, D13S317, FGA, D7S820, D18S5, GENE FLOW, GENETIC HOMOGENIZATION.

The northeastern part of India is referred to as a melting pot of Mongoloid, Australoid, and Caucasoid populations, which is exhibited in the unique sociocultural diversity of the region. It has been described as the corridor for the influx of migratory populations from Southeast Asia and neighboring Tibet, Myanmar (Burma), Thailand, etc. These populations might have settled in this region at different times and probably arrived in different batches as hordes of food gatherers, hunters, and warriors, according to prehistoric archeological evidence from the Garo hills, Meghalaya, that suggests that this region might have been inhabited as early as the Paleolithic period (Sharma 1966, 1980; Hussain 1991).

There are two indigenous and predominant tribal clusters, namely, the Khasi and the Garo, perhaps one of the few populations in India and the world that follow the system of matrilineal descent and matrilocal residence. Although the Khasi, who occupy the central and eastern regions of Meghalaya, are the only Mon-Khmer Austro-Asiatic speakers located as a pocket amid the ethnic majority of Asian populations of Sino-Tibeto-Burman origin in the entire northeastern region, the Garo, who inhabit the western parts of Meghalaya, are Tibeto-Burman speakers. Given the hypothesis that northeast India served as a major corridor for early human migrations into and out of the Indian subcontinent and particularly because of the presence of the Austro-Asiatic Khasi, the study of populations from this region assumes special significance in answering questions about the peopling of India and routes of migration. Some recent studies involving molecular genetic markers (Basu et al. 2003) speculate that Austro-Asiatic tribes were the earliest settlers in India.

Comprehensive genome diversity studies of different ethnic, regional, and linguistic groups of India are needed to find unequivocal answers to some of the issues concerning the history and peopling of this region and also to test some of the current anthropological hypotheses. This need prompted us to initiate studies among different regional and linguistic populations of India. These samples are being analyzed for different sets of DNA markers: mitochondrial, Y chromosome based, and autosomal. Here, we report findings based on the analyses of nine amplified-fragment-length short-tandem-repeat (AmpFLSTR) loci among the tribal populations of Meghalaya in the northeastern part of India. We examine the nature and extent of genetic diversity and relationships among the nine tribal populations of Meghalaya, which represent both the linguistic and the geographic heterogeneity of the state in relation to other Indian and continental populations. Further, we probe whether the matrilineal system prevalent among these populations has a role in the observed pattern of diversity and relationships among the Meghalaya populations, thus exploring the influence of social system on the genetic structure of tribal populations of India.

Microsatellite loci have been widely used to study genetic relationships among human populations on the continental (Bowcock et al. 1994; Cavalli-Sforza et al. 1994; Deka et al. 1995a, 1995b; Nei and Takezaki 1996; Perez-Lezaun et al. 1997; Jorde et al. 1997; Eller 1999), regional (Parra et al. 1999;