

OCKHAM:

Fostering DL Community Interoperability through Reference Models and Lightweight Protocol Networks

Martin Halbert

Director for Library Systems
Emory University
Atlanta, GA 30322
mhalber@emory.edu

Abstract: The OCKHAM initiative working group has been sponsored by the Digital Library Federation to investigate digital library architectural approaches that might foster greater interoperability and affordability of such infrastructures. The group has studied benefits of the strategies of reference models and lightweight protocols. The resulting approach is intended to be applied experimentally to the NSDL and other federated digital libraries.

1. Introduction

A working group of interested systems librarians, vendors and researchers was convened in early 2002 by the Digital Library Federation “to look for generalizable technical solutions and for an architecture or framework within which such solutions can be implemented, shared, and discussed.” [1] This presentation will articulate an approach for coordinated DL development formulated by the working group and describe some potential experiments for this approach to evaluate its benefits.

Participants in the working group discussions have included representatives from many digital library endeavors at institutions such as Emory, Virginia Tech, the California Digital Library, Notre Dame, University of Arizona, OCLC, and the University of Windsor. The related questions of how to simultaneously increase interoperability and decrease complexity (and therefore cost) of DL development efforts is very much on the minds of everyone involved in digital libraries today. The working group discussed the success of many recent DL approaches, including open standards, component-based architectures, and the importance of seeing DL in the larger environment of knowledge applications (a category which we see as potentially including courseware systems, for example). For reference purposes, the working group eventually termed its focus OCKHAM, for Open Component-based Knowledge Hypermedia Applications Management. The name was intentionally chosen to emphasize the concept of reusing existing technologies.

2. Key Strategies for Digital Library Coordination

There have been a variety of attempts to coordinate the development of digital libraries in recent years. These attempts have many motivations. One cluster of motivations centers on efficiency or affordability, sometimes articulated as “Don’t reinvent the wheel.” Groups working on digital library development efforts are usually not capitalized to the same degree as commercial endeavors, and lack revenue streams to fund large teams of developers. At the same time, digital libraries are very complex infrastructures to develop, evolve rapidly with developments going on in the software industry, and often emphasize complex interactions with other DL systems. This last point grows from another long-established set of motivations focusing on the benefits of effectively aggregating DL content and services through interoperability, with the aim of creating so-called “virtual” libraries, or extended consortia of disparate DL systems that act as a unified whole.

There are an enormous number of efforts that could be cited in this regard. We will here cite only a few examples that are emblematic of strategies that appear to have been successful in advancing the linked causes of interoperability and affordability.

2.1 Reference Models: OAIS and DNER

The concept of “reference models” has gained prominence in recent years. Two well-known examples that the OCKHAM working group examined closely are the OAIS [2] and DNER [3], the former originating in the United States and the latter from the UK.

The reference model for OAIS (or Open Archival Information System) was developed by NASA as (p. iii) “a technical Recommendation for use in developing a broader consensus on what is required for an archive to provide permanent, or indefinite long-term, preservation of digital information. This Recommendation establishes a common framework of terms and concepts which comprise an Open Archival Information System (OAIS). It allows existing and future archives to be more meaningfully compared and contrasted.” This is a relatively clear and straightforward summary of the purpose of a reference model. A reference model does not provide details of implementations or interoperability protocol specifications. Rather, it simply provides a shared vocabulary and generalizable scenario of interrelating entities so that groups engaging in discussions concerning emerging technological services do not waste time repeatedly trying to come to consensus on such issues as participants come in and out of such collaborative projects. This function of stabilizing discussions is far from trivial, as the lack of such shared vocabularies and scenarios greatly hinder groups in trying to pursue common good goals. The OAIS reference model has demonstrated the utility of this function in several ways. Although developed for the NASA community, it has fruitfully been adopted by many groups pursuing digital library and digital archives projects, an example being the DSPACE project. [4] The relevant finding is that clearly articulated reference models are a critical preliminary step to mobilizing group efforts, and ideally must occur before detailed technical specification work begins.

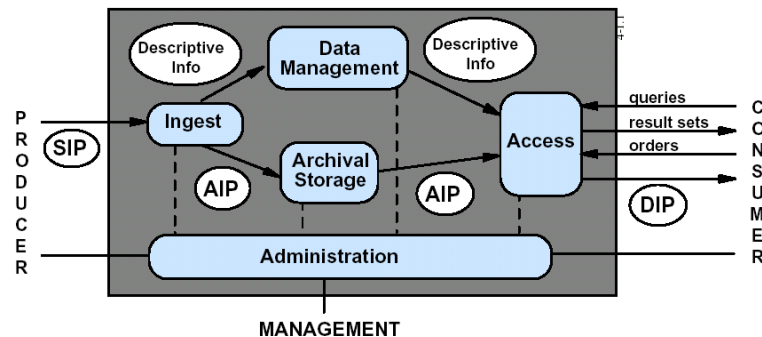


Figure 1. Example of Reference Model Entity Relationship Chart from OAIS

THE OAIS has served an effective function of coordinating efforts of separate organizations in the United States even without being officially commissioned by many of these organizations. This demonstrates that reference models can effectively be adopted informally by disparate, unconnected agencies. Reference models can also be effective as explicitly commissioned documents in bureaucratic structures in order to coordinate efforts. The DNER is an example of this situation.

The DNER (or Distributed National Electronic Resource) has served a somewhat similar purpose in the UK, clarifying abstract concepts to mobilize efforts in UKOLN (the UK Office for Library and Information networking). Individuals who have worked in the context of UKOLN confirm the importance of such reference models to coordinate work in agencies centrally funded and commissioned.

Our conclusion was that the development of a reference model document is critical as an early step when organizing any collaborative digital library effort between organizations.

2.2 Lightweight Protocols: OAI-PMH and Web Services

Reference models are not sufficient by themselves as a means of mobilizing collaborative efforts. Reference models do not provide specifications that can be developed. To begin programming functioning

applications, detailed protocol specifications must be available. The now large body of literature on communication protocols and their development suggests several things to us:

- Communication protocols almost always in fact proceed from reference models, although this is not always obvious, as the model is frequently only documented concomitantly with the protocol specification or assumed as informally and separately understood from the protocol specification.
- Simplified protocols appear to be more quickly and widely adopted than fully featured, complex protocols. This is unsurprising because such lightweight protocols have fewer requirements for programmers to add to existing infrastructures.
- Modular approaches that allow participants to selectively implement and/or recombine protocol components are more widely adopted than monolithic systems. This is a well established pattern, and the main reason that development projects do not uniformly implement protocols and software products as components is that there is often some additional overhead associated with componentizing programming efforts.

The most successful recent example of a DL protocol that has fostered many new interoperability approaches is the OAI-PMH. As a lightweight protocol, it is easy to graft onto existing systems. It lacks an explicit reference model, but in fact embodies the features of a reference model in the definitions which it sets out. Clearly, lightweight protocols are an effective means of mobilizing development efforts in loosely-coupled DL federations.

3. OCKHAM Methodology

After studying many approaches and trends in coordinating DL development efforts, the OCKHAM working group spent time synthesizing the key strategies identified above into a conceptual process. My presentation will elaborate the summary below.

3.1 Situations Relevant to the OCKHAM Strategy

The OCKHAM strategy is relevant to situations in which a group of separate organizations wish to collaborate in developing a federated digital library infrastructure. In such situations, the participating organizations typically already have an extensive DL infrastructure of their own, and are unable to extensively re-architect this existing infrastructure. The question becomes one of how to create a new virtual DL as an added layer of abstraction on top of the disparate infrastructures of the participating organizations. This situation lends itself to a combination of reference models and lightweight protocols.

3.2 OCKHAM Process

The main steps in the OCKHAM process are simply a combination of good project management steps and the collaborative development strategies of reference models and lightweight protocols:

- A. Identify Community and Needs.** In this phase, a group of like-minded organizations coalesces around a shared conception of an identified community and a set of needs that the organizations can address through collaboration. The collaborations that OCKHAM is most interested in are federations of digital libraries. This phase will often be initiated by informal contacts among individuals who comprise stakeholders in the proposed community.
- B. Develop Reference Models.** In this phase, the collaborating institutions develop a set of reference models that document the scenarios proposed for the federated digital library (FDL).
- C. Architect the FDL from Lightweight Protocols.** The protocols should ideally be selected from existing reusable and extensible standards. Members of the OCKHAM working group has expressed a preference for Web Services protocols that disseminate information as relatively portable XML streams.
- D. Iteratively Test and Assess.** The steps above should enable the collaborators to rapidly produce a model for the overall system and a minimal working version of the core components of the FDL. The FDL that results should be iteratively developed in a modular fashion, with each participating organization taking responsibility for one portion of the entire FDL.

4. Potential Experiments

The OCKHAM working group realized quickly that our methodology needed to be tried experimentally to assess what aspects were sound and which aspects needed further work. We have tentatively identified two potential projects to try out our methodology. The first project forms the basis for a proposal to NSF to establish a peer-to-peer network of services mediating the NSDL and traditional libraries. The second potential project would be an application of the OCKHAM methodology to the proposed DLF Distributed Open Digital Library (DODL). My presentation will outline our thoughts on each of these potential projects.

References

1. Greenstein, D, Martin Halbert: Frameworks and Forums for Evolving Digital Library Architectures (<http://www.diglib.org/architectures/ockham.htm>).
2. CCSDS 650.0-B-1: *Reference Model for an Open Archival Information System (OAIS)*. Blue Book. Issue 1. January 2002. (<http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>).
3. Powell, A., Liz Lyon: The DNER Technical Architecture (<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/dner-arch.html>).
4. Bass, M. et al.: DSpace – Internal Reference Specification Technology & Architecture (<http://www.dspace.org/technology/architecture.pdf>)