Editorial

Andreas Rauber · Birte Christensen-Dalsgaard

Published online: 13 July 2007
© Springer-Verlag 2007

Digital Library Research throughout the years has been very successful in providing methods and solutions to store and manage increasing amounts of data; to search and process heterogeneous types of digital objects, in short, to allow users various ways to conveniently access and use distributed digital content.

However, as the number of successfully operating digital library systems increases, a new challenge is emerging and quickly turning into a mission critical problem for them: Digital preservation.

The core task of Digital Libraries is to provide convenient access to authentic digital objects. Yet, these digital objects are more fragile than common belief holds. Simply storing bit-streams safely over long periods of time is a challenge in its own. In addition to that, and contrary to conventional books, photos, or print-outs, digital objects always require specific programs to turn the bit-stream into a useable digital object, to be able to view and use a digital document. These programs, in turn, require a specific operational environment—operating system, libraries—to run, which, in turn, depend again on a given hardware infrastructure. All of these components are bound to change in relatively short periods of time, putting digital objects at peril.

Digital preservation aims at mitigating the risk of losing digital objects by taking actions to ensure long-term access and usefulness. For a useful and comprehensive introduction

to the basic challenges and principles of digital preservation, see [1].

This special issue collects a number of papers dealing with various aspects of digital preservation challenges.

In their paper “*An intelligent decision support system for digital preservation*” Miguel Ferreira and his colleagues move digital preservation from ad-hoc decisions by experts to a process guided by an intelligent decision support system. Based on web services, a recommendation is formed based on the evaluation of the quality of migration actions, the suitability of the format and other criteria in a formal evaluation. A Migration Broker is used to carry out format migrations. Subsequently, an Object Evaluator is used to evaluate the performance of the migration steps against a series of criteria. This information is further extended by information on the suitability of file formats for long-term preservation provided by a Format Evaluator. Based on this information, well-founded advice on the suitability of a certain migration action for an archive can be provided.

A case study on the application of such preservation strategies in a digital archive is reported in “*The Florida digital archive and DAITSS: a working preservation repository based on format migration*”. In this paper Priscilla Caplan presents the DAITSS (Dark Archive in the Sunshine State) repository management system with respect to the OAIS reference model, detailing how the various functional entities and information models are implemented. Specifically, experiences from running the presented system in the first year are presented, showing potential bottlenecks and problem solutions for an operational system.

Dongwong Lee addresses another aspect that severely affects the practical operation of digital preservation systems, namely the evolution of metadata over long periods of time. In his paper, entitled “*Practical maintenance of evolving metadata for digital preservation: algorithmic solution and*

A. Rauber (✉)
Department of Software Technology and Interactive Systems,
Vienna University of Technology, Favoritenstr. 9–11,
1040 Wien, Austria
e-mail: rauber@ifs.tuwien.ac.at

B. Christensen-Dalsgaard
Statsbiblioteket, Universitetsparken,
8000 Århus C, Denmark
e-mail: bcd@statsbiblioteket.dk

system support” he investigates algorithmic solutions that allow the identification of evolving metadata in large-scale systems, and that later-on allow these evolving metadata to be incorporated into the system as such. Fixing conflicting metadata is achieved, for example, by identifying appropriate operators such as linear transformations between specific metadata attribute values, or either their splitting into several novel attribute values or the merging of several different ones into a single novel value. These may further be used to enhance search functionalities within the system.

While most digital preservation initiatives rely on thoroughly established and maintained institutional repository settings for the long-term storage of digital objects, Michael Nelson and his team propose a radically different approach in their article “*Using the Web infrastructure to preserve web pages*”. Rather than collecting all information into a single central archive, web caching as well as newsgroup postings and email attachments are utilized to provide redundant storage for information, while lexical signatures are used as a finding aid. While these approaches obviously cannot and are not supposed to replace conventional archiving solutions, the authors show for example by analyzing the storage times of web search engine caches that the infrastructure provided by the web may serve specific purposes in the archiving of information. Several scenarios from “lazy preservation” to active installment of web resources into other services that are being used as a “shared infrastructure” are discussed as potential models for information that is not taken care of by institutional repositories.

A different topic is addressed in the last paper of this special issue, presenting a system to enhance the indexing and

cataloging of cultural heritage resources in a digital archive based on watermark information. During conventional digitization approaches, only the content information printed on the paper is being used. In their paper “*A system for segmenting and extracting paper-based watermark designs*”, Hazem Hiary and Kia Ng use image processing techniques to enhance and filter watermarks embedded in the paper. By using background lightening, the information is captured during the scanning process, and—contrary to conventional scanning removing background noise—the foreground inference is being removed, with subsequent sophisticated image processing being applied to the background image. The final watermark may be utilized in several different ways, ranging from additional metadata for the digitized object, to re-embedding into the digital recreation.

The topics introduced above show the breath and constitute and solve only parts of the digital preservation challenges. Numerous additional topics are being researched in a number of recently started DP projects world-wide. We hope that the papers collected in this issue will not only be of interest to researchers active in DP, but will also influence and inspire the whole field of researchers and developers active in Digital Library Research, to insure that the systems we design will remain useful and their content accessible over long periods of time.

References

1. UNESCO, Information Society Division. Guidelines for the preservation of digital heritage. URL <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>, accessed on October 2003