



RODA - Repository of Authentic Digital Objects

Position paper for PresDB'07

Francisco Barbedo
<frbarbedo@iantt.pt>

José Carlos Ramalho
<jcr@di.uminho.pt>

Luís Corujo
<lcorujo@iantt.pt>

Luís Faria
<lfaria@iantt.pt>

Miguel Ferreira
<mferreira@dsi.uminho.pt>

Rui Castro
<rcaastro@iantt.pt>

March 28, 2007

Context

The National Archive of Portugal (IAN/TT) doesn't currently have the needed infrastructures to support the processes of ingestion and management of digital objects produced by the public administration (PA).

The eGovernment's initiatives establish the need of information and communication technologies support, to improve the efficiency, productivity and quality of their public services. In this scenario, it is clear that the number of digital objects produced by these institutions will grow, and that their legal value and authenticity should be assured. Furthermore, these digital objects will be the testimony of public organizations activity and will constitute the Portuguese social and patrimonial memory.

On that account, the IAN/TT should undertake the endeavor of developing processes, tools and resources capable of answering the

needs of the public administration institutions in terms of preservation of the digital items being produced.

The goal of this project is to provide technical solutions to digital preservation, applicable to a national archive. A prototype is expected to be produced, which will be the basis for the development of a fully functional solution, capable of ingesting and managing digital objects.

Objectives

The RODA project has, as its primary goals, the development and the definition of:

- Functional requisites for a digital archive, its clients and applications;
- Digital archive's conceptual, logic and data models;
- Structural, technical and organizational metadata requisites;

- Digital archive prototype for the preservation of digital objects that are susceptible of definitive conservation;
- Creation of a tool capable of coupling with the existing PA's document management systems and assure digital preservation functions in a administrative management perspective.

The digital archive prototype guarantees all the Open Archival Information System's ([fSDS02]) functionalities: ingestion, access and administration, detailed on the Interpares models. This prototype will have a limited support on the number of file formats which is capable of ingesting and preserve. Three object classes were chosen to demonstrate the prototype features:

- Structured text
- Still Images
- Relational databases

The project also has some secondary goals:

- Definition of an archive policy for digital objects produced by the PA (evaluation and selection);
- Definition of a preservation policy for the digital archive;
- Financial models that would support the digital archive;
- Definition of a taxonomy of significant properties for each object class to consider, i.e. still images, structured text documents and relational databases.

RODA and Databases

The Digital Archive prototype will be capable of preserving relational databases in a XML

dialect called DBML [HLRH02]. An instance of this type of document can store the structure and data of a relational database. It was also created a tool that can export Microsoft Access and Microsoft SQL Server databases to a DBML document. Although DBML is the *Archival Information Package (AIP)* for Relational Database object class, it is not suitable for dissemination purposes, because the manipulation of a very large XML files, like those expected from the PA, is not efficient enough for practical use. Therefore, it was created a disseminator which, given a DBML document, recreates the original database, inside another database engine. Then, consumers can browse the preserved database content with a variety of browsing tools.

Team

The development research team is constituted by Luís Faria and Rui Castro.

The archival coordination is of the responsibility of Francisco Barbedo from the IAN/TT. The technological coordination is of the responsibility of José Carlos Ramalho and Miguel Ferreira from the University of Minho.

References

- [fSDS02] Consultative Committee for Space Data Systems. *Reference Model for an Open Archival Information System (OAIS) - Blue Book*. National Aeronautics and Space Administration, Washington, 2002.
- [HLRH02] Marta Henriques, Giovanni Libretto, José Ramalho, and Pedro Henriques. Bidirectional conversion between xml documents and relational data bases. *International*

conference on CSCW in design,
2002.