

DISTRIBUTED ARCHIVING NETWORK – DISTARNET

Ivan Subotic, Simon Margulies, Dr. Lukas Rosenthaler
Imaging & Media Lab
University of Basel
Bernoullistrasse 32, CH-4056 Basel/Switzerland
Phone +41 61 267 04 88, Fax +41 61 267 04 85
[ivan.subotic, simon.margulies, lukas.rosenthaler]@unibas.ch

1. Introduction

The growing production of digital data (digital-born or digitized) challenges archiving institutions with new problems and with new needs for a secure preservation of the cultural heritage of our time. A system for archiving digital data was presented at the IS&T Archiving 2004 conference [1]: Distributed Archiving Network (Distarnet). In the mean time the project 'Distarnet' [2] has been funded by the Swiss National Science Foundation and the system is being implemented.

Its main subject is to define a protocol of a distributed system for the long-term preservation of digital data. On one hand the various problems of archiving digital data are analyzed, on the other hand a tangible, implemented and tested open-source solution will be presented. At this the special needs of archives, museums and libraries are taken into particular consideration being the holders of preservation and distribution of historical source material.

The design team consists of a historian and an economist, both contributing the special requirements of their disciplines to the system behavior of Distarnet. Therefore closest attention is paid legal and economic issues of archiving processes and to preservation of source material for future scientific researches. Rethinking the original concept has shown the need for additional features to guarantee a secure, successful and usable archiving solution: Question arose about suitable metadata management, security measures, routing and searching abilities.

The following conference contribution wants to depict our solution but also to highlight the important features needed in general for an distributed archiving network.

2. Distarnet

Because of the continuing and unstoppable technical innovation, archiving institutions are forced to migrate data from one data carrier to another, respectively from one data format to another, in order to guarantee the future readability of the data. To render this possible, high data security and data integrity must be guaranteed. The risk of data loss must be minimized and the data independence concerning future technical development must be achieved. Distarnet accomplishes these goals by automatic self-recovery of lost or corrupted data and by archiving the assigned metadata. Distarnet is the communication protocol of a distributed system, which saves data in high redundancy and security. In the course of the project the protocol will be verified by a reference implementation in Java. Both the protocol and the implementation will be accessible to the general public as open source in 2007.

2.1 Migration and Data Security

The architecture of the distributed system is designed according to the peer-to-peer network paradigm, offering distributed redundancy and automatic self-recovery. If a data-loss of any network member occurs or one of the members ceases to be available, Distarnet restores the former state by reestablishing the defined redundancy through redistribution of identical copies of the lost data to the member who lost its data, respectively by distributing identical copies among its other members. This way the risk of data loss is minimized and the data migration automated. This kind of behavior consequently renders the current practice of intensive manual migration almost a non-issue. To facilitate this

kind of behavior, it is necessary to implement an overlay network which allows for efficient routing and searching among its data members. Another important issue is security notably the security of the communication between data members and the security of the archived data. This is being resolved through the use of public key encryption and digital signatures methods.

2.2 Metadata Integration

The secure preservation is the precondition of archiving data, but offers neither a guarantee for its readability nor its usability for future scientific interpretation. To fulfill these needs, different types of metadata must be preserved along with their primary data. Through administrative, technical and descriptive metadata, the find-ability, the technical and content-interpretation and consequently readability and scientific usability are made possible. The loss of only one type of metadata can bring along the loss of information about the data and consequently the loss of its readability and usability.

To face these needs, Distarnet proposes an overlaying ontology [3] for all different types of metadata. It will be possible to add individual schemas and metadata of any participating archiving institution. This way Distarnet offers a platform for an overall schema-independent research into information objects supported by intelligent user agents.

3. Summary

Distarnet is the protocol of a peer-to-peer network that offers automated migration and secure preservation of digital data. By integrating different types of metadata the future interpretation of the archived data is guaranteed. More information can be found on <http://www.distarnet.ch>

References

[1] L. Rosenthaler, R. Gschwind: DISTARNET – A Distributed Archival Network. IS&T's 2004 Archiving Conference San Antonio, Texas; April 20, 2004; p. 242–248; ISBN / ISSN: 0–89208–251–8

[2] Distarnet: <http://www.distarnet.ch>

[3] S. Margulies, I. Subotic, L. Rosenthaler: Data Description and Archives for Scientific Research in the Future. Paper proposal for IS&T Archiving 2006. http://www.distarnet.ch/papers/is_paper2006.pdf

Authors

Ivan Subotic studied Business and Economics at the University of Basel. He is working on his PhD in History in the field of archiving digital data with emphasis on legal and economic issues, i.e. the tension between legal obligations and economical possibilities of digital archiving solutions and their implication on Distarnet.

Simon Margulies studied History and Computer Sciences at the University of Zurich. He is working on his Ph.D. in History in the field of Archiving digital Data. He analyzes the impact of digital data as source material for the historical research, develops Distarnet and advises various projects and companies on data modelling, metadata and data retrieval.