**PETER — Improving**

<0 images>

**Improving Compliance with Evolving Standards Using Computed Transformation of Digital Collections**

Peter, C., Granville, D., Eveleigh, A., Decker, E. and Henriot, C.

Factors adversely affecting the sustainability of digital collections have multiplied during the last decade. Many early implementations were inextricably connected with online projects, making them vulnerable to rapid technical evolution and ensuing obsolescence. There has been considerable research effort directed at decoupling Internet presentation from collection management and at exploiting network-distributed and virtualized collection deployment. Further, contemporary collection ecosystems have adopted internationally recognized metadata standards to gain a level of independence from deployment technologies. However, these trends have only shifted the focus of concern to the stability of such standards themselves: little research attention has focused on the portability of metadata infrastructure, which continue to evolve rapidly and grow in complexity. The proliferation of metadata definition activities and frequent revision of standards has made it difficult to develop implementation roadmaps, and standards has also become tainted by the competing interests of territorial industry associations and government agencies. Additionally, research funding by collections remains committed predominantly to new metadata development work, with only marginal consideration of long-term migration as standards evolve.

A related and potentially more serious problem arises from the growing recognition that *annotation* metadata must be structured independently of *object* metadata. As early as 1992 Bearman observed, ‘Descriptive standards for libraries and archives have historically been much more thorough in specifying how to describe the things within a collection than in specifying how to describe the context surrounding those things’. By 2009 the Open Annotation Collaboration was more specific, claiming, ‘The importance of annotating as a scholarly practice [and] limitations of existing tools supporting annotation of digital content has had a retarding effect on the growth of digital scholarship’.1 While progress has been made in sustainability of collection deployments and of object metadata, significant benefits and long-term effectiveness of annotation metadata—set out comprehensively in Lee’s ‘Framework for Contextual Information in Digital Collections’ paper (2011)—remain on the horizon. Even if standards-based metadata elements have been employed in creating a digital collection, there is little alternative today but to incorporate annotation information ad hoc into the ‘miscellaneous’ mechanisms provided by object metadata containers. This practice leads to both difficulty of use of annotation in research projects and the frequent loss of investment in annotation when collections must be migrated to new deployment technologies or metadata standards. The Open Annotation Data Model (OADM)2 Phase III US roll-out commenced in April 2013,3 with Europe following a few months later, and OADM remains the leading activity addressing the ‘creation, capture and preservation of contextual information’. However, OADM is natively RDF-based,4 and its effective integration with predominately XML-based collection ecosystems is currently challenging. Only after development of annotation tools that overcome this hurdle can sustainable metadata standards enter widespread use that address annotation as well as object metadata in a coherent manner. Significantly, such a breakthrough will itself create pressure for re-implementation of existing digital collection metadata solutions. Effective long-term accessibility of annotation metadata, even by current standards-based schemes, will require redistribution of annotation in existing collections and its encoding using structured annotation methodologies such as OADM. In turn, this will influence the future development of standards for object metadata encoding.

Gains in both accessibility and sustainability made through adoption of metadata standards are therefore contingent, for a multiplicity of reasons, on periodic re-implementation of collection metadata infrastructures as standards evolve. Strategies requiring human intervention on a per-asset basis to achieve this rapidly become untenable in large collections—even in the absence of funding restrictions. Developing alternate strategies to enable collections repeatedly to migrate metadata infrastructures is critical, not only to remain compliant with emerging standards, but also to gain the real functionality benefits sought from structured annotation by Bearman and Lee and, ultimately, for long-term sustainability. This paper presents progress with techniques for automating the transformation of metadata in response to evolution of standards—thereby making such migration possible.

Our approach arose in collaboration between the University of Westminster (UOW), the University of Heidelberg (UHEI), and the Institut d’Asie Orientale CNRS-Lyon (IAO-CNRS) on managing collections of contemporary Chinese imagery. In particular, after developing software migration tools for reclaiming existing digital collections at UOW that had become difficult to maintain because of obsolescence of support technologies, we elected to create new collections compliant with the VRA Core4 standard.5 This was achieved by building a mapping-driven exporter for our NoSQL migration platform (*freizo*),6 which generated VRA asset-by-asset. In this way, the results of the transformation could be viewed using the collection management interface and the exporter mapping updated to tune the VRA profile that was implemented—providing an iterative transformation cycle that, when approved, could be applied to the complete legacy collection. The resulting test collection—of Mao-era posters—was transferred to the UHEI Tamboti7 collection ecosystem to validate its VRA implementation. We report subsequent creation of a VRA-compliant digital collection combining the assets of both the UOW and UHEI poster collections. During this work, however, we noted the necessity of representing annotation (relatively mature in the case of the UOW collection, which was based on reclamation of an Internet project dating from 2001) within VRA description containers. The transformation approach enabled us to produce multiple versions of the poster collection automatically—implementing different annotation solutions in order to support research into linking both annotation tools and virtual research instruments with the collection. Consideration of potential response to VRA Core4 being superseded by future versions was also evaluated in terms of revision of the transformation mapping.

Parallel work in collaboration with IAO-CNRS, concerning historic photographs of Shanghai—part of the large-scale Virtual Cities8 project initiated by IAO—had highlighted similar concerns about sustainability of annotation using VRA description mechanisms. IAO has developed a range of innovative virtual research instruments able to exploit extensive geo-location information, which forms part of the annotation of this collection. Creating a VRA version of the collection using *freizo* transformation created potential vulnerabilities when interfacing the virtual research tools to the VRA collection; such tools might cease to operate correctly in the event of VRA Core4 being superseded were geo-location information to be encoded in miscellaneous VRA description containers. In contrast, we report results with an OADM-based annotation solution implemented using IIIF annotation extensions generated through an alternate *freizo* transformation of the collection.

**Notes**

1. http://www.openannotation.org/about.html (accessed 23 February 2015).

2. <http://www.openannotation.org/spec/core/> (accessed 23 February 2015).

3. <http://www.openannotation.org/phaseIII.html> (accessed 23 February 2015).

4. <http://www.w3.org/RDF/> (accessed 23 February 2015).

5. <http://chinaposters.westminster.ac.uk/zenphoto/> (accessed 23 February 2015).

6. http://www.data-futures.org/freizo.html (accessed 23 February 2015).

7. <http://www.asia-europe.uni-heidelberg.de/de/hra-portal.html> (accessed 23 February 2015).

8. <http://www.virtualshanghai.net/Photos/Albums> (accessed 23 February 2015).

**References**

**Bearman, D. A.** (1992). Documenting Documentation. *Archivaria*, **34**: 33–49.

**Lee, C. A.** (2011). A Framework for Contextual Information in Digital Collections. *Journal of Documentation,* **67**(1): 95–143, doi:10.1108/00220411111105470.