# National Repository of Grey Literature – second year

**2nd Seminar on Providing Access to Grey Literature**

**October 22, 2009 – Brno, Czech Republic**

**Petra Pejšová**

**petra.pejsova@techlib.cz**

The aim of this contribution is to inform about a progress of a project of building the National Repository of Grey Literature. There will be mentioned the most important specifications of the project and short results overview of the year 2008. The main part of the contribution will concentrate to a solution in the year 2009, which will be presented in detail in subsequent contributions of these conference proceedings. In the end will be introduced a plan for the next year.

According to its status the National Technical Library (henceforth the NTL) 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). At the beginning I would like to mention the most important information about this project. It is planned for period of four years, commencing in 2008 and ending in 2011. The NTL participates in this project together with the University of Economics in Prague. The project goals are first of all working pilot application, which shall form the basis of the NRGL. Further goal is evaluated and tested technology and methodology, which will formulate standards and recommendations for institutions establishing their own local grey literature repositories. The standards and recommendations will include project team experience as well as rules and methodology; namely recommended metadata format, interchange formats and templates, samples of license agreements and other relevant legal issues and methodology of protection, archiving and publishing of 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.

I would like to repeat, what was accomplished in the year 2008, in a short overview. We have acquired the hardware for the repository. These are the servers SUNXFIRE 4500 and magnetic tape library for backup maintenance. We have also defined software specification and first working version of metadata format, analyzed open source systems for digital libraries, specified a document typology, chosen suitable persistent identifier, created Czech version of project web pages and held first seminar on the access to grey literature. All outputs and the proceedings from the seminar are available in electronic form on the project web pages at http://nusl.techlib.cz .

Now I want to focus on the solution in the year 2009. We prepared beta version 0.2 of NRGL metadata format and its abstract representation in XML form. We prepared a conversion between NRGL metadata format and MARC format, which is an internal format of NRGL software. In more details I will devote to the NRGL software later.

We ordered for the NRGL a legal expertise, which is focused on possibilities of data acquisition, storage and accessing grey literature in the Czech Republic. 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.

The legal expertise is available at the project web pages and more information about a grey literature in general in following contribution by Radim Polčák "Licensing the Use of Grey Literature under the Czech Law".

We created an English version of the project web pages this year, which is a copy of the Czech web pages version. English version is available at http://nrgl.techlib.cz .


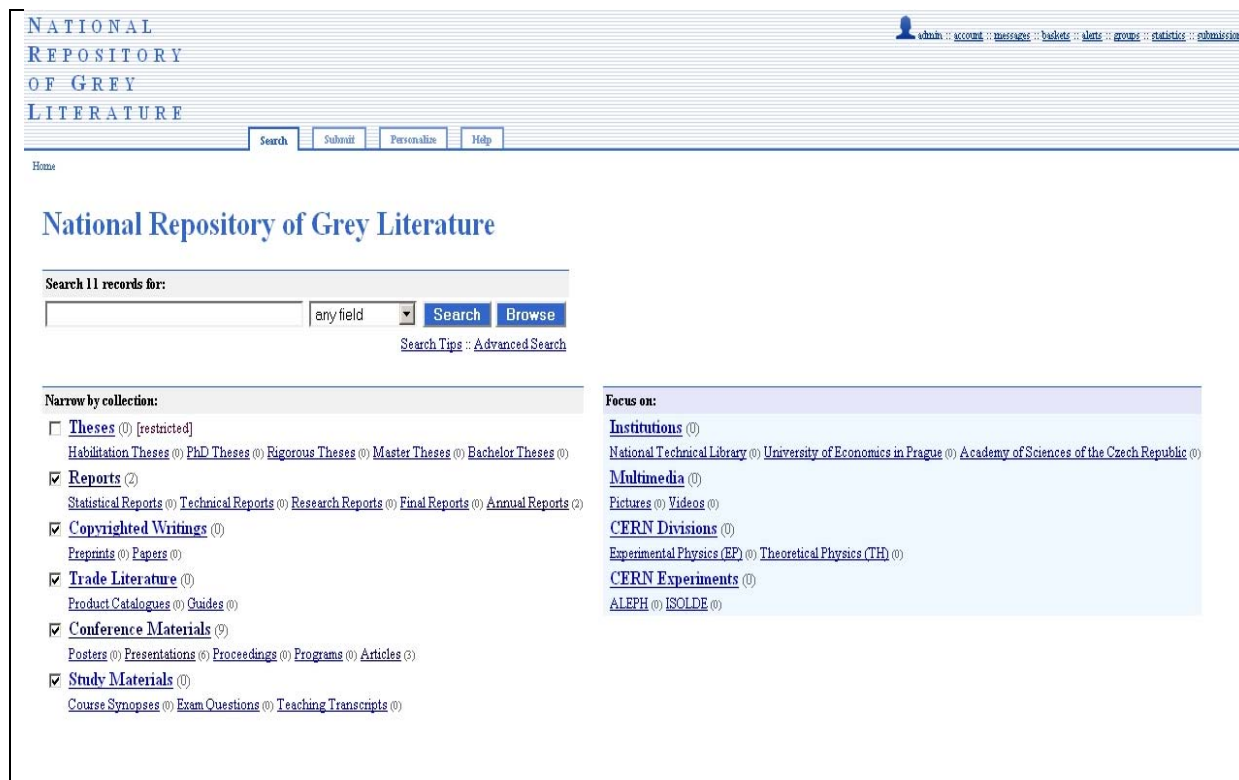*Figure 1: English version of the project web pages*

The software solution for the NRGL is composed of two parts - CDS Invenio and EPS Fast. For digital repository we chose an open source software system CDS Invenio and for user interface commercial software EPS Fast. The ESP Fast is in details presented in a following contribution by Pavel Kocourek "User interface of the National Repository of Grey Literature".

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. In the case of interest offers CDS Invenio also formalized technical support, when installed and during operation, for a consideration on the basis of an agreement.

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 .

*Figure 2: The input screen of CDS Invenio*



Technical possibilities of collaboration with data producers are available in four ways. First possibility is joining local database using ESP FAST directly to 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, they are only available through the NRGL user interface. Second possibility for data producers is harvesting their local database into the NGRL = central CDS INVENIO and taking advantage of depositing and archiving data in the NRGL. Third possibility is using CDS INVENIO as local database, which will be distributed by the NTL in a default version. Fourth possibility is entering records and fulltexts directly into the NRGL = central CDS INVENIO. This option is highly recommended for data producers, which have lesser data volume. All this scenarios are directed at an unitary user interface, which will be run at http://www.nusl.cz in November 2009.

*Figure 3: User interface of the NRGL*

Organizational aspects of collaboration between the NRGL and local data producers in the Czech Republic are following. We start with organizations, which agreed to collaborate with NRGL in the questionnaire we conducted in the year 2008. At first we got in contact with Institutions of the Academy of Sciences and in first half a year 2010 we will get in contact with Czech Universities. Continuously other organization can contact us and negotiate the cooperation.

We are preparing collaboration with foreign systems of grey literature, too. One of these is European system for grey literature OpenSIGLE. The OpenSIGLE has arisen from a former system SIGLE and is a fundamental for a project, which is focused on a creation of feasibility study for the creation of a cohesive pan-European e-infrastructure for digital grey literature repositories. This project is planned together by organizations GreyNet from Netherlands, INIST from France, CABI from UK, NTK from Czech Republic and University of Lille from France. This project mentions Dr. Dominic J. Farace, director of GreyNet, in his following contribution ""Grey publishing" and GreyNet's initiatives in this area". Another system, which NRGL plans to join, is DRIVER (Digital Repository Infrastructure Vision for European Research). DRIVER represents an European network of digital repositories offering sophisticated functional services to researchers and academics as well as to the public. More information about DRIVER you can find in a following contribution

"Open Access and digital repositories: the role of the DRIVER project" by Sophia Jones from Centre for Research Communications University of Nottingham, UK.

For the next year we plan first of all to cooperate with data producers, to test used technology and methodology, to finalize setting of a functional pilot application, to consider the NRGL eligible for a trusty repository, to publish standards and recommendations for building local repositories of grey literature in paper and electronic version in Czech and English languages as well and to continuously actualize Czech and English project web pages.

Next year will be held the 3th Seminar on Providing Access to Grey Literature in the NTL 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 NTL.

More information about the NRGL and conference proceedings:

In Czech http://nusl.techlib.cz

In English http://nrgl.techlib.cz

The project contact e-mail nusl@techlib.cz