Collections Interoperability - BTP Proposal - 6 July 2010

---

4 - Project Overview

[...]

4.4 Collections Interoperability

To support a broad range of scholarship in the humanities, scholars desperately need for distinct digital collections of research materials to become interoperable [1]. Interoperability must extend beyond support simply for resource discovery; scholars must be able to deploy tools and services across widely distributed collections without needing to be expert in every digital format used and every brand and version of repository software extent in academia today. Digital object descriptions must be rich enough and precise enough to support scholarly reference and allow the implementation of transformation and remediation tools and services that can facilitate digital information resource reuse and recombination, while simultaneously maintaining resource provenance adequate for scholarship.

Despite a variety of technical and policy challenges, we believe that Bamboo can make a significant contribution in this area, by defining standard methods for making digital content available to web services. Where existing protocols, practices or ontologies can be leveraged, we will do so, extending and profiling current community standards as required to meet the rigorous requirements of scholars. Simultaneously we will identify gaps in existing standards, and define new technical approaches, application profiles, and best practices as necessary. We will thus develop, adopt, and publish a set of guidelines, protocols, and specifications that will help content providers enhance interoperability by taking advantage of the Bamboo platform. We will also develop services that will gather usage data from collections. In this way, we can use the platform to track scholarly activity and the ways in which scholars use content collections. This in turn will allow us to understand where the efforts of Bamboo should be focused after the initial three-year building projects.

We will work with libraries, museums, and special collections, and we will work closely with such initiatives as Fedora / DuraSpace, campus enterprise content management activities, Hathi Trust, and CollectionSpace. Many of the Bamboo partners have extensive experience in meeting challenges in this arena, and we believe they will be able to integrate content collections with various content-management technologies in new and valuable ways. Further details about the ways in which Bamboo will fit into a larger ecosystem of humanities technology may be found in section 5.1 below.

---

1. In workshop 1 of the Bamboo Planning Project, at which participants were asked about current and future needs in the digital humanities, a significant number of participants raised the issue of collections interoperability and the related theme of content and tool interoperability. Because of the importance of this issue, the Bamboo Planning Project team created a strategic focus area on "Content Interoperability Partnerships" as one of the 11 major elements of Bamboo's 7-10 year program document, which was presented at workshop 4 of the Bamboo Planning Project. See: https://wiki.projectbamboo.org/x/jQGK. At workshop 4, approximately 30 institutional teams formally voted on which elements of this Program document Bamboo should focus on in the short term and which elements institutions were will to lead. In both rounds of voting "Content Interoperability Partnerships" was ranked in the top 2-3 categories. See https://wiki.projectbamboo.org/x/xYCR.

[...]

5 - Technical Approach

[...]

5.6. Collections Interoperability Services

Collections interoperability services deployed on the Bamboo Services Platform will deliver capabilities that are likely to fall in the following categories:

- metadata interchange using metadata interoperability and complex content description standards, such as RDF,1 OAI-PMH,2 OAI-ORE,3 ATOM Publishing Protocol,4 OAC,5 etc.
- facilities that exploit and support use of persistent resource URIs at appropriate levels of granularity and adhering to the principles and emerging best practices of the W3C Linked Data Initiative6
- proxied access to, aggregation of, or aggregated indexing of content stored in repositories that expose a CMIS-compliant 7 web services interface
- facilities to support updating and/or augmenting, in place or virtually, existing collections and/or collection metadata as transformations and new scholarly derivatives of content are generated
- transformation between broadly-used content formats, where need exists and opportunity permits, especially with respect to transformations that are prerequisite to consuming scholarly services deployed on the Bamboo Services Platform

The first three categories in this list address content metadata – the 'lower hanging fruit' in Collections Interoperability. Standards in these areas include those mentioned above as well as metadata semantic schemas, such as DCMI ("Dublin Core")8, MODS ("Metadata Object Description Standard"),9 and LIDO ("Lightweight Information Describing Objects"),10 and a spectrum of approaches to handling persistent references to digital content.11 - -While these standards and related practices are reasonably well-understood by humanists, librarians, and information technologists, experience suggests that profiling and more detailed guidelines and best practices will be required to support the level of content exchange and interactivity envisioned for Bamboo.

The fourth category in the list above addresses the increasingly dynamic nature of digital scholarly content. As scholars interact with distributed digital resources through Bamboo, they will create new relationships between content objects, new transformations of content and new derivatives of content. Bamboo will develop services that will gather usage data from these interactions and will help provide the basic information required to promulgate updates to resource metadata and the potential deposit of new resource derivatives in ways that maintain provenance and intellectual cohesion.

In approaching problems that require interoperability of content at a structural and semantic level, the fruit grows higher on the tree. TEI, for example, addresses structure of textual documents in a manner that enables some algorithmic operation across textual collections or corpora. Yet TEI's diverse variants, which have in many cases evolved to address concerns of real import to adopting scholars, gives rise to much frustration among textual researchers in the arts and humanities – and more so when complicated by uneven mapping of semantic meaning to structural markup. Interoperability hampered by differences in point-of-view is even less tractable (is the Royal Pavillion in Brighton, England situated on Le Terrain Crétacé or in the County of Sussex?12).

Proposed Collections Interoperability standards selection and adoption work will address these areas. To support selection, adoption, profiling, and development of digital object identifier and metadata interoperability standards and practices, some limited experimentation and interaction with both the Work Spaces and Scholarly Services tasks will be necessary to inform this process. We start with somewhat less of a foundation in regard to our understanding of the specific ways that user and service interactions with content and collections can extend and augment content metadata and sometimes the resources involved directly. We anticipate that substantive work on this issue will extend into the second 18 months of the project. The final category in the list above is a nod to initial service development progress possible in the near term on the more difficult set of interoperability problems, charting an achievable, high-value course between utopian goals ("everything will be interoperable") and solutions that apply only to a single collection or small groups of corpora.

---

1. RDF (Resource Description Framework): http://www.w3.org/RDF/

2. OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting): http://www.openarchives.org/pmh/

3. OAI-ORE (Open Archives Initiative Object Reuse & Exchange): http://www.openarchives.org/ore/

4. http://tools.ietf.org/html/rfc5023

5. OAC (Open Annotation Collaboration): http://www.openannotation.org/

6. http://www.w3.org/standards/semanticweb/data

7. CMIS (Content Management Interoperability Services): http://docs.oasis-open.org/cmis/CMIS/v1.0/os/cmis-spec-v1.0.html

8. DCMI (Dublin Core Metadata Initiative): http://dublincore.org/

9. http://www.loc.gov/standards/mods/

10. http://www.lido-schema.org/

11. Cf. "Persistent Identifiers, Considering the Options," (Emma Tonkin, Ariadne, Issue #56, July 2008); http://www.ariadne.ac.uk/issue56/tonkin/

12. Cf., for example, "Season on the Chalk," John McPhee, *The New Yorker*; March 12, 2007; http://www.newyorker.com/reporting/2007/03/12/070312fa_fact_mcphee

This right-side column is an artifact of "old-style" navigation. **Please remove sections, columns, and content of the right-side column when this page is next edited.** Cf. the Wiki navigation changes - July 2011 page for a 'how-to' on removing old-style page markup.