Representation Patterns for Cultural Heritage Resources

Richard J. Urban
Representation Patterns for Cultural Heritage Resources

Richard J. Urban
School of Library and Information Studies
College of Communication and Information, Florida State University
rurban@fsu.edu

ABSTRACT
The universe of available cultural heritage metadata schemas grows more complex every day. Existing schemas are optimized for use in the library, archive, or museum domains and to fit the needs of shared services and applications. Emerging Linked Data approaches introduce additional challenges for metadata designers and creators responsible for implementing these standards. In other domains, design patterns are used to clearly articulate problems, their contexts, and available solutions. This poster introduces preliminary research to identify such patterns in cultural heritage metadata standards using content analysis and a participatory design methodology.

Keywords
Cultural heritage, metadata, Linked Data, design patterns

INTRODUCTION
Representative surrogates that stand in for library, archive, and museum (LAM) resources are an essential component of information organization systems. How these representations are constructed is a fine balance between the domain being modeled and the objectives of the information system. Currently, librarians, archivists, and museum professionals can choose from a large universe of representation standards (Riley & Becker, 2010). Each of these standards exhibits various strengths and weaknesses based on the problems they were engineered to address. In environments where a single representation standard is required, this can mean making compromises to find the best fit. Emerging Linked Data approaches enable metadata designers to readily mix vocabularies through common semantic models, such as the Resource Description Framework. However, this presents new challenges to understanding which parts of standards offer optimal solutions to particular representation problems.

Other domains, such as architecture, software and ontology engineering, and interface design have addressed similar problems through the introduction of design patterns. The Linked Open Data for Libraries, Archives, and Museums Patterns (LODLAM Patterns) project seeks to apply the concept of design patterns to the problems described above. By developing a representation pattern library, we will provide the cultural heritage community a navigational tool that is organized around problems and solutions rather than standards and specifications.

BACKGROUND
Libraries, archives, and museums have created multiple interconnected standards that govern construction and use of surrogate descriptions used in information organization systems (Riley & Becker, 2010). A major function of standardization is to enable the interoperability of metadata records through exchange and reuse. Despite these specifications, cultural heritage metadata often suffers from quality problems that are in part due to local variances in interpreting and applying shared standards (Park & Childress, 2009; Park, 2009; Shreeves et al., 2005).

Linked Data is emerging as an important new paradigm for defining representation standards for cultural heritage materials (Bizer, Heath, & Berners-Lee, 2009). Unlike top-down standards definitions, Linked Data takes a grassroots, bottom-up approach to modeling and sharing structured data. In moving away from record-centered approaches commonly used in the LAM sector, Linked Data opens the field to representations that better suit the complexities of cultural heritage resources.

The development of metadata standards – whether using traditional methods or emerging Linked Data approaches – is fundamentally a sociotechnical process (Bowker & Star, 1999; Lampland, 2009). As Linked Data grows in importance there is an opportunity to shape how we leverage current standards while taking advantage of new approaches to constructing representations. Elizabeth Churchill (2012) has called for increased participation of human-computer interaction (HCI) researchers in defining how data is represented, aggregated, and understood by both humans and computers.

Design patterns are an example of useful HCI concepts that have not yet been applied to the problems of cultural heritage metadata. Traditionally used in architectural and manufacturing settings, design patterns also have proven...
useful for software engineering and interface design (Dearden & Finlay, 2006; Gamma, 1995). In addition to providing ready access to problems and their solutions, design patterns can facilitate debates about best practice by providing a common technical lexicon (Dearden & Finlay, 2006). Because patterns can make problems and solutions explicit, they can serve as important educational tools for students (Chatzigeorgiou, Tsantalis, & Deligianni, 2008). Design patterns have also proven useful in more relevant conceptual areas, such as knowledge engineering, ontology design, and Linked Data (Blomqvist, Gangemi, & Presutti, 2009; Dodds & Davis, 2011; Gangemi, 2005; van Harmelen, ten Teije, & Wache, 2011). Although design patterns have been applied to interface challenges in LAMs, they have not been broadly applied to the development of metadata standards.

METHODO
The LODLAM Patterns project is a new initiative that seeks to identify and publish a pattern library. Our initial work has subjected contemporary metadata standards, specifications (Dublin Core, VRACore, CDWALite, LIDO, Europeana Data Model, MODS, Bibframe, etc.) to a content analysis process in order to identify potential patterns. Previous content analysis research has taken metadata schemes as whole units in order to classify their domains, objectives, and architecture (Greenberg, 2005). Here, we are looking inside of metadata standards to identify the solutions they present to particular representation problems. The following questions are asked of the standards under review:

- what representation features are present in the standard?
- what problem does this feature try to solve?
- what contexts/forces make this feature relevant?
- how does this feature resemble other features observed in other standards? (does it constitute a new pattern or is it an exemplar of one that has already been defined?)

Because the Linked Data community values openness, transparency, and broad participation, this research blends the content analysis of standards with distributed participatory design methods to solicit contributions (Danielsson, Naghsh, Gumm, & Warr, 2008). To capture the input of this LODLAM community, a “proto-pattern” wiki (http://www.lodlampatterns.org/protopattern) is enabling:

- iterative development of patterns
- an evolving classification/hierarchy of patterns
- identification of problems that need solutions
- organization of found examples that can inform pattern development

Although not required, the wiki also provides a template (see Table 1) adapted from Gamma, et al. (1995) and Meszaros and Doble (1998). Once proto-patterns reach a sufficient level of maturity, they will be published as a more formal pattern library.

<table>
<thead>
<tr>
<th>Surrogate Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem:</strong> How can I distinguish metadata about an original resource from metadata about a surrogate resource?</td>
</tr>
<tr>
<td><strong>Context:</strong> Cultural heritage repositories contain surrogate representations of resources they hold in their collections (i.e., a digital image that depicts a painting). Some document-based data management patterns may conflate these entities, resulting in confusing, incoherent metadata (Hutt &amp; Riley, 2005).</td>
</tr>
<tr>
<td><strong>Solution:</strong> Use One Graph Per Resource and properties that indicate a surrogacy relationship between the original and the surrogate.</td>
</tr>
<tr>
<td><strong>Related Pattern:</strong> • One Graph Per Resource (Dodds &amp; Davis, 2011)</td>
</tr>
</tbody>
</table>
| **Examples:** • DCMI Abstract Model/1:1 Principle (Powell, Nilsson, Naeve, Johnston, & Baker, 2007).
• Use one dcm:Description for original resources, and another dcm:Description for surrogates. Associate these descriptions by using URIs that identify each kind of resource as the subject of dcterms:source and dcterms:hasFormat statements.
• EDM Requirement 1: “maintain distinction between ‘provided objects’ (painting, book, movie, archaeology site, archival file, etc.) and their digital representations” (Europeana, 2011, p. 8).
• Use edm:ProvidedCHO for original resources and edm:webResource for surrogates.
• CCO: Clearly distinguish between Work Records and Image Records (Baca, Harpring, Lanzi, & Whitside, 2006, p. 2).
• Use a vra:Work node for metadata about original resources and vra:Image for metadata about surrogates.
• Use cdwalite:objectWorkTypeWrap elements to describe original resources and cdwalite:resourceWrap elements to describe the surrogate resource (J. Paul Getty Trust & ARTstor, 2006).

**Table 1: Partial Example of Surrogate Identity Pattern**

DISCUSSION AND FUTURE WORK
This poster presents initial work to identify and extract problems and solutions found in current cultural heritage metadata standards, organize and publish them as Linked Data design patterns, and to begin building an infrastructure that will facilitate participatory debate and discussion of patterns for cultural heritage data (Heidt & Solivan, 2012; Park et al., 2012). LODLAM Patterns aims to create a new way to organize our debates and discussions about representation standards for cultural heritage resources that is oriented around problems and solutions rather than standards and schemas. As LAM professionals navigate the paradigm shift from traditional representation methods to Linked Data and Semantic Web contexts, LODLAM Patterns will also provide a useful tool to crosswalk current domain knowledge.
REFERENCES


