

## **eSciDoc - a flexible infrastructure for management and storage of cultural heritage data**

Tschida Ulla, Bulatovic Natasa

Max-Planck Digital Library

Abstract for Poster session at TEI Members' Meeting, November 2008

The eSciDoc project (<http://www.escidoc.org>), developed by Max-Planck Digital Library (MPDL) and FIZ Karlsruhe provides an open source infrastructure for the scientists to store, preserve, retrieve and work with research-based data. Since the wide variety of research topics demands a broad set of reusable, generic and flexible components, it is designed and developed as a service-oriented architecture. Moreover, support of disciplin- and data specific solutions to be built with these services must be guaranteed. The heterogeneity of research questions, tools, workflows and primary data as well as traditional forms of publications required us to focus on supporting multiple content models and descriptive metadata formats, together with common functionalities such as persistent identification, adequate versioning and management of primary data, aggregation of data, annotations, access control etc. The content models are formal representations of discipline-specific data models and their respective validation rules. Relations between content objects are expressed by semantic relations, defined with respective relation ontologies which are applicable to any kind of content - independent from the content structure. The underlying Fedora repository, enriched with the eSciDoc core services layer, allows for management and structuring of data on different levels of granularity - from basic items to complex aggregations of data into containers. Consequently, the eSciDoc infrastructure and its services can be used for cooperative working environments and management of content, as well as a mere archive solution for content managed in external systems. In addition, some services can be integrated by external applications independent from the overall eSciDoc infrastructure.

In addition, the MPDL develops discipline-specific solutions for specific research scenarios, based on requirements from various Max-Planck Institutes. By poster and live demonstration, we would like to present some of these solutions and our experiences with diverse content (publications, digitised books, image collections) and their respective content models.

### **Example: Solution ViRR (Virtueller Raum Reichsrecht)**

ViRR is a collection of digitised law books and documents from the period of The Holy Roman Empire. The content model for ViRR resources describes each law book as a *container* of scanned page *items*, each described with MODS metadata and associated with the respective scanned image *component*. To each scanned page *item*, a transcription *item* with descriptive metadata and a *content component* (e.g. TEI XML) is related. These resources are structurally modelled with a set of content models:

- **BookModel:** a law book is represented as a top-level *container* that aggregates scanned pages of the book
- **BookPageModel:** a scanned page of the law book is represented as an *item* which is a member of the book container
- **BookChapterModel:** a chapter is represented as a *container* that is a member of the book container and aggregates a subset of selected scanned pages
- **TranscriptionModel:** a transcription is represented as an *item* that is related to an instance of the BookPageModel *item*

eSciDoc aims to reuse any of above mentioned models for description of other textual resources and in addition benefit from other services of the eSciDoc infrastructure (persistent identification, versioning, searching, access control, long-term preservation etc.). Depending

on the requirements, one can associate adequate descriptive metadata based on standard schemas such as Dublin Core or MODS, or define resource-specific schemas which are not covered by existing metadata standards.

Being service-oriented and based on open source development, the eSciDoc infrastructure, the solutions and single services can be further reused and extended by other institutions. More information on participation in the eSciDoc community will be presented in November.