Metadata Structures in the ElkaD Project

Sari Järn, Milla Roine and Jani Salminen
Mikkeli Polytechnic
Mikkeli, Finland

Introduction

In the ElkaD project we have defined metadata as structural information which describes features of a certain resource, in which case the resource could be either a traditional paper document or a digitally created document or as in the ElkaD project's opening phase, digitised audiovisual material. In the DESIRE I report, created by the EU, metadata formats are divided into three different categories:

Reasons for Metadata
- To offer better possibilities to find catalogues describing documents
- To offer better accessibility and usability of documents
- Systematic gathering of documents in order to form and control collections
- To recognise their authenticity and integrity of documents

In our project, definitions of metadata are mainly based on the Dublin Core metadata format. One of the most important goals in the ElkaD project has been the assessment of the Dublin Core format's suitability for archiving purposes. A challenging aspect in dealing with Dublin Core is that although it is a standardised format, practical experience from its adaptability in archiving are virtually non-existent in Finland. We have also investigated Finnish metadata standards for Archives and for public sector.

In order to successfully establish a digital repository, collecting and managing metadata is absolutely fundamental. Even though archiving is done electronically, for example, in some sort of intranet, the repository's functions remain the same, as in traditional paper oriented repositories. So ingestion, handling, screening and retention of material as well as offering material for research are also imperative functions for a digital repository. Metadata is the most important tool for successful implementation of the aforementioned functions. In document management metadata is essentially bound in searching, locating and retrieval of information. Metadata is information about a document's registration, its actual content and data about its appearance and technical properties. Metadata also relates to a single document at all levels of archive hierarchy.

Groundwork of the ElkaD Project's Metadata Structure

Dublin Core format is the core of the ElkaD project's metadata structure. Dublin Core 1.1 version was standardised in Finland in 2002.

In archiving, information about the document's context is absolutely essential. Dublin Core is based on RDF-schema, which is somewhat crucial for the ElkaD project's success, because currently RDF is considered to be the best schema for specifying relationships between different resources. The RDF schema is able to provide meanings to elements or semantic structure, although the present version of RDF is still rather limited in this aspect, but this should be improved in the RDF's latest version, which is under development by the W3C XML Schema Working Group.

Fields marked with (DC) are original Dublin Core elements and sub-elements, whose contents are basically the same in the ElkaD project but not all of them are used. So far metadata elements defined in ElkaD have been used mainly in describing audio and video files, which are provided by Elka (The Central Archives for Finnish Business Records in English). ElkaD 's metadata structure can also be applied in describing paper documents, still images, moving images or originally digital documents, including text files and audiovisual representations.

At the moment we have five different levels in the archive's hierarchy, which are: Archive creator, Archive, series, sub-series and single document/file. The metadata structure is devised so that one can describe each file that the archive contains, with the desired details. Also the less-detailed description can be done only at the series level.

In the ElkaD project we have gathered experiences of detailed descriptions for audio recordings. Depending on the quality and the length of the recording one can say, as a rule of thumb, that the detailed description, which includes a table of contents and some sort of abstract, takes roughly at least twice the time of the original recording. So during one working day one can expect to be able to describe two 90 minute recordings. In dealing with voluminous material such as photo series, digitised or digital text documents it appears reasonable to describe material only at the series level.

Links

National Library of Finland maintains the Dublin Core platform, http://www.lib.helsinki.fi/cgi-bin/dc.pl
Dublin Core Metadata Elements, http://dublincore.org/documents/dces/