



From digital libraries to digital preservation research: the importance of users and context

From libraries to
preservation
research

207

Gobinda Chowdhury
University of Technology, Sydney, Australia

Received 1 January 2009
Revised 10 July 2009
Accepted 17 July 2009

Abstract

Purpose – The purpose of this paper is to point out the commonalities of research in digital libraries and digital preservation with regard to the issues of users and context of information.

Design/methodology/approach – The paper's approach is a review of selected literature and reports of research projects focusing particularly on digital preservation research.

Findings – It is noted that just like the digital library community the digital preservation research community is also confronted with the challenges of capturing, storing and making use of the information related to users and context.

Practical implications – The paper points out some current research in digital preservation that aims to handle the users and context information for building future digital preservation systems. It highlights some major challenges in these areas.

Originality/value – The paper reports on the state of the art research in digital preservation.

Keywords Digital libraries, Information science, Collections management, Information management, User studies

Paper type Literature review

Introduction

A review paper on digital library research that appeared in this journal exactly ten years ago (Chowdhury and Chowdhury, 1999), observed that digital library research was then at its infancy but was growing fast. Over the past decade it has come to adulthood, not a very long time compared to the lifespan of library and information science research, but a reasonably long time from the perspectives of the rapid changes in the world of the internet and web. During the first few years of its origin and development the field of digital library research has evolved and changed rapidly, with continuing discussions and debates on the definition and connotation of the term digital library. Gradually with the maturity of the field, and sharing of ideas among digital library researchers originating from many different fields such as library and information science, computer science, and engineering, psychology, linguistics, etc. an agreement with regard to the definition of digital libraries seems to have been reached. The DELOS Network of Excellence on Digital Libraries (The DELOS Digital Library Reference Model: Foundations for Digital Libraries, 2007) envisages a digital library as:

... a tool at the centre of intellectual activity having no logical, conceptual, physical, temporal or personal borders or barriers on information.

It is important to note that a digital library has been considered here as a tool facilitating intellectual activities across spatial, temporal and personal boundaries. The DELOS characterization of digital libraries (The DELOS Digital Library Reference



Model: Foundations for Digital Libraries, 2007) further states that the field of digital library has:

... moved from a content-centric system that simply organises and provides access to particular collections of data and information to a person-centric system that aims to provide interesting, novel, personalised experiences to users, and consequently.

... its main role has shifted from static storage and retrieval of information to facilitation of communication, collaboration and other forms of interaction among scientists, researchers or the general public on themes that are pertinent to the information stored in the Digital Library.

Two important points about the nature of emerging digital libraries should be noted here: first a digital library is becoming a person-centric system as opposed to a generic collection and service, and second its goal is now to facilitate communication, collaboration and interactions, and not just providing access to digital information.

So, a modern digital library is a space – a centre of intellectual activities – with content, available in different forms and formats in a distributed network environment, as well as tools and facilities for user-centric access, use, interactions, collaborations and sharing. Thus, as opposed to the early stages of digital library research the focus has shifted from system and content to the users and interactive use and sharing in a networked environment. This in a way echoes the conclusion made in the *JDOC* 1999 review paper where the authors concluded that in order to build and live in a true digital library world, “we have to change our generation-long habits and have to get used to our new shoes. This will take some time, because it will involve a paradigm shift in our habits of the creation, distribution and use of information” (Chowdhury and Chowdhury, 1999).

So, has there been a paradigm shift “in our habits of the creation, distribution and use of information” within the past ten years, since the publication of that work? The answer is perhaps both yes and no: yes in the sense that indeed most users have changed their habits to a great extent in the way they access and use information in the digital world; and no in the sense that content producers (publishers, database service providers, etc.) perhaps are still following the same paradigm of content creation and distribution, or are trying to replicate the old practices within the context of the digital world, without taking any revolutionary steps and breaking away from the old practices of content creation distribution and access.

Early research studies of digital library users (Greenstein and Thorin, 2002) reveal that:

- users want seamless access to heterogeneous information resources irrespective of where, by whom, or in what format they are managed; and
- users prefer somewhat personalised service in a networked information environment that meets their specific needs.

User study has become an integral part of digital library research over the past decade or so. A recent survey of digital library literature for the past 11 years (1997-2007) reveals that usability and user studies cover over a third of the published literature (34.5 per cent or 199 out of 577), and that major areas covered in those studies include: usability, interface interaction/design, HCI/user interface and accessibility.

The importance of digital library as a workspace, and the importance of user annotations of digital content, has been discussed by some researchers. For example, after studying the user (in this case teachers) behaviour in the National Science Digital Library (NSDL), it was noted that there is a “potential to exploit teacher annotations of digital library resources to support knowledge enhancement and context-building within the emerging Fedora architecture”.

Moving towards the field of digital preservation, one may note that research in digital libraries and digital preservation has progressed side by side for nearly two decades both with the same broad mission: to make digital information accessible and useable to the user community of the present and future. Both the fields of research are facing a number of major challenges. Some of these are technological while the others are related to users and context. This paper looks into the issues of context and users, discusses some of the major challenges, and highlights some recent research activities aimed at resolving some of the problems especially within the field of digital preservation research. This is not a comprehensive review of literature on digital libraries or digital preservation. Instead, based on a selected set of literature and reports on digital libraries and digital preservation with special reference to users and context of information, this paper highlights some common problems and challenges facing both the communities and indicates some approaches to possible solutions.

From digital library to digital preservation research

The digital library review paper of 1999 (Chowdhury and Chowdhury, 1999, p. 434) observed that:

... the rapid developments of technology have a negative impact: technology becomes outdated too fast ... This will continue to happen, probably faster, in future. Therefore, we have to be very careful in preserving digital information resources; and this looks to be a continuous problem.

The importance of digital preservation has been emphasized in many publications, and even in the new definition of digital libraries provided in the DELOS Digital Library Reference Model (The DELOS Digital Library Reference Model: Foundations for Digital Libraries, 2007), which states that a digital library is:

... an organisation, which might be virtual, that comprehensively collects, manages and preserves for the long term rich digital content, and offers to its user communities specialised functionality on that content, of measurable quality and according to codified policies.

This definition includes preservation as one of the main functions of a digital library, along with the provision of a specified set of functionality for the user to access and use quality information within a set of agreed policies.

Indeed, once data and information is made available in digital form – either through digitisation or by creating information in the digital form in the first place, often called born digital – the most obvious question that appears is how to preserve this information so that it can be accessed and used in future when the current technology that has been used to create and access the information will not be available any longer. The question of preservation was also important in the traditional information or the printed world, but the problem became severe because of the very short lifespan of digital information compared to printed information. As we moved towards the more portable and compact form for recording of information through generations, the

life span of the recording medium became significantly reduced – from several hundred or even thousand years in case of stone carvings or tablets to just a few weeks in the digital world. According to the director of the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP), the estimated life span of a web site is only 44 days (Library of Congress, 2008).

Quite rightly digital library researchers and funding agencies realised the importance of digital preservation, and a shift in the focus on research in digital preservation seemed a natural progression. Of course this does not mean that digital preservation research began very recently. In fact, it started almost at the same time, and the need for convergence of both the research fields, or the fact that digital preservation is an essential part of digital library research and development activities, has been emphasised in a number of reports and initiatives. For example, the European Union's digital libraries initiative sets out "to make all Europe's cultural resources and scientific records – books, journals, films, maps, photographs, music, etc. – accessible to all, and preserve it for future generations" (i2010, 2008).

So, creation of a digital library system takes us to, among others, the complexities of digital preservation the aim of which is to make sure that the stored information and data can be accessed and used in future. According to a simple definition, provided by the Digital Preservation Policies Study report, digital preservation is "the process of active management by which we ensure that a digital object will be accessible in the future" (Beagrie *et al.*, 2008). Jantz and Giarlo (2005) define digital preservation as the managed activities that are necessary:

- for the long term maintenance of a byte stream (including metadata) sufficient to reproduce a suitable facsimile of the original document; and
- for the continued accessibility of the document contents through time and changing technology.

Traditionally library and information services have played a key role in the continuum of knowledge: it has captured and organised the information resources created in the past, so that they can be accessed by the user community of the present; they also have played a key-role in preserving the information sources to facilitate access by the future generation of users. Of course the later part has largely been taken care by the usual long life of conventionally printed and published materials within a controlled climate of temperature and humidity, etc. for storage of the resources. In other words, preservation of information was not the problem facing every library, and definitely not in the shorter term. Preservation activities within the libraries were primarily reserved for materials that are reasonably old – document age being measured often in terms of centuries.

Turning to the digital library scenario, however, the problem is much more severe because in the digital world information created even a decade ago may be considered too old to be accessed and used by the rapidly changing technology and tools. Nevertheless, big initiatives are underway, for example, i2010 (2008), the digital library initiative of the European Union aims to make all Europe's cultural resources and scientific accessible to all, and preserve it for future generations focusing particularly on:

- cultural heritage – creating electronic versions of the materials in Europe's libraries, archives and museums, making them available online, for work, study, or leisure, and preserving them for future generations
- scientific information – making research findings more widely available online and keeping them available over time.

So, digital preservation is now a major concern for all institutions that deal with any kind of information or data. Anderson and Mandelbaum (2008) comment that “preservation, once the near-exclusive concern of libraries, is now a universal concern. Questions of what to save, how to best preserve it into the future, and how to finance the effort are now nearly universally asked of a much broader set of stakeholders”.

Policy issues

The recently published report by The Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2008) notes five major problems of building a sustainable digital preservation program, namely:

- (1) inadequacy of funding models to address long-term access and preservation need;
- (2) confusion and/or lack of alignment between stakeholders, roles, and responsibilities with respect to digital access and preservation;
- (3) inadequate institutional, enterprise, and/or community incentives to support the collaboration needed to reinforce sustainable economic models;
- (4) complacency that current practices are good enough; and
- (5) fear that digital access and preservation is too big to take on.

Emphasizing on the issue of urgency, the report further states that:

In the analog world, the rate of degradation or depreciation of an asset is usually not swift, and consequently, decisions about long-term preservation of these materials can often be postponed for a considerable period. The digital world affords no such luxury, digital assets can be extremely fragile and ephemeral, and the need to make preservation decisions can arise as early as the time of the asset’s creation (The Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2008, p. 9).

A number of studies have taken place in the recent past to prepare a set of policies for digital preservation in general, or for specific types of institutions. The Digital Curation Centre (DCC, 2008) and Digital Preservation Europe (DPE, 2007) have recently released the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) toolkit which is intended “to facilitate internal audit by providing repository administrators with a means to assess their capabilities, identify their weaknesses, and recognise their strengths” with regard to digital preservation (DRAMBORA Interactive, 2008).

Beagrie (2006) comments that digital preservation solutions are partly technical, and partly organizational and procedural, and therefore digital preservation relies on the interaction between the digital preservation environment and wider organisational objectives and procedural issues.

The Digital Preservation Policies Study report (Beagrie *et al.*, 2008), a JISC funded study, provides an outline model for digital preservation policies for Higher and Further Education Institutions in the UK. The report proposes that an institution can take one of the following two preservation strategies:

a life-cycle approach by going through “each implementation stage in the following order: selection, conversion, receive, verify, determine significant properties, ingest, metadata, storage, preservation techniques, and access”; or

the OAIS (ISO 14721, 2003) approach that includes: “Preservation Planning, Ingest, Archival Storage, Data Management, Administration, Access, Deletion, and possibly a description of the different archival packages: Archival Information Package, Submission Information Package, and Dissemination Information Package”.

The digital preservation activities at The National Archives of UK (TNA) is based on two sets of activities: “passive preservation, which provides secure storage, and active preservation, which ensures the continued accessibility of the stored records over time, and across changing technologies” (Brown, 2007, p. 5).

It may be noted that digital preservation research has always focused on meeting one main objective: to make sure that the information can be used in future. But use by whom and in what context? In the digital age this is a major question. A given content (information) may have different types of potential users, each with a different characteristic, need and expectation, and the same content may be viewed and used by different types of users differently. Libraries have always played a key role in handling this sort of problem in the printed world by acting as an intermediary between content producers/providers and users, and adding value in the process based on an understanding of the user community as well as the context. For example, with knowledge of the university and its missions, etc. in case of an academic library, or nature and composition of the society vis-à-vis the mission and targets of the government and the local community in case of a public library, and so on. How can this role be simulated in a digital library environment, and how this can be passed on to the future generation of users through the preservation system, remain key questions.

Progress in digital preservation research

The goal of any digital preservation system is that “the information it contains remains accessible to users over a long period of time” (Rosenthal *et al.*, 2005). This has been the general view of the digital preservation community. Moore (2008) comments that “the concept of preservation can be characterized as communication with the future”. He further suggests that in order to enable us to communicate with the past data in a future time, “the preservation environment will need to incorporate new types of storage systems, new protocols for accessing data, new data-encoding formats, and new standards for characterizing provenance”.

Researchers in the digital library community have also worked towards this end. For example, Mischo (2005) mentions that for years, information providers have focused on developing mechanisms to transform the myriad of distributed digital collections into true digital libraries with the essential services that are required to make these digital libraries useful to and productive for users.

Several research and development activities focusing on different aspects of digital preservation have taken place over the past few years. An excellent review of digital archiving activities for the past ten years, in the context of Australia, has been provided by Cunningham (2007), while the portal PADI (2008) is an excellent source of information on all kinds of information and resources on digital archiving and preservation around the globe. The Digital Preservation Coalition (DPC, 2002) site provides a table listing the digital preservation projects undertaken by the DPC members and partners around the world. Various other institutional and research

group pages also provide valuable information on the past and present research projects on digital preservation, see for example, Library of Congress (2008), British Library (2008) and OCLC (2008).

However, progress in digital preservation research has been slow, as may be noted from the following observation of a recently published research report funded by the European Commission:

After almost two decades of setting digital preservation research agendas there is little evidence of actual progress in the development of solutions (DPE, 2007)

A number of challenges are associated with the current digital preservation systems that range from the increasingly large volumes of data to the underlying hardware, data formats, metadata and the various management practices used by these systems. However, some researchers propose that the focus of digital preservation research needs to be shifted from systems to users. In her keynote address at the 2008 iPRESS conference, Lynne Brindley, chief executive of the British Library argued that the term “digital preservation” is means-focussed and should be dropped in favour of the more end-focussed term “digital access forever” (Ball *et al.*, 2008). This brings the question of information access and information services, which are the ultimate goals of all the activities, associated with digital library and digital preservation systems.

Cunningham (2007) comments that “the concept of preservation can be characterized as communication with the future”, and that such communication with the future “corresponds to moving records onto new choices of technology”. Anderson and Mandelbaum (2008) believe that “the science of managing and preserving them for future generations of scientists is library science, and libraries will continue to lead and sustain networks for these future generations”.

They further comment that:

... library communities have developed and refined the standards for the capture, recording, validation and exchange of the essential metadata elements. These types of metadata, accompanied by community practices that support them, will be crucial to the maintenance of the massive data stores being built by scientists. Metadata can pass forward, to future custodians, information about file formats, provenance, access rights, retention periods and scores of other critical elements (Anderson and Mandelbaum, 2008).

Watry (2007) recommends that a futuristic digital archiving system should be based on a theory of preservation that can blend technology and context together within an integrated information management system where various stakeholders of a digital preservation system can seamlessly work together and make specific contributions to the development and management of a persistent digital archive.

Context in information seeking and retrieval

Kari and Savolainen (2007) comment that ‘Context is defined as all those things which are not an inherent part of information phenomena, but which nevertheless bear some relation to these’. Indeed context in information seeking and retrieval research has been defined by many different parameters such as the demographic, social, professional, educational and behavioural characteristics of users, specific tasks of users, place and time of an inquiry, and so on. In information seeking and retrieval research, the notion of context is important because it allows one to study “realistic information-seeking behaviour taking into account both specific cases and their influence on the

information seeking process itself, to the choice of search strategies, information sources, methods of evaluation of information quality, reliability, and relevance”.

Hundreds of research papers in journals and conference proceedings, especially in the ISIC conference series and in the *Information Research* online journal, have appeared discussing and debating the importance of context from the perspectives of user information behaviour, and information seeking and retrieval. Some of these papers discuss general issues of context and information behaviour while others report on studies on context and information behaviour of specific categories of users, in specific domains, and so on. In essence these studies have shown that users in a specific context have some typical characteristics, which influence their information seeking and retrieval behaviour, and consequently the same information may be accessed, perceived and used differently by users in different contexts.

This notion of context is extremely important in digital libraries. As discussed earlier in this paper, modern digital libraries tend to be person-centric with the mission of allowing users to perform various activities, and communicate and share information across individual, institutional and geographical boundaries. Quite obviously context plays a key role in such cases. Thus in order to meet their objectives, digital libraries should have a way to preserve the context which will facilitate access, understanding, interpretation and use of information.

Context in digital preservation

Context is used in two different ways in digital preservation: the technological context within which a digital document needs to be studied, and semantic context that is required to access, interpret and use information. Watry (2007) argues in favour of developing a theory of preservation that “extends the concept of digital preservation from one that is focused on sending the records (metadata) into the future to one that can also send into the future a description of the environment that is being used to manage and read records”. In this case the notion of context is mainly limited to the technological environment which can be noted through the following statement of Watry (2007) where he suggests that the true test of a digital preservation system is whether:

... it describes the entire preservation information context sufficiently well that the records can be migrated into an independent preservation environment without loss of authenticity or integrity. This requires migrating not only the records, but also the characterizations of the preservation environment context. The new preservation environment would have to apply the same management policies, the same preservation processes, use the same logical name spaces, and manage the same persistent state information.

The same view is expressed by Moore (2008) who argues that:

by demonstrating that the preservation environment controls the information context needed to preserve the ability to apply preservation procedures, we can create a theory of preservation in which the information content of the records and the information context of the preservation environment are communicated into the future.

Some of the ideas behind this are already represented in the literature, for example, Moore and Smith (2007) describe a rule-driven approach that enables all preservation processes (not just metadata) to be migrated onto new technologies.

Management of contextual information – technological information as well as semantic information – is possible within the OAIS system (ISO 14721, 2003) through

the RI (representation information). RI is defined as “information that maps a Data Object into more meaningful concepts” (Giaretta, 2007). In effect, RI should contain everything that is needed to make the preserved content (the Data Object which is a collection of bits) understandable and usable. Representation information comprises “any information that is required to render, process, visualize and interpret data, and includes: file formats, software, algorithms, standards and semantic information” (Patel and Ball, 2008).

However, a single RI is not enough for making sense of a given content; often a network of RI is required to understand, use and interpret a given content. Imagine that we are asked to understand and interpret the CV of a person; in understanding each item of information on the CV we need to interpret it by referring to a knowledgebase; for example, to understand the qualification of the person, we need access to a large knowledgebase of academic institutions, their various degrees and qualification levels, and how they map onto various other institutions in a country or in the world, etc. Similarly to interpret the address of the person, we need to have access to geospatial datasets to interpret the location and its various characteristics, etc; for understanding the professional background of the person, we need yet another very large knowledgebase of various institutions, their job titles and job specifications, etc. The list may go on and on depending on the items of information on the CV, and level of understanding and interpretation required.

In fact, to build and manage a network of RIs for understanding and interpreting every bit of information for every user community in every possible domain will be next to impossible. In order to set a limit to the extent of the RI network, OAIS proposes the concept of a Designated Community comprising an identified group(s) of people who will be able to understand a particular set of information within a given context. Moore (2008) comments that “the representation information defines the structures present within a record and their semantic labels. A designated community is defined that maintains the ability to interpret the semantic labels”. However, a specific Designated Community may comprise multiple user communities each having and using different knowledgebase(s) to understand and interpret information objects for their specific activities.

So, even for a designated user community the nature of the RI networks may be quite complex. Three types of RI have been identified: structural (such as file formats), semantic (providing additional meaning to the content through data dictionaries, ontologies, thesauri, etc.) and other (such as software, algorithms and standards, etc.). An AIP (Archival Information Package) in OAIS comprises both RI and PDI (Preservation Description Information) and thus it is “a form of encapsulation collecting together all the information relevant to the preservation, interpretation and reuse of digital data” (Patel and Ball, 2008).

Preserving context within digital preservation systems

Watry (2007) proposes a theory of preservation “through the application of a “digital ontology”, which can be used to represent the structural, semantic, spatial, and temporal relationships inherent within a record (e.g. the context relative to its production)”. Thus Watry proposes the use of “digital ontologies” for creating a network of representation information (RI) in the OAIS model that organizes, with the semantic labels and structure used within a specific community, to derive and interpret the meaning and context of information.

There may be six different kinds of relationships among information attributes of digital bits in a preservation environment that can represent the semantic meaning of content (the preserved information) within a specific context (Watry, 2007):

- (1) logical that can be interpreted with the help of a rule-based system;
- (2) temporal that represents a time dimension that can be used to understand and interpret the information;
- (3) spatial that represents a spatial dimension that can be used to understand and interpret the information;
- (4) procedural that may help understand a procedure or a workflow associated with the information;
- (5) functional that may represent the outcome of the application of an action or transformation; and
- (6) epistemological that may represent a systemic property of the preservation environment.

Capturing context with content: a common problem for various communities

Capturing and recording context for better access and use of information have remained a challenge in many communities. Capturing and representation of context have been an area of study of early information researchers and these are manifested in many library classification and subject indexing schemes. Indeed, the notion of context based on the semantic content of documents, intended user communities, specific context of application of information contained in documents, etc. can to some extent be captured and represented through library classification and pre-coordinate subject indexing schemes, such as the Library of Congress Classification and Subject Headings List. However, the notion of context cannot be represented in conventional term-based information retrieval systems which are based on term matching, and the task of determining the context and hence the suitability and usefulness of a given information resource in a specific context is left entirely up to the user to decide.

The importance of understanding the nature of the content and its intended use in the context of digital preservation has been highlighted by Liz Madden, Office of Strategic Initiatives, Library of Congress, as follows:

An understanding of the intellectual nature, intended use, and the relationship between the intellectual and the digital is critical to the preservation and presentation of digital content. An understanding of the intellectual value or nature of content helps inform decisions about digitization and presentation (Madden, 2008).

This practice has been commonly followed in the Records Management community, as may be noted from the following statement of Cunningham (2007):

The peculiar challenge of archiving is devising and implementing strategies for preserving the evidential meaning of records by capturing and preserving records in context. This is achieved through complex, dynamic, interlocking and finely engineered metadata regimes. Recordkeeping metadata is fundamentally different to and infinitely more complex than resource discovery metadata and preservation metadata. It is event oriented metadata in an object-oriented world.

User-centric and context-based digital library design has also been a major area of research as reported in the literature (see for example, DELOS Summer Schools (n.d.); Meyyappan *et al.*, 2001, 2004; Theng *et al.*, 1999); and several examples of user-centric and context-based digital libraries can also be found. For example, the American Memory (2008), in addition to providing search and browse facilities for all users, provides a special service, targeted for teachers, for use of the American Memory service in the classroom; The Health Information + (2008) service of the NHS Scotland provides health information suitable for various stages of the “patient journey” – from diagnosis to treatment, hospitalization, discharge and recovery – and also provides information with different levels of details, chosen by the users; Intute (2008), in addition to providing information services in various disciplines, provides some specific services that may be useful for certain categories of users, viz. academics, librarians, researchers, students and teachers, while the *MyIntute* is a service designed to help individual users.

However, capturing the changing nature of the users, and adjusting the digital library services accordingly is a very challenging as well as resource intensive job. Nevertheless, there is now a general consensus that the major challenge facing a digital library as well as a digital preservation program is that it must describe its content as well as the context sufficiently well to allow its correct interpretation by the current and future generations of users.

Several research projects are now underway to address this issue. For example, the EU project CASPER (2006) aims to find out how digitally encoded information can be understood and used in the future when the software, systems, and everyday knowledge will have changed. In order to find the solutions the project is specifically looking into the OAI framework (OAI, 2002) and creation of RI networks.

Patel and Ball (2008) discuss how RI is used for preservation of context in two different disciplines viz. Crystallography and Engineering, while Patel and Coles (2007) discuss the creation of a Registry and Repository of RI (RRoRI) in the context of JISC funded eBank-UK project. The EU funded SHAMAN (n.d.) project also, among other things, looking into the ways of preserving context along with the content and metadata in order to facilitate information access and re-use by the future generation of users (Watry, 2007).

The importance of temporal and spatial information

Meaning and interpretation of information depend much on the time and space dimension of the information contained in a document. In a recent article Mestl *et al.* (2009) discuss that time may have influence on several aspects of information, such as definitions, names, semantics such as meaning of terms, access rights, laws and politics, data models and metadata, formats and their metadata. Space is also equally important and may have similar influence on different aspects of information. Hence capturing and recording both temporal and spatial information will be extremely important for proper understanding and use of information. It should be emphasized that time and space in this case do not only refer to the time of creation or alteration of a digital document, or the space where it was created, and so on. It is much more than that. Time and space provide additional points of reference for interpretation of information and for putting it into proper context. For example, in order to properly understand a news report from Saudi Arabia or China, one may need to have an

understanding of the culture, society, politics, people and their practices, and so on, in those countries at the given point in time. Similarly, in order to understand a literature, a painting or an engineering design, one may need several reference points that can be drawn from the relevant temporal and spatial information.

Importance of time and space facets in content representation has been studied for a long time in the field of information studies, especially in the area of subject classification and indexing. For example, the five fundamental categories, and more importantly the space and time facets, introduced by Ranganathan in his famous Colon Classification Scheme first published in 1933 (Ranganathan, 1989), and the importance of space and time dimension in the relational indexing scheme introduced by Farradane (1961). Role of time, in the interpretation of digital content, has been discussed recently by Mestl *et al.* (2009), Klein *et al.* (2002), and Santos and Staab (2003).

Temporal and spatial information is more important in humanities and social science disciplines, but it is also important in other disciplines, for example in technology, law; in fact to a greater or lesser extent it is important in every discipline. It would therefore be ideal, if we could attach spatial and temporal information to every digital document so that when it would be retrieved by users in future they could also get a reference framework for interpretation and use of that information. For example, interpretation and use of a document discussing a social event, a drama or a literature, an engineering specification or a standard, a court case or a parliamentary proceeding, and so on, say 50 or 100 years after its creation would be so much easy if we could put it into an appropriate reference framework drawn by the attached temporal and spatial information. If, on the other hand, such information was not available, it would be extremely difficult to put the information in its proper context, and thus the information could be misinterpreted and misused.

One may argue that spatial and temporal information, once identified from the content of an information resource can easily be captured and recorded using appropriate tags in the metadata scheme being used. However, the problem is far more complex than this. First of all, determining the spatial and temporal attribute of content or information in a given document is a very resource intensive task if it has to be done by an indexer. However, such information, can be easily added by creators of information provided appropriate fields, and tags are created within the mark up language being used to create an information resource. More importantly, simply by attaching a date and names of one or more places will not be useful for the future users. What one requires is a reference framework derived by the attached temporal and spatial information against which the preserved information can be interpreted, understood and used.

Conclusion

In order to make the digital past suitable for access, use and interpretation by the user community in the future, digital library and digital preservation researchers face the same challenge of capturing and representing contextual information along with the digital content. Research activities related to the OAIS representation information (RI) and RI networks are encouraging. However, creation and management of RI networks have some inherent challenges because semantic meaning and world view of a given object are often different from one community to another or even from one user group to another within the same community, and how such differences may result in a failure or misuse of RI networks among communities will remain a matter of concern. Other associated research

questions may be related to decision making, intellectual property rights and privacy/confidentiality and business interests of the user community. For example, often in some areas (e.g. engineering) interpretation and use of content is embedded in proprietary or restricted tools and software, e.g. a specific engineering design can often be used only when the service of a company or a specific product/tool is used. Turning towards the activities of memory institutions, interpretation of content may also have several similar constraints, e.g. interpretation and use of a specific content may need access to a variety of RI networks in order to get the correct information for interpretation of the content for a specific audience.

Then again the nature of a designated community, its knowledge base and its perception and interpretation of certain content, may change over time. How should such changes be monitored, captured and stored within the digital library and the preservation environment, so that future users can use and interpret a specific content within the framework of a community's structure and knowledgebase at a given point in time, and thereby researchers in a specific community can study how a particular content's interpretation and use has changed over a period of time, or in a given country, in a specific business, and so on?

It seems that the problems and issues, beyond those that are technological, remain the same for the printed as well as digital world, and also for the digital libraries and digital preservation world, and they always centre-around the users and context. We still are confronted with the complexities of identifying and representing the complicated and changing nature of users and the context that represent, and also influence, the information behaviour. The challenges are manifold, for example:

- How do we create an environment – tools, techniques, standards and the appropriate ICT infrastructure – so that the huge volume and variety of content can be taken from the past to the future, along with their specified users and the intended use, on a global scale?
- Will it be possible to create a network of representation information (in the OAIS terms) that can be used universally, or shall we have to resort to domain-specific RI networks, and therefore domain-specific preservation environments?
- Is it possible to capture the changing context along with the content of each information resource, because as we know the use and importance of a specific content (a piece of information) changes significantly with time and changing nature of the society?
- Will it be possible to re-interpret the stored digital content in the light of the changing context and user community, and thereby re-inventing the importance and use of the stored objects?
- Will alternative players and economy appear in the digital world – Google is just one example – that will take care of the previously-mentioned problems and therefore the general users will remain shielded from the complexities and costs of these activities, and at the same time data and information will be preserved for use, and re-use, by generations of users on a long-term basis?

Tanner (2006) rightly mentions that as we move from management of containers (such as various library approaches to managing bibliographic items), to content (digital content per se rather than the metadata of the container), to context at each stage the

volumes of data and the complexity of the information domain grow exponentially. Perhaps what we need is a whole new approach to indexing and tagging digital information resources for better preservation.

Thus we need to build domain-specific spatial and temporal reference frameworks and preserve them along with digital content. This obviously has to be domain- and community-specific. For example, the role of temporal and spatial information in drawing a reference framework for an architectural drawing or a design standard will be different from the reference framework to understand and interpret a specific social event or a piece of creative work. Research and time will tell us how best to do this, but there may be several alternatives. For example, for specific subjects and disciplines, one may be able to prepare ontologies showing lists of terms/phrases in the field, how they map onto other fields, how the terms have been used and changed over time and so on. Some help may be obtained from existing tools and technologies, for example the ISI Web of Knowledge (www.isiwebofknowledge.com/) for determining terminologies and descriptors used in scholarly communications, or in Google *Insights for Search* (www.google.com/insights/search/#) facility that generates various reports on the use of a given term over a period of time, across different countries/regions, and so on. It may also be possible to identify specific facets of subjects that have specific implications for temporal and spatial dimension, and thus one may be able to create an ontology that will automatically trigger specific temporal and spatial reference point as soon as a given term is used for indexing, search and retrieval. Some form of examples of such spatial and temporal facets are available in Bibliographic classification schemes like Dewey Decimal Classification, and early examples facet analyses with space and time facets can be found in Colon Classification. The inherent structure of the FRBR model (IFLA, 1997) may also be useful in creating a network and establishing relationships among various RIs.

These are just some examples, and one cannot use them in their present form. However, some of these principles may be useful in building new tools and techniques. Similar tools may be built to capture and store characteristics, typical tasks and information requirements of various user communities – for example, student users, engineers, lawyers, etc. – that may be stored for specific time periods and specific geographic locations, and can be used as reference points for understanding and interpretation of information.

With such temporal and spatial indexing and the associated tools, users will be able to refer to the appropriate reference framework in order to interpret, understand and use retrieved information at a given instance. It will be like giving a manual to a student to enable him to interpret the results of a lab test, or giving the map of a country to a user to enable him to interpret and locate a specific building in relation to a specific activity or event, say.

Digital libraries in future, if we want to continue with this term in future that is, need digital librarians (again many believe the term is an oxymoron) and such tools and technologies can play the role of experienced librarian or information personnel in helping users access, understand, interpret and use digital information in its proper context. This is a huge challenge and significant amounts of research efforts and resources are needed to develop the appropriate solutions. Both the digital library and digital preservation community need to work hand in hand to solve these problems.

References

- American Memory (2008), *The Library of Congress*, American Memory, Washington, DC, available at: <http://rs6.loc.gov/>
- Anderson, M. and Mandelbaum, J. (2008), *Planning for the "long term"... in library time, Digital Archive Preservation and Sustainability (DAPS) Workshop*, MSST2008 25th IEEE Symposium on Massive Storage Systems and Technologies, Baltimore, MD, September 22, available at: www.digitalpreservation.gov/library/resources/pubs/docs/anderson_mandelbaum_daps2008.pdf
- Ball, A., Day, M. and Patel, M. (2008), "The Fifth International Conference on Preservation of Digital Objects (iPRES)", *The International Journal of Digital Curation*, Vol. 3 No. 2, available at: www.ijdc.net/index.php/ijdc/article/viewFile/89/60
- Beagrie, N. (2006), "Digital curation for science, digital libraries, and individuals", *The International Journal of Digital Curation*, Vol. 1 No. 1, available at: www.ijdc.net/ijdc/article/view/6/5
- Beagrie, N., Semple, N., Williams, N.P. and Wright, R. (2008), *Digital Preservation Policies Study. Part 1*, final report, JISC, available at: www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf
- (The) Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2008), *Interim Report*, available at: http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf
- British Library (2008), *National and International Collaboration*, British Library, London, available at: www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/natintcol/index.html
- Brown, A. (2007), "Developing practical approaches to active preservation", *The International Journal of Digital Curation*, Vol. 2 No. 1, available at: www.ijdc.net/ijdc/article/view/37/42
- CASPER (2006), *Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval*, available at: www.casparpreserves.eu/
- Chowdhury, G.G. and Chowdhury, S. (1999), "Digital library research: major issues and trends", *Journal of Documentation*, Vol. 55 No. 4, pp. 409-48.
- Cunningham, A. (2007), "Digital curation/digital archiving: a view from the national archives of Australia", paper presented at DigCCurr 2007 International Symposium on Digital Curation, Chapel Hill, NC, April 18-20, available at: www.ils.unc.edu/digccurr2007/papers/cunningham_paper_7.pdf
- DCC (2008), *Digital Curation Centre*, Digital Curation Centre, Edinburgh, available at: www.dcc.ac.uk/
- (The) DELOS Digital Library Reference Model: Foundations for Digital Libraries (2007), *DELOS Network of Excellence*, EC project no. 507618. Thematic Priority: IST-2002-2.3.1.12, available at: www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf
- DELOS Summer Schools (n.d.), *DELOS Summer Schools*, available at: www.delos.info/index.php?Itemid=268&id=517&option=com_content&task=view
- DPC (2002), *Digital Preservation Coalition*, Digital Preservation Coalition, Heslington, York, available at: www.dpconline.org/graphics/join/projects.html
- DPE (2007), *Digital Preservation Europe Research Roadmap*, EC project no. 034762. available at: www.digitalpreservationeurope.eu/publications/reports/dpe_research_roadmap_D72.pdf
- DRAMBORA Interactive (2008), *Digital Repository Audit Method Based on Risk Assessment*, available at: www.repositoryaudit.eu/
- Farradane, J. (1961), "Relational indexing", *The Indexer*, Vol. 2 No. 1, pp. 127-33.

- Giaretta, D. (2007), "The CASPER approach to digital preservation", *The International Journal of Digital Curation*, Vol. 2 No. 1, available at: www.ijdc.net/index.php/ijdc/article/viewFile/29/18
- Greenstein, D. and Thorin, S.E. (2002), *The Digital Library: a Biography*, Digital Library Federation, Council on Library and Information Resources, Washington, DC, available at: www.clir.org/pubs/reports/pub109/pub109.pdf
- Health Information+ (2008), *NHS Scotland*, available at: www.healthinfoplus.co.uk/cmspi/involve/index.aspx
- IFLA (1997), *Study Group on the Functional Requirements for Bibliographic Records. Functional Requirements for Bibliographic Reports*, IFLA, The Hague, Final report. available at: www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf
- Intute (2008), "User support", Intute, Manchester, available at: www.intute.ac.uk/help.html
- ISO 14721 (2003), *Space Data and Information Transfer Systems – Open Archival Information System – Reference Model*, available at: www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683
- Jantz, R. and Giarlo, M.J. (2005), "Digital preservation: architecture and technology for trusted digital repositories", *D-Lib Magazine*, Vol. 11 No. 6, available at: www.dlib.org/dlib/june05/jantz/06jantz.html
- Kari, J. and Savolainen, R. (2007), "Relationships between information seeking and context: a qualitative study of internet searching and the goals of personal development", *Library & Information Science Research*, Vol. 29 No. 1, pp. 47-69.
- Klein, M.C.A., Kiryakov, A., Ognyanov, D. and Fensel, D. (2002), "Finding and characterizing changes in ontologies", *Proceedings of Twenty-first International Conference on Conceptual Modeling, Tampere, LNCS, Vol. 2503*, Springer-Verlag, Berlin, pp. 79-89.
- Library of Congress (2008), "Digital preservation: national digital information infrastructure & preservation program: a collaborative initiative of the Library of Congress", Library of Congress, Washington, DC, available at: www.digitalpreservation.gov/
- Madden, L. (2008), "Applying the digital curation lessons learned from American Memory", *The International Journal of Digital Curation*, Vol. 3 No. 2, /article/viewFile/92/63
- Mestl, T., Cerrato, O., Olnes, J., Myrseth, P. and Gustavsen, I-M. (2009), "Time challenges: challenging times for future information research", *D-Lib Magazine*, Vol. 15 Nos 5/6, available at: www.dlib.org/dlib/may09/mestl/05mestl.html
- Meyyappan, N., Chowdhury, G.G. and Foo, S. (2001), "Use of a digital work environment (DWE) prototype to create a user-centred university digital library", *Journal of Information Science*, Vol. 27 No. 4, pp. 249-64.
- Meyyappan, N., Foo, S. and Chowdhury, G.G. (2004), "Design and evaluation of a task-based digital library for the academic community", *Journal of Documentation*, Vol. 60 No. 4, pp. 449-75.
- Mischo, W.A. (2005), "Digital libraries: challenges and influential work", *D-Lib Magazine*, Vol. 11 Nos 7/8, available at: <http://dlib.org/dlib/july05/mischo/07mischo.html>
- Moore, R. (2008), "Towards a theory of digital preservation", *The International Journal of Digital Curation*, Vol. 3 No. 1, pp. 63-75, available at: www.ijdc.net/ijdc/article/view/63/82
- Moore, R. and Smith, M. (2007), "Automated validation of trusted digital repository assessment criteria", *Journal of Digital Information*, Vol. 8 No. 2, available at: <http://dspace.mit.edu/handle/1721.1/39091>
- OAIS (2002), "Reference model for an open archival information system (OAIS) blue book", available at: <http://public.ccsds.org/publications/archive/650x0b1.pdf>

- OCLC (2008), "Digital collection and preservation solution", OCLC, Paris, available at: www.oclc.org/digitalpreservation/why/default.htm
- PADI (2008), *Preserving Access to Digital Information*, PADI, Canberra, available at: www.nla.gov.au/padi/index.html
- Patel, M. and Ball, A. (2008), "Challenges and issues relating to the use of representation information for the digital curation of crystallography and engineering data", *The International Journal of Digital Curation*, Vol. 3 No. 1, available at: www.ijdc.net/index.php/ijdc/article/viewFile/64/43
- Patel, M. and Coles, S. (2007), *eBank-UK Phase 3: WP4*, September 2006-June 2007, Report final version (revised), 7 September, available at: [www.ukoln.ac.uk/projects/ebank-uk/curation/eBank3-WP4-Report%20\(Revised\).pdf](http://www.ukoln.ac.uk/projects/ebank-uk/curation/eBank3-WP4-Report%20(Revised).pdf)
- Ranganathan, S.R. (1989), in Gopinath, M.A. (Ed.), *Colon Classification*, 7th ed., Sarada Ranganathan Endowment for Library Science, Bangalore.
- Rosenthal, D.S.H., Robertson, T., Lipkis, T., Reich, V. and Morabito, S. (2005), "Requirements for digital preservation systems: a bottom-up approach", *D-Lib Magazine*, Vol. 11 No. 11, available at: www.dlib.org/dlib/november05/rosenthal/11rosenthal.html
- SHAMAN (n.d.), *Sustaining Heritage Access through Multivalent Archiving*, available at: <http://shaman-ip.eu/>
- Santos, J. and Staab, S. (2003), "Engineering a complex ontology with time", *Proceedings of the Intl' Workshop on Ontologies and Distributed Systems (in conjunction with IJCAI), Mexico*.
- Tanner, S. (2006), "Managing containers, content and context in digital preservation: towards a 2020 vision", *Archiving 2006*, The Society for Imaging Science and Technology, Ottawa, pp. 19-23, May 23.
- Theng, Y.L., Duncker, E., Mohd-Nasir, N., Buchanan, G. and Thimbleby, H. (1999), "Design guidelines and use-centred digital libraries", in Abiteboul, S. and Vercoestre, A. (Eds), *Proceedings of the ECDL'99*, p. 167.
- Watry, P. (2007), "Digital preservation theory and application: transcontinental persistent archives testbed activity", *The International Journal of Digital Curation*, Vol. 2 No. 2, pp. 41-68, available at: www.ijdc.net/ijdc/article/view/43/50
- i2010 (2008), *Digital Libraries Initiative: Europe's Cultural and Scientific Riches at a Click of a Mouse*, available at: http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm

Further reading

- Library of Congress (2008), *Preservation Newsletter*, September, available at: www.digitalpreservation.gov/news/newsletter/200809.pdf

Corresponding author

Gobinda Chowdhury can be contacted at: gobinda.chowdhury@uts.edu.au