Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002: 91-96
© Firenze University Press

Building Digital Books with Dublin Core and IMS Content Packaging

Michael Magee
Netera Alliance
magee@ucalgary.ca

D'Arcy Norman, Julian Wood, Rob Purdy, Graeme Irwin
University of Calgary
dnorman@ucalgary.ca, woodj@ucalgary.ca, rpurdy@ucalgary.ca, irwing@cpsc.ucalgary.ca

Abstract

The University of Calgary Learning Commons has been developing solutions for the digitization and dissemination of educational digital assets for two years. The most recent work focused on creating digital books that can be accessed online and assembled from digital components. Several theoretical and technical issues were examined and resolved. The U of C worked with existing partners to modify their educational object repository software solution to meet these needs. The software was developed to deal with the workflow of assembling the numerous digital components of a book into a cohesive whole and an online browser was built to view the constructed digital books. The digital books were created in an XML-based IMS container package of Dublin Core metadata and manifests of all the components that were used to create the online digital books.

Keywords: *Dublin Core, Digital Books, IMS, Metadata, XML*

Introduction

The efficient management of information is a driving force behind modern society. Canada is not immune from this trend. The Federal government looks upon the Internet as an opportunity to "fulfill its responsibilities in the generation and dissemination of information in a more effective and timely manner" [4]. This has led to a number of recommendations to enable the vision of creating an accessible body of digital content for Canadians and the rest of the world.

A large part of this mandate has been the digitization or re-purposing of existing content. The University of Calgary has been working with the CAREO (Campus Alberta Repository of Educational Objects) and the BELLE (Broadband Enabled Lifelong Learning Environment) projects to take existing educational content and place it online. As the body of existing content undergoes review it becomes obvious that a framework needs to be in place to deal with decisions of what content will be chosen for digitization and how it needs to be re-organized to work in an online environment. The spectrum of media that is a candidate for movement into the digital realm is considerable. These include text, video, film, photographs and a host of new and emerging multimedia formats. Books represent one of the oldest and one of the most challenging of these formats.

The intellectual cull of the vast herd of literature that currently exists is not an easy decision. As books are deigned worthy of preservation the ones that do not meet the criterion are lost to future generations forever. Other decisions may not affect the preservation of a resource but it will affect its accessibility to the general public, as more and more dependence is placed on online resources alone. The decisions are therefore not to be taken lightly as they may represent an intellectual and sociological bias that will affect our worldview.

Our project was not initially focused on the decisions about what books would be chosen for digitization. We had to examine how to create digital books so that they would be easy to find, accessible and of great utility. We saw the opportunity to explore a new approach to the creation of digital books. The IMS content package is an abstract container designed to describe a large, complex hierarchical data structure as well as its component media. It was designed to allow the movement of complex objects between systems and give them the ability to communicate with other software. The solution we chose is only one of many options available in the world of information technology and will need to be evaluated to determine its appropriateness as a new tool. It represents a solution that will combine the approaches currently being used in the world of educational technology and library science. The solution has a number of benefits that will create a digital book that is not only readable but has considerably more utility that just the book on its own.

This is an important consideration as there is a movement towards efficiency in the library world that will attempt to increase effectiveness by ending the physical book and replacing those anachronisms with pure data [5]. There need to be as many technological options as possible for the book to be properly evaluated. There is no doubt that many books will lack utility in some areas but will be extremely useful in others. The largest number of options available to those archiving and sharing those volumes will provide the most justice to the choices being made about which volumes will chosen.

Previous Research

There was a considerable amount of research that had occurred to examine the issues surrounding the search, retrieval and organization of educational digital assets online. This work became the basis for development of the mechanisms to deal with more complex organization of objects.

CAREO

The CAREO project is involved in the research and development of both a provincial and a national educational object repository [2]. Educational objects can be defined in a variety of ways but there are a few common elements. They are fundamentally small, digital instructional components that can be reused a number of times in different learning contexts and delivered over the Internet [10].

The system was created to address the problems of the explosion of online, digital educational content and the increasing difficulty in locating and utilizing that content.

As a result of this research CAREO developing a networked repository system that displayed XML document records based on IMS metadata, an educational metadata set. Although the focus was on educational applications the architecture was kept as flexible as possible so that any metadata standard based on an XML-schema could be stored.

The CAREO application software is designed to allow the search, retrieval and display of IMS metadata records in a web browser. These records are linked to educational objects located online. The CAREO repository is designed to be a modular component of a larger system (Figure 1). It has a built in communication layer based on XML-RPC that allows other repositories and tools to search and utilize the features of the software. In its current implementation, CAREO is integrated with the ALOHA (Advanced Learning Object Hub Application) metadata server which provides additional functionality in the role of a middleware layer between the CAREO application and the user's browser application.

For the user, the CAREO application acts as an educational portal or website, providing a central point of reference for educators and students when looking for information and resources to support teaching and learning. By providing a set of tools to enable such activities as resource discovery (searching and browsing), publication, aggregation, and sharing, CAREO is able to provide meaningful and immediate access to online materials.

By implementing the IMS Content Packaging specification, CAREO has been able to extend its suite of tools to enable its users to create compound aggregations of learning resources. These compound aggregations may range from simple collections of images into a single package, to electronic representation of physical books, to highly structured online courses.

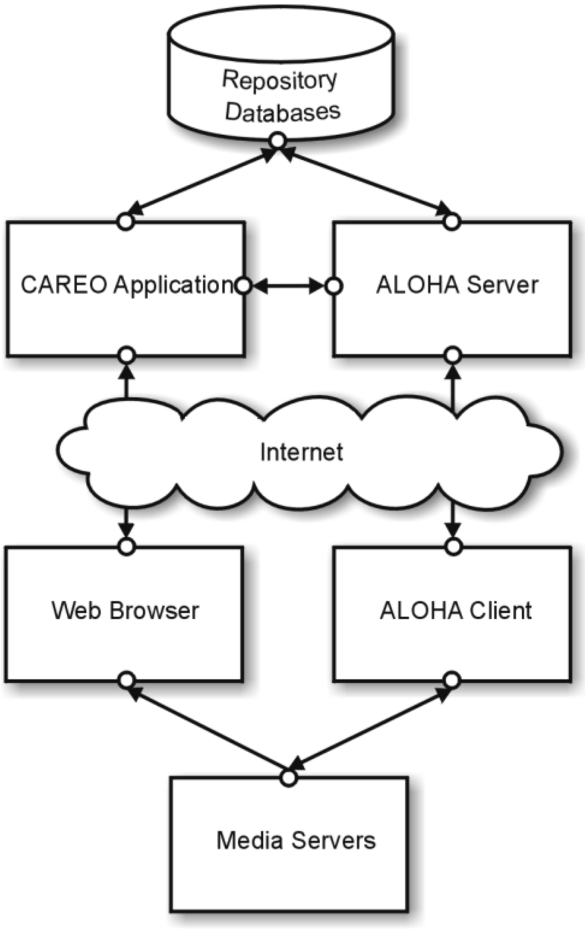
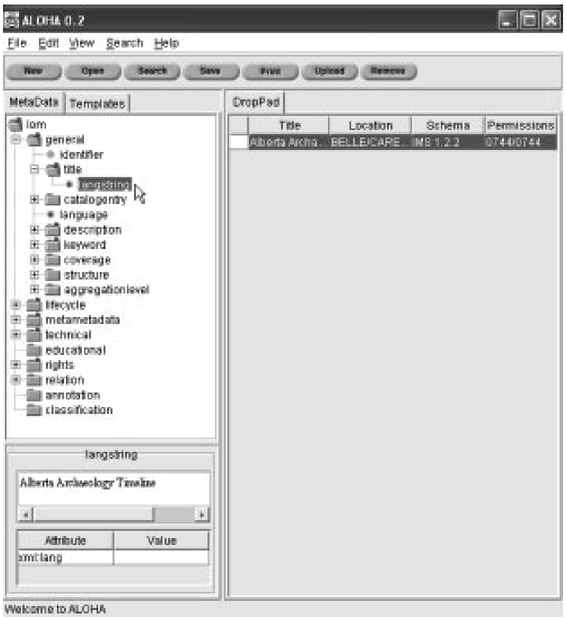


Figure 1. CAREO/ALOHA Architecture

ALOHA

The ALOHA java client was developed as the result of research early in the CAREO project. It indicated that the time and effort required to create metadata records for the individual educational objects was much too long for most of the projects and academics involved. A tool to streamline the upload of metadata and the media itself was therefore required to improve the workflow of objects into the system. The ALOHA tool is a java client that was designed to allowing indexing using any metadata standard. The decision to use Java was based on the power and flexibility Java has demonstrated in interacting with the World Wide Web [6]. The tool ingests a valid XML schema, creates a data entry interface based upon the schema and allows both amateur users and professional indexers all the simplicity and sophistication they require. It is easy to create, share and customize indexing templates and forms (Figure 2).

It also integrates drag and drop functionality that can automatically extract metadata from over 200 files types. It makes marking up IMS, or other forms of metadata, much easier. Administrative tools managing workflow issues with multiple indexers including the librarian, the educator, and the media developer are available. This supports the idea of modularity where different users can index objects in context specific ways and share their metadata with other users and metadata schemas [3]. Once the indexing is complete the media and metadata can be uploaded simultaneously with the touch of a button to an appropriate media-server, handling the job of an FTP program.



	Title	Location	Schema	Permissions
	Alberta Archa...	BELLE/CARE...	IMS 1.2.2	0744/0744

langstring

Alberta Archaeology Timeline

Attribute	Value
xml:lang	

Welcome to ALOHA

Figure 2. ALOHA Interface

The Problem

A search and retrieval system had been successfully implemented for educational objects but as more complex objects were examined it became apparent that a greater degree of sophistication would be necessary to deal with large constructs. Books were one of the most obvious assets the system would be unable to handle. The current system could handle a simple description of the book as a single entity. Unfortunately books, like many types of complex media, are composed of a large number of organizational structures such as sections and chapters that combine to make up their whole structure.

It was necessary to retain the search and retrieval features of the metadata but there was also a need to describe the actual structure of the object so that it could be assembled for online viewing in a meaningful way.

The Solution

The first physical asset chosen to test structural metadata was a book. Simplistically, books can be thought of as hierarchical arrangements of content. Words are aggregated into sentences, sentences into pages, pages into chapters, and chapters into books. In constructing digital books it was convenient to follow this pattern. The CAREO project wanted to continue to develop and support IMS metadata technology and therefore looked at IMS Content Packages as a solution. This involved treating the pages, chapters and books of the digital books as IMS content packages. There have been other solutions to the problems of digital books. The Library of Congress has taken a similar approach using METS (Metadata Encoding and Transmission Standard) [7]. The Open eBook standard has also been created to describe digital books in a standardized way [8]. In all cases structural metadata was created to describe the separate digital files that could make up a book

IMS Content Packaging

The IMS Content Packaging Specification 1.1.2 was designed to assist in the creation of complex educational content [1]. Basically, a content package allows numerous assets to be brought together, organized, and described. The educational objects within the package can have several organizational structures so that one content package can place them in a number of different contexts, educational or otherwise.

At the top level of the IMS content package is the Manifest (Figure 3). A Manifest consists of Resources, Organizations, Metadata, and optional sub-Manifests. The Resources are the actual digital files and/or links that makeup the package content.

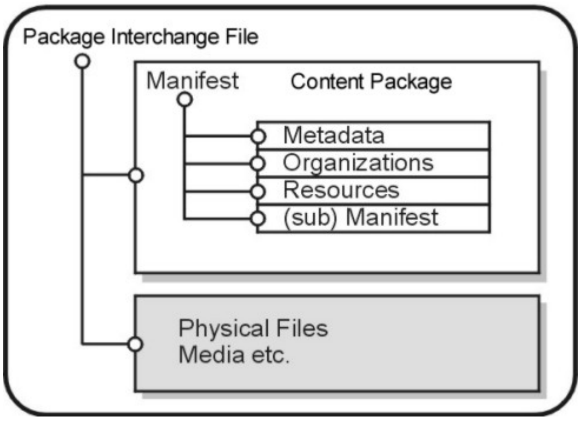


Figure 3. IMS Package Interchange File

The Organizations section is a hierarchical arrangement of the Resources. This is where the content is given order and context. The Metadata section provides an area to add descriptive metadata for the content package. Additionally, metadata can be optionally attached to nearly any part of the Manifest. It is important to note that this metadata can follow any metadata standard. Although IMS recommends that its own IMS metadata specification be used it is not required

An example of how a content package might be used to construct a book follows. First, a content package representing a chapter could be created. The pages that belong to that chapter are then added as Resources. Each page is put in proper order using the Organizations section and metadata could be added to each page to further describe it. This process can be repeated for each chapter in the book. Then, a content package representing the book could be created and each chapter added, ordered, and annotated.

The treatment of each part of the book as a separate package allows each section to be independently manipulated, searched, and created. The specification would also allow the XML file describing the manifest and all the physical files used in the package to be bundled together into a Package Interchange File that could be compressed into a single file and moved between systems. As the standard was published a number of vendors were creating software that could both build, package and exchange these Package Interchange Files. In the educational world this offered the opportunity to share not only atomic educational objects but also large aggregations that could be formed into lessons, courses and entire programs.

Test Case: Canadian Local Histories

The "Our Roots/Nos Racines" project is a project that was initially undertaken by the University of Calgary and Laval University. The project is doing an inventory and assessment of all media associated with Canadian cultural heritage. The goal is to get as much of that material into an online venue that is accessible to all Canadians [9]. Initially the library projects wanted to take the thousands of pages of local histories it had digitized and place them online as complete books for online access by genealogists and researchers. The first phase of the project created digital, online versions of local histories by scanning the books into graphic files and uncorrected OCR files. The combination of the two allowed for rough text searching and display of the actual page in a web browser.

Under the direction of Tim Au Yeung at the University of Calgary the initial proof of concept was successful but it became apparent that the system was going to need to scaled up to accommodate a number of other repositories, many more types of digital assets, larger volumes and data and increase in users. The Our Roots project consulted the CAREO project as it was involved in research and development in this area.

There were a number of similarities between the needs of the two projects. Both of them needed to research and develop ways of organizing and structuring large volumes of online content. Where they differed was in the type of metadata being used to describe the content packages. The library project was using Dublin Core metadata and it's own proprietary extensions but both systems had metadata as a focus if a search and retrieval system as a common element. The use of Dublin Core in the IMS Content Package did not represent a difficulty as the container package was designed to be generic and flexible enough that it could contain many kinds of descriptive metadata.

The limitation of the IMS Content Package in describing books was the generic nature that made it so useful in the first place. Other digital book standards were designed explicitly to describe books while the IMS standard could describe any combination of digital assets. This required that the structure had to be carefully defined while the books were being assembled. It was critical that the packages identify themselves as pages, chapters and books as there was no pre-defined slot for those elements.

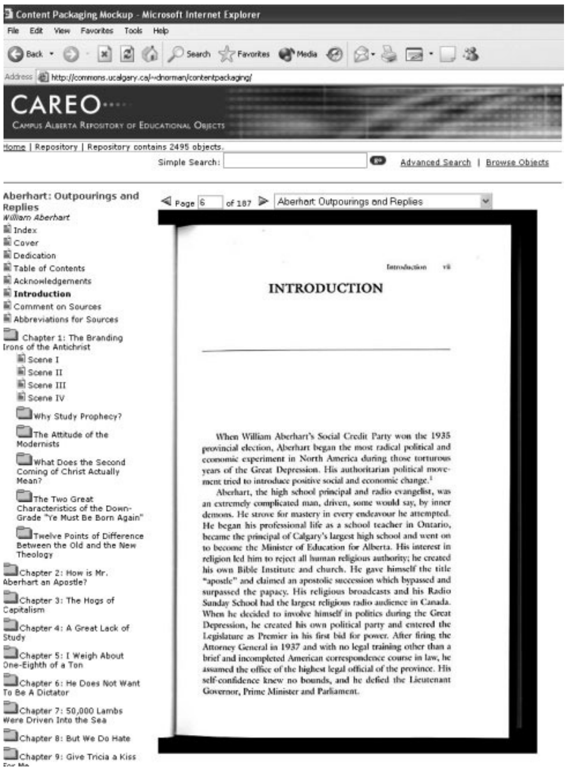
The IMS Content Packages were assembled in the ALOHA software from scanned components of the book. These included the text from the OCR and several sizes and formats of digital images of the page itself. ALOHA would treat them like digital assets and allow the users to organize them into pages, chapters and books. Once organized, the files were moved online. The digital files representing the pages of the books were moved to the media servers and the Dublin Core metadata representing the description of the various components of the book was moved to the metadata server. The CAREO software was used to display the digital books online.

CAREO used a browsing structure that allowed the books to be browsed, searched and read online (Figure 4). The search and retrieval aspect of the Dublin Core metadata used to describe the various components of the book would allow searching down to the level of page in the book.

Conclusion

IMS Content Packing presents one of a number of XML-based container package standards for digital books. The advantages of the IMS standard come from the generic nature of the content package. A book can be one of several media types all within the same package. This allows a book, chapter or page to be part of a larger, more complex multimedia presentation. This opens books up to a larger realm of opportunity than just the library.

The work of education is demanding the creation of many new Learning Management Systems based on IMS standards. These systems will be able to import and export IMS Package Interchange Files and present the IMS Content Packages to the students and teachers. When books are described using this standard these systems will be able to ingest a page, a chapter or a whole book as part of a course or a lesson. The packages could expand to include lessons and tests specific to a book and target audience. The packages will also have the ability to communicate with a LMS through a standardized API. This will allow instructors to track progress through the book and the test and score results of a student working through the book online. As more and more vendors create software and tools that can work with the IMS standard content that is described in this way will gain access to a greater level of utility and interoperability


Content Packaging Mockup - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
http://commons.ucalgary.ca/~dnorman/contentpackaging/
CAREO
Campus Alberta Repository of Educational Objects
Simple Search:
Advanced Search | Browse Objects

Home | Repository | Repository contains 2495 objects.

Aberhart: Outpourings and Replies
William Aberhart
Index
Cover
Dedication
Table of Contents
Acknowledgements
Introduction
Comment on Sources
Abbreviations for Sources
Chapter 1: The Branding Irons of the Antichrist
Scene I
Scene II
Scene III
Scene IV
Why Study Prophecy?
The Attitude of the Modernists
What Does the Second Coming of Christ Actually Mean?
The Two Great Characteristics of the Down-Grade "Ye Must Be Born Again"
Twelve Points of Difference Between the Old and the New Theology
Chapter 2: How is Mr. Aberhart an Apostle?
Chapter 3: The Hogs of Capitalism
Chapter 4: A Great Lack of Study
Chapter 5: I Weigh About One-Eighth of a Ton
Chapter 6: He Does Not Want To Be A Dictator
Chapter 7: 50,000 Lambs Were Driven Into the Sea
Chapter 8: But We Do Hate
Chapter 9: Give Tricia a Kiss For Me

Page 6 of 187 Aberhart Outpourings and Replies

Introduction vii

INTRODUCTION

When William Aberhart's Social Credit Party won the 1935 provincial election, Aberhart began the most radical political and economic experiment in North America during those torturous years of the Great Depression. His authoritarian political movement tried to introduce positive social and economic change.[1]

Aberhart, the high school principal and radio evangelist, was an extremely complicated man, driven, some would say, by inner demons. He strove for mastery in every endeavour he attempted. He began his professional life as a school teacher in Ontario, became the principal of Calgary's largest high school and went on to become the Minister of Education for Alberta. His interest in religion led him to reject all human religious authority; he created his own Bible Institute and church. He gave himself the title "apostle" and claimed an apostolic succession which bypassed and surpassed the papacy. His religious broadcasts and his Radio Sunday School had the largest religious radio audience in Canada. When he decided to involve himself in politics during the Great Depression, he created his own political party and entered the Legislature as Premier in his first bid for power. After firing the Attorney General in 1937 and with no legal training other than a brief and incompleted American correspondence course in law, he assumed the office of the highest legal official of the province. His self-confidence knew no bounds, and he defied the Lieutenant Governor, Prime Minister and Parliament.

Figure 4. CAREO Book Browsing Interface

The same ability that allows content packages to be exported as interchange files would make it possible to easily move digital books between libraries. This movement could be just the organization of the book linked to its online components or a complete package that included its organizational structure and all its digital assets. The generic nature of the metadata used to describe contents of the package allows the use of Dublin Core as well as other metadata standards to describe the components of the package. It would be possible to add additional metadata sets as well so that a book that was using Dublin Core for search and retrieval metadata could also add a section of IMS metadata that would describe its educational context.

The storage of the digital book IMS Content Package Information in XML provides opportunities to move the data between standard digital book formats. There is also a degree of similarity in the structure of other digital book packaging standards that could eventually allow a degree of interoperability between books stored in the various formats. The potential to move books between these formats and allow them a large venue of exposure in many different contexts is a definite avenue for future research in this area.

References

[1] Anderson, Thor & McKell, Mark
IMS Content Packaging Best Practice Guide: Version 1.1.2 Final Specification, http://www.imsproject.org/content/packaging/cpv1p1p2/imscp_bestv1p1p2.html, 2001.

[2] CAREO Project, CAREO project website, http://careo.netera.ca, 2002.

[3] Duval, E., Hodgins, W., Sutton, S., and Weibel, S., Metadata Principles and Practicalities. *D-Lib Magazine*, Vol. 8, No. 4, http://www.dlib.org/dlib/april02/weibel/04weibel.htm, 2002.

[4] Federal Task Force on Digitization, *Towards a Learning Nation: the Digital Contribution: Recommendations Proposed by Federal Task Force on Digitization*, Canadian Government Publication, Ottawa, 1997.

[5] Hannah, S.A. and Harris, M.H, *Inventing the Future: Information Services for a New Millennium*, Ablex Publishing Corporation, Stamford, Connecticut, 1999.

[6] Jones, P, Java and Libraries: Digital and Otherwise. *D-Lib Magazine*, http://www.dlib.org/dlib/march97/03jones.html, March 1997.

[7] METS Metadata Encoding and Transmission Standard, *Metadata Encoding and Transmission Standard Official Web Site*, http://www.loc.gov/standards/mets/, 2002.

[8] Open eBook Forum, http://www.openebook.org/, 2002.

[9] Our Roots, Nos Racines, http://ahdptest.lib.ucalgary.ca/CDLHS/, 2002.

[10] Wiley, D.A., Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy, In D. A.Wiley (Ed.), *The Instructional Use of Learning Objects: Online Version*, http://reusability.org/read/chapters/wiley.doc, 2000.