národní
úložiště
šedé
literatury

Grey Literature in the Czech Republic

Pejšová, Petra; Pfeiferová, Martina
2008

Dostupný z http://www.nusl.cz/ntk/nusl-42884

Grey Literature in the Czech Republic

Petra Pejšová, Martina Pfeiferová

State Technical Library, Mariánské nám. 5, P.O.B. 2006, 110 01 Prague 1, Czech Republic

p.pejsova@stk.cz, m.pfeiferova@stk.cz

Abstract

Our paper summarizes and describes activities concerning grey literature in the Czech Republic. The managing organization of the activity is the State Technical Library (henceforth the STL); in the past, it was the STL, which was collecting, publishing and submitting grey literature data into the SIGLE system. The STL was the representative of the Czech Republic in the EAGLE. Now, EAGLE being extinct, there is no coordinated collection of grey literature on the national level since 2005.

The STL complements the role of the Czech National Library, which under the National Digital Library project, aims at accessibility of widespread published documents ("white literature"). On the other hand, the STL intends to deal with literature not acquired through normal bookselling channels (grey literature) and initiated a project for grey literature retrieval.

The project the National Repository of Grey Literature (henceforth the NRGL) is supported by the Ministry of Culture. Its main objective is formation of a digital repository of grey literature in the Czech Republic. The project aims at gathering metadata and possibly full texts of grey documents in the field of education, science and research. The NRGL shall solve

the typology of documents collected as well as metadata formats, persistent identifiers, intellectual property issues, SW and HW support, formation of network of collaborating institutions etc.

Close collaboration with representatives of Czech universities has been established. They face the issue of storing university qualification theses, which is one of the segments of typology of documents collected by the NRGL. The National Registry of Theses shall become a component of the NRGL. The STL has also got in touch with further producers of grey literature in the Czech Republic, in particular research institutes of the Academy of Sciences the Czech Republic, the institution covering a major part of production of grey literature in the segment of research and development.

The NRGL should facilitate to research the data on grey literature in the Czech Republic at one place with a single interface, as well as to retrieve the information on the owner of the document and - if possible - the full text of the document, either in electronic form or via the contemporary network of libraries (interlibrary loan, Document Delivery Service etc.).

The NRGL project does not assume retrospective digitalization of grey literature documents. However, we intend to certify the NRLG a trustworthy repository. The aim of the project is to provide services not only to the NRGL contributors, but also to the widest public. The STL has based the project on practice of universities, which had already experience with local repositories. The entire the NRGL project is consulted with the National Library of the Czech Republic as a part of the Czech Digital Library project.

Introduction

Our paper summarizes and describes activities concerning grey literature in the Czech Republic. We present projects addressing grey literature and especially the project the National Repository of Grey Literature (henceforth the NRGL), which solves the State Technical Library[1] (henceforth the STL).

About the State Technical Library

The STL is the central professional library under the governance of the Ministry of Education, Youth and Sports of the Czech Republic. It was founded in 1718; it provides library and information services to corporate and individual public, especially those in higher education, research and development. The STL is also a public library dedicated to science and technology, as such it collects and administers state-funded collections of czech and foreign literature and other information resources and sources pertaining to technology and applied natural and social sciences associated with technology.

The STL collections contain over 1.2 million volumes - books, journals, newspapers, scholarly studies, trade information, electronic documents and other publications and texts from the field of technology and applied natural and technology-related social sciences. Historical collection (books and journals published between 1500 and 1920) contains 22 965 volumes.

The STL is centre of digital document delivery system – the Virtual Polytechnic Library working on the basis of multifunctional centre of information services together with co-operative development of decentralized collection of periodicals specialized by subject and

---

[1] http://www.stk.cz/cs/

document type. The system has special union catalogue of collection of 50 Czech libraries and facilitates ordering by registered users from the Czech Republic.

The STL is host of the Czech National ISSN Centre (ČNS ISSN), which assigns ISSN numbers and registers continuing resources (serial publications) published on the territory of the Czech Republic.

History

The idea of the NRGL originated in the STL in 2005. The STL aims to collect a grey literature and to complement a role of the National Library Czech Republic[2] (henceforth the NL CR), whose main task is to collect and preserve "white" literature. The NL CR doesn't collect and doesn't plan to collect grey literature. The STL was motivated by the termination of the SIGLE system, formerly implemented by EAGLE, European Association for Grey Literature Exploitation. There were two participating members from the Czech Republic in EAGLE: the STL and the Library of the Academy of Sciences of the Czech Republic. To support the participation and collaboration, in 1994 the STL formed a specialised system, Co-operative System for Grey Literature. The system, based on bilateral agreements with grey literature producers, collected bibliographic records on grey literature, namely dissertation theses, from participating czech universities. Metadata collected were converted into a special SIGLE data format. Besides, the STL was processing grey documents from its own collections into the SIGLE format. Universities had an option to make a preliminary relevance-based choice of data submitted. Hence, this procedure did not cover the entire production of dissertation theses in the Czech Republic. Some universities sent to the STL

---

[2] http://www.nkp.cz/_en/index.php3

printed full texts of theses, this collection now contains over 4 000 theses. The task of the Library of the Academy of Sciences of the Czech Republic, i.e. to collect and file grey literature produced by institutes of the Academy of Sciences of the Czech Republic, was not fulfilled during the entire period of existence of the SIGLE. Thus, the STL was the only active contributor and national coordinator of the SIGLE activities in the Czech Republic. After the termination of the SIGLE, the STL began with an initiative to collect grey literature at the national level. The implementation of this initiative has begun in 2008 thanks to the support of the Ministry of Culture of the Czech Republic in the framework of research and development projects.

The NRGL Project

The 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. There are two participants: the STL and the University of Economics, Prague. The Project shall provide a working pilot application, which shall form the basis of the NRGL. Having evaluated and tested technology and methodology, we shall formulate Standards and Recommendations for institutions establishing their own local grey literature repositories. The Standards and Recommendations shall include our experience gained during the Project as well as rules and methodology; namely recommended metadata format, interchange formats and templates, samples of licence agreements and other relevant legal issues, methodology of protection, archiving and publishing of digital data, and web interface for communication with producers of grey literature. The Standards and Recommendations

shall be published both in Czech and English in printed form and on project web pages. The NRGL will hold grey literature records from research and development areas, Civil Service, entrepreneurial sphere, education and open access. We negotiate about co-operation with representatives of systems covering these areas.

Conception of the Czech Digital Library and the NRGL

The NL CR builds a concept of the Czech Digital Library[3] (henceforth the CDK) via a scheme. The core of the CDK is the National Digital Library (henceforth the NDK), which is at the centre of the scheme. The NDL is concerning just for "white" literature in framework of preservation of cultural heritage of the Czech Republic.
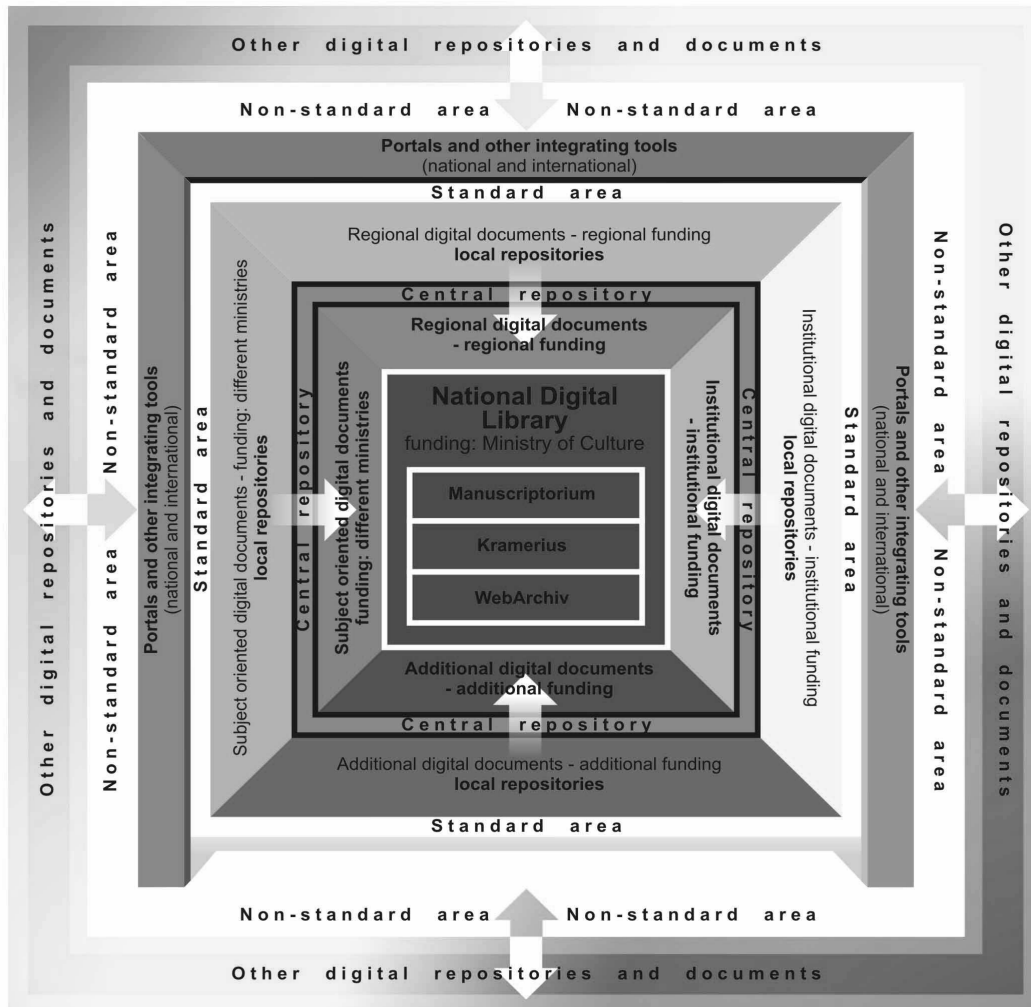
---

[3] http://www.ndk.cz/

Czech Digital Library

Other digital repositories and documents

Non-standard area   Non-standard area

Portals and other integrating tools
(national and international)

S t a n d a r d   a r e a

Regional digital documents - regional funding
local repositories

C e n t r a l   r e p o s i t o r y

**Regional digital documents
- regional funding**

National Digital Library
funding: Ministry of Culture

| Manuscriptorium |
| Kramerius |
| WebArchiv |

**Additional digital documents
- additional funding**

C e n t r a l   r e p o s i t o r y

Additional digital documents - additional funding
local repositories

S t a n d a r d   a r e a

Non-standard area   Non-standard area

Other digital repositories and documents

Other digital repositories and documents

Non-standard area

Portals and other integrating tools
(national and international)

S t a n d a r d   a r e a

Subject oriented digital documents - funding: different ministries
local repositories

C e n t r a l   r e p o s i t o r y

**Subject oriented digital documents
funding: different ministries**

Institutional digital documents - institutional funding
local repositories

C e n t r a l   r e p o s i t o r y

**Institutional digital documents
- institutional funding**

S t a n d a r d   a r e a

Portals and other integrating tools
(national and international)

Non-standard area

Other digital repositories and documents

Figure 1: Scheme of the Czech Digital Library

The NRGL will comprehend grey literature. In the basic functional scheme of the CDK, the
NRGL belongs to the Standard area, which is covering databases and aggregate catalogues of
digital documents; the NRGL shall be such a database. The NRGL itself shall collect
metadata and full texts from local repositories, both institutional and branch-oriented, and
other sources in the Standard area. The responsibility for local repositories (including

financial) will be born by particular institutions, and by their authorities, i.e. government departments and agencies.

Within the research project "Development of Mutually Compatible Information Systems for Access to Heterogeneous Information Resources and their Coverage by Uniform Information Gateway" NL CR defined a new sub-objective "Developing Digital Libraries and Repositories in consideration of their possible integration under the Uniform Information Gateway and other national information portals". In order to address this task, a specialised body, Working Group for Coordination of Development and Exploitation of Digital Repositories, originated. The participants in the NRGL project are members of this Group.

The National Registry of Theses and Plagiarism-Tracing System

The National Registry of Theses and Plagiarism-Tracing System is a project involving 20 Czech and Slovak universities. The project has two main parts. The first part of the National Registry of Theses (henceforth the NRT) gathers metadata on university qualification theses (i.e. metadata). The data is accessible to the public. The NRT contains over 30 000 records of theses now. The other part, Plagiarism-Tracing System, serves for detecting plagiarism in scholar texts. The system will help educators to unveil shady documents and decide, if the document in question (or its part) is or isn't plagiarism. The system can be found on the web page http://theses.cz.

From the outset of the project, the NRT has been considered an important source for the NRGL. Such documents form an important sector of grey literature to be filed by NRGL. The STL will collect metadata from the NRT into it's the NRGL repository using the OAI-PMH

protocol. Retrieving of full documents must be based on bilateral agreements between the NRT and particular universities. This is why we have prepared standards of licence terms. The compatibility is guaranteed by the participation of University of Economics, Prague.

The STL intends to collaborate with all universities and colleges in the Czech Republic (including private colleges) that are not participating in the NRT project, in order to obtain data on their qualification theses (and also - based on licence agreements - full texts as well).

In addition, the NRGL takes into consideration normative documents of "Electronic Theses and Dissertations Working Group, Association of Libraries of Czech Universities". At present, the metadata format of the NRT is being incorporated into the format of the NRGL.

The Register of Publication Activity implemented in the Academy of Sciences of the Czech Republic

The Academy of Sciences of the Czech Republic[4] (henceforth the AS CR) is a public non-university institution formed by a system of research institutes focusing on basic research. The AS CR defines the policy of scientific research; it is involved in both national and international research programmes, it supports collaboration with application sphere and it promotes the development of education. The AS CR consists of 54 research institutes.

The AS CR does not monitor its production of grey literature separately. Grey literature is included in a general system of monitoring of output and results of research (primarily for the purposes of research funding). These results are gathered by the Register of Publication Activity implemented by the Library AS CR[5].

---

[4] http://www.cas.cz/en/

[5] http://www.lib.cas.cz/cs

The STL collaborates with the Library AS CR. Metadata from the Register of Publication Activity of the AS CR[6] shall be collected into the NRGL repository using the OAI-PMH protocol. Full-texts retrieval from the AS CR institutes has to be treated in the same way like with universities, i.e. by licence agreements.

The Project Schedule

The Project is planned for period of four years. It is scheduled into three phases. In the First Phase currently running (April 2008 – June 2009), we shall define the requirements and documentation for the model application. This involves metadata specification, a choice of relevant persistent identifiers, analysis of typology of documents, software specification for the model application, a proposal of licence agreements for producers of grey literature and creation of project web pages. In the Second Phase, from July 2009 to September 2010, the model application shall be implemented, tested and evaluated. The Third Phase (July 2010 – December 2011) shall include formulation of standards, recommendations and methodology, verified in the Second Phase, which shall be published on the web as well as in a printed form. During all three phases, results and conclusions shall be published on the project's web pages. Seminars on an access to grey literature and electronic university qualification theses will take place annually. In addition, research results will be presented to the professional community.

---

[6] http://www.library.sk/i2/i2.entry.cls?ictx=cav&language=3

The NRGL metadata format

We have defined our own metadata format for the NRGL system. There are some important requirements that the format shall fulfil: simplicity, few obligatory attributes, consistency with Dublin Core[7], accessibility for OAI-PMH and possibility of creation of elementary indices. As a basis we used metadata formats of the two principal the NRGL participants: the NRT and the Register of Publication Activity of the AS CR[8]. We also took in account the format of the Information Register of Research and Development Results[9] (henceforth the Register R&D). We do not incorporate this format to be able to download data from the Register R&D directly, rather to enable local repositories, that created a record into the Register R&D with the possibility to provide the NRGL with the same record without any further conversion. If the OpenSIGLE will be re-opened, we plan to incorporate this metadata format into the NRGL metadata format in order to be able to submit records into the OpenSIGLE. We also respect Dublin Core format. At present, we are defining hierarchic relations among records and rules for particular identifiers.

Questionnaire survey

At the beginning of the project, we addressed total of 77 producers of grey literature via questionnaire while we focused on research institutes of the AS CR and public (state-funded) czech universities. The aim was to obtain contact information on persons responsible in individual institutions together with their intention to cooperate with the NRGL. In addition, we tried to determine the method of registration, collection and access to grey literature. Total

---

[7] http://dublincore.org/documents/dcmi-terms/

[8] http://www.iach.cz/knav/database_en.htm

[9] http://www.vyzkum.cz/FrontClanek.aspx?idsekce=1028

of 39 institutions (from 47 responding to the questionnaire) declared their willingness to cooperate with the NRGL. However, their approval depended on the licensing conditions that would be negotiated for this cooperation. The STL will commence the cooperation with these institutions in the second half of 2009 in accordance with the project schedule.

The NRGL Document Typology

When determining types of documents for the NRGL we started with typology of the GreyNet[10], the Register of Publication Activity of the AS CR and the Register R&D[11]. During the analysis of the document types we found that typologies concerned reflect different aspects such as events (manifestation, arrangement, organization,), form (presentation – notification, processing – translation, output), content (policy documents), places (domestic, foreign), type of format (e-text), etc. Therefore, we decided to divide our typology according to various aspects. Therefore, we decided to create typology, which will be structured in several levels so that we could express the various aspects better via a mind map, see picture below. We also anticipated that the NRGL would contain only "digital-born" documents.

---

[10] http://www.greynet.org/greysourceindex/documenttypes.html

[11] http://www.vyzkum.cz/FrontClanek.aspx?idsekce=29415
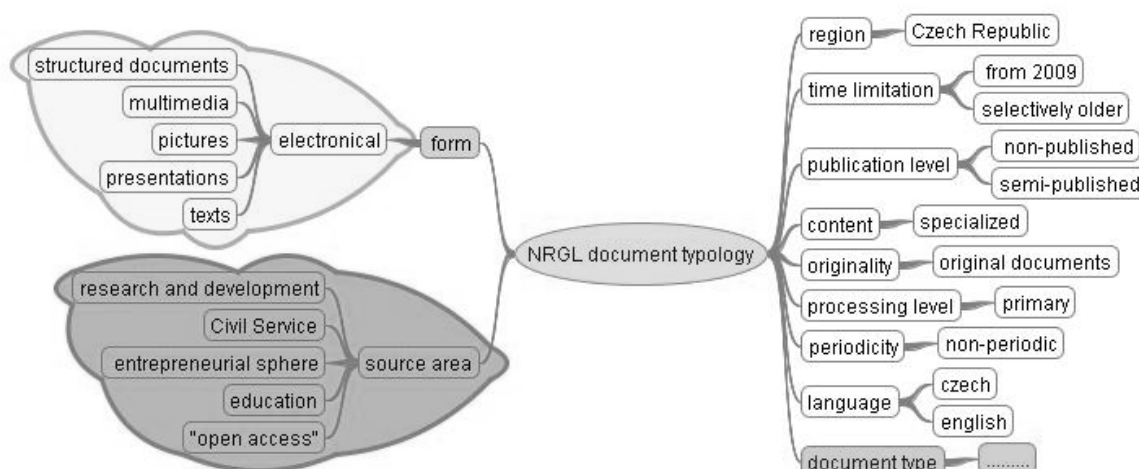
Figure 2: Mind map of the NRGL Document Typology

Software for the NRGL

During 2008, we were able to specify requirements for the SW functionality for the NRGL and follow-up services. These requirements were specified on two levels. In the first level requirements for pilot implementations of the system are specified, while on the second level requirements for further development of the basic system functionality are specified. During the process of SW selection, taking place in 2009, we will take into account the requirements of both the first and second level. Requirements of the second level are important for selection of such modern technology that will ensure compatibility of digital libraries with modern trends in information technology. Since we're considering the use of open source solutions, we compare SW requirements with available open source software for digital libraries. In this analysis we included the following open source SW: DSpace, Fedora, CDS Invenio, Eprint, and Greenstone. We plan to take a close look at LOCKSS Program and some Web 2.0 based tools, too.

During this year we also deal with the problems of persistent identifiers. We carry on analysis of available persistent digital object identifiers in the Czech Republic. Our decision will also depend on the phase of implementation of NBN (National Bibliographic Numbers) nationwide variant by the National Library of the Czech Republic.

The Project Web Pages

Within the project we have created a website which can be found at http://nusl.stk.cz/. Web pages are created in the Media Wiki[12] tool and WordPress[13] graphic style modified in accordance with the STL graphic manual. Media Wiki application supports interactive interface, which in our case represents the shared content creations within the working group and discussion forums opened to the public. The website contains information on the project, which are continuously updated, while we continue to open discussions on individual articles. In addition, there is information on the grey literature from the Czech Republic as well as the rest of the world, info on the legislation related to the grey literature in the Czech Republic, links to sites supporting R&D in the Czech Republic, on-line proceedings of the "Seminar on the access to grey literature" and grey literature information resources. In 2009 we plan to translate project web pages into English. For the moment, we have translated at least basic information about the project in the article "About project in English" available on Czech web pages[14].

---

[12] http://www.mediawiki.org/wiki/MediaWiki

[13] http://www.redaktionundalltag.de/

[14] http://nusl.stk.cz/index.php/About_project_in_english

Figure 3: Sample of The Project Web Pages

Seminar

In relation to the project, this year we organized for the first time "Seminar on the access to grey literature 2008", held on October 8th, 2008. It followed the seminar "Systems of access to electronic university qualification theses 2008", held on October 7th, 2008. At the seminar we presented the project itself, the present state of grey literature systems around the world, PDF format and ISO standards for long-term archiving, standardization of open archives focused on description and exchange of aggregated web sources via OAI-ORE, Creative Commons license copyright issues and the state of implementation of persistent identifiers in the NL CR.

All presentations and full texts of lectures are available in the online proceedings from the seminar on the Project web pages. Further seminars on the access to grey literature will be held annually in the coming years.