Enhancing Content Visibility in Institutional Repositories: Maintaining Metadata Consistency Across Digital Collections

Ahmet Meti Tmava and Daniel Gelaw Alemneh, Ph.D.

Introduction

With the development of new digital technologies in the last decade a growing body of the scholarly work produced by universities exists primarily in digital format. Academic institutions have increasingly recognized that an institutional repository (IR) is an essential infrastructure of scholarly dissemination. Thus, the number of IRs worldwide has increased exponentially in the last decade.

Institutional Repositories

The goal of IR is to collect, preserve, and make persistently accessible a variety of scholarly materials. With the exploding popularity of IRs, academic libraries are grappling with how to effectively integrate them into their services and collections. Generally IRs have a wide range of content and each institution has to make its own decision on what items and types of materials are to be included in an IR. Therefore, an IR is not simply a fixed set of software and hardware. It can take many forms and hold different content, use different IR software platforms and metadata schemas.

In light of the ever-changing information environment and user needs, digital curators and other stakeholders that are actively involved in supporting IR development try to deal with the task of maintaining consistency and the creation of and adherence to institutional-specific policies for metadata quality.

Consistency in Metadata Use

Metadata plays a key role in describing, accessing, and managing digital objects of different formats and media. The purpose of metadata is to offer the user multiple access points (e.g., author, title, subject, etc.). However if the metadata is incorrect, the resources in IR are not adequately represented and will remain invisible to the users. Shreves et al. (2005) explains that metadata problems can occur not only when elements are not applied properly but also when values are not consistently or accurately recorded. Literature review reveals five categories of metadata problems, “incorrect values, incorrect elements, missing information, information loss, and inconsistent value representation” Yasser (2011).

The majority of open access digital repositories have implemented the Open Archive Initiative-Protocol for Metadata Harvesting (OAI-PMH) as a mechanism to achieve interoperability in the exchange of meta-information with other systems. Metadata interoperability is in essence, to facilitate the exchange of information between repositories and to enable cross-collection searching. Accordingly, IR systems must be able to support interoperability in order to provide access via multiple search engines.

However, according to Chapman et al. (2009) because metadata is coming into the repository from many different streams, including directly from researcher themselves, it is difficult to enforce consistent use of metadata and entry of metadata values. This mixed metadata environment means that IRs face a number of challenges that more controlled environments do not.

The usage of Open Access Repository Software Worldwide

Recognizing the importance of quality and the strategic benefit of maintaining metadata consistency across the entire repository, many institutions employ a number of metadata quality assurance procedures and tools. Based on the University of North Texas Libraries’ experiences, Alemneh (2008) identified factors influencing metadata quality (such as: errors, omissions, ambiguities, etc.) and emphasized employing various quality assurance mechanisms (machine and/or manual solutions).

Summary

Consistency is the key factor in successfully managing different scholarly works in IRs and in making information more widely and easily available to users. Regardless of the repository software use or guidelines/standards metadata consistency plays a key role in describing and managing digital objects of different formats to effectively integrate the contents of IR into the existing services and collections. The greater consistency in metadata will lead to more faceted searching by sophisticated open source search engines. Indeed, it is the rich metadata that is consistently encoded that makes the digital items more discoverable.

References


