

Submission and repository management tools for digital libraries, with WWW interface: a Demo proposal

SARANTOS KAPIDAKIS* GREG KARVOUNARAKIS†

April 18, 1999

1 Overview

Digital library objects can be quite complex, with a lot of metadata, possibly in many languages, and data in different digital forms. In order to reduce the overhead of adding, or modifying objects to a digital library we made tools that simplify the operation of submitting new objects, so that the object creators can directly send them to the library, where appropriate. Moreover, our tools simplify the numerous actions that a non computer-expert librarian may have to perform to maintain such a library.

These tools provide an easy way to input or modify the metadata describing a submission and also to upload the digital documents, as well as inspect, commit or modify new submissions or old objects. They also comply with multilingual and heterogeneous collection extensions, where the metadata fields describing each collections may be different. Finally, to be able to be widely used and to suit most needs, these tools are fully configurable (including paths, languages, fields per collection or even messages, colors etc) while at the same time their minimal configuration is very simple, especially when used with DIENST([2]), by reading directly the DIENST configuration information.

We made these tools easy to install and at the same time powerful and configurable. The tools, have matured through requirements and comments of many users and are already in use in the digital library of the University

*Institute of Computer Science, Foundation for Research and Technology – Hellas, Heraklion 71110, Greece. E-mail: sarantos@ics.forth.gr and Massachusetts Institute of Technology Operations Research Center Cambridge, MA 02139-4307 E-mail: sarantos@mit.edu (contact person)

†Institute of Computer Science, Foundation for Research and Technology – Hellas, Heraklion 71110, Greece. E-mail: gregkar@ics.forth.gr

of Crete, with great acceptance. System security is enforced mainly through web-server access control mechanisms. For more information, see [1].

The tools are normally sharing configuration files with DIENST, so that no redundant configuration information need be given. Nevertheless, they do not depend on the running of a DIENST server, and can be used for submission for other digital library systems, with the appropriate configuration. They provide functionality that will cover even rare requirements and offer many optional features, through configuration choices, depending on the sophistication of the environment. For example, confirmation messages may be desired for destructive actions, or notification messages may be sent on specific asynchronous.

2 Demo of the tools

In our demo, we will show the operation of these tools with WWW interface in usual digital library actions, such as the submission and manipulation of the repository of a digital library and the way in which these tools can be used to perform these actions with minimal effort. The functions of submitting an object to a digital library and managing the objects which form the library can be subdivided into three distinct but also self-complementary categories: metadata manipulation, data (digital format) handling, and repository management.

Metadata are a basic part of any submission, because the digital library operations are based on them. Consequently, each digital object is accompanied by its metadata description [3]. No data will be accepted, unless their metadata are already given. This description can also be given and edited in the WWW interface. Some metadata fields (like *dates*) may get default values, and other (like *title*) may be mandatory. The tools handle multilingual metadata, providing filling entries for all supported languages, and heterogeneous collections

The digital library objects may be represented in several digital formats in the digital library. For example, a picture may be in different resolutions, and possibly according to different image standards, or a document may be accompanied by many translations. The file that holds the digital format of the document is specified (usually by browsing), possibly repeatedly, for submitting multiple digital formats [4]. The object creator can add new formats to previously submitted objects or overwrite an uploaded format with a newer version.

The tools can even handle complex digital formats like ones that are physically formed by a hierarchy of files. A degenerate such hierarchy is the scanned pages of a paper document. Apart from the validity of each file separately, we also check the ability to put these files in a logical order, and for missing pieces, in order to display them correctly. We handle cases of files that have an *obvious*, unambiguous order.

When one person is authorized to add objects in a collection, he will normally use the administrator interface. For collections that the objects are contributed

by many users, the creator interface will be used. To avoid polluting the digital library with inappropriate material, either intentionally or unintentionally, the creator interface may have restrictions on modifications of some fields, and will normally place all new submissions in a temporary repository with the same structure as that of the permanent repository, rather than adding them directly to the permanent repository,

The collection administrator can use the administrator interface to submit new object directly, with some more privileges than that the creator, or to browse, inspect and modify all new submissions, and also to reject or approve some submissions, to commit them to the permanent repository. He also has full modification and deletion privileges on permanent repository entries.

References

- [1] Greg Karvounarakis, Sarantos Kapidakis: *Submission and repository management tools for digital libraries, with WWW interface*
`ftp://crete.csd.uch.gr/pub/thesis/diplom/gregkar/TR.ps`
- [2] Carl Lagoze and Jim Davis. *DIENST: An Architecture for Distributed Documents Libraries*, Communications of the ACM, 38(4), April 1995, p. 47
- [3] R. Lasher, Stanford, D. Cohen, Myricom, June 1995, *RFC 1807: A Format for Bibliographic Records*
- [4] E. Nebel, L.Masinter, Xerox Corp, November 1995, *RFC 1867: Form-based File Upload in HTML*,
`http://rfc.fh-koeln.de/rfc/html/rfc1867.html`