A linked-data vocabulary of grey literature document types: Version 1.0

Petra Pejšová
National Technical Library, Technická 6/2710, 160 80 Praque, Czech Republic. Email: petra.pejsova@techlib.cz.

Abstract
The role of grey literature in today society was well expressed by Vaska: “It is therefore remarkable how efforts over the past two decades have altered originally preconceived notions of grey literature, resulting in a deeper understanding of how the formation, retrieval, and use of non-traditional material becomes a fundamental companion in one’s research endeavors. The advent and use of new technologies, a prevalent trend in 21st century society, has created increasing awareness of the need to incorporate the grey alongside the white, a boundary that is becoming less distinct and which may become translucent in the not-too-distant future.” (Vaska, 2010)

The poster presents the first version of the vocabulary of the types of grey literature in RDF (Resource Description Framework), which is based on an analysis of grey literature typologies and a draft version of vocabulary presented on 12th International Conference on Grey Literature. The draft version of the vocabulary of the types of grey literature was reviewed by 14-member international working group appointed to this purpose. For cooperation work we chose the methodology of Google Code, a free collaborative development environment for open source projects, available at http://code.google.com/p/grey-literature-typology/.

Vocabulary of the types of grey literature is a hierarchically structured typology and is distributed as a controlled vocabulary in machine-readable format. The vocabulary is exposed in RDF/XML and Turtle. Both of the formats are available on project website:
1. RDF/XML (http://code.google.com/p/grey-literature-typology/source/browse/gltype.rdf)

Keywords: grey literature, linked data, document types, structured data, Resource Description Framework

Introduction

“Grey Literature is a field in library and information science that deals with the production, distribution and access to multiple document types produced on all levels of government, academics, business, and organization in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body.” (GreyNet International 1992 – 2012, 2012)

The vocabulary of grey literature document types is based on an analysis of six existing grey literature typologies, which was presented at the Twelfth International Conference on Grey Literature (GL12) in 2010.

“The aim of the analysis was to create, define, and implement a current credible grey literature document typology, in order to open discussions in the grey literature community, leading to a means of collecting grey literature from reputable events and producers rather than relying on social networking tools or Wiki contributions. While the later types of sources can assist researchers, scientists, and teachers with their information-seeking pursuits, documents of this nature need to be evaluated on a regular basis.” (Pejsova, 2010)

The draft typology was built upon the already available typologies for grey literature, creating an intersection of their features that were found to be important, while excluding duplicates and merging the same types with distinct labels. The draft version of vocabulary of grey literature document types was first presented in the end 2010. The draft version was reviewed by 14-member international working group appointed to this purpose. For cooperation work was chosen a free collaborative development environment for open source projects Google Code, available at http://code.google.com/p/grey-literature-typology/.

Sections
Vocabulary
The typology of grey literature is a controlled vocabulary in RDF expressed as SKOS concept scheme. Each type is provided with a definition and most of them are exemplified with a prototypical example of a document for which it can be used. By design, the typology is focused on the description of types.

Linked data
The vocabulary is published as linked data. Each of its types is identified by URI and the vocabulary is interlinked and mapped to other datasets.

Development
“The typology is developed collaboratively as a Google Code project. It is based on a distributed version control system and it includes documentation Wiki. This repository provides a feedback mechanism for the development via a reporting issue or adding comments which supports iterative development with a long-term evolution in mind, and a regular half-year release cycle for the versions of the typology.” (grey-literature-typology, 2012)

The draft version was reviewed by 13-member international working group appointed to this purpose. In the year 2011, 42 comments and suggestions were accepted. Changes mainly redefined some collections, some document types were merged and others excluded. The major changes concerned the following collections: monographs, informative materials, datasets and research plans. The first version is available for use from November 2011 on project website in machine-readable format. The vocabulary is exposed in RDF/XML and Turtle. Both of the formats are available on project website:
1. RDF/XML (http://code.google.com/p/grey-literature-typology/source/browse/gltype.rdf)

Example of the type:
gltype:conference-proceedings a skos:Concept ;
  skos:inScheme gltype: ;
  skos:prefLabel "Conference proceedings"@en, "Sborník z konference"@cs ;
  skos:broader gltype:conference-material, gltype:proceedings ;
  skos:definition "Conference proceedings are the collection of academic papers that are published in the context of an academic conference. They are usually distributed as printed books or electronic version either before the conference opens or after the conference has closed. Proceedings contain the contributions made by researchers at the conference."@en ;

Usage example:
<> dc:type gltype:conference-proceedings .

References and Citations
Curriculum Vitae

Currently Petra Pejšová is a manager of the National repository of Grey Literature in the Czech Republic. Its goal is a systematic collection, long-term archival and provision of access to specialized grey literature, especially to do with research and development, civil service and education, as well as from the business sphere and "open access" at the national level. She runs an international working group for development of the Vocabulary of Grey Literature Types. She works in the National Technical Library as a head of the Digital National Technical Library, where she also leads a digitalization projects. Through lectures and publications she provides information on grey literature in the Czech Republic. She is the editor of the book Repositories of Grey Literature, published in 2010. She is also an author of the certified methodology for processing, long-term preservation and accessibility of grey literature in the Czech Republic.

More at LinkedIn: http://cz.linkedin.com/pub/petra-pejsova/2a/8a9/b31