

Title: Collecting Grey Literature – Institutional Repository versus National Aggregator

Authors: Petra Černošková and Hana Vyčítalová, National Library of Technology, Czech Republic

Abstract

The Czech National Library of Technology (NTK) provides two digital repositories – the National Repository of Grey Literature (NRGL) and the NTK Institutional Digital Repository (IDR). NRGL's primary is providing access to grey literature as well as long-term archiving and preservation of grey literature from various institutions in the Czech Republic. The IDR was created for collecting and archiving of employee-generated content and other documents, including grey literature, connected with the library and its services. Our poster highlights the differences between collecting grey literature at the institutional level (through the institutional repository) and at the national level. What commonalities and differences do they have? What problems do they solve? Differences include not only overall conceptions and document types, but also methods for collecting, legal issues and standards as well as functionality and options. Thanks to our experiences in managing both types of repositories, we define general differences, obstacles, and development possibilities. Information presented here, including a mode for cooperating at the institutional or national level, is useful for all institutions planning to start collecting (not only) grey literature at the institutional or the national level even at cooperating institutional model/level.

Introduction

The poster represents the two most common streams for how the grey literature in electronic form is presented nowadays - by institutional repositories and aggregators. The contribution highlights the differences between collecting grey literature at the institutional level and at the national (or international) level. The major difference is that an institutional repository usually provides access to and preservation of the institutional publications, whereas an aggregator gathers results from multiple resources such as databases, repositories, digital libraries, or webpages.

The Czech National Library of Technology (NTK) runs two repositories – the National Repository of Grey Literature (NRGL¹, since 2009) and the NTK Institutional Digital Repository (IDR², since 2011). NRGL's primary aim is providing access to grey literature at the national level as well as long-term archiving and preservation of grey literature from various institutions in the Czech Republic (NRGL Project, 2017). The IDR was created for collecting and archiving of employee-generated content and other documents, including grey literature, related to the library and its services.

Based on our experiences, we define seven of the most important topics necessitating discussion when establishing a new repository – general conception, document types, collecting methods, participation, legal issues, functionalities, and accessing. The topics are briefly discussed in the rest of the contribution.

General conception is the starting point for all the topics.

General conception

The general conception of the institutional repository is quite clear compared to an aggregator. Institutional repositories usually collect publishing activities of the institution and internal documents such as directives or reports of business trips. Nowadays there is increasing number of repositories,

¹ <http://nusl.cz/?language=en>

² <http://repozitar.techlib.cz/?ln=en>

including also datasets (131 out of 2952 institutional repositories³). The general conception of aggregators differs from one to the other. For example, the conception of the NRGL is to collect all grey literature in the Czech Republic. The aggregators very often are multi-disciplinary or have multiple purposes. Based on the data from OpenDoar, we have created a chart dividing aggregators by their focus/profile (Chart 1).

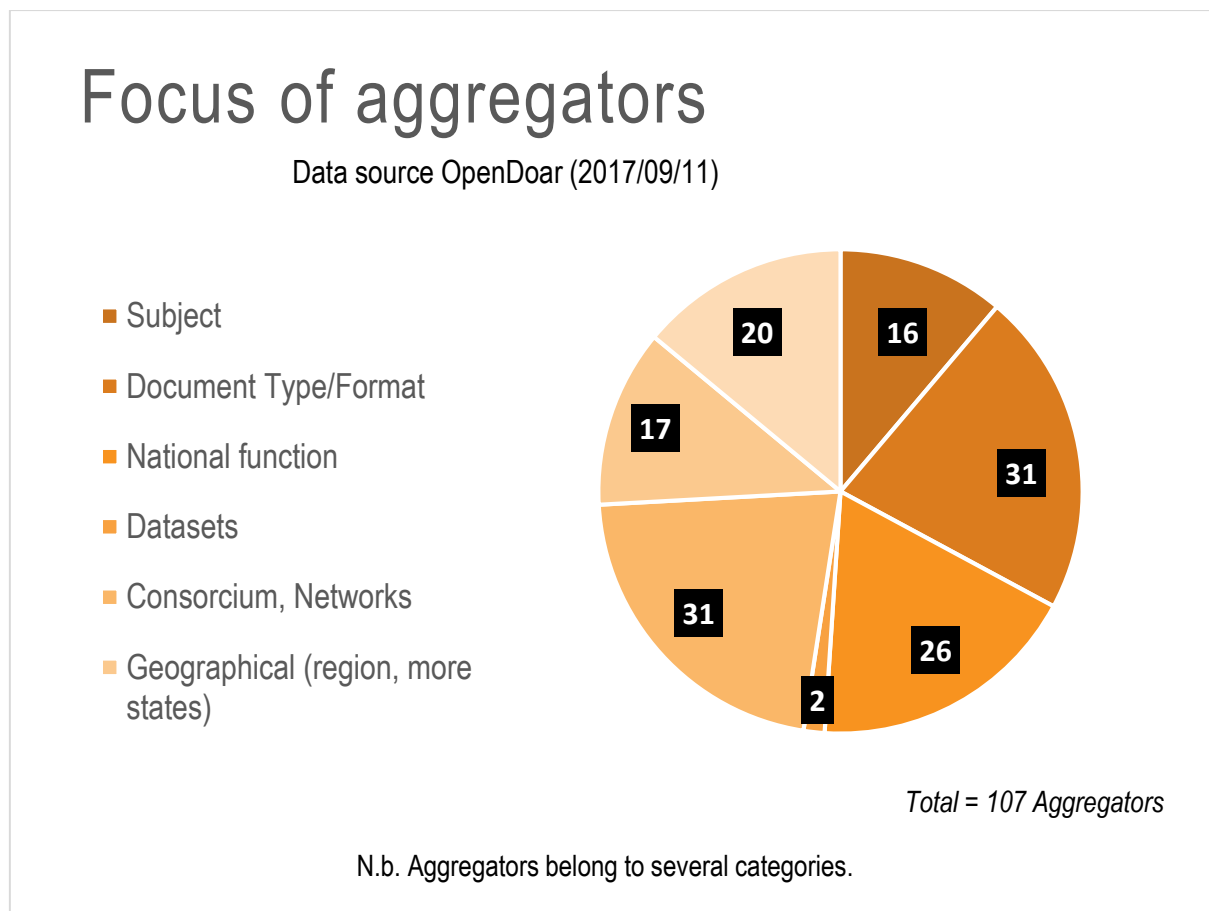


Chart 1: Focus of aggregators (data source: OpenDOAR <http://www.opendoar.org/>)

Document types

There are plenty of document types that could be collected by any repository. At the beginning, it is necessary to analyse or map which types of documents are published in the institution(s) and in which formats. Regarding formats, it is important to verify which are appropriate for long-term preservation, and which are more suitable for access to end users. Any repository must be flexible enough to deal with the addition of another type of documents than have been determined at the beginning and fulfil needs of the institution(s). Because each repository and institution has different needs and publication activities, we distinguish different document types for the IDR and the NRGL (see Table 1; types in italics are common for both typologies). These typologies also have been changing on account of changing needs of the institution(s).

³ OpenDoar, up to 17th October 2017

Table 1: Comparison of NRGL typology and IDR typology

NRGL	IDR
Academic theses (ETDs)	Academic theses (ETDs)
<i>Bachelor's theses</i>	<i>Bachelors thesis</i>
<i>Master's theses</i>	<i>Master's thesis</i>
Doctoral theses	Others
Habilitation theses	
Rigorous theses	
Analytical and methodological materials	
<i>Analysis</i>	
<i>Methods</i>	
Studies	
Author works	Author works
<i>Monographs</i>	<i>Monographs</i>
<i>Preprints</i>	Scholarly Articles
Reviews	Post-prints
Thematic collections	<i>Preprints</i>
Conference materials	Conference Materials
<i>Posters</i>	<i>Posters</i>
<i>Programs</i>	<i>Programs</i>
<i>Papers</i>	<i>Papers</i>
<i>Proceedings</i>	<i>Proceedings</i>
Promotional and educational materials	Promotional Materials
Brochures	Photographs
Flyers	Leaflets
Exhibition catalogues	Monitoring
Exhibition guides	Help
<i>Press releases</i>	Awards
	Invitations
	PF
	Videos
	<i>Press releases</i>
Reports	Reports
<i>Business reports</i>	<i>Business reports</i>
Grant reports	<i>Progress reports of the project</i>
<i>Progress reports of the project</i>	<i>Annual reports</i>
Statistical reports	<i>Final report of the project</i>
Technical reports	Work placement reports
Research reports	
<i>Annual reports</i>	
<i>Final reports of the project</i>	
Status reports	
Survey reports	
Trade literature	Informative Documents
Trade print	<i>Analysis</i>
Product catalogues	Conceptions
	<i>Methods</i>

Gazettes	Normative documents
	Statistics
	Legal Documents
	Study Materials
	Presentations of training

Table 1: Comparison of NRGL typology and IDR typology

The OpenDoar differentiates following 12 content types: Journal articles, Theses and dissertations, Unpublished reports and working papers, Books, chapters and sections, Conference and workshop papers, Multimedia and audio-visual materials, Learning Objects, Bibliographic references, Datasets, Other special item types, Software, and Patents (OpenDOAR, 2016). You can see in Chart 2 that the collected documents are very similar but the intensity and the order is slightly different.

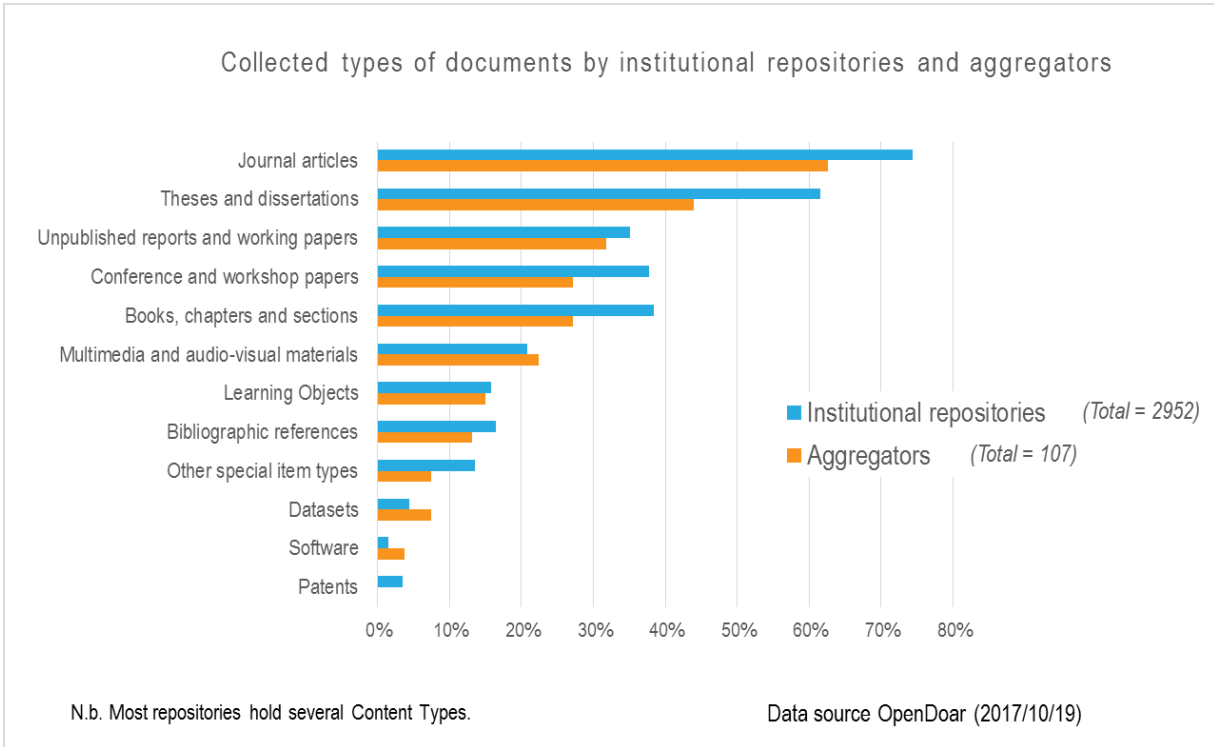


Chart 2: Types of documents collected by institutional repositories and aggregators (data source: OpenDOAR <http://www.opendoar.org/>)

Collecting methods

An institutional repository can use the same collection methods as an aggregator. However, they can have very often different priorities and possibilities. Nevertheless, they should agree on one strategy – get the records and full-text quickly and easily.

This strategy could be trickier for the institutional repositories if there is not any other system(s) from which they could harvest the documents via OAI-PMH or export and then import to the repository. Exclusion of these solutions moves the possibilities to the self-archiving or manual submission by a person in charge of the repository, very often a library staff member.

Aggregators often combine the aforementioned methods. The most suitable is harvesting via OAI-PMH which allows collecting the data automatically from other repositories or library catalogues at regular intervals; for example, once a week. There are many metadata formats that can be used, therefore it increases the importance of the metadata schema selection for your repository. Based on analysis of the resources that could be harvested, keep in mind that selection of the same schema as other

institutions use means less extra work. The most common metadata format within repositories is Dublin Core and, within library catalogues, modifications of MARC.

Participation

At the institutional level, all employees should participate and submit their employee works to the repository. It is usually determined by institutional mandates or norms signed by an organizational director. In cases of universities, these mandates mostly include also theses and dissertations. Nevertheless, for implementing mandates, authors, all departments with publishing activities, and the repository team (often part of the library) must participate.

Participation in the aggregator's network is usually voluntary. It is related to the willingness of the document/data producers such as universities or research centres to share their publishing outputs, grey literature, or data. This willingness can be affected, in case of manual submissions, by the fact that the depositor often does the submissions as an extra additional task with lower priority. There could be some exceptions when the participation is determined by the law – e.g. results of the research.

It could seem that an institutional repository will more easily reach high numbers of submissions, but this is not always the case. If there is no penalty for missing submissions, the publishing activity of the institution or the content collected, the repository is never going to be completed. On the other hand, it cannot be forgotten that copyright legislation will always provide some limits and specific conditions.

Legal issues

Legal issues are very important topic and sometimes very specific in the context of grey literature. Even grey documents are authorial works and are protected by copyright legislation in many countries. Managers of all repositories must be aware of this and respect this.

In the case of institutional repositories, there are usually employees' works and emerge out of the status the employer has rights to archive these works in its repository. It is not necessary to ask the authors for their agreement, but many institutions do so to be safe⁴.

Aggregators are the opposite – since participation in an aggregator is voluntary, repositories cannot collect anything without previous agreement or contract with the collected organizations or all authors. It can be classical paper contract or only an electronic confirmation checked during submitting of the document. For our repositories we prefer paper contracts. Then there is an option of using any type of open licenses, e.g. Creative Commons. Documents marked with Creative Commons' symbol are possible to archive or share freely under the terms of the license.

It is very important to keep in mind that copyright regulations are different in every country, especially if part of planned management is the involvement of repositories in other countries. See Chart 3 to notice the stratification of the repositories around the world.

⁴ For further reading: “Breaking It Down: A Brief Exploration of Institutional Repository Submission Agreements” by Amanda Rinehart and Jim Cunningham, available from <https://doi.org/10.1016/j.acalib.2016.10.002>

Repositories by continent

Data source OpenDOAR

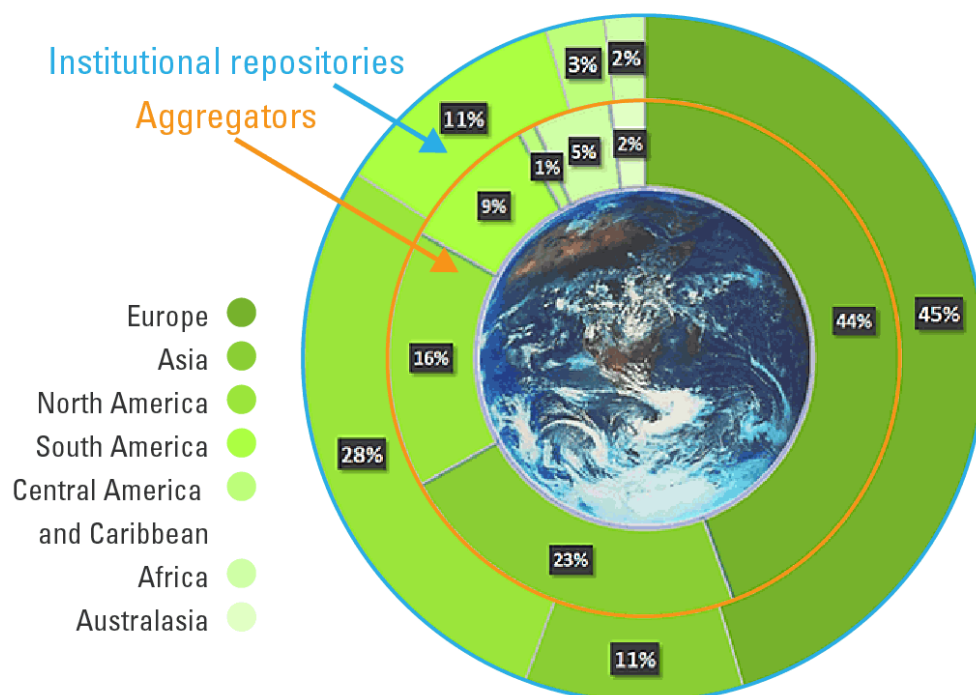


Chart 3: Repositories by continent (data source: OpenDOAR <http://www.opendoar.org/>)

Functionalities

The primary function of an institutional repository is long term preservation of its documents. Secondary functions could be considered a digest of publishing activities of the employees or a presentation of the institution to the public (Charvátová, 2016). However, it is expected that the employees will use it the most often. Search functionality could be limited if there is higher emphasis on archiving. As mentioned above, the main user groups are employees and - at universities - professors and students, based on this fact the easy management of electronic documents is required to simplify their work.

The aggregator's most important functionality is searching over many different sources from one access point together with very robust advanced searching options. The aim is save time to the end user who would otherwise have to search all these sources one by one. The aggregators usually have access to less full-texts, but they capable provide information on full-text availability and link end users directly to the full-text. Secondly, the goal is to collect documents with similar aspects or topics and, in some cases, this could have equal or even higher priority than search functionality. For example, at the NRGL case, goals - to collect all Czech grey literature and to enable search over many resources – are going side by side.

Accessing

The institutional repository serves mainly to the particular institution and its internal policy decides about access to documents. An institutional repository can be open access or only for internal use, or a mix of both options. An embargo period also very often appears regarding to the journal articles, documents from research area, or project documents. In the IDR, access to internal documents is limited to the IP addresses in the building, but access management could also be managed using employee user accounts.

The aggregator's provider usually cannot decide easily about accessing of the documents. He or she must respect copyright regulations and particularly licenses and conditions given by the producers of the documents (authors or institutions) in some agreements. There could be some exceptions too; e.g., when the existence of the aggregator is established by the law or by some authority (Ministry, European Commission). However, in many cases the aggregator provides access to metadata and only selectively to full-texts (e.g., if there is some agreement between the producer and the provider of the aggregator or if the full-text is open access or under any open licence). Metadata records without full-text should always contain the information about full-text availability. See the schema below for a detailed comparison of institutional repositories and aggregators in point of view of accessibility.

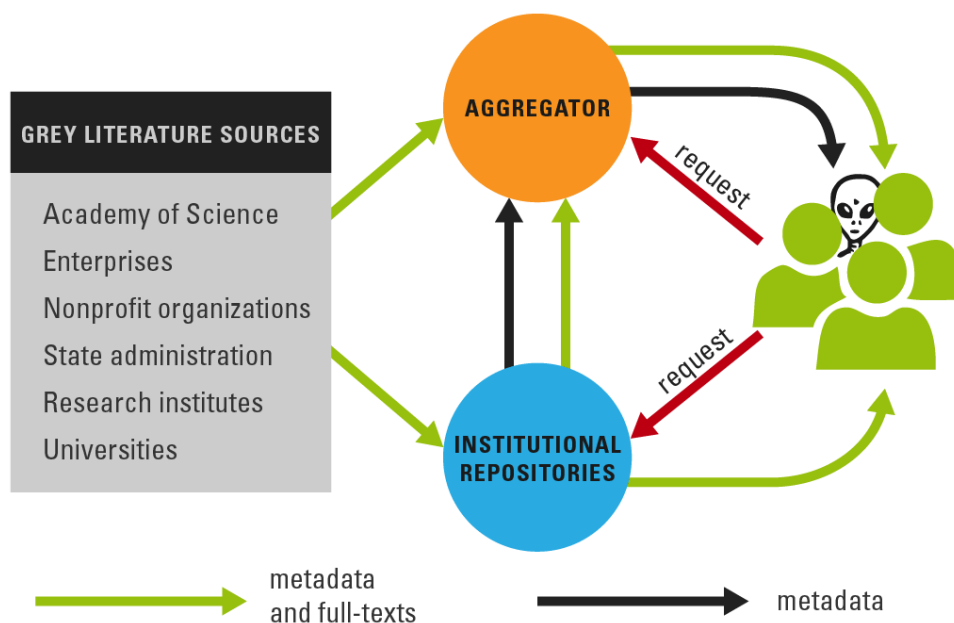


Chart 4: Schema of aggregator workflow (source: The National Library of Technology)

Conclusion

Merging all the discussed points and factors – the general conception, the collected document types, the collection and preservation methods, participation and legal points of view, manners of presentation of results and dissemination tools – could help in the selection of the best system for a new repository (institutional or aggregating type) and even at the beginning of a project or when rethinking the current situation of your repository.

References

CHARVÁTOVÁ, Michaela, 2016. Institucionální digitální repozitář NTK. *Čtenář* [online]. Středočeská vědecká knihovna v Kladně, 2016, 68(11). ISSN 1805-4064. Available from: <http://ctenar.svkkk.cz/clanky/2016-roc-68/11-2016/tema-institucionalni-digitalni-repozitar-ntk-163-2372.htm>.

NRGL Project, 2017. National Repository of Grey Literature [online]. Praha: National Library of Technology [cit. 2017-12-01]. Available from: <https://nrgl.techlib.cz/nrgl/nrgl-project/>

OpenDOAR: The Directory of Open Access Repositories, 2016 [online]. Nottingham, UK: University of Nottingham, UK, 2014 – 2016 [cit. 2017-12-01]. Available from: <http://www.opendoar.org/>