

Title: A linked data Vocabulary of the Types of Grey Literature: Version 1.0

Authors: Petra Pejšová, Tereza Simandlová, Jindřich Mynarz, National Technical Library, Prague, Czech Republic

Annotation

The aim of this article is to introduce the Vocabulary of the Types of Grey Literature, the development of which was initiated at the National Technical Library in Prague, Czech Republic. The vocabulary is created in a way compliant with the principles of linked data. The first version of the vocabulary is available from the project website <http://code.google.com/p/grey-literature-typology>

Keywords

grey literature, controlled vocabulary, typology, RDF, SKOS, linked data

Vocabulary overview

The Vocabulary of the Types of Grey Literature (henceforth vocabulary) is a controlled vocabulary in RDF (Resource Description Framework) data format expressed as a SKOS (Simple Knowledge Organization System) concept scheme.

The vocabulary is based on [an analysis of six existing grey literature typologies](#)¹, 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 GL 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, 2011)

The vocabulary can be seen as a formalization of the outputs of the systematic examination done during this analysis. This means that the typology is 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 vocabulary has a loose structure with hierarchical relations. *“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. Other documents’ attributes, such as content or format annotations, are excluded from the vocabulary.”* (Grey-literature-typology, 2010)

Moreover, the vocabulary is published as linked data. Linked data is a publication model for exposing structured data on the Web in a way that uses links between datasets to

¹ GreyNet, (the Grey Literature Network Service), OpenSIGLE, (the System for Information on Grey Literature in Europe), and the Registry of Open Access Repositories (ROAR), as well as focusing on national schemata in the Czech Republic, namely ASEP ([Register](#) of Publication Activity of the AS CR), NRGL (National Repository of Grey Literature), and RIV (Information Register of R & D Results)

create a network of interlinked data. Each of its types is identified by URI and the vocabulary is interlinked and mapped to other datasets.

Working group

The Working Group for Grey Literature Typology², established to support the further development of the vocabulary, consists of a set of experts from various domains, ranging from grey literature and knowledge organization systems, to semantic web technologies.

The core part of working group is formed by the National Technical Library (NTK) project team, which is the informal body coordinating the vocabulary's development. The members of the working group are responsible for building the vocabulary and its updates, maintaining the project's website, incorporating changes, determining the work time schedule, publishing new versions of the typology, providing support in the case of implementation of the vocabulary, and presenting the project and the vocabulary.

NTK project team's members are (all from the National Technical Library, Czech Republic):

- Petra Pejšová
- Jindřich Mynarz
- Tereza Simandlová

Other members of the working group are responsible for providing feedback on the vocabulary's development via comments or by filing issues, addressing the lists of proposed changes on the call by the project owners, presenting the vocabulary and supporting the vocabulary's implementations in the systems of grey literature.

The members of the extended working group are:

- Marcus Vaska – University of Calgary, Canada
- Vojtěch Svátek – University of Economics, Czech Republic
- Anne Gams Steine Asserson – University of Bergen, Norway
- Claudia Marzi – CNR Institute for Computational Linguistics, Italy
- June Crowe – Information International Associates, Inc., U.S.
- Keith G Jeffery – STFC Rutherford Appleton Laboratory, United Kingdom
- Marta Dušková – Slovak Centre of Scientific and Technical Information, Slovakia
- Joachim Schöpfel – Charles de Gaulle University of Lille 3, France
- Dominic J. Farace – GreyNet, Netherlands
- Daniela Luzi – IRPPS-CNR, Italy

Development

The typology is developed collaboratively as a Google Code project. As we mentioned above, its draft version, based on analysis of the six existing grey literature typologies and the result, was presented at GL12 in 2010. The draft version of the vocabulary was published in January 2011 and being reviewed by the 13-member international working group until July 6th, 2011. The members of the working group primarily analyzed the structure of collections and document types definitions and commented directly to <http://code.google.com/p/grey-literature-typology/issues/list> .

NTK's team gathered 42 comments and suggestions that were accepted and incorporated into the first version of vocabulary. Changes mainly redefined some collections, when some document types were merged and others excluded. The major changes concerned

² http://code.google.com/p/grey-literature-typology/wiki/Project_members

the following collections: monographs, informative materials, datasets, and research plans. The first version of the vocabulary is available for use from November 2011 on the project website <http://code.google.com/p/grey-literature-typology>. Hierarchical structure of the current version is consisting of 16 main types, which we called *collections*, and 55 subtypes in these collections. Draft and first version of the vocabulary is available for the purposes of comparison in appendix 1 and 2.

The development cycle of the vocabulary was planned to be bi-annual, but it proved to be unrealistic, and it was re-assessed to release a new iteration of the vocabulary once a year. All versions will be issued via the Google Code repository and will be persistently linkable.

Example:

```
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 ;
  skos:editorialNote "Taken from Wikipedia, the Free
  Encyclopedia: http://en.wikipedia.org/wiki/proceedings,
  retrieved May 23, 2011"@en .
```

Usage example:

```
<> dc:type gltype:conference-proceedings .
```

Benefits

The vocabulary is the first functional general model that can directly implemented in grey literature repositories. Unlike the other analyzed typologies, the vocabulary is available in a machine readable format. Each grey literature type has a definition and has a globally unique identifier, and as such can be unambiguously a persistently referenced. Another benefit of the vocabulary is explicit mapping to other vocabularies, e.g., Dublin Core Metadata Initiative Types, or DBPedia.

Future

Working group expects that the vocabulary will find a specific implementation. The first interest in implementation was expressed by the Slovak Centre of Scientific and Technical Information for their national system for collecting grey literature. Effort of the working group has not been stopped, the first version is prepared for further comments and suggestions. These will be included in updated versions. Now, only 10 % of document types are provided with an example. We plan to accompany all of the grey literature document types

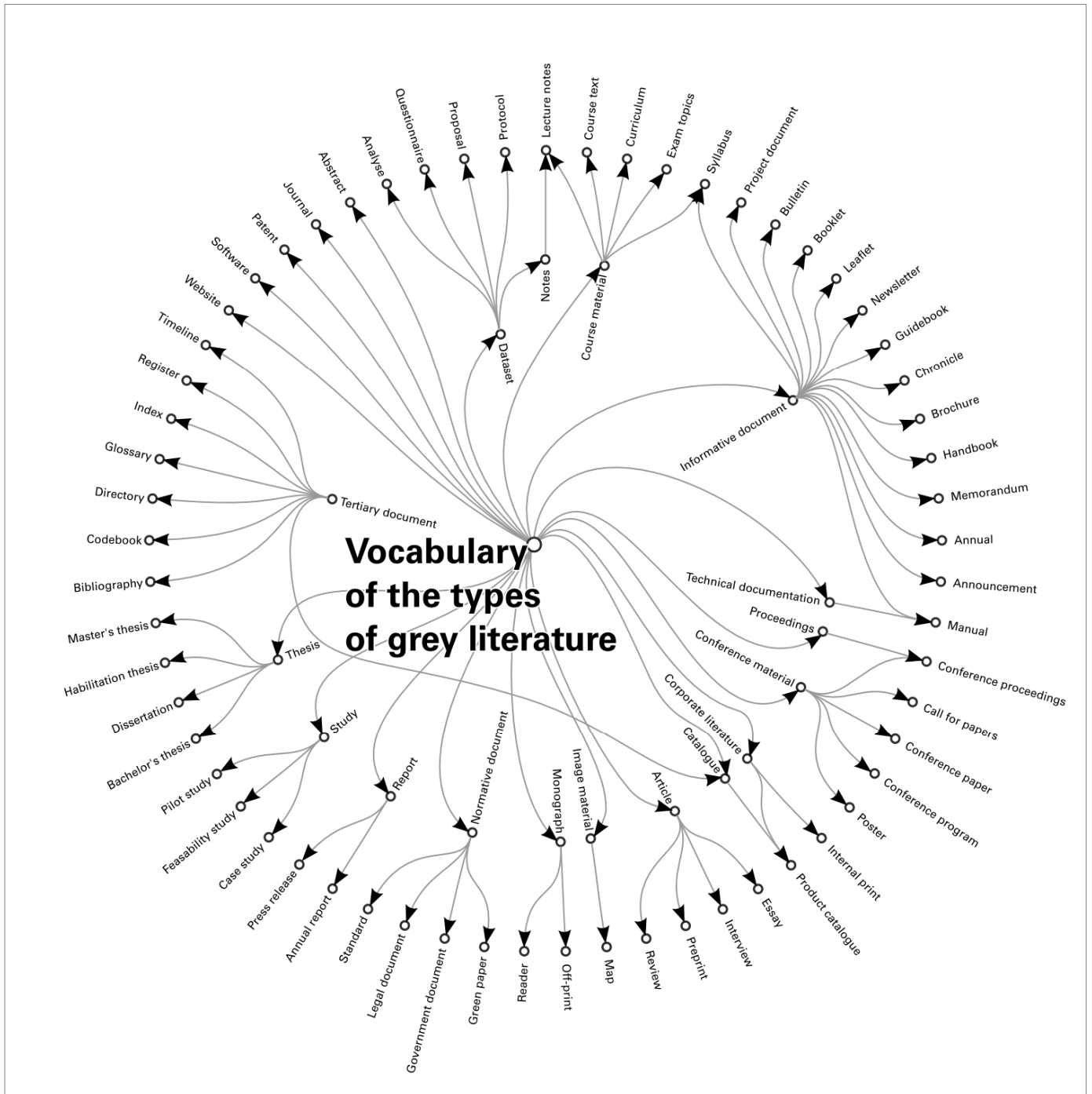
with examples in 2012. And last but not least, the vocabulary is available only in Czech and English. We are looking for experts for translation to other languages.

References

grey-literature-typology (2010). [Vocabulary of the Types of Grey Literature](#). Retrieved November 20, 2010 from <<http://code.google.com/p/grey-literature-typology/>>.

PEJŠOVÁ, Petra; VASKA, Marcus. An Analysis of Current Grey Literature Document Typology. In *Twelfth International Conference on Grey Literature : Transparency in Grey Literature, 6-7 December 2010*. Amsterdam : TextRelease, 2011. ISBN 978-90-77484-16-6. ISSN 1386-2316.

Appendix 1: Draft version of the Vocabulary of the Types of Grey Literature



Appendix 2: The version 1.0 of the Vocabulary of the Types of Grey Literature

