Metadata Diversity in the Cultural Heritage Repositories

D

Sumeer Gul
University of Kashmir, India

Shahkar Riyaz Tramboo
University of Kashmir, India

Humma Ahangar
University of Kashmir, India

INTRODUCTION

Metadata is a term that only entered archival glossaries in the 1990s to refer to all or any of the various traces and statements (or 'data about data') that are made by or about records and recordkeeping structures, processes and contexts, especially in digital recordkeeping, online description, and digitization (Gilliland, 2011).

Metadata is increasingly recognized as key infrastructural component and knowledge management tool that encompasses far more than the descriptive information that is created by archivists or other information and heritage professionals (Gilliland, 2011). With the rapid growth of Internet, research on digital libraries and digital museums dealing with heritage collections has received worldwide attention (Bekaert, Ville, Rogge, Strauven, Kooning, & Walle, 2002; Chen, Chen, Chen, & Hsiang, 2002). Digital information is now an integral part of our culture and heritage. Cultural heritage encompasses all contemporary demonstrations, when intangible, and past evidences, in the case of tangible artefacts, of human creative activity that are inherited from previous generations and considered by communities, groups or society at large to be of value, and therefore maintained in the present and transmitted to future generations for their benefit (Roders & Oers, 2011). Tangible cultural heritage includes monuments, groups of buildings, sites and cultural landscapes (United Nations Educational, Scientific and Cultural Organisation (UNESCO), 1972), while intangible cultural heritage includes the practices, representations, expressions, knowledge, skills of communities and groups, and sometimes individuals, as well as the instruments, objects, artefacts and cultural spaces associated therewith (UNESCO, 2003). Cultural institutions at the local, regional, national and international levels now actively digitise the cultural and heritage resources in order to stabilise and protect those resources so that they will be permanent and durable besides being retrievable, readable and usable overtime. Making resources available is important, but ensuring accuracy in resource discovery is vital for future reference (Manaf, 2006) and resource discovery with the aid of metadata has resolved this problem to a greater extent including the ones related to culture and heritage artefacts.

Different types of metadata schemes represent heterogeneous digital assets. Digital objects showcasing culture and heritage objects that are currently receiving an important place in the society are described by some specific surrogates which are somewhat different from the general types of digital objects. Various types of schemas evolved from time to time to describe the digital artefacts representing the culture and heritage in one form or the other which Baca (2003) justifies by saying that, there is no *"one-size-fits-all"* metadata scheme.

BACKGROUND ABOUT THE METADATA SCHEMAS STUDIED

Dublin Core Metadata Element Set (DCMES)

The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. The elements are intended to be the most significant pieces

DOI: 10.4018/978-1-4666-5888-2.ch178

of information by which a user might seek an electronic resource. The fifteen elements are also known as simple (unqualified) Dublin core. These consist of contributor, coverage, creator date, description, format, identifier, language, publisher, relation, rights, source, subject, title and type (Dublin core metadata initiative, 2013). Miller (2011) reveals that Dublin Core has grown far ahead of the 15 element set and can be qualified and extended to meet the requirements of a wide variety of communities. Qualifiers could be added to the 15 elements to further refine their meaning and enrich the DC metadata scheme. Greene and Meissner (as cited in Fear, 2010) reveal that Dublin core is OAI-PMH compliant that makes it attractive for use in libraries and archives also. A study on Dublin core metadata element set beyond the element set by Harper (2010) reveals that despite many advantages, one of the major weaknesses of Dublin core schema is that the standard does not offer the richness and specificity required for resource description. Urban (2012) in his study on *"Principle Violations Revisiting the Dublin Core 1:1"* reveals that in order to distinguish between records describing original resources and records describing surrogates, Dublin core metadata initiative introduced the *"one to one principle" (DCMI 1:1).* In practice, however large number of cultural heritage institutions creating metadata for digital collections today fail to adhere to this principle. Instead, they commonly mix elements representing two manifestations together in a single metadata record, most often elements describing both an original analog resource and digitized version of that resource and thus, presenting many problems for resource discovery. Furthermore, the Dublin Core Element Set itself does not have a way to distinguish between metadata that applies to different manifestations of a resource when it appears within a single record (Miller, 2010). Caplan (as cited in Parnell, 2011) elucidates the problems caused by the simplicity and flexibility of the Dublin Core Schema.

Metadata Object Description Schema (MODS)

MODS an acronym for Metadata Object Description Schema, was developed by the Library of Congress Network Development and MARC Standards Office in 2002 for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. Moreover, as a metadata schema, MODS is richer than Dublin core and is simpler than the full MARC format (Metadata Object Description Schema, 2010). Guenther (2004) ascertains that MODS derives from MARC21, uses language based tags and is an XML encoded schema. As an XML descriptive standard, it provides the flexibility to be combined with other XML-based standards such as METS to satisfy needs for the digital library environment. McCallum (2004) reveals that the XML environment gives MODS to take advantage of XML for accommodating variable length data, explicit data tagging to multiple levels, hierarchical structure (even better than MARC), and all possible characters through Unicode. The author points out that unqualified Dublin Core has 15 elements and qualified Dublin core has twenty eight sub elements, whereas MODS has 20 top twenty elements and 47 sub elements. As a result MODS provides richer description. Furthermore, Dublin Core does not allow for distinguishing between types of name (i.e., personal, corporate) whereas, MODS allows for a type associated with the name element. Bountouri and Gergatsoulis (2009) in their study reveals that MODS is one of the most widely implemented metadata schemas for the description of (digital) material in the field of cultural heritage. Furthermore, MODS is recursive, thereafter it can include multiple hierarchies inside a single MODS record.

Visual Resource Association (VRA) Core (4.0)

The VRA Core is a data standard for the cultural heritage community. It consists of a metadata element set (e.g. title, location, date, etc.), as well as an initial blueprint for how those elements can be hierarchically structured. The element set provides a categorical organization for the description of works of visual culture as well as the images that document them (VRA Core 4.0, 2007). Various versions of VRA core schema have been released and the latest version is VRA core 4.0. This Core contains significant changes from the previous version, Core 3.0. The structural changes that have been made to VRA Core 4.0 were largely done to facilitate the development of the XML (extensible mark up language) expression of the Core. The VRA Core 4.0 XML schema has nineteen elements and twenty-three sub elements. Some of these elements