The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0264-0473.htm

Digital repository in ceramics: a metadata study

Chandana Patra
Library and the Data Bank, Central Glass and Ceramic Research Institute, Kolkata, India

Received 19 July 2007
Revised 22 August 2007
Accepted 19 September 2007

Abstract

Purpose – This paper seeks to study the metadata requirements for setting up a digital repository in ceramics resources that would provide researchers and ceramic art professionals with access to the information as per their requirements.

Design/methodology/approach – The paper first reviews and analyzes various metadata standards and formats already available. Open software (Greenstone) is used to develop the repository and the paper discusses its metadata provisions. Thereafter, the paper focuses on ceramics resources and attempts to determine the metadata elements required to describe and organize ceramic resources. Existing controlled vocabularies to standardize content metadata of the repository are also reviewed.

Findings – The paper finds that selected metadata elements of Dublin Core and Categories for the Description of Work of Art can be used to describe and organize the ceramics resources. Local qualifiers are added when necessary to describe the resources. As Categories for the Description of Work of Art metadata standards are not provided in Greenstone, these were defined using GEMS to describe and organize ceramic art works. It also found that existing controlled vocabularies are not sufficient to standardize the content metadata of the repository.

Research limitations/implications – A digital repository should also contain information resources such as video and audio-video information resources. The study has not considered studying metadata requirements to describe such information resources.

Originality/value – This paper could be useful for others who want to develop their repositories in various disciplines.

Keywords Ceramics, Digital libraries, Information management

Paper type Research paper

Introduction

A digital repository/library is collection of digital objects. It is either a local, institutional, or central (e.g. subject- or discipline-based) archive for depositing and providing access to digital contents[1]. A digital object may be held locally or accessed

The author is greatly indebted to Shri Dibyendu Paul, Senior Lecturer, University of Kalyani, for his valuable guidance and deep involvement in carrying out this study. The author also wishes to express her deep sense of gratitude and indebtedness to Dr Jian Quin, Associate Professor, School of Information Studies, Syracuse University, for helping to carry out the study. The author is grateful to the Director of Central Glass and Ceramic Research Institute, Kolkata, for his kind permission to place this paper for publication, and is also very grateful to Dr K.K. Phani, Chairman, Library Advisory Committee, Scientist-in-Charge, Library and the Data Bank, for his kind support. The author is also grateful to the members of the Greestone user list for their support in solving problems. Last but not least, the author would like to thank all those who helped her to carry out this study, both directly and indirectly.

The Electronic Library
Vol. 26 No. 4, 2008
pp. 561-581

0264-0473
DOI 10.1108/02640470810893792

remotely via computer networks. Digital repositories/libraries have the potential to store much more information, simply because digital information requires very little physical space to contain it. Digital repositories/libraries provide access to much richer content in a more structured manner, i.e. we can easily move from the catalog to the particular book then to a particular chapter and so on. The user is able to use any search term belonging to the word or phrase of the entire collection. Digital repositories/libraries can provide very user-friendly interfaces, giving clickable access to its resources.

According to Duguid and Atkins (1997), digital libraries must work with a highly diverse range of collections of digital objects, assembled on different principles by numerous contributors and continuously changing as more content and value are added to them. Equally, they must work with users who will be as diverse as society itself, with ever-changing needs and expectations. They must be useful to different communities for different purposes, at different times (Duguid and Atkins, 1997). Therefore, scalability, interoperability, extensibility, federation, and composability are major issues for digital library systems.

In the internet era, digital repositories represent completely new information infrastructures and knowledge environments. Integrating and utilizing the newest computer and communication technologies and digital content, the digital repository builds huge extendable and interoperable collections. Rachel Heery and Sheila Anderson[2] have shown that a digital repository is differentiated from other digital collections by the following characteristics:

- content is deposited in a repository, whether by the content creator, owner or a third party;
- the repository architecture manages content as well as metadata;
- the repository offers a minimum set of basic services (e.g. put, get, search, access control); and
- he repository must be sustainable and trusted, well-supported and well-managed.

In a digital repository digital objects of all kinds must be described, organized, and indexed in ways that allow users to locate and browse them in useful ways. The information required to accomplish this is often called metadata, or structured, descriptive data about an object (a book, a photograph, a movie, a letter, etc.). In order to manage the massive scale of these digital collections effectively, establishing a metadata model and application profile has become a fundamental part of any digital repository project. A number of metadata initiatives provide detailed and descriptive information about a digital resource to facilitate discovery by users. Resource description is essentially about describing information resources using a standard framework or set of principles. But because of the specific nature of heterogeneous digital resources, describing digital resources in a consistence fashion may not be an easy task, and in some cases, it is a complex process.

Those concerned with digital information management all regard metadata as an essential component of the evolving networked information environment. Metadata supports the browsing, indexing, and relationships between digital objects and collections. Metadata must be carefully designed to support the user interfaces for

browsing, searching and displaying digital objects. Now, let us see first the need of a digital repository in ceramics before finding out the requirements of metadata elements to organize ceramic resources.

Need for developing a digital repository on ceramics

Ceramics have developed around the world through the centuries as a result of how human beings think and act within their unique cultural environments. The development, manufacture and use of ceramics are a social activity as much as a technological process. The history of ceramics is the history of the human beings who created and used technical processes and products. Even before the electronic age of rapid communications, glass and ceramic scientists and industrialists spanned the globe with their ideas and research collaborations, creating ceramic materials and products that would serve society. Ceramics is one of the primitive artistic productions of early man. It had a respectable place in all the major ancient civilizations. Ceramics manufacturing companies exhibit various types of lifestyles of ancient people and the presence of ceramics can be seen in various domestic articles such as tiles, glassware, glass, dinner sets, sanitary ware, etc.

From countries around the world, scientists and engineers, educators and industrialists, historians and museum curators as well as artists are interested in sharing international knowledge and information of the ceramic world. This would be possible with the development of a digital library on ceramics. The overwhelming need for the creation of such a digital repository in ceramics is to provide students, scientists, artists and industrial communities with an open and inter-operable platform for the following activities:

- to help facilitate research and education;
- to promote global cooperation;
- to foster economic development, including rural development; and
- to help facilitate archaeological research.

Some other important benefits could also be achieved, such as:

- providing perpetual access to collections;
- acting as a preservation archive for digital material in the field of ceramics;
- promoting the study of local and regional collections on ceramics;
- promoting an understanding of ceramic materials among the user community and the public at large;
- nurturing scientific ideas;
- honoring ceramics as a product of human minds and human beings, regardless of nationality; and
- providing a superb opportunity to bridge the gap in understanding and diminishing the boundaries between societies and technologies, represent the cultural heritage of ceramics of various countries, and understand the culture and societies of others (Patra, 2006).

It is a paradox that, under circumstances where economic activity takes place at an increasingly global level, individual countries must give increased attention to their

performance at the national level so that they can find a favored position within the world community. Therefore, each country may bring together all their national activities related to ceramics, developing a digital institutional repository as well as a community repository on ceramics. The premier institutes responsible for carrying out research activities in the field of ceramics of the various countries need to take steps to design and develop a suitable infrastructure for the digital repository. It is also necessary to study the requirements of metadata to describe, locate, evaluate and manage the ceramics information resources of the repository or to solve the problem of organizing the information resources of the digital repository of ceramics. There are various metadata standards and formats and these are reviewed below to determine the metadata requirements to describe, locate, evaluate and manage the ceramic information resources of the repository.

Review of metadata standards and formats

Metadata in its broadest sense is data about data. The familiar library catalogue record could be described as metadata in that the catalogue record is "data about data" – details about a book, for instance. Similarly, database records from abstracting and indexing services are metadata (with a different variation on location data). However, the term "metadata" is increasingly being used in the information world to specify records referring to digital resources available across a network. By this definition a metadata record refers to another piece of information capable of existing in a separate physical form from the metadata record itself. Metadata also differs from a traditional catalogue data in that the location information is held within the record in such a way as to allow direct document delivery from appropriate application software; in other words the record may well contain detailed access information and the network address(es).

There is a great diversity of perspectives on various aspects of metadata issues. For instance, librarians have used machine-readable cataloguing (MARC) since the 1960s to identify, describe and provide access to their collections. However, what worked well for libraries may not work in other environments. Similarly, the basic metadata required for describing an image or work of art or non-text objects will bear a strong resemblance to the metadata that describes traditional print documents. However, some significantly different extra elements will be required for a complete description of non-text images and multi-media resources. In light of this, some formats of metadata have been developed specifically for use in certain fields of study or type of information source.

Metadata standards come from various professional community efforts to support many needs in the digital environment. The literature reveals that different communities view metadata in significantly different contexts. No single metadata standard can be expected to accommodate the needs of all communities. Although some projects, such as Dublin Core (DC), have tried to develop a coherent set of metadata schemes that can work for wide range of communities, they have not yet provided a complete description or solution for all types of digital information resources.

Several metadata format and syntax standards have been created and are in use throughout the world. In order to identify what types of metadata exist and how they are used, a survey of the web and library literature was conducted. Many of the major metadata projects initiated recently have been designed to address specific problems or needs in classifying and cataloging digital resources. Creators of these new metadata standards had specific formats or subjects in mind. A large number of metadata

schemas and standards have been developed, some quite simple in their description and others quite complex and rich. These schemes support an extremely wide range of activities. Metadata schemas generally specify names of elements and their semantics. Optionally, they may specify rules for how content must be formulated (for example, how to identify the main title), representation rules for content (for example, capitalization rules), and allowable content values (for example, terms must be used from a specified controlled vocabulary). Many metadata schemas are being developed in a variety of user environments and disciplines. Some of the most common ones are given in Table I with an analysis of their appropriateness to develop a digital repository on ceramics.

Metadata elements to develop repository

From Table I, the metadata elements which are utilized to develop the ceramic repository are now discussed in detail.

Dublin Core

The Dublin Core metadata is described as an efficient and simple metadata for electronic articles and digital objects. The Dublin Core elements describe a journal article, movie, image, etc. The Dublin Core Metadata Element Set represents a simple resource description record. Elements are designed to be used by content creators. The Core contains just 15 metadata elements:

(1) subject and keywords: the topic addressed by the work;

(2) title: the name of the object;

(3) author or creator: the person(s) primarily responsible for the intellectual content of the object;

(4) publisher: the agent or agency responsible for making the object available;

(5) description: textual description of content;

(6) contributor or other agent: the person(s), such as editors and transcribers, who have made other significant intellectual contributions to the work;

(7) date: the date of publication;

(8) object type: the genre of the object, such as novel, poem, or dictionary;

(9) form: the data representation of the object, such as Postscript file or Windows executable file;

(10) identifier: string or number used to uniquely identify the object;

(11) relation: relationship to other objects;

(12) source: objects, in either print or electronic, from which this object is derived, if applicable;

(13) language: language of the intellectual content;

(14) coverage: the spatial locations and temporal duration characteristics of the object; and

(15) rights management: a rights management statement, an identifier that links to a rights management statement, or an identifier that links to a service providing information about rights management for the resource.

Table I.
Metadata format and standards

Name of metadata schema/standards	Type of metadata schema/standard	Developer	Description and website	Appropriateness to digital repository in ceramics
Machine-Readable Cataloging (MARC)	Bibliographic description	Library of Congress	A long established standard within the library community for exchanging cataloging information. MARC supports the Anglo-American Cataloging Rules and is maintained by the library community. Over the last several years, MARC has been enhanced to support descriptive elements for electronic resources www.loc.gov/marc/	This standard has nothing specific to describe ceramic resources
Metadata Object Description Schema	Bibliographic description	Library of Congress	An XML schema for descriptive metadata compatible with the MARC 21 bibliographic format www.loc.gov/standards/mods/	This standard has nothing specific to describe ceramic resources
The Text Encoding Initiative (TEI)	Bibliographic description	The US National Endowment for the Humanities, the European Union Language Engineering Programme, and the Canadian Social Science and Humanities Research Council	An international standard for representing all kinds of literary and linguistic texts for online research and teaching. Its primary goal was to define a set of recommendations for the encoding of literary and linguistic textual materials in electronic form, in order to standardize existing work, and to facilitate the development of good practice in a rapidly developing field. In addition to specifying how to encode the text of a work, the TEI Guidelines for Electronic Text Encoding and Interchange also specify a header portion, embedded in the resource, that consists of metadata about the work www.tei-c.org/	This standard has nothing specific to describe ceramic resources
Dublin Core	Bibliographic description	The Dublin Core (DC) project grew out of a workshop sponsored by OCLC and the National Center for Supercomputing Applications in 1995	A simple generic element set applicable to a variety of digital object types. Dublin Core has been adapted by a number of communities to suit their own needs (such as the CIMI application profile for the museum community), and incorporated into several domain-specific metadata schemes http://dublincore.org/	Used for all types of electronic resources or subject areas. Therefore, this scheme is appropriate to build various types of literature collections on ceramics
Categories for the Description of Works of Art (CDWA)	Visual object	CDWA was developed by the US-based Art Information Task Force (AITF), with funding from the J. Paul Getty Trust, National Endowment for the Humanities (NEH), and the College Art Association (CAA)	Produced by the Art Information Task Force (AITF), the Categories for the Description of Works of Art are guidelines for formulating the content of art databases. They articulate an intellectual structure for descriptions of objects and image. In this sense they constitute a schematic representation of the requirements and assumptions implicit in the practice of the discipline of art history. By providing a single, encompassing framework for descriptive information about works of art, the Categories are intended to enhance compatibility between diverse systems that wish to share art information www.getty.edu/research/institute/standards/ cdwa/index.html	Contains information categories for the description of ceramic art works

(*continued*)

Table I.

Name of metadata schema/standards	Type of metadata schema/standard	Developer	Description and website	Appropriateness to digital repository in ceramics
VRA Core Categories version 3.0	Visual object	The Visual Resources Association	The VRA core group of data categories has been developed for describing surrogate images of art and architecture in visual resources collections and for sharing that information electronically. The core is a level between minimal and full and suggests which data elements are required in order to describe an item in a visual resources collection in a shared environment www.vraweb.org/vracore3.htm	Contains information categories for the description of collections of ceramic art works
Archives: EAD (Encoded Archival Description) DTD	Archival information	EAD originated in 1993, at the University of California, Berkeley. EAD is currently administered and maintained jointly by the Society of American Archivists and the United States Library of Congress	A set of rules for the representation of the intellectual and physical parts of archival finding aids. Often expressed in XML or SGML so that the information can be searched, retrieved, displayed, and exchanged www.loc.gov/ead/	This standard has nothing specific to describe ceramic resources
Content Standards for Digital Geospatial Metadata (CSDGM)	Geospatial data	The Federal Geographic Data Committee	The Federal Geographical Data Committee's Content Standards for Digital Geospatial Metadata (FGDC) specifies the information content of metadata for a set of digital geospatial data. The purpose of the standard is to provide a common set of terminology and definitions for concepts related to this metadata. Geospatial metadata allows researchers to locate relevant data sets for use in geographic information systems (GIS). The FGDC standard is a complex format with over 334 different elements, 119 of which exist only to contain other elements www.fgdc.gov/metadata/contstan.html	This standard has nothing specific to describe ceramic resources
The INDECS Project	E-commerce	The project was established at the end of 1998, with support from the European Commission's Info 2000 Programme	Created to address the need, in the digital environment, to put different creation identifiers and their supporting metadata into a framework where they could operate side by side, especially to support the management of intellectual property rights. The main focus of < indecs > is on the use of what is commonly (if imprecisely) called content or intellectual property The INDECS project: www.indecs.org/ The < indecs > metadata framework: principles, model and data dictionary: www.indecs.org/pdf/framework.pdf	This standard has nothing specific to describe ceramic resources
ONIX (Online Information Exchange)	E-commerce	ONIX was developed and is maintained by EDItEUR (an international group coordinating development of the standards infrastructure for electronic commerce in the book and serials industries) jointly with the book industry	The ONIX for Books Product Information Message is the international standard for representing and communicating book industry product information in electronic form. It has elements to record a wide range of evaluative and promotional information as well as basic bibliographic and trade data www.editeur.org/onix.html	This standard has nothing specific to describe ceramic resources

(*continued*)

Table I.

Name of metadata schema/standards	Type of metadata schema/standard	Developer	Description and website	Appropriateness to digital repository in ceramics
Learning Object Metadata scheme: the IEEE Standard for Learning Object Metadata (1484.12.1-2002)	Educational purpose	IEEE	Learning Object Metadata is used to describe educational resources in course management systems and learning management systems. The main standard is the IEEE Standard for Learning Object Metadata (1484.12.1-2002), also called the LOM. However, the LOM has been incorporated into a number of other standards, including the IMS Global Learning Consortium's Meta-Data Specification, which is freely available from the IMS	This standard has nothing specific to describe ceramic resources
MPEG-4	Media-specific	Moving Picture Experts Group	A standard for multimedia for the fixed and mobile web www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm	Contains elements to describe multimedia for the fixed and mobile web (not utilized)
MPEG-7	Media-specific	Moving Picture Experts Group	MPEG-7 is a multimedia description and indexing system that combines XML-based content description with non-textual indexing of physical features (color, movement, shape, sound, etc.) via processing of the media bit stream for multimedia information – audio, video, and images. Part 5 of the standard (ISO/IEC 15938-5) provides descriptive, technical, and usage metadata www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm	Contains elements to describe video, and audio-video information resources on ceramics (not utilized)
CIDOC Conceptual Reference Model (CRM)	Heterogeneous cultural heritage information	International Committee for Documentation (CIDOC) of the International Council of Museums	The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modeling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. Briefly, the CRM consists of 62 *entities* or *classes*, each of which has attributes (which themselves are often expressed as links to other entities)	Contains metadata elements for the description of ceramic art works

Each element is repeatable and optional, and the entire set has been defined as extensible. Each Dublin Core metadata element can also have a sub-type and sub-scheme information. For example, if an existing scheme is being used for subject and keywords, such as the Library of Congress Subject Headings (LCSH), then this information can also be attached to the element name (El-Sherbini, 2001).

For additional needed information about a resource, the Dublin Core Working Group has developed a "qualified" Dublin Core, which consists of the element and its qualifiers. The idea is that the basic elements may be further enhanced by use of these qualifiers, with the purpose of informing the user on how to view or interpret the content of the element. These qualifiers are defined as modifiers and schemes.

Categories for the Description of Works of Art (CDWA)

The CDWA was formulated for the needs of those who record, maintain, and retrieve information about art information, including the academic researcher and scholar. The categories and subcategories that are indicated as *core* are those that the task force agreed represented the minimum information necessary to identify uniquely and unambiguously and to describe a particular work of art or architecture. However, which categories are considered core can and indeed should vary depending upon the end-users whom the particular art information systems are intended to serve, the mission of the specific institution, and a number of other factors. The categories which are used to develop the digital repository in ceramics are discussed below:

- *Object/work*. An identification of the type and number of works described.
- *Titles or names*. The titles or names given to a work of art, architecture, or group, as well as the type of title, and the dates when the title was valid.
- *Measurements*. Information about the size, shape, scale, and dimensions of a work of art or architecture.
- *Materials and techniques*. The substances or materials used in the creation of a work of art or architecture, as well as any production or manufacturing techniques, processes, or methods incorporated in its fabrication. This information includes a description of both the materials used to create the work and the way in which they were put together.
- *Physical description*. A description of the appearance of a work expressed in generic terms, without reference to the subject depicted. This includes the names of any recognizable patterns, motifs, or textures used in the decoration of the work.
- *Subject-matter*. The subject-matter of a work of art (sometimes referred to as its content) is the narrative, iconic, or non-objective meaning conveyed by an abstract or a figurative composition. It is what is depicted in and by a work of art. It also covers the function of an object or architecture that otherwise has no narrative content.
- *Person/corporate body authority*. Information about artists, architects, and other individuals and corporate bodies responsible for the design and production of works of art and architecture. This authority may also be used to store information about patrons and other people or corporate bodies important to the

record for the work. It is used as Artist to develop the ceramic art works collection.

- *Creation date*. A concise description of the date or range of dates associated with the creation, design, production, presentation, performance, construction, or alteration of the work or its components, presented in a syntax suitable for display to the end-user and including any necessary indications of uncertainty, ambiguity, and nuance. It is used as Date to develop the ceramic art works collection.
- *Styles/periods/groups/movements*. A description of a work of art or architecture that associates it with a defined style, historical period, group, school, or movement whose characteristics are represented in the work. It is used as Periods to develop the ceramic art works collection.
- *Current location (repository name/geographic location)*. The name and geographic location of the repository that is currently responsible for the work, or, for monumental works and architecture, the geographic location of the work. It is used as Location to develop the ceramic art works collection[3].

Open source software for developing a digital repository in ceramics

Setting up a digital library/repository in ceramics is also not a trivial task for the institutions. Many institutions and organizations are setting up open access digital repositories using open source software that follows open standards. The Open Access and Open Source Software movements have gained rapid momentum the world over. Various free open source software is available for developing digital libraries, the most important being Greenstone Digital Library Software (from the University of Waikato, New Zealand), Eprints (from Southampton University), DSpace (from MIT), CDSWare (from CERN), and FEDORA (from Cornell and University of Virginia). Selecting suitable digital library software or systems is challenging. The process involves "evaluating the strengths and weaknesses of the various systems and selecting one that satisfies or most closely matches your goals" (Di Lauro, quoted by Zhang and Gourley, 2005).

The Greenstone Digital Library Software (version 2.72) has been employed here for building the digital repository in ceramics. Many projects around world are using it – Lesk (2005) points out that "Greenstone open source system is particularly important for digital libraries". Goh *et al.* (2006) in their study developed a checklist consisting of 12 categories of items that digital library software should possess. Using these, Greenstone was found to be the best performer, followed by CDSware, Fedora and EPrints. In fact, Greenstone was the only software package that consistently fulfilled the majority of the criteria in many of the checklist categories (Goh *et al.*, 2006). It provides a great deal of flexibility at the hands of the user. It attaches content specific and highly descriptive metadata such as descriptors or keywords to describe each item in the collection. Therefore, searches can produce more useful results, save time and effort in searching, and in the best of cases browsers may directly access the text or multimedia content for which searches were executed. The Greenstone developer team also publishes promotional articles in various publications, which indicates their continued interest and commitment to the software. Z39.50 is supported by Greenstone, which provides a new way of organizing information and publishing it on the internet

or on CD-ROM. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and is developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source, multilingual software, issued under the terms of the GNU General Public License.

The aim of the Greenstone software is to empower users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries[4]. Greenstone has been used to make many digital library collections in different countries, from different kinds of library, and with different sorts of source material. Greenstone can be deployed with minimal configuration so librarians can build digital collections without a great deal of information technology support. This includes a user interface for presenting Greenstone functionality, such as browsing and searching collections and user preferences for searches and displays of search results. Basic customizations include the definition of indexes for searching, "classifiers" for browsing, and format statements for defining web page layout. Greenstone provides a widely used standard (Dublin Core) but also allows collection-builders to use their own metadata scheme either by extending an existing one in an *ad hoc* manner or by defining an entirely new one using a metadata set editor. Collections built in Greenstone can be exported to DSpace and *vice versa*. Ceramic information resources are of various types and different metadata are required to describe, organize and retrieve them effectively, which cannot be covered using a single standard. Based on its features, therefore, the Greenstone digital library software was preferred to create the digital repository on ceramics.

Metadata in Greenstone

In Greenstone, one or more metadata sets are associated with each collection. There are a few pre-prepared sets, of which Dublin Core is one. Modifications to existing sets and new ones can be defined using an auxiliary Greenstone application called GEMS (Greenstone Editor for Metadata Sets). One important set is the extracted metadata set, which contains information extracted automatically from the documents themselves (e.g. HTML title tags, meta tags, or built-in Word author and title metadata). This is always present behind the scenes, though it may be hidden from the user. The system keeps metadata sets distinct using namespaces. For example, documents can have both a Dublin Core Title (dc.Title) and an extracted Title (ex.Title); they do not necessarily have the same value. Behind the scenes, metadata in documents, and metadata sets themselves, are represented in XML.

In order to expedite the manual assignment of metadata, the Librarian interface allows metadata to be associated with document folders as well as with individual documents. This means that users can take advantage of existing document groupings to add shared metadata in one operation. Within the interface users can organize the document hierarchy by dragging items around and creating new sub-hierarchies, which may expedite joint metadata assignment. Metadata values assigned to a folder remain with that folder and are inherited by all files nested within it. If the user subsequently selects a file and changes an inherited metadata value, a warning appears that doing so will override the inherited value. (Of course, these warnings can be turned off since, for experienced users, they soon become annoying.)

Metadata in Greenstone can be a simple text string (e.g. title, author, and publisher). Or it can be hierarchically structured, as with hierarchical classification values, in

which case new values can be placed in the classification tree. In addition, it is multi-valued: each element can have more than one value. This is used, for example, for multiple authors. The Librarian interface allows existing metadata values to be reused where appropriate, encouraging consistency in metadata assignment by eliminating the need to retype duplicate values (Witten and Bainbridge, 2005).

Ceramic resources

The literature or information resources on ceramics are many (Patra, 2006). The works of great men of the field are recorded and published. It is the literature or information resource that provides a permanent record of what has been done in the past, usually with sufficient details to avoid repetition (unless this seems desirable for some special reason). Much of the information one requires will be found some where in the literature or information resources. The various types of information resources are discussed below.

Primary literature/information resources

This contains the first and sometimes only publication of specific information. These are articles published in journal and conference proceedings, reports, newspaper articles, commercial literature, patent, standards, theses/dissertations, and government publications.

Secondary literature

This consists of republication of material distilled from primary resources, generally after a substantial period of time has elapsed. These are books, reference materials, reviews and serials.

Ceramic art works

The above information resources are sources of the type of formal information usually sought by the scientists, researchers and students working in the field. There is another important resource which is not the formal type, which is ceramic art works that provide information to the artist as well as scientists developing a new product. Ceramic art works are very good information resources for students and researchers for studying art works.

Use of metadata for ceramics resources

It is a great problem to define metadata for different types of ceramic information resources. Standardization of metadata values is another important issue. Currently MARC 21 specifications define nearly 2,000 fields and subfields available to library cataloguers working to crate catalog records. While working on the Institute of Museum and Library Services (IMLS)-funded project to establish a Z39.50 interoperability test bed, Moen and Miksa (2005) found that very few of these fields were actually being used. In fact, they discovered that only 36 of the available MARC fields accounted for 80 percent of all utilization. These preliminary findings have important implications for library catalogues and other library and information science professionals (Moen and Miksa, 2005). A number of commentators (e.g. Moen, 2001; Besser, 2000; Sutton, 1999) are optimistic that the core element set will be as minimal as possible. Thus, the core element set meanings will be easy for most users to understand

and the element set will be flexible enough for description of diversified resources in a wide range of subject areas (Alemneh *et al.*, 2002).

The objective of Dublin Core is to define a simple set of data elements so that authors and publishers of internet documents can create their own metadata records with no extensive training. The Dublin Core approach is to have the level of bibliographic control midway between the detailed approaches of MARC and "structured" TEI, and the automatic indexing of locator services such as Lycos. It is acknowledged that the Dublin Core is a minimal set, and that many "publishers" or metadata producers may wish to augment this simple set with more specialized data. Because of this, therefore, Dublin Core has been selected as the metadata schema to organize ceramics information resources – namely articles of journals, reports, standards, newspaper clippings, books, etc. – and the use of all 15 elements for all types of collections was not enforced. Local qualifiers are added when necessary to distinguish metadata that is refined beyond the core elements.

Another metadata set was defined using GEMS to organize ceramic art works. The metadata elements for this set were taken from the metadata standard "Categories for the Description of Work of Arts". So, metadata elements of the Dublin Core and Categories for the Description of Work of Arts are used to develop the digital repository on ceramics along with locally added metadata. Table II gives metadata details for the different types of resources covered in the repository. In addition, chapter and section title metadata are used for book, report and thesis collections wherever required.

Screenshots of the repository home page and various collections (book collection, article collection, report collection, thesis collection, standard collection, patent collection and ceramic art works collection) are provided in Figures 1-9. Creating an index of values to describe digitized objects is a vital component of the digitized process. Therefore, search indexes and browsing facilities are decided from the metadata.

Review of controlled vocabularies and thesauri

Good metadata uses standard controlled vocabularies to reflect the what, where, when and who of the content. Controlled vocabulary lists and thesauri offer consistency in terminology for use in elements like Subject (where the metadata creator wants to indicate what the resource is about). The more consistency that can be applied to this procedure, the more fruitful searches will be, both within one set of metadata records and across records held by different organizations. If multiple organizations describe their collections consistently by using terms from a controlled list, the common approach will reap great benefits during searches.

Examples of controlled vocabularies, include standard subject heading lists (e.g. Library of Congress Subject Headings), thesauri (e.g. the Art & Architecture Thesaurus) and taxonomic lists (e.g. TRITON, Taxonomy Resource and Index to Organism Names). Locally defined vocabularies, where appropriate, can be utilized. Classification systems (e.g. Dewey Decimal Classification) can also be used to provide subject access. There are also controlled lists for terms within particular disciplines. These are produced by authoritative bodies, and are often available online. The National Monuments Record Type (NMR), Humanities and Social Sciences Electronic Thesaurus (HASSET), The Art & Architecture Thesaurus

Table II.
Details of metadata used for different types of ceramic resources

Metadata	Articles	Patent	Standard	Reports	Thesis	News clippings	Books	Ceramic art works
	Ceramic resources							
Title	✓	✓	✓	✓	✓	✓	✓	✓
Creator/artist	✓	✓	✓	✓	✓	✓	✓	✓
Subject and keywords	✓	✓	✓	✓	✓	✓	✓	✓
Publisher	✓	NU	✓	✓	NU	NU	✓	NU
Description	✓	✓	✓	✓	✓	✓	✓	✓
Date	✓	✓	✓	✓	✓	✓	✓	✓
Type	✓	✓	✓	✓	✓	✓	✓	NU
Identifier	✓	✓	✓	✓	✓	✓	✓	NU
Source	✓	NU	NU	NU	NU	✓	NU	NU
Patent no.	NU	✓	NU	NU	NU	NU	NU	NU
Country	NU	✓	NU	NU	NU	NU	✓	NU
Standard no.	NU	NU	✓	NU	NU	NU	NU	NU
Approving body	NU	NU	✓	NU	NU	NU	NU	NU
Report no.	NU	NU	NU	✓	NU	NU	NU	NU
Guide	NU	NU	NU	NU	✓	NU	NU	NU
University	NU	NU	NU	NU	✓	NU	NU	NU
Award year	NU	NU	NU	NU	✓	NU	NU	NU
Object/work	NU	NU	NU	NU	NU	NU	NU	✓
Measurements	NU	NU	NU	NU	NU	NU	NU	✓
Materials and techniques	NU	NU	NU	NU	NU	NU	NU	✓
Volume	✓	NU	NU	NU	NU	NU	NU	NU
Issue	✓	NU	NU	NU	NU	NU	NU	NU
Pages	✓	✓	✓	✓	✓	NU	✓	NU
Contributor/editor	NU	NU	NU	NU	NU	NU	✓	NU
Series	NU	NU	NU	NU	NU	NU	✓	NU
Period	NU	NU	NU	NU	NU	NU	NU	✓
Location	NU	NU	NU	NU	NU	NU	NU	✓

Notes: NU, not used; ✓, available

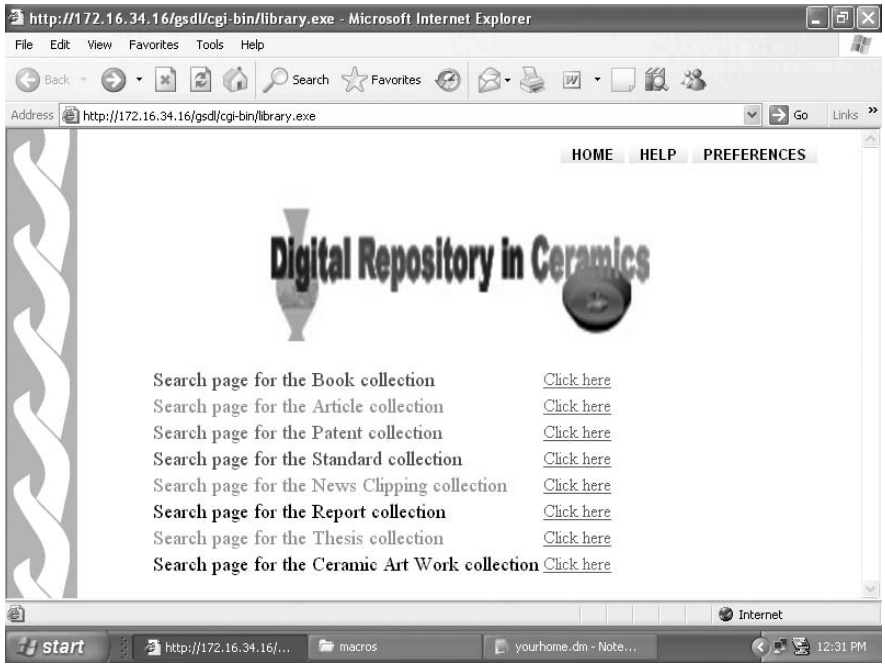

Figure 1.
Home page

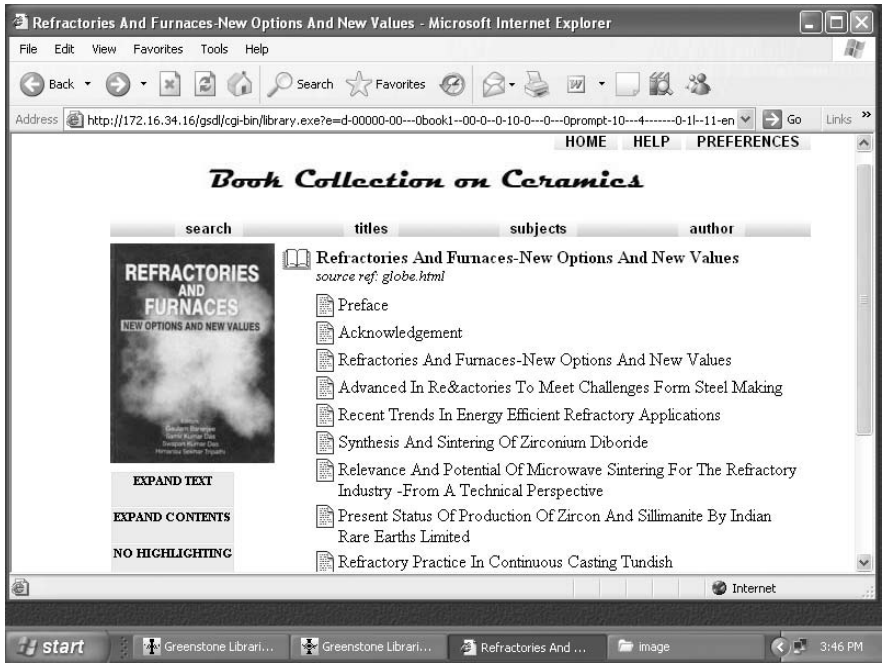

Figure 2.
Screenshot of sample "Book collections on ceramics"

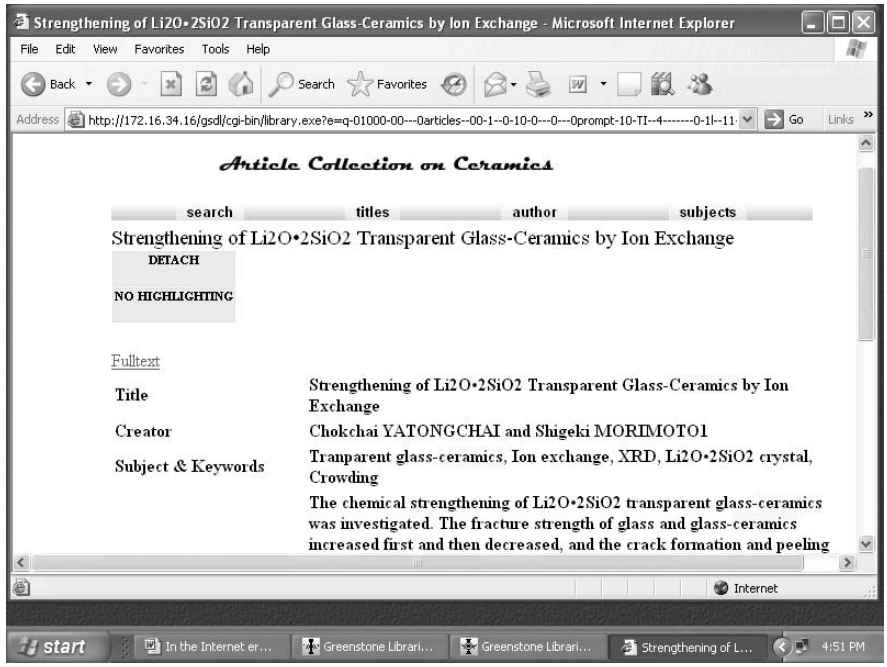

Figure 3.
Screenshot of sample "Article collections on ceramics"

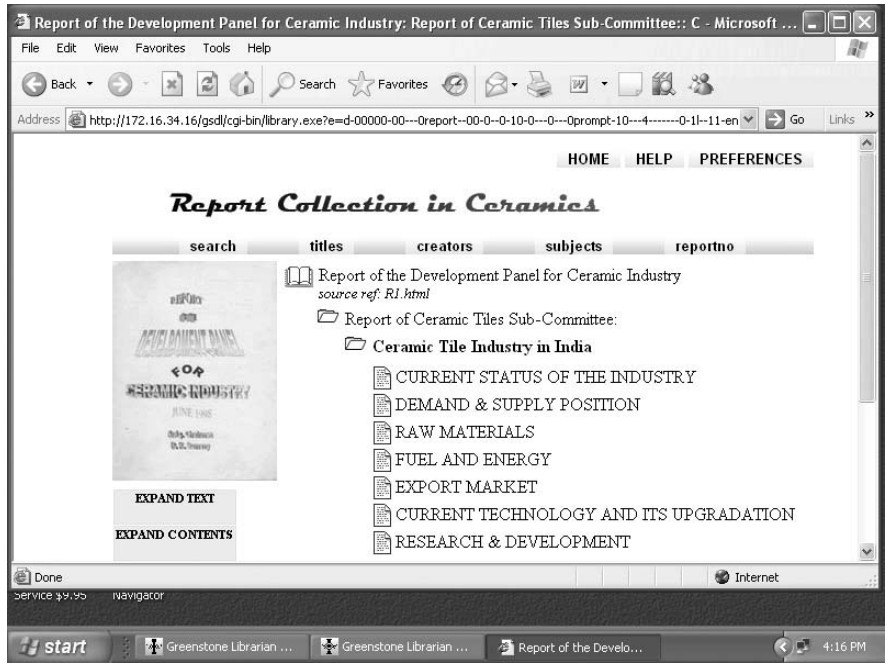

Figure 4.
Screenshot of sample "Report collections in ceramics"

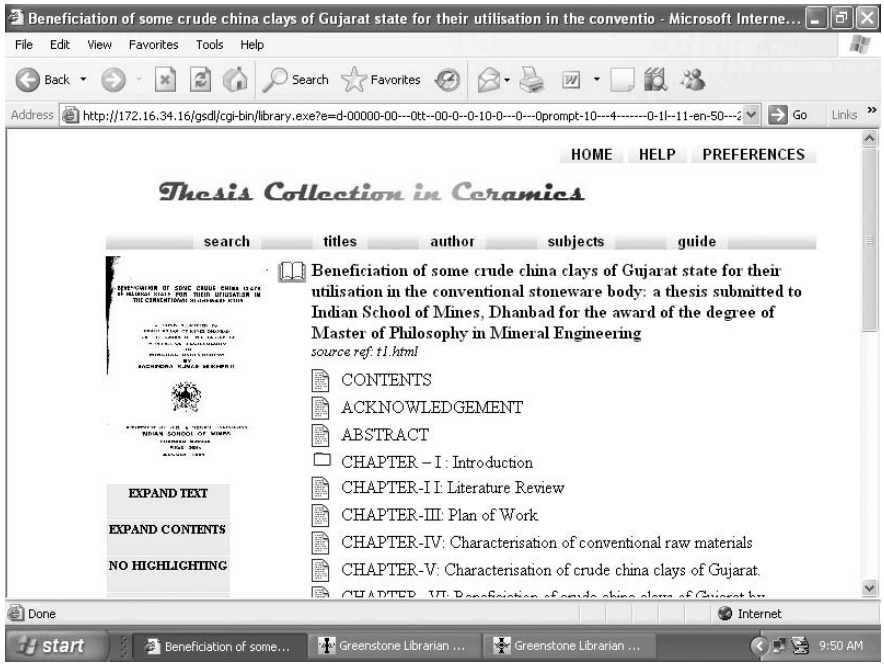

Figure 5.
Screenshot of sample "Thesis collections in ceramics"

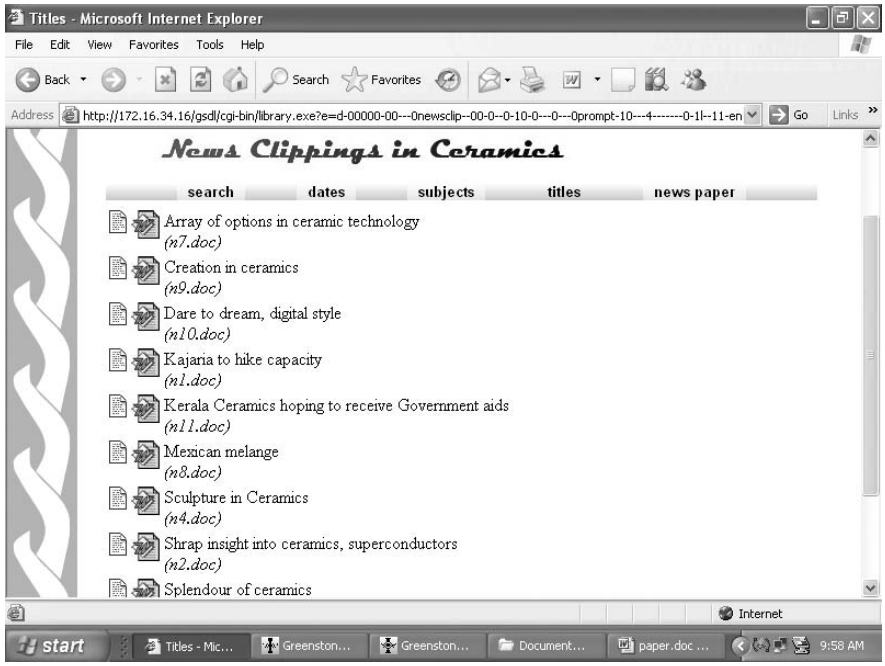

Figure 6.
Screenshot of "News clippings in ceramics"

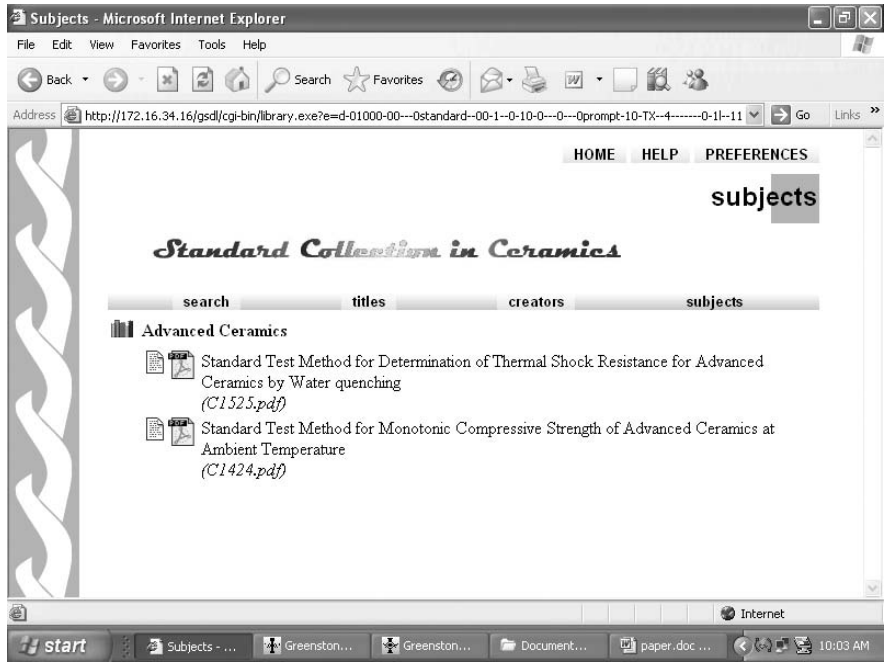

Figure 7.
Screenshot of "Standard collections in ceramics"

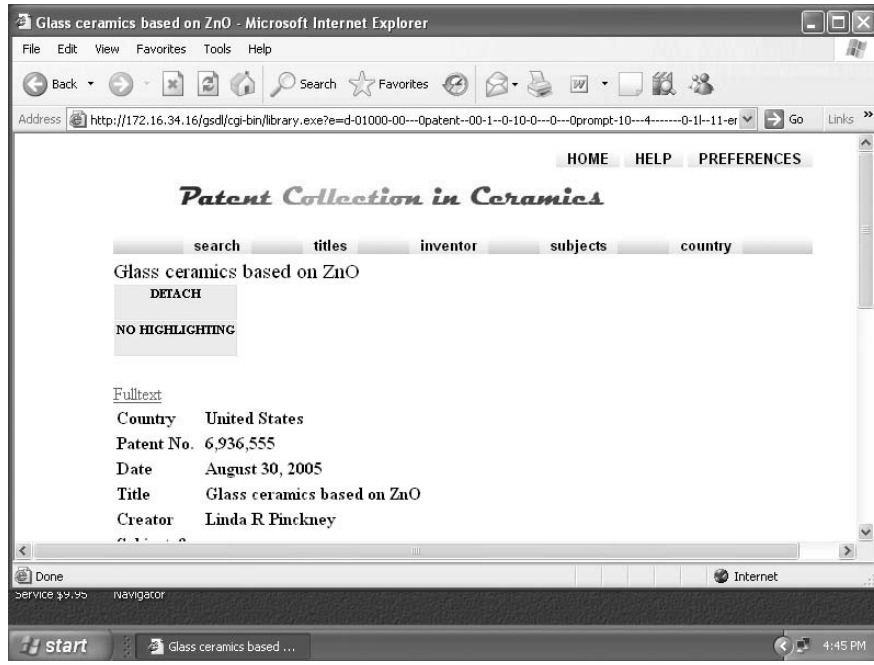

Figure 8.
Screenshot of "Patent collections in ceramics"

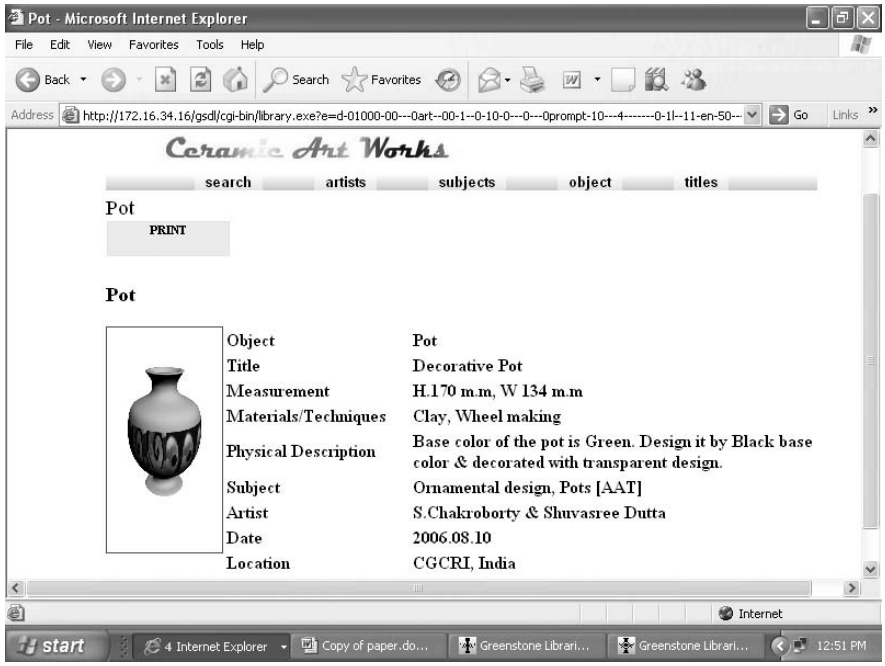

Figure 9.
Screenshot of sample "Ceramic art work" collection

(AAT), Union List of Artists Names (ULAN) and Thesaurus of Geographic Names (TGN) are all examples of these. The last three were all developed at the Getty Research Institute in California, which promotes innovative scholarship in the arts and humanities[5]. The WWW lists several dozen web-accessible controlled vocabularies by subject area[6]. The High Level Thesaurus Project (HILT) is a clearinghouse of information about controlled vocabularies, including related resources, projects, and an alphabetical list of thesauri[7]. Some other controlled vocabularies are Medical Subject Heading List[8] Thesauri of Graphic Materials I[9] Thesauri of Graphic Materials II[10] and the Thesaurus of Graphic Names[11]. In addition to these, there are also established standards for expressing elements like date, type and language, such as ISO 639 for language abbreviations and RFC 2045 and 2046 for Media (MIME) types.

For the ceramics repository, the standard terms are collected mostly from the Art & Architecture Thesaurus and Library of Congress Subject Headings. For example, if a group of resources on vase all attach the Subject Heading "vases" (which comes from Getty Art & Architecture Thesaurus; AAT), as opposed to making up their own headings (e.g. vase), individual records will not slip through the net during a search. It is important also to state which list a term is selected from (e.g. by putting [AAT] in brackets after the heading). Similarly, the term "pots" needs to be used for the word "pot" as per the AAT thesaurus. It is felt that existing controlled vocabularies are not sufficient to standardize the content metadata of the repository. Some words simply

did not exist in the existing controlled vocabularies. Domain-specific metadata will describe the content of ceramics objects, enabling a more precise description.

Conclusion

Good metadata should be appropriate to the materials in the collection. Metadata is simply a method used to describe an electronic document (data) and it can be done at the point of creating the document or afterwards. Although many metadata formats and sets have been developed, it is difficult to choose appropriate format to organize and describe ceramic resources. It is also agreed by the developer of various metadata communities that minimum metadata elements can be taken to uniquely and unambiguously identify and describe a particular work. However, which elements can be taken indeed should vary depending upon the end-users whom the particular information system are intended to serve, the mission of the specific institution, and a number of other factors. Dublin Core is one of the best known metadata sets created to describe electronic resources. Categories for the Description of Works of Art was developed for art and architecture specialists and provides a structure for describing works of art and electronic images of them. Therefore, selected metadata elements of Dublin Core and Categories for the Description of Work of Art are used to describe and organize the ceramics resources. Content metadata are standardized on the basis of existing controlled vocabularies and it is found that these are not sufficient to standardize the content of the repository.

Notes

1. See http://en.wikipedia.org/wiki/Digital_repository
2. See www.ukoln.ac.uk/repositories/digirep/index/What_is_a_Digital_Repository%3F
3. See www.getty.edu/research/institute/standards/cdwa/index.html
4. See www.greenstone.org
5. See www.getty.edu/research/institute/vocabulary/index.html
6. See www.lub.lu.se/metadata/subject-help.html
7. See http://hilt.cdlr.strath.ac.uk/Sources/index.html
8. See www.nlm.nih.gov/meshhom.html,
9. See http://lcweb.loc.gov/rr/print/tgm1/
10. See http://lcweb.loc.gov/rr/print/tgm2/
11. See www.gii.getty.edu/vocabulary.tgn.html

References

Alemneh, D.G., Hastings, S.K. and Hartman, C.N. (2002), "A metadata approach to preservation of digital resources: the University of North Texas Libraries experience", *First Monday*, Vol. 7 No. 8, available at: www.firstmonday.org/issues/issue7_8/alemneh/index.html

Duguid, P. and Atkins, D.E. (1997), "Report of the Santa Fe Planning Workshop on Distributed Knowledge Work Environments: digital libraries", available at: www.si.umich.edu/SantaFe/

El-Sherbini, M. (2001), "Metadata and the future of cataloging", *Library Review*, Vol. 50 No. 1, pp. 16-27.

Goh, D.H., Chua, A., Khoo, D.A., Khoo, E.B., Mak, E.B. and Ng, M.W. (2006), "A checklist for evaluating open source digital library software", *Online Information Review*, Vol. 30 No. 4, pp. 360-79.

Lesk, M. (2005), *Understanding Digital Libraries*, 2nd ed., Morgan Kaufmann, San Francisco, CA.

Moen, W.E., Miksa, S.D. and Marsh, C. (2005), "Examining present practices to inform future metadata use", *D-Lib Magazine*, Vol. 11 No. 6, available at: www.dlib.org/dlib/june05/06inbrief.html#MOEN

Patra, C. (2006), "Developing a digital library on ceramics", *The Electronic Library*, Vol. 24 No. 4, pp. 453-69.

Witten, I.H. and Bainbridge, D. (2005), "Creating digital library collections with Greenstone", *Library Hi-Tech*, Vol. 23 No. 4, pp. 541-60.

Zhang, A.B. and Gourley, D. (2005), "Building digital collections using Greenstone digital library software", available at: http://gourley.wrlc.org/IRSQ-article/

About the author

Chandana Patra is currently working as Technical Officer at the Library and the Data Bank of the Central Glass and Ceramic Research Institute, Kolkata, India. She has a B. Lib. and Inf. Sc. degree from Vidyasagar University, India and an Associateship in Documentation and Information Science degree from the Documentation Research and Training Centre, Indian Statistical Institute, India. She has an experience of more than 15 years in various roles – teacher, professional and administrator – in library and information science and during this time she has witnessed many changes and challenges in her libraries, including introducing internet services, e-journal services, and the automation system Libsys for the library. She is a member of the Editorial Committee of her Institute's newsletter. She has also presented papers at national conferences and seminars and has published several articles in national and international publications. Her major areas of research interest are digital libraries, management of e-resources, internet services, information services to users, and patent information services. Chandana Patra can be contacted at: cpatra@cgcri.res.in