

## **Audit of the National Repository of Grey Literature (NRGL) in the NTK using the DRAMBORA tool**

Second audit, 2010

The first audit of NRGL as a trustworthy digital repository using the tool and methodology DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) was performed at the end of 2009 as an integral part of creating a digital repository of grey literature in the National Technical Library (NTK). The outcomes of the audit and the experience from its course were summarized in the final report and during 2010 they were published in the book of the NTK *Repositories of Grey Literature*.

The most important outcome of the audit was the set of identified risks related to the NRGL repository and endangering its operation, quality, image and other features. The NRGL team was trying to eliminate or moderate these risks during the whole year of 2010. The main principle of the DRAMBORA audit and, at the same time, its main contribution, is its iteration, i. e., its repetition after a certain time period in new conditions when the original risks are reassessed, the measurements adopted for their solution are assessed and new risks are identified.

The second audit of the NRGL digital repository was performed after one year, and, again, it was run internally, only with the help of an external consultant, who brings an independent view of the issue from outside and who meant a valuable reinforcing of the NRGL team especially at the year's end. The complete NRGL team participated in the preparatory phase and the acquisition of relevant data. A meeting of the team with an external consultant took place. On this occasion, the present state of the repository was assessed, the progress reached during 2010, new risks and possible ways to their elimination or moderation of their impact were identified by means of brainstorming. During 2010, the NRGL documentation, the description of the project, of its procedures and processes and related documents advanced significantly and thus, they represented a very good basis for the audit.

All acquired data were processed using on-line DRAMBORA tool (<http://www.repositoryaudit.eu>). For this second audit, the incremental method was chosen, i. e., including new data into the saved data of the existing audit in the form of comments and adding new records into the existing database. In order to maintain consistency, the same format was selected for this audit as for the one in 2009 and the most important information on the DRAMBORA tool and on the audit methodology were repeated in the new report to avoid the necessity of re-reading the previous report.

The very work with the DRAMBORA tool went on without great surprise and problems using the experience of the previous audit. On the basis of the experience from 2009, we communicated with the authors of the tool and methodology and we proposed some improvements and modifications. As a result, the web tool DRAMBORA appeared more stable after the one-year's pause; however, the most unpatently expected modifications have not been introduced yet, especially elimination of the unpleasant fact that the on-line version of the DRAMBORA tool does not support languages from the code page of iso-8859-2/windows-1250, neither the Czech language. Therefore, we decided to continue using English for the same reasons as last year, mainly because using Czech without diacritics is unnatural, furthermore, because the tool and the methodology are available only in English and, last but not least, since the audit results will be presented in the international field, in the area of grey literature projects and hopefully in the area of the audit of the DRAMBORA tool.

Notably, the careful mapping of the repository using the DRAMBORA tool may already give an overview of what is finished and what is not, which important documents, procedures, tools and measurements are missing and where the most critical risks reside, which endanger the existence and success of the repository, how the individual parts of the repository relate, which documentation is available etc. The data entered into the audit database may be linked to external sites by URL or any documents may be attached. Thus, the database may be considered as a centre of important information on the audited repository.

After all necessary information is entered the Reporting Centre function helps to create output reports on the identified risks for the repository, on their gravity, relations and solutions. Two types of output reports are available, those in .pdf format and those in .html format. Other saved descriptive information cannot be exported easily enough; however, with some effort they may be copied from the saved snapshot of the audit page. These output reports and copied information may be found below in this file or in the attachments.

Besides the mapped repository and its relevant environment, the producers of the methodology and the tool consider the most important output to be the analysis of identifiable risks endangering the repository, its quality, readiness, reputation and position in the eyes of both specialists and ordinary users. For those seriously interested, the complete audit is available at <http://www.repositoryaudit.eu>. Since the DRAMBORA tool does not provide read-only access, we cannot allow free access to the audit. In the following text, the identified risks and resulting conclusions will be analyzed. The complete audit may be found in the attachments and in the DRAMBORA database.

### **Analysis of identified risks**

As mentioned above, during the audit we assessed the risks identified in the last audit, especially the progress in their elimination, and we identified new risks. The NRGL repository is still in the pilot project stage; however, it is run on full software versions and real data are stored. Identified and reassessed risks mainly refer to the description of activities and procedures of the repository, the state and development of the staff, project funding, hardware and software sources including their backup and related NRGL environment.

16 originally identified risks were assessed and 8 new risks were identified – their order is insignificant, it originated in the DRAMBORA tool:

Risk Number 1: Loss of Staff Members

Risk Number 2: Pilot Project End

Risk Number 3: Disaster Recovery

Risk Number 4: Partner Network Voluntary

Risk Number 5: Backup Tapes Storage  
Risk Number 6: Financial Shortfall  
Risk Number 7: Budget for Services  
Risk Number 8: FAST Trial Version  
Risk Number 9: Weak Mandate  
Risk Number 10: No Ingest Policy  
Risk Number 11: Document Formats  
Risk Number 12: Software Administration  
Risk Number 13: Undocumented Policies  
Risk Number 14: Long Term Preservation Strategy not described  
Risk Number 15: Staff Skills insufficient  
Risk Number 16: Deliberate System Sabotage

Risks newly identified in 2010:

Risk Number 17: Duplicate project  
Risk Number 18: Partners do not supply fulltexts  
Risk Number 19: Slow growth of partner network  
Risk Number 20: Sample partner contract has limited usability  
Risk Number 21: Migration to new HW platform  
Risk Number 22: New CDS Invenio version  
Risk Number 23: Legal Risk - Authors Act  
Risk Number 24: Termination of legal support

## **Risk Number 1: Loss of Staff Members**

This risk considers the fact that the NRGL team is relatively small, recently created, it can hardly cover all related activities and thus, there is no substitutability if some team members are absent or leave. Thus, NRGL activity is directly endangered by any absences. Ideally, the repository team should be expanded in such a way that knowledge could be shared and substitutability could be increased; however, this is beyond the project budget. Another suitable solution would be to complete and broaden the descriptions of knowledge and activities of individual team members, possibly also in electronic form as used in Wikipedia, the free encyclopedia. Thus, any substituting team member could be quickly acquainted with his new activity.

State in 2010: During 2010, the team slightly expanded and thus the substitutability of its members increased. The activities of team members are documented and presented in regular team meetings; the documents are stored in a newly created knowledge database NRGL wiki. The team is compact since they work all together in one office.

The susceptibility to the risk has diminished; however, the risk still exists.

This risk is related to risks nos. 6, 12, 15 and optionally to risk no. 16

## **Risk Number 2: Pilot Project End**

One of the NRGL mandates follows from the project of the Ministry of Culture of the Czech Republic for the period of 2008-2011 – The Digital Library of Grey Literature – Functional Model and Pilot Realization. This project will come to a conclusion in 2011, which may cause weakening of the NRGL mandate and possible financial problems. This challenge may be met by applying for participation in another similar project or looking for other financial sources to continue and develop the NRGL project.

State in 2010: Main NRGL activities (long-term acquisition, archiving and providing access to grey literature) are going to be covered in the NTK budget after 2011.

Further development of the repository should ideally be covered by another grant project. This risk substantially diminished and its basis slightly altered; however, it still exists.

This risk is related to risks nos. 4 and 6, partially also 9.

### **Risk Number 3: Disaster Recovery**

There is a guideline for the backup of the NRGL database but no plan exists on how to proceed in the case of recovery after a potential failure or damage to the database and who shall be responsible for individual steps. This risk may lead to the prolongation of the time needed to recover after a failure or after other problems. Such a plan must be prepared including the nomination of responsible staff and connection to the backup plan, which also needs to be finalized and embodied in the NRGL wiki mentioned in Risk no. 1.

State in 2010: Along with the strategy and methodology of the backup, a plan for the recovery of the repository after a failure is being developed. This plan should be finalized and verified. The susceptibility to the risk slightly diminished; however, the risk still exists.

This risk is indirectly related to risk no. 5.

### **Risk Number 4: Partner Network Voluntary**

The network of partner organizations supplying material to NRGL is based upon volunteer participation; no law mandates its creation, all is exclusively based on the willingness of employees of partner institutions. This risk may cause difficulties in supplying material to NRGL. A possible solution consists in finding motivating factors for partner organizations (such as the benefits of a centralized repository, the international aspect of the project, etc.), and in encouraging the promotion of NRGL targeted to partner organizations and academic public. Last but not least, a motivation similar to publishing in specialized journals would be helpful at a political level – some kind of positive evaluation for organizations supplying their grey literature production to NRGL, possibly with impact on their funding.

State in 2010: Partners are offered local installation of the CDS Invenio system preinstalled for building a grey literature collection. Further support consists in methodical guidelines and instructions for NRGL partners, targeted promotion of the NRGL in specialized journals, in the conferences etc. Contracts with 16 partners were signed but this risk still exists.

This risk is related to risks nos. 2, 6 and 9.

### **Risk Number 5: Backup Tapes Storage**

Backup tapes containing copies of the NRGL database are stored in the same room as the hardware repository. If there were a fire, terrorist attack or another serious disaster, the database would be destroyed together with its backup. At the present time, under pilot operation, when the database only contains minimal records, this risk is negligible; however, in the full operational phase, an alternative storage location must be found and this fact must be specified in the finalized backup plan.

State in 2010: Backup strategy and methodology including manipulation with backup tapes and recovery plan after failure are under development. These documents must be finalized, verified and saved in the NRGL wiki. This risk slightly diminished; however, it still exists.

This risk is indirectly related to risk no. 3.

### **Risk Number 6: Financial Shortfall**

The present budget and funding sources cover the project in its existing form but they do not assure NRGL development in any relevant areas. No institution, project or company showing zero development and growth has a potential for a long-term existence.

The solution is the same as that for risk no. 2 – timely application for participation in a similar project or search for other financial sources to support the continuation and development of the NRGL project.

State in 2010: The project funding is being continuously assured, for instance for software licenses, upgrade of hardware, etc. See also risk no. 2.

This risk is related to the majority of other risks mentioned here.

### **Risk Number 7: Budget for Services**

Services supporting the FAST software system, which assure the NRGL user interface and are provided by the INCAD company, are under contract and covered by the budget until the end of 2010.

When budgets are prepared for the NTK and the NRGL for 2011, this item should not be neglected and discussions about continuing cooperation with the INCAD company should be initiated in time.

State in 2010: This item is covered in the budget until 2013. It should not be forgotten later on.

This risk is related to risks nos. 2 and 6.

### **Risk Number 8: FAST Trial Version**

The NTK is running the trial version of the FAST software system based on a contract with the INCAD company until the end of 2011. This software supports the NRGL user interface.

During 2010, discussions about the purchase, operation and support of the full version should be initiated to ensure routine operation of the NRGL. The purchase of licenses and their support must be included in the appropriate budget.

2010: Since the full version of the FAST system is currently under operation, this risk was eliminated.

This risk is related to risks nos. 6 and 7.

### **Risk Number 9: Weak Mandate**

The NRGL mandate is based on a time-limited project of the Ministry of Culture of the Czech Republic and on a brief note in the NTK status. This fact may negatively influence the reputation and position of the NRGL in the eyes of the academic public and partner organizations at home and abroad.

This problem should be solved at a political level, for example by gaining formal support from the NL CR, and AS CR, by promotion in the respective ministries, etc.

State in 2010: The topic of grey literature is included in the concept of library development in the Czech Republic until 2014. A new supplement to the status of the NTK of 12th May 2010 adds new mandates in paragraphs 3f and 3h; see the appropriate section of the audit. This risk was substantially minimized.

This risk is related to risks nos. 2, 4 and partially 6.

### **Risk Number 10: No Ingest Policy**

There is no formal procedure and methodology for handling the acquired documents, processing them, determining which checkpoints must be cleared before saving in the database, who is responsible for individual steps etc. This risk may result in ununified process of saving acquired documents and possibly in their insufficient processing and monitoring. As a result, the saved data may lack consistency. This methodology must be drawn up, including nomination of responsible staff, and must be embodied in the electronic NRGL wiki, mentioned in Risk no. 1.

State in 2010: Acquisition methodologies are being developed; a new position of the contents manager was created and filled in the NRGL team. The risk has strongly diminished.

This risk is loosely related to risks nos. 11, 13 and 14.

### **Risk Number 11: Document Formats**

Although the typology of NRGL documents is well described, formats for documents to be saved in the NRGL database are not sufficiently defined. This is also true for formalized procedures and processes involving potential migration of formats to newer versions, acquisition of new software needed to access these new versions etc. This risk may result in the inconsistency of saved documents and to problems accessing them over a long time span.

A plan and described methodology should be prepared defining which document formats will be supported by the NRGL and how their migration to upgraded versions will be done, including the software necessary to access the documents.

State in 2010: An analysis of this issue was ordered and created, along with the analysis of long-term storage of digital contents, which also contains manipulation with document formats. This analysis must be verified and drawn into practice. At the moment, this risk is very weak.

This risk is loosely related to risks 10, 13 and 14

### **Risk Number 12: Software Administration**

The administration of the NRGL repository software system – CDS Invenio – depends on a single individual within the NTK. In addition, this person is not formally controlled by the NRGL manager but is a member of the IT department. This employee is difficult to integrate into the NRGL team and his working performance is not optimal. If he is absent, there is no acceptably rapid means to substitute for him. This could seriously endanger NRGL operation.

A possible solution may be to hire another person instead this employee and thus increase substitutability in software administration, or, alternatively, to hire an external company under contract. The latter solution seems more suitable but tends to be more financially demanding. This fact should be considered during budget preparation.

State in 2010: The software administrator was included in the NRGL team, another team member was trained in the basic administration, the administration documentation is continuously created and saved in the NRGL wiki database. A weak risk.

This risk is related to risks nos. 1 and 6 and loosely to risk no. 16.

### **Risk Number 13: Undocumented Policies**

Although the documentation of the NRGL project is at a very good level, the audit identified several specific points, at which there is insufficient documentation or none at all. The audit concentrated on the most substantial facts and risks in the pilot project phase. After the audit is completed, the documentation of all basic processes must be analyzed, areas lacking coverage must be identified and a timetable for completing the relevant documentation must be created, including the electronic NRGL wiki encyclopedia – see Risk no. 1.

State in 2010: Important working procedures and processes are continuously documented, the knowledge database NRGL wiki was newly created, in which relevant documents are continuously saved. If this procedure is strictly followed further on, this risk is very weak.

This risk is related to Risks nos. 1, 10, 11, and 14.

### **Risk Number 14: Long Term Preservation Strategy not described**

A long-term strategy for storing digital documents is one of the most important aspects in operating and developing a trustworthy digital repository.

This strategy globally considers the vitality of the digital repository regarding technology (such as migration of document formats), functioning of repository team and of the organization as such, funding of repository operation and development etc. The rudiments of this technology are contained in the documentation of the NRGL pilot project; however, this strategy must be prepared as a separate document.

State in 2010: An analysis of this issue was ordered and created, along with the analysis of manipulation with NRGL document formats. This analysis must be verified and drawn into practice. At the moment, this risk is very weak.

This risk is related to risks nos. 10, 11 and 13.

### **Risk Number 15: Staff Skills insufficient**

As novel technologies and methods for building and operating digital libraries are developed, it may happen that the knowledge and capabilities of the repository staff will not keep pace with the latest developments. This risk endangers any team and is especially relevant in the case of a small, busy team with practically zero substitutability, where there is insufficient time for training and education. The NRGL management should not underestimate this problem and should try to provide space and means for regular training and education of themselves and all team members. Naturally, this should be taken into consideration when the number of team members and the NRGL budget is planned.

State in 2010: The expanding team is continuously gaining necessary knowledge by self-study, by participation in specialized conferences and training. The team members themselves state to be gaining unique knowledge. This topic must remain among the main priorities of NRGL management. The risk has been substantially minimized.

This risk relates to risks nos. 1, 6 and 16.

### **Risk Number 16: Deliberate System Sabotage**

The risk of deliberate system sabotage by dissatisfied employees or employees on notice, damage to the database, long-term disruption of data consistency and

other factors should never be excluded or ignored. Similarly to risk no. 15, such risks are especially relevant in the case of a small, busy team with practically no mutual substitutability and with one member formally belonging to another department; moreover, if no formal procedure exists for the database management.

The NRGL management should try to stimulate positive motivation in the team and encourage their professional development. On the other hand, functional backup should be implemented, along with control mechanisms such as periodic control of data consistency in the NRGL database.

State in 2010: The repository contents and the system are daily backed up, important procedures are documented, the role of the contents manager was created, the substitutability of team members increased along with their knowledge and, consequently, with their satisfaction. All these facts are resulting in minimization of this risk.

This risk is related to risks nos. 1, 6, 12 and 15.

## **Risks newly identified in 2010**

### **Risk Number 17: Duplicate project**

Masaryk University in Brno is showing activity in order to gain funding, support and mandate for creating an alternative national repository of grey literature. The NRGL repository is in the pilot project stage and as such it is not fully coined among its partners and academic public; the partner network is growing slowly. This risk may weaken the position of the NRGL and thus partially degrade the effort and costs for its building.

This risk is related to risks nos. 4, 9 and 19.

### **Risk Number 18: Partners do not supply fulltexts**

The contract partners of the NRGL repository do often not provide full texts of documents to be stored, which complicates their accessibility, automatic indexing of records and their subsequent internationalization for those linked to the international network of grey literature organizations and digital repositories. It is necessary to perform dissemination and persuading activities.

This risk is related to risks nos. 4 and 19.

### **Risk Number 19: Slow growth of partner network**

Despite a large scale of services and promotion events for NRGL partners, the partner network is growing slowly and it is very difficult and time demanding to gain a new partner, for instance for the following reasons:

- insufficient support of management in the potential partner organizations
- potential partners state to have lack of grey literature
- insufficient knowledge and capacities in the potential partner organizations

The promotion effort should be intensified, a case study with a good partner should be created, support by a specific person in the potential partner organization should be sought for, etc.

This risk is related to risks nos. 4 and 20.

### **Risk Number 20: Sample partner contract has limited usability**

Considering the generally complicated Czech legal environment and the low willingness to collaborate, the partners offer require a great number of changes in the sample partner contract to seal their cooperation, which prolongs and complicates gaining new partners. Without the collaboration with the Faculty of Law of Masaryk University in Brno this risk would be even more pronounced. It is necessary to analyze the most frequently required changes, to modify the sample partner contract, or, alternatively, to prepare several variants of the contract.

This risk is related to risks nos. 4, 19 and 24.

### **Risk Number 21: Migration to new HW platform**

At the beginning of 2011, the NRGL should be transformed to a new platform. Such a step always brings about some risk concerning i. a. data security, drop-out of services and accumulation of work in the related areas. Such a transformation must be well prepared, the employees should get trained in the use of the new platform and specific risk areas such as back-up and reparations in the case of problems should be assessed in advance.

### **Risk Number 22: New CDS Invenio version**

At the beginning of 2011, a new version of the CDS Invenio software system will be released, on which the NRGL repository is based. Such a step always brings about some risk concerning i. a. data security, drop-out of services, accumulation of work and modifications in the related software areas. The transfer to the new version must be well prepared; employees should get acquainted with the new software version, especially in the areas related to the NRGL repository, and specific risk areas including backup and restoration in the case of problems.

### **Risk Number 23: Legal Risk - Authors Act**

Authors Act is a generally complicated legal standard, Czech legal environment is still not sufficiently stable and legal awareness and attention devoted to this topic is rarely at the necessary level on the side of the NRGL partners. As a result, there is still a potential risk that the material provided by the partner and subsequently processed, stored and offered to the public by the NRGL repository may be burdened by unclear author rights, for example concerning coauthorship, or in case the author gave exclusive rights to the third party. This situation potentially threatens with legal disputes.

NRGL partners should be continuously educated in this respect; a document or a presentation on this topic should be prepared.

This risk is related to risk no. 24.

### **Risk Number 24: Termination of legal support**

In 2010 the NRGL repository team used legal service of two lawyers from the Faculty of Law of Masaryk University (JUDr. Radim Polčák a Mgr. Jaromír Šavelka). Their cooperation is quite financially profitable and it is based on the personal interest of both lawyers in the topic of the repository. Since this cooperation is based on particular persons and their personal interest, there is a potential risk that it may be terminated. For this reason, it would be suitable to find some alternative possibilities of potential substitution of this cooperation.

This risk is related to risks nos. 19, 20 and 24.

Generally, the risk elimination is much easier in case the respective area is fully under control and in charge of the NRGL management and team. If the risk relates to the cooperation within or even outside the NTK, the situation is considerably more complicated. The creation of a knowledge database NRGL wiki means a great progress; this database should be further developed and strictly used because the continuous documentation of procedures, activities and results of the NRGL team is of crucial importance for the elimination or minimization of the impact of most risks. A great progress may be seen in the development of the NRGL repository since the last audit and new activities and goals have been added. Therefore, the documentation of activities and analysis of risks are the more important. A large portion of risks reflect the topic of building the NRGL

partner network, i. e., the partner network of providers of the repository content. Importantly, this area should be of priority especially as for promotion and education. In relation to NRGL partners, a sufficient attention should be devoted to legal issues connected to the Author Act.