Czech National Repository of Grey Literature

Petra Pejšová

petra.pejsova@techlib.cz

Abstract:
The National Technical Library runs a project of building the National Repository of Grey Literature. 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. The National Technical Library is developing a functional model and pilot application, inform professionals and support scientific discussion concerning grey literature to reach this goal.

Keywords:
grey literature, project, digital repository, user interface, typology, National Technical Library, Czech Republic

The aim of this contribution is to inform about a progress of a project of building the National Repository of Grey Literature in the Czech Republic. According to its status the National Technical Library (henceforth the NTK) runs – among others – a project of building the National Repository of Grey Literature (henceforth the NGRL). This project is supported by the Ministry of Culture of the Czech Republic and its full title is "The Digital Library for Grey Literature – Functional Model and Pilot Implementation" (henceforth the project). It is planned for period of four years, commencing in 2008 and ending in 2011. The NTK participates in this project together with the University of Economics in Prague. Its main goals are the 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. To support this goal, the NTK created a functional network of partner organizations, a functional model and a pilot application which shall form the basis of the NRGL. In addition, on the basis of verified technology and methods defined under the project, recommendations and standards are created for other institutions electing to build their own digital grey literature repositories. Recommendations and standards consist mainly in a recommended metadata format, exchangeable formats and templates, examples of licensing models and of legal issues resolved, preservation methodology, archival and the provision of access to digital data. The standards and recommendations will be published both in Czech and English languages in printed and electronic forms. Continuous outcomes are project web pages and holding annually seminar on the access to grey literature.

Typology

Before collecting grey literature using the NRGL system, a typology had to be defined to clearly specify the subject of collection. Most typologies, including GreyNet, focus solely on document type. In the case of NRGL, a specific typology was created that also took other criteria into account. The NRGL typology is structured on two levels, to better capture individual aspects. The basic typology (level 1), depicted in the mind map (Figure 1), describes all document criteria adopted by NRGL. The basic criterion consists in the collection of documents in electronic form (i.e., documents originally created in digital form or subsequently digitized) since 2009, with older documents processed on a selective basis. Metadata on all accessible grey literature records in the Czech Republic is collected without restriction. Grey literature is supplied by partner organizations in the area of education, research and development, Civil Service, the business sphere and verified open access sources. Grey literature collected in the NRGL system must have originated in the Czech Republic and its contents must be original and specialized. Preferred languages for items are Czech, English, Slovak and other European languages on a selective basis.
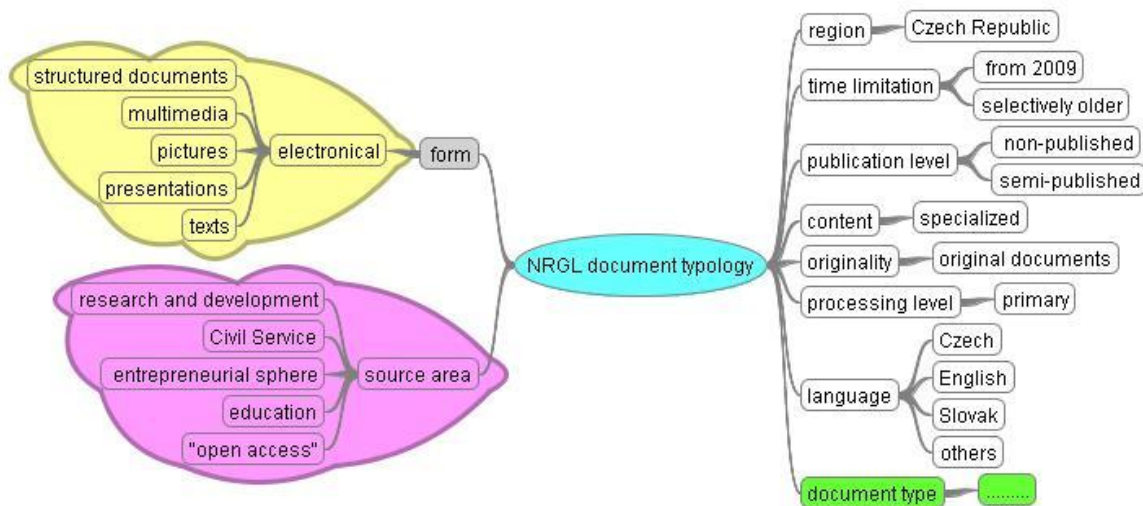
Figure 1: Basic NRGL typology (level 1).

Level 2 of the NRGL typology, concerned with document types, is based upon level 1. The main document groups consist of ETDs, course materials, work by authors, corporate literature, conference materials, etc. Individual groups contain specific types of documents. Another project aim is to verify which types of documents are more numerous, and to extend the typology by adding new document types needed for science and education.
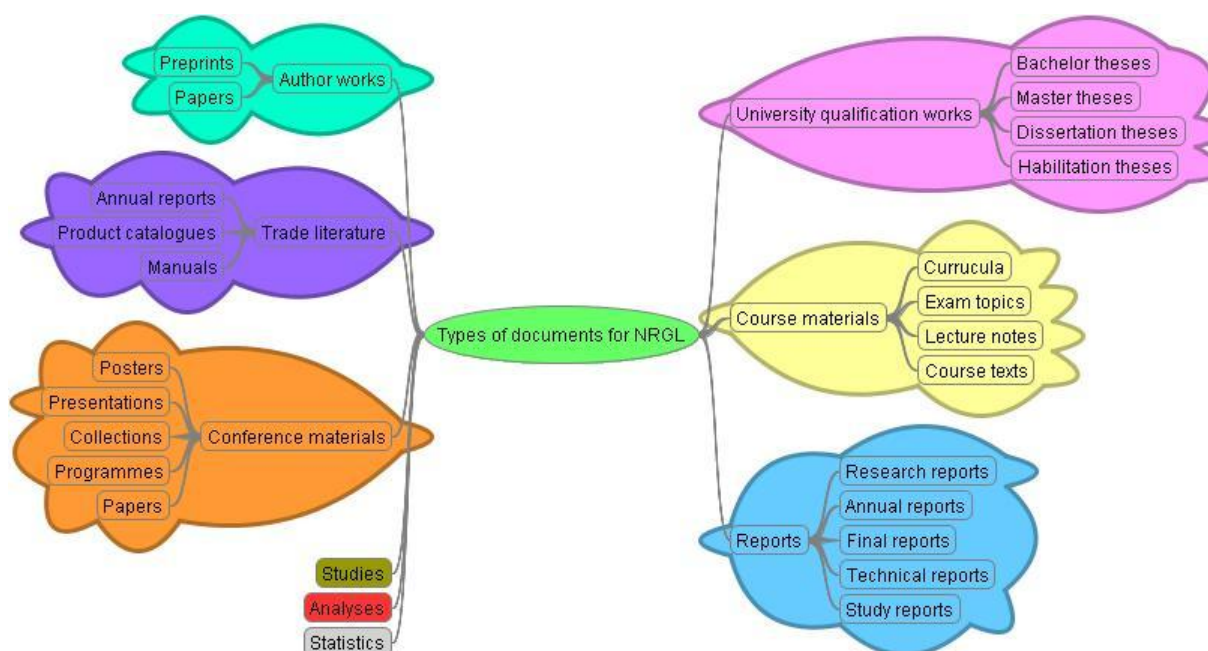
Figure 2: Typology according to the type of document (level 2)

Software solution

The software solution for the NRGL is composed of two parts - the CDS Invenio[1] system for digital repositories and indexing, and the FAST ESP[2] search system for the user interface. The solution led to a user interface being constructed for the digital repository which enabled easy, enjoyable search and data access, as well as functionality integrating the platforms of independent repositories. The same architecture has been used successfully for several years at CERN in Switzerland.

CDS Invenio is an open source software licensed under the GNU GPL, developed by the CERN organization and EPFL (École Polytechnique Fédérale de Lausanne) for general administration of documents, institutional repositories or a large-sized library systems. CDS Invenio is free for installation, usage and adaptation. CDS Invenio has been localized to eighteen (18) languages including Czech and Slovak languages. The CDS Invenio system is very elaborated and flexible, and it can be customized in several ways for various uses. It enables the user to define own metadata schema, having the internal naming convention of MARC21 as its underlying standard. It supports OAI-PHM standard.

From general technical information is important to know, that CDS Invenio is possible to install on GNU/Unix system, it uses MySQL database server and Apache/Python web server. Apart from the system itself we can also expect that some of the services are localized to Czech, or Slovak.

Access to CDS Invenio as of the NRGL is available at http://invenio.ntkcz.cz .

---

[1] http://cdsware.cern.ch/invenio/index.html

[2] http://download.microsoft.com/download/1/4/8/1483939B-15B8-4DD3-B06D-204D03EC8A1E/Fast_ESP_Prod_Guide.pdf

Figure 2: The input screen of CDS Invenio

The FAST ESP indexing and search system provides user-friendly, easy data access. The FAST ESP interface provides a secure, scalable and relevant search, including declension of Czech terms. The characteristics of the FAST ESP solution make it possible to identify the context and purpose of the query, find the corresponding terms both in the metadata in the documents and return a response in context. Users benefit from precise results and contextual dynamic navigation for further searches for related information. The FAST ESP is used for central user interface, which is in progress since the end of 2009. All institutions in the NRGL network are gradually integrated into this interface. The interface is available at www.nusl.cz .

Figure 3: User interface of the NRGL

Support cooperation

The acquisition of grey literature in the Czech Republic is aimed at the systematic collection, long-term archival and provision of access to grey literature for users in science, research and education. The credibility of information thus acquired is considered paramount. This is ensured by cooperation with credible producers of grey literature of the type in question. As a result, unverified grey literature such as blogs, phone calls, discussions, and e-mails is not subject to acquisition.

Focused on possibilities of data acquisition, storage and accessing grey literature in the Czech Republic was developed a legal expertise for the NRGL. The whole legal expertise is for both producers of data and database administrators. Especial part is dedicated to dissertation theses, because this kind of grey literature has other legal framework than the rest of grey literature in the Czech Republic. In support of collaboration with Czech universities a part of the legal expertise is oriented to a recommendation for universities, how to conclude license agreements with students. Together with the legal expertise we prepared examples of license agreements for collaboration with data producers or administrators.

Technical possibilities of collaboration with data producers are available in four ways. First possibility is joining local database using ESP FAST directly to central user interface. In this case it is necessary for data producers to have local database, which is available through OAI-PMH protocol. In this case data are not deposited and

archived in the NRGL digital repository (CDS Invenio in NTK), they are only available through the NRGL central user interface. Second possibility for data producers is harvesting their local database into the NGRL digital repository and taking advantage of depositing and archiving data in the NRGL digital repository. Third possibility is using CDS INVENIO as local database which will be distributed by the NTK in a default version and harvesting it as other local bases to the NGRL digital repository. Fourth possibility is entering records and digital documents directly into the NRGL digital repository. This option is highly recommended for data producers, which have lesser data volume. All this scenarios are directed at a central user interface at http://www.nusl.cz in Czech and English version.
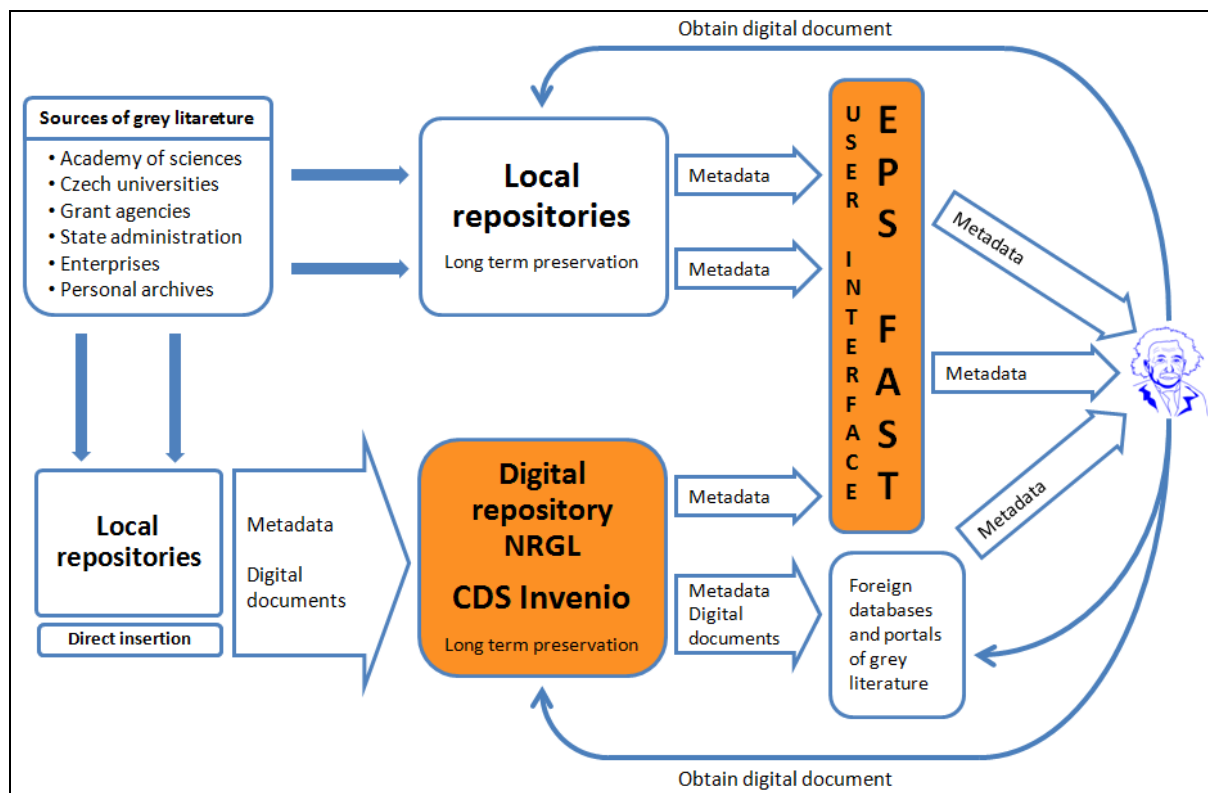

Figure 13: Scheme of the NRGL cooperation

The Project websites

In 2009, the English project websites came into being. English version follows the Czech original. These websites feature continuously updated information on the NRGL project and its outcomes. Furthermore, these websites support collaboration between NRGL and grey literature producers, for example in terms of legislation touching on grey literature. They also offer a downloadable software solution for local installation of a digital repository. To enable broader connections, selected grey literature information sources are cited, along with references to portals supporting research and development in the Czech Republic.

The website may be found at http://nrgl.techlib.cz/. The Media wiki application and the WordPress graphical style, modified to follow NTK graphical guidelines, were used to create the site. The media wiki application supports participation in content creation within the working group and the opening of discussion groups on various topics with public participation.

Figure 4: Screenshot of the project homepage

Workshops

Once a year since 2008, the *Seminar for the Access of Grey Literature* has been organized to inform and support scientific discussion concerning grey literature. In the workshop, updated results of the NRGL project are presented and topics relating to grey literature are discussed such as the national and worldwide status of systems dealing with grey literature, formats and standards for long-term archival, standards for describing and exchanging sources and copyright. All presentations and the full text of lectures are available in online proceedings on the project website. The workshop languages are Czech and English, with proceedings held in English. The reason for this is to approach the worldwide grey literature community.

This year will be held the 3th Seminar on Providing Access to Grey Literature in the NTK in Prague in the Czech Republic on December 8, 2010. The seminar will pick up on Twelfth International Conference on Grey Literature on December 6 – 7, 2010, witch will be held in the NTK.

More information and links NRGL in English:

The project webpage: http://nrgl.techlib.cz

User interface: www.nusl.cz (select English)

Digital repository: http://invenio.ntkcz.cz (select English)

The project contact e-mail: nusl@techlib.cz

REFERENCES:

Seminar to access of grey literature [online]. 2008- . Prague : National Technical Library. Retrieved March 15, 2010 from www: http://nusl.techlib.cz/sbornik/. ISSN 1803-6015.