4B: Research Paper

Metadata in Museum Collections

Jessica Dadisman

Morgridge College of Education, University of Denver

LIS 4010: Organization of Information

Dr. Sarah Werling

Abstract

This research paper will be examining innovations in the field of museum collection metadata. Various projects and initiatives have been started to positively change the way metadata is handled in museums, most striving to create a standardized framework that turns the metadata into linked open data and makes collections more accessible to both museum staff and outside users. Another study will be examined in which the efficiency of human-generated subject keyword tags is compared to machine-generated tags, which could have implications for the future of museum collection records. Although metadata is treated differently in museums, efforts for reform are very similar to those in the library field.

Introduction

As stated in the abstract, this paper will be examining several innovations in the field of museum collection metadata. More specifically, the initiatives being discussed are the Natural Europe Project, the DOLMEN project, and the MOSC Project. Additionally, this paper will examine a case study on the Metropolitan Museum of Art's human-generated subject keyword tags compared to the machine-generated tags produced by Google Cloud Vision, Amazon Rekognition, and IBM Watson. Finally, an initiative at the Seattle Interactive Media Museum to create a metadata schema for video games will be investigated. These efforts point toward a desire to overhaul and reform how metadata is handled in museums and created a standardized framework that not only spans museums, but libraries and archives as well.

The Natural Europe Project

The Natural Europe Project is a European initiative to create a standardized metadata framework that allows cultural heritage object (CHO) metadata to be published as linked data, making it vastly more accessible both inside and outside European museums. The initiative was created to address several barriers that prevent end users from accessing the content available in European Natural History Museums (NHMs) (Skevakis et al., 2013). These barriers include the "lack of interconnection and interoperability between the management systems used by museums, and lack of centralized access through a European point of reference like Europeana, as well as the inadequacy of current content organization and the metadata used" (Skevakis et al., 2013, p. 1). Europeana is a centralized digital library, archive, gallery, and museum that provides online access to cultural heritage collections across Europe, and it was the hope of the Natural Europe Project to integrate it into their model as a point of reference.

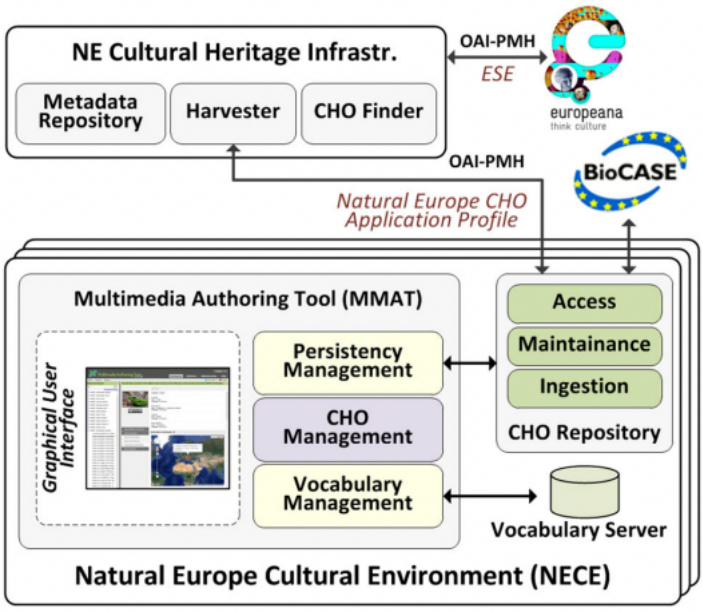

Fig. 1 *The overall architecture of the Natural Europe Project*

A CHO Application Profile was created for the Natural Europe Project as "a superset of the Europeana Semantic Elements (ESE) metadata format" (Skevakis et al., 2013, p.2). The application profile describes three main element categories for each CHO: the Cultural Heritage Object metadata category (for analog objects), the Digital Object metadata category (for digital objects), and the Meta-metadata category (includes the creator of the record and the languages in the metadata) (Skevakis et al., 2013). It is also of note that the Meta-metadata category "describes the history of the record during its evolution in the MMAT [Multimedia Authoring Tool], including the operations and entities that affected it" (Skevakis et al., 2013, p.2). Preserving the history of a record's edits in the metadata is an uncommon practice but could be invaluable to the accessibility and quality of a collection's metadata as a whole.

The Multimedia Authoring Tool (MMAT) is an integral part of the Natural Europe Project, as it is what allows interconnectivity between digital collections and Europeana, as well as BioCASE/GBIF, a transnational network of biological collections. It also creates

interoperability between NHMs and allows for the "seamless ingestion of legacy metadata" (Skevakis et al., 2013, p.3). MMAT supports three types of users: administrators, curators, and simple users. These three types of users are in a hierarchal format designated by how each user is allowed to interact with the data. Administrators have the ability to manage user accounts and the application overall, curators handle the administration of CHO records, and simple users have the singular ability to inspect the data (Skevakis et al., 2013). The client side of MMAT follows the Model-View-Presenter (MVP) design pattern and is "responsible for the interaction with the user, the presentation of the information, as well as the communication with the server when needed" (Skevakis et al., 2013, p.3). The server side of MMAT is comprised of three layers: the service layer (controls communication between the client and server logic), the business logic layer (contains the business logic of the application and keeps it separate from the other layers), and the data layer (external systems used for data storage) (Skevakis et al., 2013). The data layer contains the CHO Repository and the Vocabulary Server, which handle content/metadata and vocabulary/authority files respectively.

Following completion, The Natural Europe Project created a packaged version of the entire infrastructure, so that the MMAT, the CHO Repository, and the Vocabulary Server can be hosted in any museum on any server (Skevakis et al., 2013). As of 2013, the infrastructure has been implemented in the six NHMs who participated in the project. These museums have provided feedback on the operability of the infrastructure, and continued improvements have been made both to the user interface and the internal components of the package. Curator workshops were also held in which participants from various professions could interact with the MMAT and review their own CHO collections.

The DOLMEN Project

The DOLMEN project was essentially Canada's version of the Natural Europe Project. However, it was solely a research project, and did not continue into the development or implementation stages. This project also sought to transform museum collection metadata into linked data and create easier ways for end users to access collections. DOLMEN, a loose acronym for Linked Open Data: Museums and Digital Environment, was a project that proposed to "examine the fundamental elements for the description of museum objects and model them by using linked open data" (Fortier and Ménard, 2017, p. 486).

The DOLMEN project analyzed the metadata in online museum collections across Canada and created an inventory of the elements found and the frequency in which these elements were used. Researchers looked for any form of standardization across these elements, and then created ways to combine and simplify the myriad of elements in their inventory.

Metadata	Definition
Maker/Creator/Artist/ Manufacturer/Brand	Refers to the entity that creates, manufactures or produces the object or the work of art.
Material or medium	Refers to what the work is comprised of, made with. This represents materials based on their composition or origin.
Subject	Contains identification, description (or interpretation) of what is depicted by a work or image. It may include a concept, place, an activity, an event, a person, etc.
Object name	Contains the common name of the object or work of art.
Geographical origin	Contains the name of the general regions, continents, countries and further subdivisions where the object or the work of art is created.
Cultural origin	Indicates the historical, social, economic, religious or other cultural origins of the object or the work of art.
Style and period	Provides the names of distinct historical periods, broad cultural region styles and periods, art and architecture movements and groups and schools that are represented in the object or work of art.
Technique	Represents the processes, methods and means used to produce an object or a work of art.
Copyright	Contains the name of the entity that holds the copyright of the digital image of the object or the work of art and the copyright date.
Credit	Contains the credit line or acknowledgment to be used with the digital image of the object or the work of art.

Fig. 2 *Example of simplified elements created by DOLMEN*

Results showed that there was inconsistency and a lack of standardization in the museum metadata they examined (Fortier and Ménard, 2017). This created issues with access both on the internal level and the external. It was the aim of the DOLMEN project to analyze the systems these museums already had in place and find ways to simplify them into open linked data. The first phase of the project, outlined in the Fortier and Ménard article (2017), was completed and produced "a comprehensive picture of what museums already offer in terms of metadata associated to their online collections" (Fortier and Ménard, 2017, p. 485). It is now the hope of the researchers that this information can be used to produce a new, tangible system to be used by museums across Canada.

The MOSC Project

The Music of Social Change (MOSC) Project was an Institute of Museum and Library Services grant-funded initiative created by the MetaScholar Initiative of Emory University Libraries, in collaboration with the Center for the Study of Southern Culture, the Atlanta History Center, and the Georgia Music Hall of Fame. Their aim was to "develop a new model for library-museum-archives collaboration" (Roel, 2005, p.22) through the use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Specifically, this project focused on materials related to music and musicians who are associated with social-change movements. This protocol was chosen because it can function with unqualified Dublin Core (DC) metadata as well as any metadata format expressed by Extensible Markup Language (XML) (Roel, 2005). The OAI-PMH does not have many requirements for content, so it is optimal for contributors who have minimal metadata-encoding practices. On the other hand, OAI-PMH is also ideal for data formats that are highly structured. No fields in OAI-PMH are required by DC, so institutions can pick and choose what metadata is expressed if they have privacy concerns (Roel, 2005). Because

OAI-PMH has such a wide range for data formats, it was the hope of the MOSC Project to "maximize participation from fairly disparate kinds of organizations, with equally disparate kinds of metadata cultures and practices" (Roel, 2005, p. 23). OAI-PMH also eliminates the need to strip down metadata objects due to the cost of including them in aggregated systems (Roel, 2005).

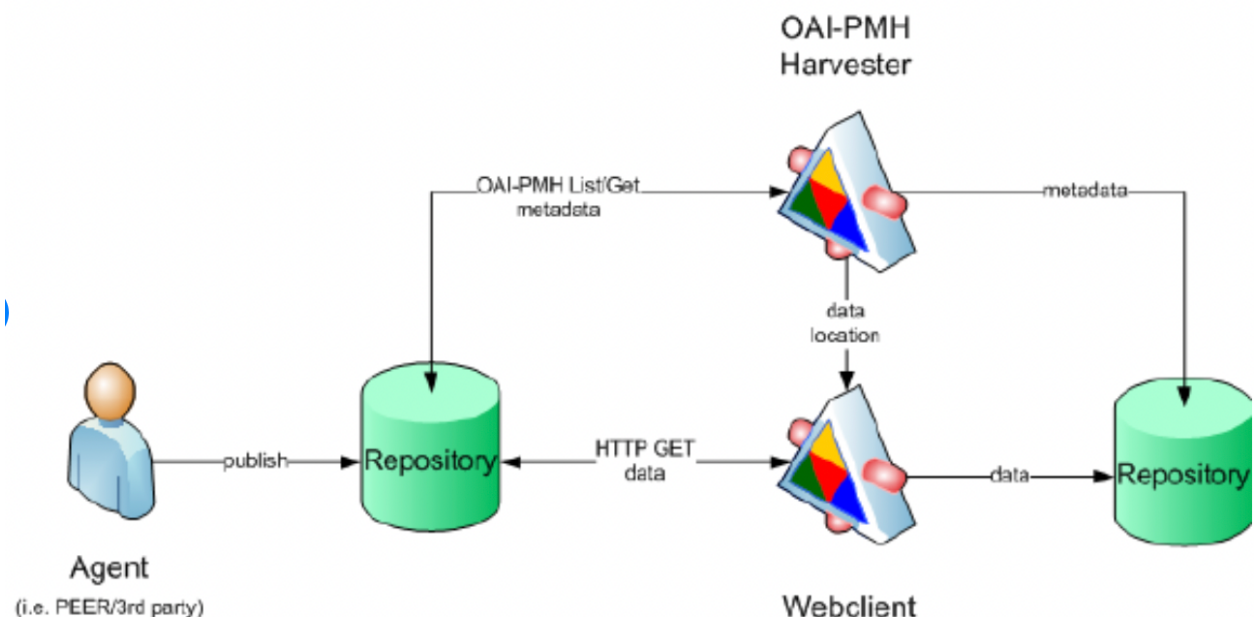

Fig. 3 *Flowchart illustrating an OAI-PMH data harvest*

Comparing Human and Machine-Generated Tags

Elena Villaespesa and Seth Crider (2021) conducted a case study based in the Metropolitan Museum of Art to compare the efficiency of human-generated subject keyword tags on museum objects and the tags produced by three different computer vision systems (Google Cloud Vision, Amazon Rekognition, and IBM Watson). These systems can "identify objects, text, colors, people and other subjects in images" (Villaespesa and Crider, 2021). The study looked at a sample of 1,323 objects from The Met's collection and brought to light a

number of similarities and differences between the human and machine-generated tags (Villaespesa and Crider, 2021).

Fig. 4 *The computer vision systems used in The Met case study*

The three computer vision systems returned a total of 3,145 tags and 31,065 records for the data set. These tags were then matched with the controlled vocabulary of the Getty's Art and Architecture Thesaurus, and results showed that most of the computer-generated tags fell under the "object" category, the most common facet represented by the museum's collection (Villaespesa and Crider, 2021). Furthermore, the computer vision systems produced a higher volume of tags for each object (~8 per object) as compared to the 2.68 tags produced via manual tagging (Villaespesa and Crider, 2021). The computer vision systems also produced a higher volume of unique tags. However, only a third of the tags manually produced by the museum were part of the computer vision system vocabularies (Villaespesa and Crider, 2021). Furthermore, the computer vision systems were only able to identify "what" was depicted in the artwork (e.g. person, dog, water) and not what the subject was doing (e.g. walking, swimming,

reading). The systems also could not identify the names of any actual people depicted in the artwork (e.g. George Washington in *Washington Crossing the Delaware*). These tags were only present in the records manually applied by the museum (Villaespesa and Crider, 2021).

The findings of the study have implications for the application of computer vision systems in museum record creation. These systems are capable of processing a high volume of objects "and complement the current taxonomies and vocabularies inputted in the collection database" (Villaespesa and Crider, 2021). Computer vision systems show much promise in the future of museum metadata practices, but they are not perfect. For the time being, there may be a great deal of value in combining both manual and AI-based collection tagging.

Creating a Video Game Metadata Schema

Jin Ha Lee, Joseph T. Tennis, Rachel Ivy Clarke, and Michael Carpenter make up a group of researchers who collaborated with the Seattle Interactive Media Museum to create a formal metadata schema for video games. Their aim was to create a schema that "can capture the essential information about video games and interactive media in a standardized way" (Lee et al., 2013, p.106) for the purpose of making navigation of game collections easier for end users. With the rise in popularity of video games and how deeply they are embedded in human lives, it becomes increasingly important to create a way to access this form of media in a scholarly way (Lee et al., 2013).

This team faced a number of challenges in their development of this schema. First, it is historically difficult to apply LIS description standards to non-book objects. Video games do not contain title pages, where a large amount of metadata is normally pulled. It is also becoming increasingly common for video games to be published in a strictly digital form ("born digital"), which creates another obstacle when trying to apply the standards of physical objects to them (Lee

et al., 2013). To address these challenges, the team borrowed and modified aspects of FRBR and CIDOC-CRM for purposes of interoperability. They then decided on 16 CORE elements for their schema, and proceeded with testing its usability (Lee et al., 2013).

At the direction of the Seattle Interactive Media Museum, the team described each game at the Manifestation level of the FRBR model, or, in other words, the "edition" of the game. However, as they developed more elements, they once again hit a snag with games that were "born digital." These games do not come with a box, so they need to be purchased and played in order to access any usable metadata (Lee et al., 2013). They found themselves turning to websites such as Amazon, GameFAQs, GameSpot, and Wikipedia for metadata information to use in the records. This especially came in handy when they needed to fill in the "Features" category, which would have been time-consuming for catalogers to describe if the information was not readily available (Lee et al., 2013). They also encountered some issues with using copyright dates when the release date is unknown, especially with games that are part of a series, as the copyright date only applies to the first manifestation of the game and does not apply to any of the sequels. The team ran into other challenges such as determining the publisher vs. developer, the differences in names based on region, and dealing with game series wherein each manifestation is not connected by story (Lee et al., 2013).

Despite these challenges, most of which were unique to video games, the team recognized that it only bolstered the importance and need for a standardized schema for this type of media (Lee et al., 2013). Now that the first phase is complete, the team hopes to develop a second version of this schema that will contain a larger number of CORE elements as well as controlled vocabularies. Their efforts will be useful not only to media museums, archives, and libraries, but to game developers, manufacturers, and distributors as well (Lee et al., 2013).

Table 1 The relative importance of metadata elements for the five personas

	Player	Parent	Collector	Academic	Designer	Curator
Title	●	●	●	●	●	●
Edition	●		●		●	●
Platform	●	●	●	●	●	●
Format	●	●	●		●	●
Developer	●	●	●	●	●	●
Publisher	●	●	●	●	●	●
Retail release date	●	●		●	●	●
Number of players	●	●	○	●	●	
Genre	●	●	●	●	●	●
Style	●				●	
Series/franchise	●	●	●	●	●	●
Region	●	●	●	●	●	●
Rating	●	●		○	●	
Language	●	●	●	●	●	●
UPC			●		●	●
Features		●	●	●	●	
Characters	●	○	●	●	●	
Theme	●	●		●	●	
Setting	●			●	●	
Summary	●	●			●	○
Images/screenshots	●	●	○	●	●	●
Gameplay video	●	●				
Awards	●	●		○	●	●
Achievements/trophies	●			●	●	
Credits	●				●	●
Packaging	●		○		●	●
Similar games	●		○			
Related media	●		○			○
Manual	●				●	●
Difficulty	●	●			●	
Price	○	●	○			○
System requirements	●	●	●		●	●

Fig 5. *A chart displaying the relative importance of metadata elements created for the schema*

Conclusion

A vast number of innovative ideas, initiatives, and studies have been conducted in the field of museum metadata. It seems that the general atmosphere across the globe is a desire to switch collections over to linked open data. After examining the three linked open data projects in Europe, Canada, and the U.S., it is clear that this endeavor will require great effort and collaboration, but, ultimately, could change the nature of museum metadata in a very positive way. Access to collections by end users can be simplified and streamlined, helping to close the gap between information and the public. Additionally, the future application of AI vision systems in object tagging is promising, but for now may still require manual tagging techniques.

Organization methods used in one of the first video game metadata schemas is also among the transformative ideas currently being discussed in the field. This could inform the way museums handle cataloging as well as include forms of media that have previously not had any kind of standardized framework.

References

Fortier, A., & Ménard, E. (2017). Laying the Ground for DOLMEN: Offering a Simple Standardization Starts with Understanding What Museums Do. *Knowledge Organization*, *44*(7), 485–493. https://doi.org/10.5771/0943-7444-2017-7-485

Lee, J., Tennis, J., Clarke, R., & Carpenter, M. (2013). Developing a video game metadata schema for the Seattle Interactive Media Museum. *International Journal on Digital Libraries*, *13*(2), 105–117. https://doi.org/10.1007/s00799-013-0103-x

Roel, E. (2005). The MOSC Project: Using the OAI-PMH to Bridge Metadata Cultural Differences across Museums, Archives, and Libraries. *Information Technology & Libraries*, *24*(1), 22–24. https://doi.org/10.6017/ital.v24i1.3360

Skevakis, G., Makris, K., Kalokyri, V., Arapi, P., & Christodoulakis, S. (2014). Metadata management, interoperability and Linked Data publishing support for Natural History Museums. *International Journal on Digital Libraries*, *14*(3/4), 127–140. https://doi.org/10.1007/s00799-014-0114-2

Villaespesa, E., & Crider, S. (2021). A critical comparison analysis between human and machine-generated tags for the Metropolitan Museum of Art's collection. *Journal of Documentation*, *77*(4), 946–964. https://doi.org/10.1108/JD-04-2020-0060