International Journal on Digital Libraries (2006) 6(2): 113–114
DOI 10.1007/s00799-005-0127-y

REGULAR PAPER

Michael L. Nelson · Herbert Van de Sompel

IJDL special issue on complex digital objects: Guest editors' introduction

Published online: 24 January 2006

Keywords Digital objects · Metadata · XML · Oais

1 Introduction

Although the Web of today looks very different from the Web of 10 years ago, digital library (DL) file formats still look much the same. Using terminology from the Open Archival Information System (OAIS), the submission information packages (SIPs) and archival information packages (AIPs) were frequently the same application specific formats of the dissemination information packages (DIPs).

Driven by the proliferation of enabling technologies such as XML, Resource Description Format (RDF) and Dublin Core, there has been increasing interest in the DL community to use *complex digital objects*: objects that aggregate data, metadata and sometimes services into a single, logical digital entity. For example, instead of just a bare PDF file, a complex object could aggregate the PDF, the descriptive metadata (perhaps in multiple formats), provenance and rights statements and links to format conversion services. While the end-user continues to receive the PDF, DLs would ingest and exchange the complex digital object, building extended services based on the additional encapsulated and aggregated data.

In much the same way that metadata formats arise from different communities, so have complex digital objects. In the realm of scientific and high-performance computing, data formats such as NetCDF [1], HDF [2], SmartFiles [3] and ELFS [4] have been in widespread use. In the commercial content provider industry, complex objects can be traced to the persistent stores of object-oriented databases [5, 6] and multimedia codecs (e.g. Apple's QuickTime). The evolution of this lineage is the MPEG-21 Multimedia Framework. Within DL community, the single biggest catalyst has been what is colloquially known as the Kahn–Wilensky Framework (KWF) that shaped the thinking of DL community and influenced the Metadata Encoding and Transmission Standard (METS).

Published as a Web page in 1995, it began as the architecture for the Computer Science Technical Report (CS-TR) project [7]. The CS-TR project later merged with the WATERS project [8] to form the basis for the Dienst protocol and the NCSTRL project [9]. Lessons learned in the implementation and deployment of Dienst later influenced the design of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [10]. The KWF influenced some of the thinking in the Dublin Core community, resulting in the Warwick Framework [11], which was later extended with "distributed active relationships" [12], which itself later evolved into Fedora. The KWF formed the basis for a prototype implementation for the Library of Congress National Digital Library Program [13]. The representation of metadata in digital objects in the NDLP influenced the Making of America II project [14], which gave rise to METS. The KWF also employed a URN implementation known as handles [15]. The handle system has been quite successful in its own right, and is the key technology in the implementation of digital object identifiers (DOIs) [16].

2 In this issue

The contributions of the KWF are specifically promoting handles as globally unique identifiers and defining the relationship between digital objects and repositories. But more generally, it has shaped the design of digital libraries for the last 10 years. Thus, this special issue begins with a reprint of the original KWF. The other papers in this special issue are best understood in this context.

There are two papers about object models in popular repositories, two papers about complex digital object

M. L. Nelson (✉)
Computer Science Department, Old Dominion University, Norfolk, VA 23529, USA
E-mail: mln@cs.odu.edu

H. Van de Sompel
Research Library, Los Alamos National Laboratory, Los Alamos NM 87545, USA
E-mail: herbertv@lanl.gov

formats themselves and two papers about how complex digital objects are used for digital preservation.

Fedora and DSpace are two popular repository systems with large user communities. Fedora has made the transition from research project to production quality system and is in use at many different sites. Although primarily known as a repository, in "Fedora: An Architecture for Complex Objects and Their Relationships," Lagoze, Payette, Shin and Wilper review the Fedora data model and describe how RDF is used to represent relationships between digital objects. While Fedora is positioned as middle-ware that requires integration into a local environment, DSpace is a ready-to-deploy repository system. "MIT's CWSpace Project: Packaging Metadata for Archiving Educational Content in DSpace", Reilly, Wolfe and Smith describe how METS and IMS-CP (another complex digital object format) are used with DSpace to support courseware.

There are two main complex digital object formats in the DL community: METS and MPEG-21 DIDL. METS comes from the Making of America II project. In "METS: Standardized Encoding for Digital Library Objects", McDonough gives the history and motivation for METS, along with working through illustrative examples. MPEG-21 DIDL, an official ISO standard, is just one part of the MPEG-21 standardization effort led by the digital entertainment community. In their paper "Representing Digital Objects using MPEG-21 Digital Item Declaration Language", Bekaert, de Kooning and Van de Sompel describe how MPEG-21 DIDL can be used in a DL environment.

The last two papers focus on digital preservation. In the first, "PANIC: An Integrated Approach to the Preservation of Complex Digital Objects Using Semantic Web Services", Hunter and Choudhury describe extensions to the METS and MPEG-21 formats for preserving Web multimedia objects. The authors also present an ontology and awareness infrastructure to disseminate local preservation knowledge. Finally, in "The Design of the VERS Encapsulated Object Experience with an Archival Information Package", Waugh gives an update about the Victorian Electronics Records Strategy (VERS) project. This is a long running project to preserve digital public records encoded in VERS Encapsulated Objects (VEO), a complex object format developed specifically for this project. Waugh gives a comparison of VEO and METS, as well as a review of VEO design decisions.

Acknowledgements We would like to thank the editorial board of IJDL for their support and especially thank Gary Marchionini for suggesting this issue. We also thank all of the referees for reviewing the submissions.

References

1. Rew, R., Davis, G.: NetCDF: an interface for scientific data access. Comput. Graphics Appl. **10**(4), 76–82 (1990)
2. Fortner, B.: HDF: the hierarchical data format. Dr. Dobb's J. **23**(5), 42–48 (1998)
3. Haines, M., Mehrotra, P., Van Rosendale, J.: SmartFiles: an OO approach to data file interoperability. In: Proceedings of the 1995 Conference on Object-Oriented Programming, Systems, Languages and Applications, pp. 453–466 (1995)
4. Karpovich, J.F., Grimshaw, A.S., French, J.C.: Extensible File Systems (ELFS): an object-oriented approach to high performance file I/O. In: Proceedings of the Ninth Annual Conference on Object-Oriented Programming Systems, Languages and Applications, pp. 191–204 (1994).
5. Zand, M., Collins, V., Caviness, D.: A survey of current object-oriented databases. ACM SIGMIS Database **26**(1), 14–29 (1995)
6. Bernstein, P.A.: Repositories and object oriented databases. ACM SIGMOD Record **27**(1), 88–96 (1998)
7. Arms, W.Y.: Key concepts in the architecture of the digital library. D-Lib Mag. **1**(1) (1995)
8. French, J., Fox, E.A., Maly, K., Selman, A.L.: Wide area technical report service: technical reports online. Commun. ACM **38**(4), 45 (1995)
9. Davis, J.R., Lagoze, C.: NCSTRL: design and deployment of a globally distributed digital library. J. Am. Soc. Inform. Sci. **51**(3), 273–280 (2002)
10. Lagoze, C., Van de Sompel, H.: The open archives initiative: building a low barrier interoperability framework. In: Proceedings of JCDL 2001, pp. 54–62 (2001)
11. Lagoze, C.: The Warwick framework: a container architecture for diverse sets of metadata. D-Lib Mag. **2**(7/8) (1996)
12. Daniel, R., Lagoze, C.: Extending the Warwick framework: from metadata containers to active digital objects. D-Lib Mag. **3**(11) (1997)
13. Arms, W.Y., Blanchi, C., Overly E.A.: An architecture for information in digital libraries. D-Lib Mag. **3**(2) (1997)
14. Hurley, B.J., Price-Wilkin, J., Proffitt, M., Besser, H.: The making of America II Testbed Project: a digital library service model. Council on Library and Information Resources Report 87 (1999)
15. Paskin, N.: Towards unique identifiers. Proc. IEEE **87**(7), 1108–1227 (1999)
16. Sun, S.X.: Establishing persistent identity using the handle system. In: Proceedings of WWW 2001 (2001)