e-ISSN 1574-180X

An International Journal on Grey Literature

Volume 17, Number 1, Spring 2021

'EXPLORING THE GREY SIDE OF OPEN SCIENCE'

www.textrelease.com Grey Literature Network Service www.greynet.org

The Grey Journal

An International Journal on Grey Literature

COLOPHON

About TGJ

The Grey Journal is a flagship journal for the international grey literature community. It crosses continents, disciplines, and sectors both public and private.
The Grey Journal not only deals with the topic of grey literature but is itself a document type classified as grey literature. It is akin to other grey serial publications, such as conference proceedings, reports, working papers, etc.

The Grey Journal is geared to Colleges and Schools of Library and Information Studies, as well as, information professionals, who produce, publish, process, manage, disseminate, and use grey literature e.g. researchers, editors, librarians, documentalists, archivists, journalists, intermediaries, etc.

CIP
The Grey Journal (TGJ) : An international journal on grey literature / Dominic Farace (Journal Editor); Jerry Frantzen (Technical Editor) ; GreyNet International, Grey Literature Network Service. - Amsterdam: TextRelease, Volume 17, Number 1, Spring 2021 – TIB (DE), DANS-KNAW (NL), CVTISR (SK), EBSCO (USA), ISTI CNR (IT), KISTI (KR), NIS IAEA (UN), NTK (CZ), and the University of Florida (USA) are Corporate Authors and Associate Members of GreyNet International.
e-ISSN 1574-180X (PDF)

Subscription Rate:
€240 institutional

Contact Address:
Back Issues, Document Delivery, Advertising, and Subscriptions:

TextRelease
Javastraat 194-HS
1095 CP Amsterdam
Netherlands
T +31 (0) 20 331.2420
info@textrelease.com
http://www.textrelease.com/glpublications.html

About GreyNet

The Grey Literature Network Services was established in order to facilitate dialog, research, and communication between persons and organizations in the field of grey literature. GreyNet further seeks to identify and distribute information on and about grey literature in networked environments. Its main activities include the International Conference Series on Grey Literature, the creation and maintenance of web-based resources, a moderated Listserv, and The Grey Journal. GreyNet is also engaged in the development of distance learning courses for graduate and post-graduate students, as well as workshops and seminars for practitioners.

Full-Text License Agreement

In 2004, TextRelease entered into an electronic licensing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The full text of articles in The Grey Journal (TGJ) can be found in *Library, Information Science & Technology Abstracts* (LISTA) full-text database.

Contents

'EXPLORING THE GREY SIDE OF OPEN SCIENCE'

EDITOR'S NOTE

NEXT GENERATION GREY

Grey literature is once again at a crossroads in its known 80 years in the vernacular. It appears that every quarter or more century this field of information faces a change in direction. In the mid-20th Century, it originated with the collection of government documents – namely war reports. In the years succeeding, it expanded to numerous other types of documents and collections produced by organizations in government as well as academics and business. This period was occupied with the acquisition and document delivery of these scientific and technical materials. It was also during this period that the many problems associated with these materials surfaced, namely their lack of indexing, translation, open access, preservation, and assessed value for science and society.

Since the final decade of the 20th Century up to the present, the field of grey literature has initiated and undertaken evidenced based research hand in hand with technological developments and sustained information management. The problems of the prior period became challenges and this led to programs of research and education in grey literature. This period coincides with the digital transformation of grey literature. While much has been accomplished in connecting the supply and demand sides of grey literature, the field once again finds itself in transition driven by the fact that the entire information landscape is itself in a period of unprecedented change and flux.

GL2021 offers the many and diverse communities of practice in grey literature a unique opportunity to collaborate in addressing and defining the next phase in the digital transformation of grey literature. Together this can be accomplished by unlocking the potential next generation grey holds for information science and society.

Dominic Farace,
Journal Editor

Open Science and the Transformation of Scholarly Communication*[1]

Jerry Sheehan, National Library of Medicine, Bethesda, Maryland, USA

Abstract

The continuing transition toward open science is fundamentally changing both the ways in which scientists communicate research findings and what they communicate. Open science places emphasis on enhanced access not only to published research findings but also to elements of grey literature: pre-published literature (preprints), underlying research data, study protocols, and other products of the research process. This presentation reviews the transformation in scholarly communication associated with open science and illustrate the ways in which libraries and other information service providers can support it, drawing on examples from the U.S. National Library of Medicine. Particular attention will be devoted to efforts to support open science as part of the ongoing response to the COVID-19 pandemic.

Introduction

A significant change is taking place in the way the scientific community works and interacts. This change can be characterized as more widespread adoption of the principles and practices of open science. Open science is not new to science, nor to the grey literature community, but it is gaining more widespread adoption. In the last year alone, considerable progress has been made around the world in advancing open science, both as a continuation of a longstanding trajectory and as a response to the unprecedented, pandemic-inspired times in which we are living. The acceleration of the open science transformation presents new opportunities for the Grey Literature community.

Illustrations of the ways open science is transforming scientific communications can be seen in the work of the U.S. National Library of Medicine (NLM).[2] NLM is a component of the U.S. National Institutes of Health (NIH). Like other parts of NIH, NLM funds and conducts research, with its focus on biomedical informatics and data science. NLM is also a library, the world's largest biomedical library by any of several measures. It is well-known for the range of information services it develops and provides. These include:

- Literature resources, such as Pubmed,[3] the database of more than 30 million citations and abstracts to the biomedical literature, and PubMed Central (PMC),[4] a digital archive of full text biomedical literature
- Scientific data resources, such as ClinicalTrials.gov,[5] a registry and results database for clinical trials, and the Database of Genotypes and Phenotypes,[6] which holds and shares data from studies of the relationship between variations in genotypes and the expression of disease or other features.
- Consumer oriented information, such as MedlinePlus,[7] which provides consumer-friendly information from trusted sources on more than 1000 diseases and conditions.

With the more than 8,000 organizational members of the Network of the National Library of Medicine,[8] NLM works to ensure widespread awareness and effective use of these resources.

---

* First published in the GL2020 Conference Proceedings, February 2021.

[2] See https://www.nlm.nih.gov/

[3] https://pubmed.ncbi.nlm.nih.gov/

[4] https://www.ncbi.nlm.nih.gov/pmc/

[5] https://clinicaltrials.gov/

[6] https://www.ncbi.nlm.nih.gov/gap/

[7] https://medlineplus.gov/

[8] https://nnlm.gov/

Together, these resources position NLM as a platform for biomedical discovery and data-powered health – and a home of open science

Open Science and Scholarly Communications

Open science is a set of principles and practices that make the products and processes of scientific research more findable, accessible, interoperable, and (re)usable (FAIR).[9] It is about making these resources FAIR for all those who can use them--scientists, innovators and entrepreneurs, students, and educators, and members of the general public--recognizing, of course, that some information resources cannot be widely shared out of respect for privacy, confidentiality, national security, and other concerns.

By making the products of research more widely available, open science can accelerate science, improve its rigor and reproducibility, spur innovation, improve health, enhance education, strengthen the economy, advance other societal objectives, and build trust in science. All of these outcomes increase the return on investments in research. Many individuals and institutions have roles to play in advancing open science, including scientists, research administrators, funding agencies, libraries and archives, publishers, and information technology professionals.

Traditionally, and certainly in the sciences, scholarly communication focuses largely, if not exclusively, on literature, in particular peer-reviewed scholarly publications that report on the results of research. Open science certainly includes scholarly publications, but also drives interest in making more of the scientific process and its outputs accessible and available for use. These outputs include:

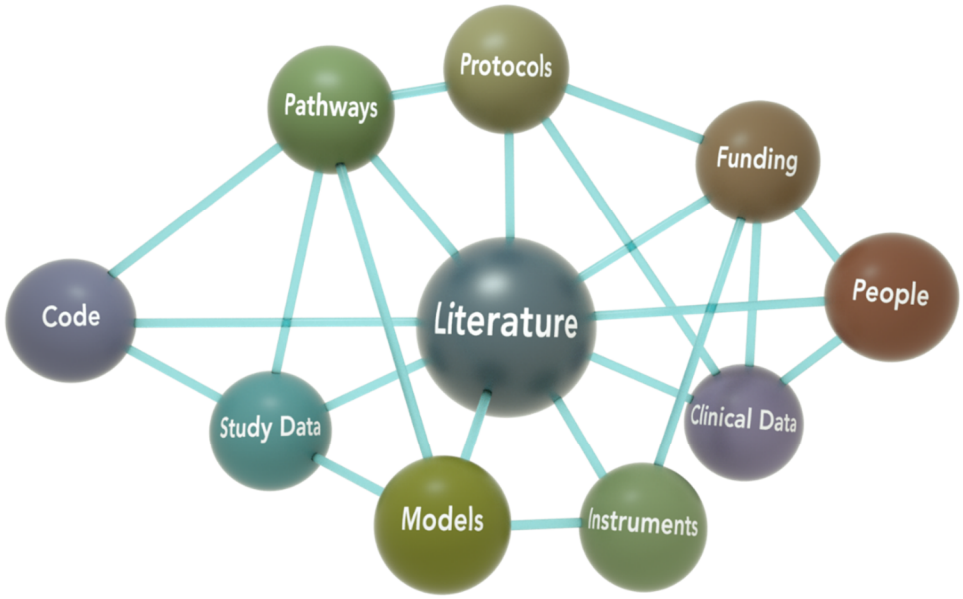
- Related study data, e.g., the data sets generated in a research study. In medical research, this includes associated clinical data that may have been a source of information for research;
- Study protocols that describe in detail the aims of the research (e.g., hypotheses to be tested), how the research was conducted, and how the data were collected and analyzed;
- Code and software that was used to analyzed data and help generate the results, which can be necessary for validating results and reusing data;
- Funding information that describes the sources of funding for a research study, which is of particular interest to understanding potential conflicts-of-interest and enables funders to track the outcomes associated with their investments;
- Information about instrumentation and other critical infrastructure used in the research, which can demonstrate the utility of those resources and support reproducibility;
- People involved in the research, including the authors of any research publication as well as those who contributed in other ways, such as managing the data and developing the code used to analyze it; and
- Analytical pathways and pipelines used to manipulate the data.

The value of increasing access to these research objects can best be achieved by connecting them together, creating the linkages that allow a user to navigate from one element to the other, ideally with the click of a button (Figure 1). With robust linkages between an among elements, users can enter a research network via any of the research objects and navigate their way to related information. Creating an interlinked network of associated research objects requires considerable effort from stakeholders across the broader research and scholarly communication enterprise to promote things like interoperability, to assign unique

[9] See https://www.nature.com/articles/sdata201618

identifiers to research objects, and help ensure the quality and trustworthiness of those research objects.

Figure 1. Network of Interlinked Research Objects



Pubmed Central as Hub For Open Science

Recent work of the NLM illustrates the interlinking that can advance open science and expand the accessibility of research objects beyond (but including) scholarly publications. One example centers around NLM's PubMed Central (PMC), the digital archive of full-text biomedical literature. PMC contains a growing collection of more than 6.5 million articles submitted by journals and by authors who deposit their manuscripts in compliance with the NIH Public Access Policy.[10] That policy requires researchers whose work is funded by NIH to ensure that their final, peer reviewed manuscripts or final published articles are submitted at the time of publication to PMC, where they are made freely and publicly accessible not more than 12 months later.

PMC is heavily used by researchers, care providers, students, educators, and the general public. Each day, some 3 million unique visitors retrieve some 6 million articles. A growing subset, now equal to more than half of the total content, is available not only for individual retrieval but also for bulk downloading to support text mining and machine learning applications. That subset includes articles published using open access mechanisms and those submitted as author manuscripts

Recognizing its high usage by a wide range of users, NLM has begun to use PMC as hub to support access to other research resources associated with available articles. Most notably, PMC is being used to improve the discoverability of associated data. This can take the form of: 1) a data citation that provides a link to a data set in an external data repository; 2) a link to a data availability statement in the article that provides a narrative description of data and how to access it; or 3) a link to data submitted to PMC as supplemental material.

[10] https://publicaccess.nih.gov/

PMC accepts small datasets, up to 2GB in size, that are associated with submitted articles. NLM issued updated guidance describing procedures for doing so.[11] If there is no existing persistent identifier for the dataset, NLM will generate and display with the article a globally unique identifier to support the citation and discovery of this content.

As of October 2020, more than 1.5M articles in PMC had an associated dataset. PMC has search filters to allow users to find articles that include associated data. Adding this feature has resulted in a notable increase the number of data sets accessed through PMC. NLM expects an increase in the availability of datasets related to biomedical journal articles as a result of the new NIH Data Management and Sharing Policy.[12] This policy, issued in October 2020, requires that all applicants for NIH funding provide with their application a data management and sharing plan that outlines the data to be collected in the study and proposed approaches for preserving and sharing the data.

NIH has developed a recommended approach for researchers to use in selecting a data repository for preservation and sharing.[13] It expresses a preference for deposit in a disciplinary or data type-specific repository, including the 80 or so repositories supported by NIH.[14] It also calls out the role of PMC, cloud storage, institutional repositories, and generalist repositories as acceptable repositories for data preservation and sharing. NIH also identified a set of desired characteristics for those repositories – characteristics such as assigning a persistent ID, attaching good metadata, having plans for long-term sustainability.

To further advance open science, NLM has also begun including certain preprint articles in PMC through a new Pilot Project.[15] This step is significant for NLM because it includes content that has not been peer reviewed in a flagship literature systems. To help address concerns about the quality of preprint literature, the pilot was launched in June 2020 with a focus on preprints resulting from NIH-funded research that are deposited in preprint servers that demonstrate a set of good practices, e.g., they offer basic screening for quality, assign identifiers, have processes for updating and linking to final articles. The first phase of the pilot focuses exclusively on COVID-19 related research, where the immediacy of access to results is paramount. In addition, PMC contains various banners, labels, links, and filters to help ensure users recognize these entries as preprints, understand that they have not been peer reviewed, and to control whether they are included or excluded from search results.

As of October 2020, PMC contained more than 1,000 preprints dating back to January 2020. About 30 percent had been published – and PMC links to the published version. The remaining 70 percent are available ahead of publication, satisfying NLM's aim of accelerating access to these important research results. NLM is evaluating the first 3 months of the Pilot to help inform future steps and possible expansions.

NLM is also working with publishers to make open peer review materials available in PMC. Not many journals have yet taken the step of adopting open peer reviews, but NLM is working with those that are doing so and helping to identify preferred approaches. NLM issued guidance a year ago about tagging peer review documents to submit with articles in PMC.[16] The preferred approach is to have peer reviews submitted as separate objects that can be independently identified and linked to the journal article, rather than having them bundled together with the article as a single object.

[11] See https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/#suppm
[12] https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html
[13] https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-016.html
[14] https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html
[15] https://www.ncbi.nlm.nih.gov/pmc/about/nihpreprints/
[16] https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-019-0063-9

Clinicaltrials.Gov as an Open Science Platform

NLM's ClinicalTrials.gov provides another example of how grey literature can become better integrated into open science. ClinicalTrials.gov contains information about clinical trials conducted in the United States and elsewhere in the world, by public and private sector organizations. Some trials are required to register and submit summary results – namely those studying drugs, devices and biological products that require marketing approval from the US Food and Drug administration and those of any type of intervention that are funded by NIH.[17] Registration provides structured information summarizing key elements of the study protocol: study design, eligibility criteria, and locations of trial sites. Summary results includes tables of information summarizing participant characteristics, outcome measures by arm of the trial, and adverse events.

As of October 2020, ClinicalTrials.gov contained registration information for more than 350,000 studies, and results information for more than 45,000 them. For clinical trials subject to regulations and NIH policy, results must be submitted regardless of whether they are formally published. Recent research has found that about one-third of registered trials did not result in a peer reviewed journal article as long as 5 years after the completion date.[18] [19]

Like PMC, ClinicalTrials.gov is becoming a hub for open science—clinical science, in particular. In addition to summary results, Responsible Parties are also required to submit their full protocol documents and statistical analysis plans (if separate). Under separate human subjects protection regulations in the United States, responsible parties must also submit the informed consent forms used to communicate information to potential participants about the benefits and risks of participating in the study.

NLM systematically connects records in ClinicalTrials.gov to associated journal publications – including those in PMC. NLM is also exploring the other elements of grey literature to include or link to: information such as conference abstracts, clinical study reports, and press releases. NLM is also considering ways to link to external data repositories that contain the participant level data underlying the summary data we receive – recognizing that additional protections are needed for that data.

Open Science and the Response to Covid-19

Platforms such as PMC and ClinicalTrials.gov have enabled NLM to contribute in important ways to the global response to the COVID-19 pandemic. In March 2020, the director of the White House Office of Science and Technology Policy and science ministers and advisors of a dozen other called upon publishers to make freely and immediately available all publications reporting on COVID-10 and the broader family of coronaviruses.[20] To ensure easy discoverability of and access to these publications, they requested that publications be made available through systems like PMC in formats that would allow automated text processing.

The response to that call to action was unprecedented. More than 50 major commercial and society publishers began working with NLM to modify workflows begin submitting articles in machine readable forms with no embargo. As of October 2020, NLM had more than 140,000 articles in its COVID-19 literature collection, dating from 1970 to present. Of these, 77,000 specifically related to COVID-19; the others to the broader family of coronaviruses. Through

[17] https://clinicaltrials.gov/ct2/manage-recs/background
[18] https://pubmed.ncbi.nlm.nih.gov/22214755/
[19] https://www.bmj.com/content/352/bmj.i637
[20] https://www.imagwiki.nibib.nih.gov/content/ostp-president-trumps-science-advisor-and-government-science-leaders-around-world-call

the end of September 2020, these articles had been retrieved a combined total of more than 80 million times—far above the average for recent literature.

NLM has made these articles available to a larger public-private partnership that has assembled an even larger collection of coronavirus articles to support text mining. The COVID-19 Open Research Dataset, or CORD-19, maintained by the Allen Institute for Artificial Intelligence, has become a resource for the artificial intelligence and machine learning community to develop and apply natural language processing algorithms.[21] It has been used in several challenges and competitions aimed at text mining the collection to gain new insight into COVID-19.[22] [23]

ClinicalTrials.gov also plays an important role in the response to COVID-19, providing information to researchers, clinicians, and the public about relevant clinical trials related. To ensure rapid dissemination of comprehensive information about COVID-19 research, NLM adapted and streamlined submission procedures and has prioritized the processing of coronavirus-related information. NLM is also providing one-on-one support to researchers during the process of submitting results information to ClinicalTrials.gov to address questions and optimize reporting. New features ensure that information about COVID-19 studies registered on the World Health Organization's International Clinical Trial Registry Platform are discoverable through ClinicalTrials.gov. As of October 2020, registration information was available for almost 4,000 clinical trials registered directly with ClinicalTrials.gov and an additional 3,000 trials registered with the WHO.

Recognizing the value of making information about COVID-related clinical trials available as rapidly as possible to the widest possible audience, NIH Director, Francis Collins issued a statement strongly encouraging the clinical research community to register their clinical trials and submit summary results information for COVID-19 and SARS-CoV-2 trials as quickly as possible and ahead of regulatory and policy deadlines.[24] This statement comes as a number of important, large scale vaccine and therapeutic trials are drawing closer to completion.

Next Steps for Grey Literature

As these examples illustrate, there is a growing opportunity associated with open science to improve the discoverability of, access to, and utilization of a diverse set of research outputs, including many outputs that would be considered grey literature. The fact that grey literature tends to be more open – less encumbered by restrictions on use – has to potential to make it easier to integrate into open science networks.

The opportunities are accentuated by the growing interest in open science resulting from ongoing experience in the response to the COVID-19 pandemic. There is greater appreciate of the value of making research outputs more readily available and accessible and of making them more interoperable and usable. The urgency around the situation has also engendered a greater willingness among all parties to experiment with new approaches to opening science and breaking with traditional models.

This all bodes well for grey literature. There is an opportunity to seize. Clearly there are challenges to address in making grey literature in its many forms more findable, accessible, interoperable and reusable. Should the community establish platforms and networks that make grey literature more FAIR, integrate grey literature into other emerging open science networks, or make grey literature more discoverable through major search engines. A

[21] https://allenai.org/data/cord-19
[22] https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge
[23] https://ir.nist.gov/covidSubmit/
[24] https://www.nih.gov/about-nih/who-we-are/nih-director/statements/nih-calls-clinical-researchers-swiftly-share-covid-19-results

combination of these approaches may help best ensure that grey literature is integrated into open science.

Plotting next steps will take collaboration among the many individuals, organizations, and stakeholders involved in grey literature and open science networks. The grey literature community must explore ways to engage with the broader stakeholder community to advance those discussions and engage more deeply in open science, further advancing the transformation that is well under way.

Czech National Repository of Grey Literature

NUSL is

a digital repository for grey literature

Free

online access

Goals

- Central access to grey literature and the results of research and development in the Czech Republic
- Support of science, research and education
- Systematic collection of metadata and digital documents
- Long-term archiving and preservation
- Cooperation with foreign repositories

Features

Provider:
National Library of Technology
Prague, Czech Republic

Records:
over 500,000 records

Collection provenance:
Czech Republic

Partners:
over 150 organizations (Academy of Science, Public Research Institutions, Universities, State Offices, Libraries, NGOs etc.)

International Cooperation:
OpenGrey, OpenAire, ROAR, OpenDOAR, BASE, WorldWideScience

What else?

Conference on Grey Literature and Repositories
https://nusl.techlib.cz/en/conference

Informative Webpages
https://nusl.techlib.cz/en/

www.nusl.cz

NTK
Národní technická knihovna
National Library of Technology

NUSL
czech national repository of grey literature

Czech Grey Literature and Research Outputs: Transformation of the NUŠL Service*

Petra Černohlávková and Hana Vyčítalová
National Library of Technology, Czech Republic

Abstract
The Czech National Repository of Grey Literature (NUŠL) has been bringing GL to science and society over 10 years. It is based on Lisbon definition from 1997, updated in 2004, and on its own document typology that was created in 2008 as one of the outputs of the same name project. Since then there have been several papers and articles calling for new definition and new typology, asking what should be still considered GL and what not anymore. (Savić, 2018; Baxter and Hilbrecht, 2020) Savić (2019) pointed out that if everything can be marked as GL then probably nothing is GL.
The NUŠL already had to change its typology few times to provide still good service and to reflect research institutions' needs. However, it is not only the document typology that influences the repository and the service, there are other circumstances which affect further development of the NUŠL such as partner institutions' needs, the National Research, Development and Innovation Policy, European Commission's guidelines on research outputs, current quantity and quality of other repositories and systems for grey literature and research results in the Czech Republic etc.
The contribution/paper will bring a brief overview on NUŠL's experience over 10 years of providing the service. Then it will be focused on a current situation in research area of the Czech Republic in detail. Finally, out of these points will conclude to the necessary changes that has to be taken by NUŠL and his provider – National Library of Technology in Prague, in order to keep the service meaningful.
Keywords: The Czech National Repository of Grey Literature; Grey literature; Research and development; Research outputs; Digital repository

Introduction

The Czech National Repository of Grey Literature (NUŠL) has been bringing grey literature to science and society for over 10 years. During this period, the fourth industrial revolution started, based on the initial concept presented in Hannover Fair in 2011. Its main principle is "smart factory" or "smart manufacture" based on cyber-physical systems, internet of things, artificial intelligence etc.. Industry 4.0 will affect every activity and every industry in some way, including grey literature (GL). "In the last few decades, developments in information technology have had an immense impact on the way we manage information in general, and on the way we create, disseminate and use GL" (Savić 2018).

Even NUŠL must react to this and to the changing Czech societal and R&D environment. NUŠL already has had to proceed with several reforms in the past to continue providing good service and to reflect the needs of research institutions, often related to its document typology, but now it faces a major change that will affect the entire service.

This paper summarizes experiences with NUŠL, GL, and its collecting processes. These experiences, together with external influences, provide the background for NUŠL's transformation and a new GL approach.

* First published in the GL2020 Conference Proceedings, February 2021.

NUŠL Service and Lessons Learned

With the NUŠL service, the National Library of Technology in Prague (NTK) followed a long tradition of collecting GL. Within the project,[1] a document typology[2], metadata standard[3], cataloging rules, data import and export possibilities, and a submission workflow were established. The repository was built using Invenio, version 1, open-source software (https://invenio.nusl.cz). A central search interface (https://nusl.cz) was created above the repository using FAST software (Pejšová 2009), nowadays on open-source Elasticsearch.

In 2012, the repository switched to regular operation as a service. By the end of 2020, it managed to extend its content up to 600,000 records on documents in various languages that have Czech provenance. Czech institutions (e.g., universities, public and private research institutions, museums, and others) that provide content on the basis of a signed cooperation agreement have two options to participate: either automatically through their systems (usually via OAI-PMH) or by manual submission directly into NUŠL made by their authorized submitters**.** Using the service is free of charge and voluntary. The only exception is the Regulation of the Minister of Culture on Certified Methodologies funded by the Ministry of Culture of the Czech Republic; full-texts of these methodologies are mandatorily submitted into NUŠL (Czech Republic. Ministry of Culture 2018).

Technical aspects of NUŠL's operations

During its entire operation, NUŠL's workflow has been continuously adjusted and new functionality has been added, such as possibilities for sharing records, exports into citation managers, and links to project records in the Czech R&D Information System[4] (hereafter, R&D System) via project number (Charvátová, Pejšová 2015). However, we have recently run into more barriers and limits to fulfilling the mission of the repository, conceptually as well as technically, and including limiting content acquisition options. Automated harvesting from other storages means the less effort for providers, but correction of potential errors in NUŠL is more complicated. Whereas, direct manual submissions are time demanding. Submitters have to deal with challenging assignments: select GL from other outputs, collect GL from colleagues, and then create records in NUŠL and attach full texts. Furthermore, submitters often have to insert the same documents into multiple systems, mostly the R&D System where submissions are mandatory each year due to legislative requirements,[5] then into library catalogs (ILS) or institutional publishing activities reports. This double workload is exhausting and needless. We would like to focus on this issue in the future and reduce the manual workload. Because of this, future interoperability with the R&D System is essential.

Since 2018, we had to start addressing problems with the repository software because Invenio v1's sustainability ended (CERN 2020) followed by, in 2020, support for Python 2.7 (Peterson 2014), the programming language in which Invenio v1 was written.

Invenio is an open-source software and its support and development depend primarily on CERN, the European Organization for Nuclear Research. 63 installations[6] of different versions

[1] Digitální knihovna pro šedou literaturu – funkční model a pilotní realizace, 2008 – 2011 [Digital Library for Grey Literature – functional model and pilot release], project no. DC08P02OUK007

[2] First NUŠL document typology (https://invenio.nusl.cz/record/111521/files/idr-266_3.pdf)

[3] NUŠL Metadata Format, version 1 (http://www.nusl.cz/ntk/nusl-111514)

[4] The Czech R&D Information System (https://www.rvvi.cz/)

[5] Zákon č. 130/2002 Sb. Zákon o podpoře výzkumu a vývoje z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu a vývoje), available from: https://www.zakonyprolidi.cz/cs/2002-130

[6] of which 28 is registered OpenDOAR

are known worldwide (CERN 2019a), of which four are in the Czech Republic[7] plus one private library (Tilsch Library[8]). This leads to human resources issues in the Czech Republic because few developers are familiar with Invenio.

In 2018, CERN released Invenio v3 (CERN 2020), no longer a classical ready-to-install repository but a framework that provides the background for building one's own final software solution. Its advantages include immense flexibility and the possibility of building any solution. Its disadvantage is increasing demands on human resources, especially developers. The "Invenio Framework" is intended primarily for large-scale repositories, which means it can handle an enormous amount of data: over 100 million records and petabytes of files (CERN 2019b); this is the main advantage of the new version. Also advantageous is a change of internal format from MARCXML to web-native JSON, which brings flexibility in metadata description but increased labor requirements. Therefore, these circumstances might lead some institutions and people operating small installations of Invenio v1 or v2 to a consideration of migrating to other software solutions.

Migrating to Invenio v3 means many challenges for NTK, but most importantly, new opportunities for enhancing the repository and its functionalities. For example, Elasticsearch, used for the current central search interface (nusl.cz), is incorporated within Invenio v3. This we recognize as a substantial improvement because there will no longer be the need of the separate central search interface. Confusion about different URLs should disappear and user experience should be substantially improved.

GL influence on the NUŠL service

NUŠL only focuses on grey literature, which has its pros and cons, and means NUŠL has a clear range of documents that can be accepted. These documents are not usually registered anywhere, and NUŠL profits from this. Generally, there are not many concerns about making GL freely available, because GL is not burdened with relations between authors and publishers. Mostly, it is about gaining the consent of co-authors or defining the institutional policies regarding employee works. Nevertheless, due to the separation of GL from other outputs or publication activities, and since white literature is given preference over GL because of science evaluation principles, GL rarely attracts attention.

The term "GL" often causes troubles of its own. It is still not well known among Czech librarians and information professionals, not to mention other researchers or the public. Moreover, not even information professionals cannot agree on what document types should be still considered "grey", under what conditions. GreyNet International[9] provides an international GL document typology, but other types still appear or are under discussion. For example, Simadlová (2012), and Gelfand and Lin (2012) describe the various content of social networks and media as being GL. Ferry et al. (2008) also found a parallel between Second Life and GL. Not only text materials can be considered GL. Video recordings (e.g., recordings from events or other scientific videos) (Drees, Plank 2018) or research data (Giannini, Deluca, Molino, Biagioni 2018) can also be considered GL. Savić (2019) pointed out that if everything can be marked as GL, then probably nothing is GL.

Defining the exact target user group for GL is also difficult, since a broad spectrum of documents can be marked as GL. Target groups of each type of GL can be diverse: students typically focus primarily on electronic theses and dissertations (ETDs) and often do not notice

[7] Institutional Digital Repository NTK (National Library of Technology) http://repozitar.techlib.cz;
National Repository of Grey Literature (National Library of Technology) https://invenio.nusl.cz;
The Institute of Theoretical and Applied Mechanics (https://invenio.itam.cas.cz);
Surveying Library of VUGTK (Research Institute of Geodesy, Topography, and Cartography) https://knihovna.vugtk.cz/

[8] Tilsch Library available from https://violet.katabrozova.cz/collection/Emanuel%20Tilsch?ln=cs

[9] Document Types in Grey Literature (http://www.greynet.org/greysourceindex/documenttypes.html)

other types of GL, whereas researchers may be more interested in conference proceedings and research reports. Current Google Analytics settings for NUŠL partially help to determinate the most searched document types, but the question of actually *who its users are* remains open. Statistics data only show the way users enter the NUŠL sites (direct, organic or referral) and how much time they spend in the repository. The most consequential unknown is if users do not stay long on the NUŠL platform because they have found what they were looking for or if this is the other way around and Google Analytics does not suffice for answering this question.

Evolution of the NUŠL typology

Over time, NUŠL's focus on GL, together with its applied typology, have themselves become hurdles to use of the repository by more partner institutions. The original NUŠL typology resulted from research into several typologies in 2008 (State Technical Library 2009a). The typology has been adjusted several times based on experiences since then (i.e., growth of each document type and dealing with partner institutions contacts regarding their document production). Figure 1 contains its current form. A vital part performed the willingness and possibilities of institutions to make documents publicly available in NUŠL (Černohlávková, Vyčítalová 2018).

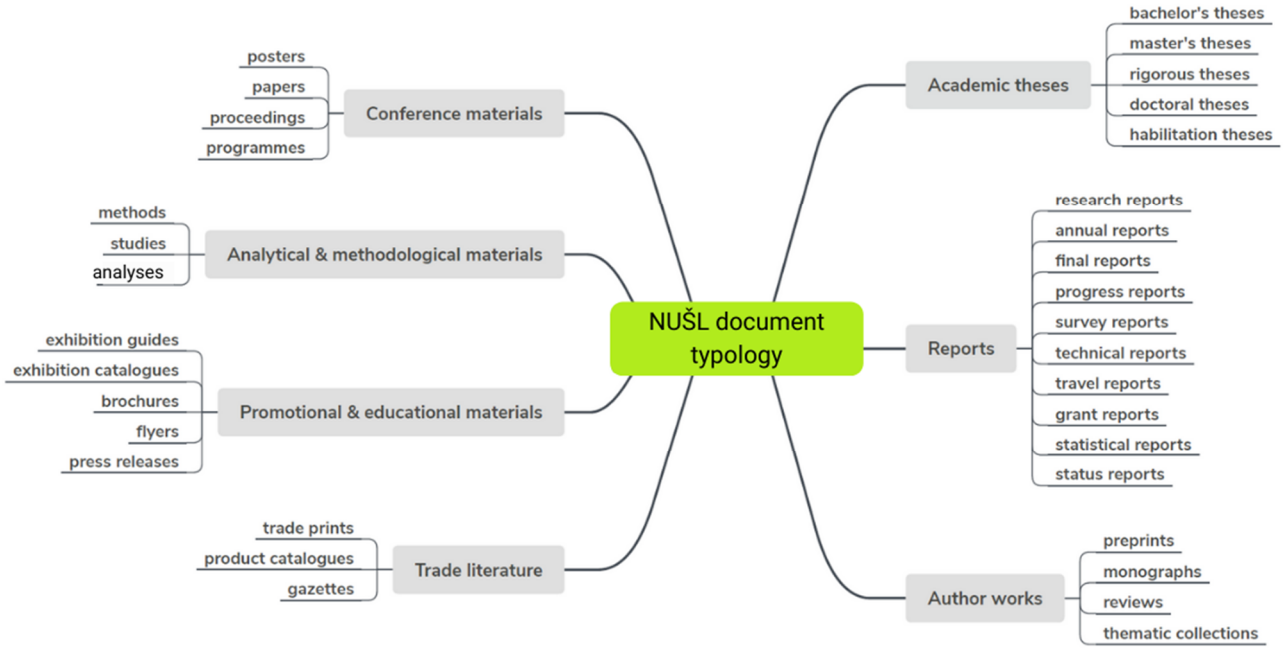


Figure 1: Current NUŠL document typology

Nevertheless, even this adjusted typology is still limiting for participating institutions, particularly those without their own repository. Submitters have to carefully sort unpublished documents from published ones in order to store them in NUŠL. Then, if they would like to store published documents they will have to look for other possibilities Overall, the question of how to face such challenges and where to direct the focus of the repository became more pressing over time.

External circumstances impacting NUŠL's operations

Over time, NUŠL the operation and further development are influenced by external circumstances such as the legislative, economic, and current situation as well as the needs of GL providers. First, except for one exception, participation is not supported legislatively. Therefore, growth of NUŠL content depends only on the institutions, their leaders, and the submitters themselves. Certain Czech Copyright Act limits apply and dealing with the property copyright in practice.

Technical obstacles on the side of providers include low levels (or non-existence of) system interoperability in the Czech Republic. During NUŠL's initial development, it was assumed that the interoperability would increase and that all publishing activities and other outputs, including GL and full texts, would be stored in them. Unfortunately, this did not happen. Czech institutional repositories are usually university systems for ETD archiving purposes, largely due to mandatory requirements included in the national Higher Education Act. Smaller institutions often lack the budget and human resources to develop and maintain their own repositories or to subscribe to fee-based services.

On the other hand, the pressure for open access to research outputs that have been publicly funded has increased, supported by conceptual documents and strategies (e.g., Plan S, Horizon 2020, followed by Horizon Europe in 2021+, the Czech Action Plan for Implementation of the National Strategy for OA to Scientific Information for 2017-2020, followed by the Czech National Research Development and Innovation Policy 2021+). Therefore, some institutions plan on or already have started with extending their repository typologies. For example, Czech Technical University in Prague[10] has started to collect articles and certified methodologies; Tomas Bata University in Zlín,[11] articles, conference proceedings, books and book chapters; Brno University of Technology,[12] conference proceedings and journals; and the University of West Bohemia[13] books, conference papers, working papers and reports, and so on.

The National R&D System only collects bibliographical records with URLs to the output's storage locations, where applicable. Recently is has been extended to provide information about the accessibility of an output in compliance with the COAR vocabulary (Council for Research, Development and Innovation of the Czech Republic 2021). So far, NUŠL was very limited in assistance to its partners with the fulfillment of availability obligations of R&D outputs due to its focus on GL. The only intersections of NUŠL's typology and the R&D System's "definition of research and development document types"[14] (typology) at present only include research report, conference paper, and certified methodology.

Transformation of the NUŠL service

The abovementioned experiences, circumstances, and software upgrade requirements strongly indicate that an upgrade and improvement of NUŠL in its current form do not suffice. It has been necessary to reevaluate the entire NUŠL concept, including its goals, service offer, as well as a further presentation of GL. A primary question in discussions about the future included whether the GL specialization would be sufficient for NUŠL stakeholders and needs of the Czech R&D community, or if a broader scope would better meet these demands.

NTK would like to support the growth of open access to any documents, i.e., enable institutions to store all their outputs of publishing activities and other research outputs, including GL. The Ministry of Education, Youth and Sports, which funds NTK, supports this idea and has incorporated it into a new grant project entitled "National Centre for Information

[10] Digital Library of the Czech Technical University in Prague (https://dspace.cvut.cz)
[11] DSpace at Tomas Bata University Zlín (http://publikace.k.utb.cz/)
[12] Digital library of Brno University of Technology (https://dspace.vutbr.cz/)
[13] Digital Library University of West Bohemia (Digital Library of UWB) (https://dspace5.zcu.cz)
[14] Available from https://www.vyzkum.cz/FrontClanek.aspx?idsekce=29415

Support of Research, Development, and Innovation (2021 - 2027).” Under the auspices of this project, NUŠL will be transformed into the Czech National Repository and it will accept both published and unpublished research documents of a textual nature. Another significant change will be the ability to scale up document accessibility in compliance with the COAR AccessRights Vocabulary.[15]

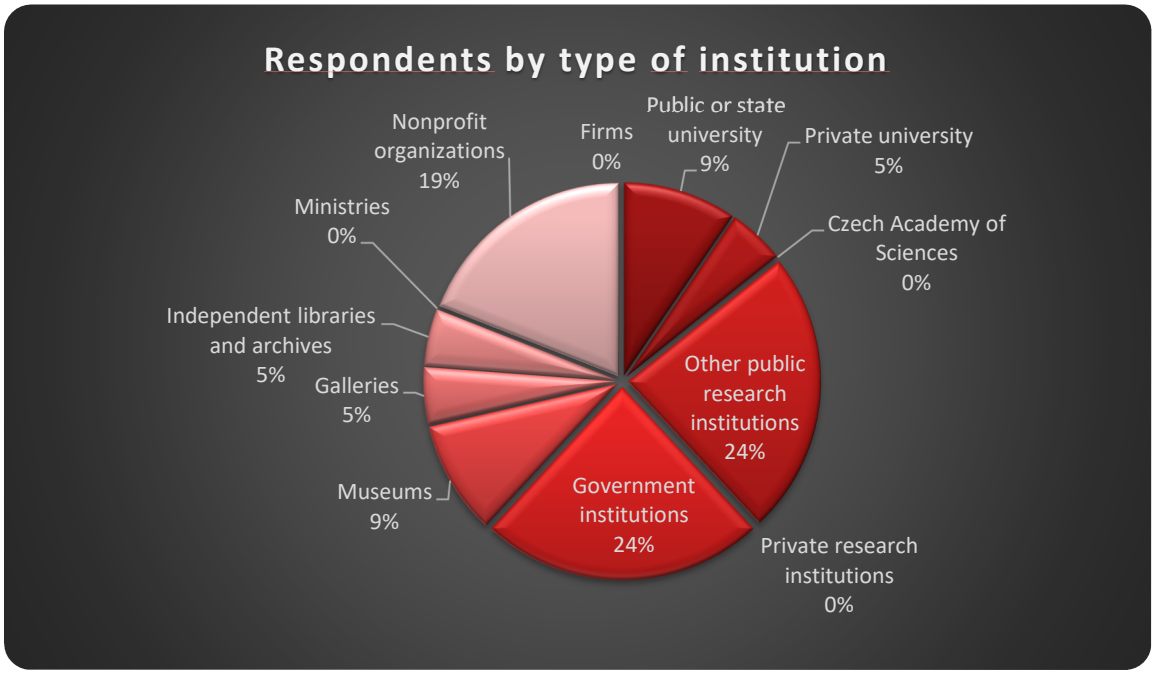
NTK, the University of Chemistry and Technology Prague, and CESNET[16] will collaborate on the technical upgrade. CESNET will also provide its infrastructure. The Council for Research, Development and Innovation of the Czech Republic is willing to cooperate to ensure interoperability with the R&D System maintained by the Council, and thus duplication of stakeholder effort will be avoided. The renewal of interoperability with international systems, which NUŠL achieved,[17] is intended.

Transformation: first steps

The government approved funding for the new project in October 2020. The project kicks off in 2021. In preparing for the project, we have already identified the needs of our stakeholders and subsequently, defined a new document typology.

Survey among stakeholders

The purpose of the survey was to collect feedback from our main stakeholders regarding typology, accessibility, file formats, and participation limits in the form of a questionnaire sent to all NUŠL submitters (87) and 9 administrators of harvested systems (96 in total). Email delivery failed in three cases. 21 respondents completed the survey; the questionnaire return rate was 22.6%. Of this total, only 2 respondents were administrators; 2 others responded by email. The representation of types of institutions participating in the survey roughly corresponded to the amount of partner institutions in NUŠL categories (see Graph 1).

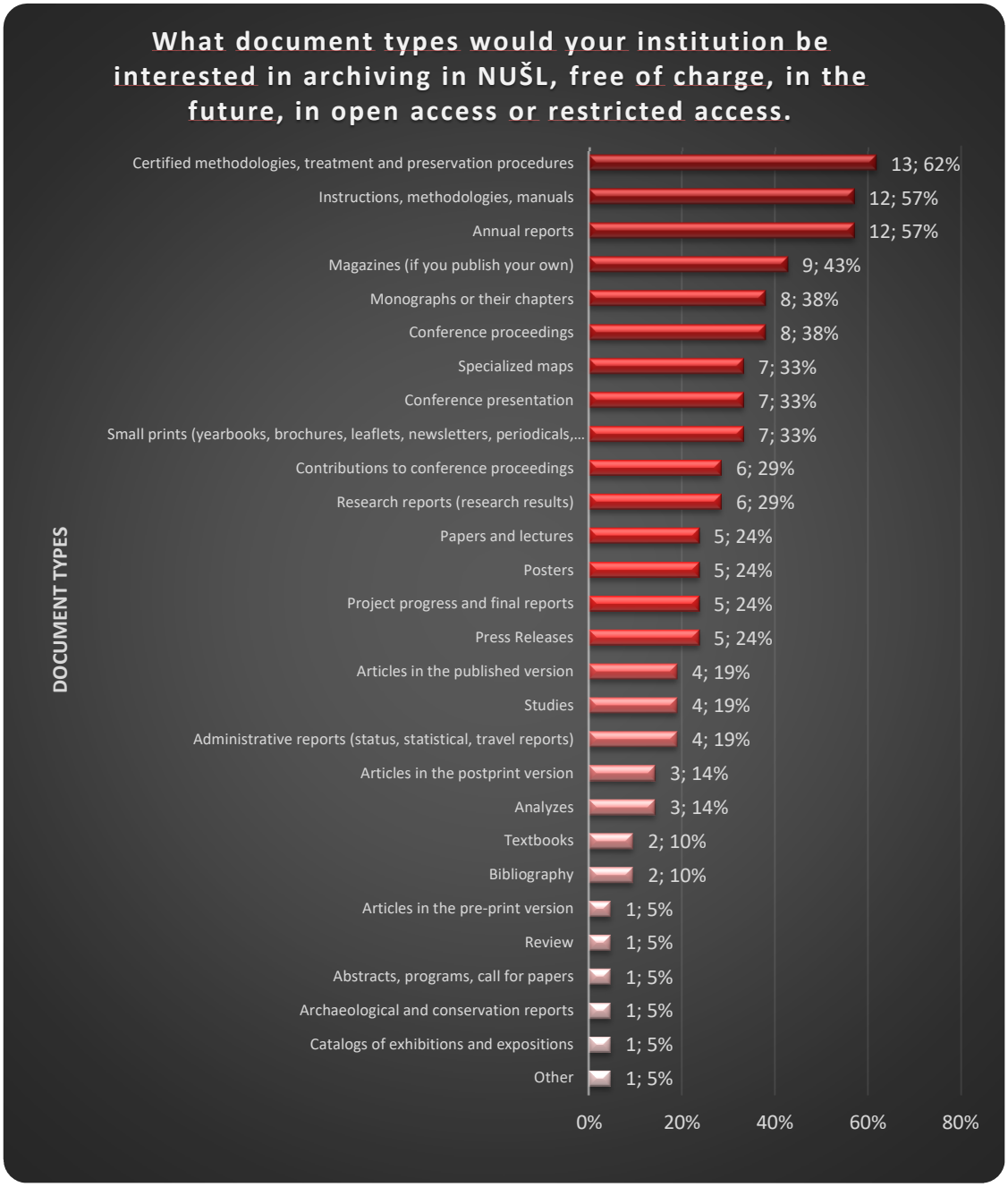


Graph 1: Types of institutions who participated in the NUŠL survey

The survey examined which documents participants would be interested in storing in the repository in the future, either openly or with some restrictions, in order to create a new

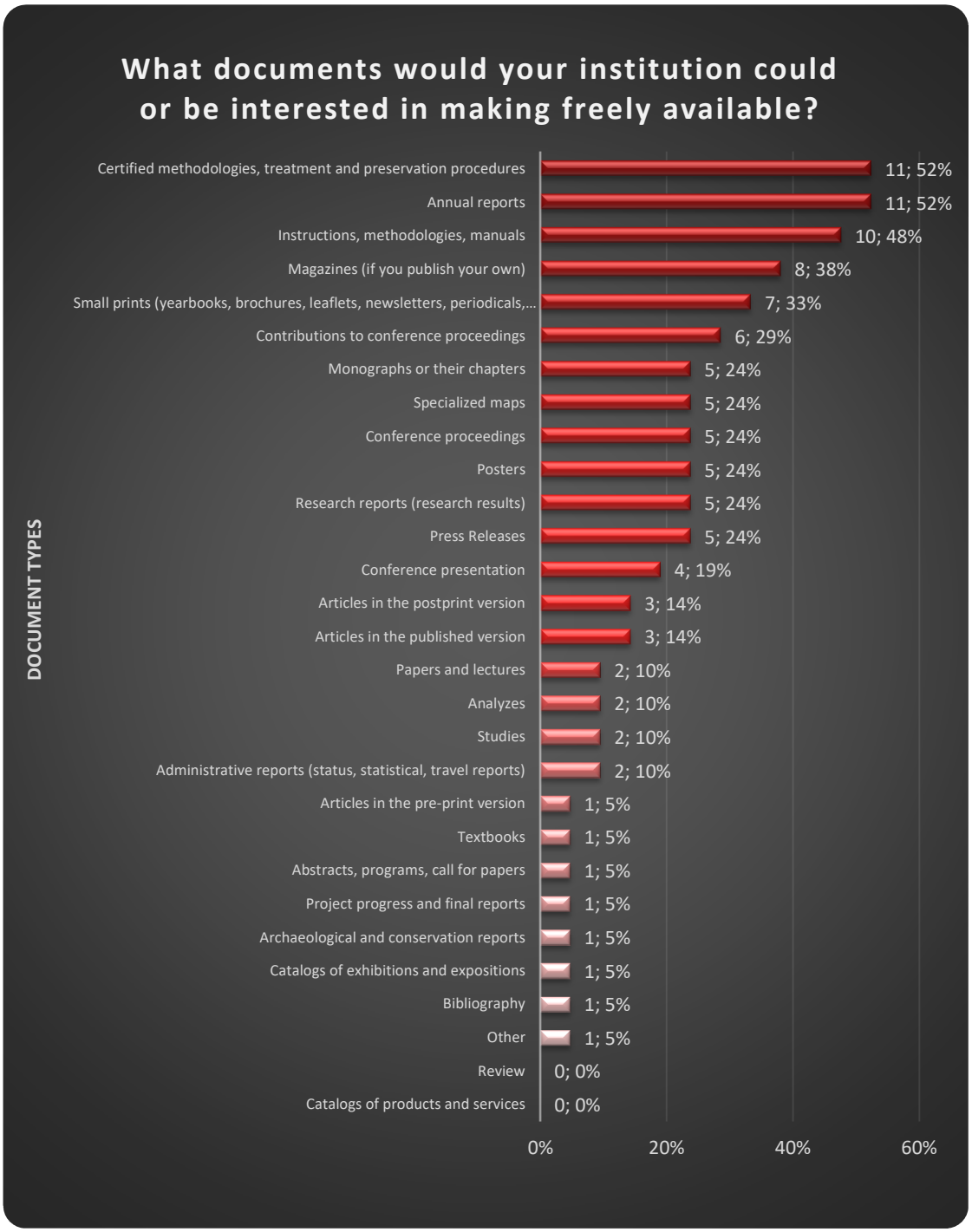
[15] Controlled Vocabulary for Access Rights (Version 1.0) (http://vocabularies.coar-repositories.org/documentation/access_rights/)
[16] CESNET is an association of universities of the Czech Republic and the Czech Academy of Sciences. It operates and develops the national e-infrastructure for science, research and education which encompasses a computer network, computational grids, data storage and collaborative environment. (https://www.cesnet.cz/?lang=en)
[17] NUŠL cooperation with international portals and registries (https://nusl.techlib.cz/en/nrgl/nrgl-and-international-cooperation)

typology (see Graph 2). Multiple response options were possible. Most often, respondents stated they would be willing to provide methodologies and procedures (certified or not) and annual reports. The fourth most common document type for submission to the repository mentioned by respondents was "journals published by the institution." However, it is not clear whether respondents meant by this real publishing activity in peer-reviewed or impact journals or if they were focused on GL and meant other materials such as bulletins, newsletters, and working papers. It is also possible that other people or departments (not the NUŠL contacts who comprised the respondent cohort) are in charge of scholarly publishing outputs of their institution. The respondents completed the survey on behalf of their institution but it is not clear, if they checked their answers with leaders and other departments or if they only expressed their assumptions. Therefore, the survey likely did not provide accurate feedback in this regard. The questionnaire was non-binding; the decision to store new kinds of documents in the repository would have to be made by institutional leaders.



Graph 2: Overview of documents that survey respondents would like to store in the repository

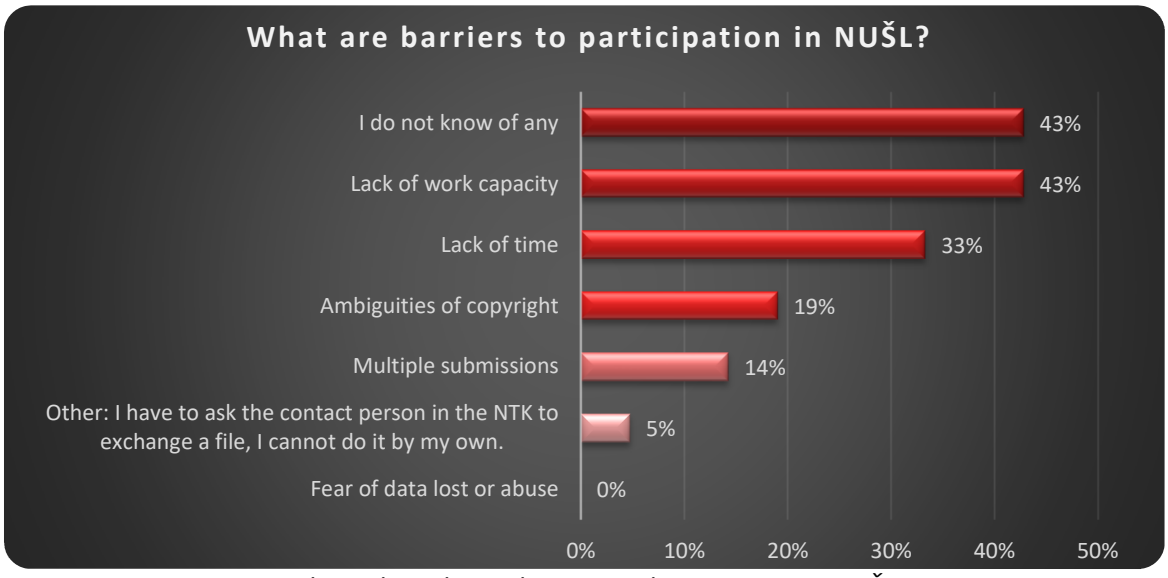
When asked about their willingness to put documents in open access (OA) mode into the repository (see Graph 3; multiple choices were possible), the order of documents mentioned changed slightly. This was most noticeable for monographs, which dropped by 3 places in comparison to Graph 2. Willingness to store OA small prints was noticeably apparent. "Journals" remained at the same level in Graph 2 as in Graph 3, which only confirms the assumption that respondents imagined newsletters rather than peer-reviewed periodicals. Overall, the number of selected document types in OA was lower than in Graph 2 (149 vs. 109), implying that institutions would welcome the possibility of closed or restricted access. However, regarding a possible time embargo, only 29% of respondents stated they would use it, mostly for journals, monographs, textbooks, studies, conference papers, and research reports.



Graph 3: Overview of documents that survey respondents would be interested in making freely available

The survey also examined how satisfied stakeholders are with the allowed file formats (pdf, jpg, mp3, and mp4) in NUŠL. 86% of respondents described them as sufficient and only 14% would welcome other formats, namely png, csv, xlsx/xls, docx/doc, ppt, zip. Thus, both formats typical for data (csv, zip and xlsx/xls) and formats not recommended for long-term archiving (docx/doc, ppt and xlsx/xls) appeared. Therefore, we will only include png from suggested to allowed formats for the Czech National Repository, because data storage is not within scope of the planned transformation. Set up processes for data sets would significantly slow down development of the repository. Focus on data is not possible now due to the need to migrate rapidly to a new version of the repository software, as described above.

The last question in the survey aimed at identifying barriers to depositing documents in NUŠL (see Graph 4); multiple responses were possible. We were surprised that 43% of respondents did not perceive any barriers to deposit in NUŠL. The other most frequently chosen options were: labor shortages (43%) and lack of time (33%). The low number (14%) of "multiple submission" limitations was also surprising. The reason may be that NUŠL submitters do may not be in charge of submissions into the R&D System or other internal institutional systems.



Graph 4: Obstacles in depositing documents into NUŠL

New Typology

In the subsequent creation of a new typology, we relied on the survey mentioned above, our experience from previous communications with partner institutions, the current representation of types collected in NUŠL so far, and a comparative analysis of the typologies applied in the following systems: Institutional Repository of the Academy of Sciences of the Czech Republic, OpenAIRE, Zenodo, the Czech R&D System, and the Charles University Digital Repository. The aim was to find a compromise connecting all of these.

Crucial to this process was the inclusion of published document types, so-called "white literature": articles (in any version), books, their chapters, and reviews. We asked for feedback on the draft version of the new typology by selected representatives of partner institutions. Based on their feedback, some categories were further generalized to facilitate more rapid orientation in the typology and to avoid unnecessary fragmentation for types which users have difficulty understanding the distinctions of such as studies and analysis or statistical and status reports.

The final typology is shown in Figure 2**.** "White literature" does not have a broader category; a user-friendly and sufficiently concise label was not found. The categories "ETDs" and "conference materials" did not change. We transformed "methodological and analytical materials" into a category entitled "methodologies and procedures" because these are often

used by NUŠL partners that created a certified methodology under the Ministry of Culture, due to the mandatory deposit requirements mentioned above. Despite major changes, the "reports" category still has more than five types. Several original types were merged or transferred because they are difficult to distinguish or are rarely used (survey reports, technical reports). New document types were added to the typology, including field and conservation reports, which could be provided by heritage institutions or institutions focused on history and archeology. Annex 1 provides a complete overview of changes in the document typology.

The last category groups the remaining GL document types. It contains five types - trade literature (now encompasses many original NUŠL types, see Annex 1), studies and analyzes, exhibition catalogs, and newly added specialized maps and educational materials. Labeling this category "other" was excluded due to its ambiguity, as was the label "grey literature" with regard to given the circumstances and facts described above (it has a similar information value for nonprofessionals as "others"). We found inspiration in the Gambling Research Exchange (GREO) in Canada, where they faced similar problem in the presentation of GL in the GREO digital library in 2017. They renamed the GL category "specialized resources." This change increased access to the digital library. (Baxter, Hilbrecht 2020) Thus, we decided to follow best practices and to present this category as "other specialized materials."

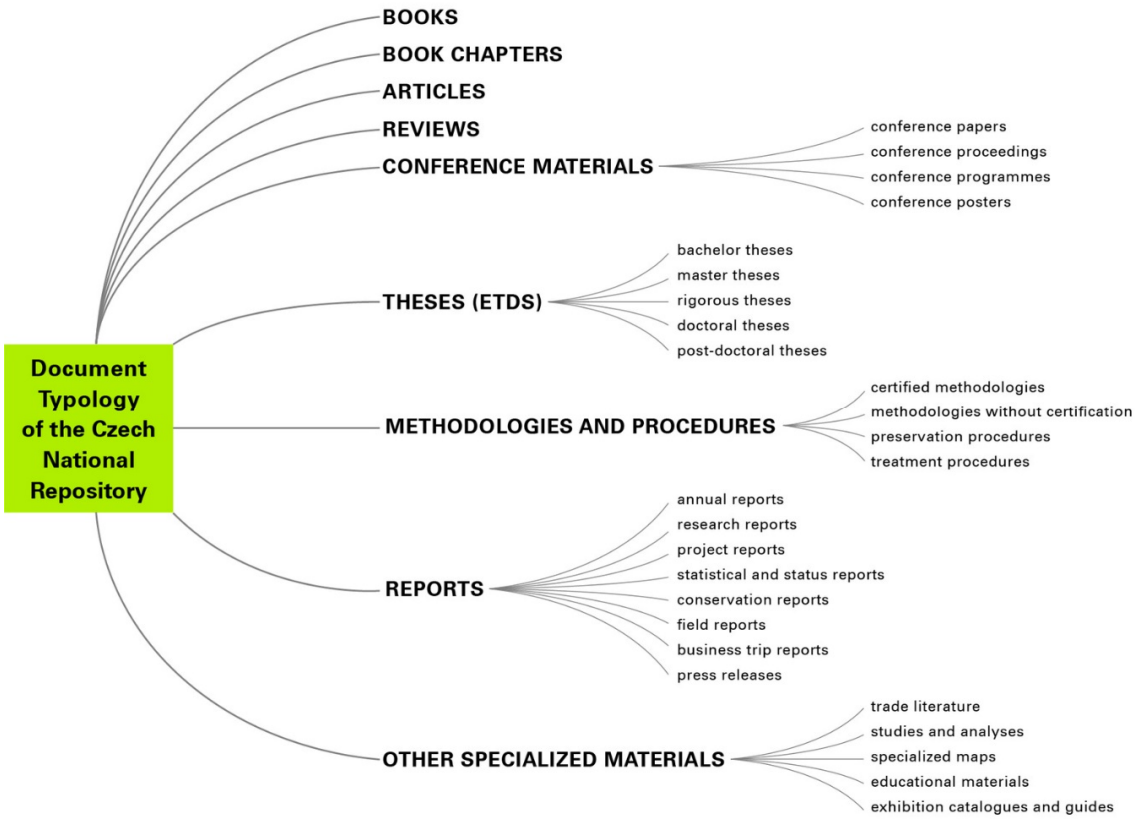


Figure 2: the Czech National Repository document typology

An analysis for the original NUŠL typology revealed that most GL databases do not include patents and that these are usually managed by other information resources (State Technical Library 2009b). In the Czech Republic, patents are typically included in the well-known Industrial Property Office databases.[18] Thus, we did not consider including patents in the new typology.

[18] Available from https://upv.cz/en/client-services/online-databases.html

We hope the new typology will enable better user orientation and decision-making regarding document types and improve the user experience of the service.

Conclusion

NUŠL was established as a special project (2008-2011) and has been in full operation since 2012. Over time, various service modifications have been made. However, Invenio software upgrade, reflections on the further direction of GL and other external circumstances have led to the transformation of NUŠL into the Czech National Repository within a national grant project entitled "National Centre for Information Support of Research, Development, and Innovation (2021 - 2027)." NUŠL's original goals will remain, GL will still be included, and assistance to smaller institutions will be priority.

The GL presentation to end users in the Czech National Repository will be different from in NUŠL. GL will disappear from the name of the repository (since "white literature" will be also collected), and the term GL will not appear in the new typology, being replaced by the designation "other specialized resources." This decision was made based on GREO best practices. End users should not encounter the term GL in the user interface, but we will continue to label documents that have not gone through the traditional publishing process as "isGL" in their metadata records. This will enable filtering and subsequent transfer of only GL to specialized databases if needed.

Work on a technical upgrade has already started. In parallel with the other changes mentioned above, this should ensure the new Czech National Repository will serve the needs of all its users: curators, submitters and end users.

References

BAXTER, D. and HILBRECHT, M., 2020. From "grey literature" to "specialized resources": Rethinking terminology to enhance grey literature access and use. *Grey Journal*. 2020. Vol. 16, no. 1, p. 49–56. ISSN 1574-1796.

CERN, 2019a. *Invenio instances around the world*. [online]. c2016-2019. [Accessed 4 January 2021]. Available from: https://inveniosoftware.org/showcase/

CERN, 2019b. *Invenio Framework: Open Source framework for large-scale digital repositories.* [online]. c2016-2019. [Accessed 4 January 2021]. Available from: https://inveniosoftware.org/products/framework/

CERN, 2020. *Maintenance Policy — Invenio 3.4.0 documentation*. [online]. c2015-2020. [Accessed 4 January 2021]. Available from: https://invenio.readthedocs.io/en/latest/releases/maintenance-policy.html

COUNCIL FOR RESEARCH, DEVELOPMENT AND INNOVATION OF THE CZECH REPUBLIC, 2021. *Předávání údajů do Informačního systému výzkumu, vývoje a inovací RIV – Rejstřík informací o výsledcích 2021* [online]. 2014. [Accessed 4 January 2021]. Available from: https://www.isvavai.cz/dokumenty/RIV21_3.0.0.pdf

ČERNOHLÁVKOVÁ, P. and VYČÍTALOVÁ, H., 2018. Collecting grey literature – Institutional repository versus national aggregator. *Grey Journal*. 2018. Vol. 14, no. 2, p. 81–86. ISSN 1574-1796. Available also from: http://www.nusl.cz/ntk/nusl-369349

CHARVÁTOVÁ, Michaela and PEJŠOVÁ, Petra, 2015. Vychytávky v NUŠL. *Searching Session 2015 aneb co se skrývá za vyhledáváním* [online]. Prague: National Library of Technology. 2015. [Accessed 4 January 2021]. Available from: http://invenio.nusl.cz/record/200703

CZECH REPUBLIC. MINISTRY OF CULTURE, 2018. *Postup certifikace metodik, památkových postupů a special. map s odbor. obsahem jako uznatelných výsledků VaVaI z podpory MK* [online]. 2018. [Accessed 4 January 2021]. Available from: https://www.mkcr.cz/postup-certifikace-metodik-pamatkovych-postupu-a-special-map-s-odbor-obsahem-jako-uznatelnych-vysledku-vavai-z-podpory-mk-862.html

DREES, Bastian and PLANK, Margret, 2018. Video is the new Grey. *International Conference on Grey Literature – Public Awareness and Access to Grey Literature* [online]. 2018. Vol. Volume 19, p. 127–131. [Accessed 4 January 2021]. ISSN 1386-2316. ISBN 978-90-77484-31-9. DOI 10.26069/GREYNET-2018-000.013-GG.

FERRY, Kristine, GELFAND, Julia, PETERMAN, D. and TOMREN, H., 2008. Virtual Reality and Establishing a Presence in Second Life: New Forms of Grey Literature? *Ninth International Conference on Grey Literature - Grey Foundations in Information Landscape, House of the Province, Antwerp, Belgium 10-11 December 2007* [online]. 2008. Vol. 9, no. 1, p. 113–118. [Accessed 4 January 2021]. ISSN 1385-2316. ISBN 978-90-77484-10-4. DOI 10.26069/GREYNET-2020-000.255-GG.

GELFAND, Julia and LIN, Anthony, 2012. Social Networking: Product or Process and What Shade of Grey? *Thirteenth International Conference on Grey Literature : The Grey Circuit, From Social Networking to Wealth Creation, Library of Congress, Washington D.C. USA 5-6 December 2011*. [online] 2012. Vol. 13, no. 1, p. 20–33. [Accessed 4 January 2021]. ISSN 1386-2316. ISBN 978-90-77484-20-3. DOI 10.26069/GREYNET-2020-000.161-GG.

GIANNINI, Silvia, DELUCA, Rosaria, MOLINO, Anna and BIAGIONI, Stefania, 2018. Grey Literature and Research Assessment exercises: From the current criteria to the Open Science models. *International Conference on Grey Literature – Public Awareness and Access to Grey Literature* [online]. 2018. Vol. 19, p. 11–25. [Accessed 4 January 2021]. ISSN 1386-2316. ISBN 978-90-77484-31-9. DOI 10.26069/GREYNET-2018-000.001-GG.

PEJŠOVÁ, Petra, 2009. System of Grey Literature in the Czech Republic. *Eleventh International Conference on Grey Literature: The Grey Mosaic, Piecing it All Together* [online]. Washington D.C.: TextRelease. 2009. [Accessed 4 January 2021]. Available from: http://invenio.nusl.cz/record/41958

PETERSON, Benjamin, 2014. PEP 373 -- Python 2.7 Release Schedule. *Python.org* [online]. 2014. [Accessed 4 January 2021]. Available from: https://www.python.org/dev/peps/pep-0373/#id4

SAVIĆ, Dobrica, 2018. Rethinking the role of grey literature in the fourth industrial revolution. *Grey Journal*. 2018. Vol. 14, no. Special Winter Issue, p. 7–14. ISSN 1574-1796. Available also from: http://www.nusl.cz/ntk/nusl-370664

SAVIĆ, Dobrica, 2019. *From Digitization and Digitalization to Digital Transformation: a Case for Grey Literature Management* [online]. Hannover, 2019. [Accessed 4 January 2021]. Available from: https://doi.org/10.5446/36545. Video record from 21st International Conference on Grey Literature (GL21).

SIMANDLOVÁ, Tereza, 2012. Šedá literatura 2.0 = Grey Literature 2.0. *Seminář ke zpřístupňování šedé literatury 2012* [online]. Prague: National Library of Technology. 2012. [Accessed 4 January 2021]. Available from: https://invenio.nusl.cz/record/126858/

STATE TECHNICAL LIBRARY, 2009a. *Document types of the National Repository of Grey Literature = Typologie Národního úložiště šedé literatury* [online]. 2009. [Accessed 4 January 2021]. Available from: http://invenio.nusl.cz/record/111521

STATE TECHNICAL LIBRARY, 2009b. *Analysis of Portals of Grey Literature whether to Include Patents into the Classification of NRGL = Analýza portálů šedé literatury z hlediska zařazení patentů do typologie dokumentů NUŠL* [online]. 2009. [Accessed 4 January 2021]. Available from: http://invenio.nusl.cz/record/111522

Annex 1: Changes in the document typology

Annex 1: Changes in the document typology

NUSL Typology

Parent category	*Document type*	*Status*	*Note*
Author works *(withdraw)*	Synopses (Referáty)	Withdraw	converted to **article in version preprint or educational materials**
	Preprints	Withdraw	converted to **article in version preprint with the GL flag**
	Thematic collections	Withdraw	converted to conference proceedings or books with the GL flag
	Monographs	Extended	for white books
Conference materials	Papers		
	Proceedings		
	Programmes		
	Posters		
Academic theses (ETDs)	Bachelor's theses	New label	
	Master's theses	New label	
	Doctoral theses		
	Rigorous theses	New label	
	Habilitation theses		
Analytical and methodological materials *(withdraw)*	Analyses	Merged	with studies
	Methods	Extracted	in a separete category and distinguish among certfied and non certified, procedures added
	Studies	Merged	with analyses
Reports	Annual reports		
	Final reports	Merged	with progress reports into project reports
	Grant reports	Withdraw	converted to project reports
	Progress reports	Merged	with final reports into project reports
	Research reports		
	Statistical reports	Merged	with status reports
	Status reports	Merged	with statistical reports
	Survey reports	Withdraw	converted to field reports
	Technical reports	Withdraw	converted to research reports
	Travel reports	New label	
Trade literature *(transformed into a type)*	Trade print	Withdraw	converted to trade literature
	Product catalogues	Withdraw	converted to trade literature
	Gazettes	Withdraw	converted to trade literature
Promotional and educational materials *(withdraw)*	Brochures	Withdraw	converted to trade literature
	Flyers	Withdraw	converted to trade literature
	Exhibition catalogues	Merged	with Exhibition guides
	Exhibition guides	Merged	with Exhibition catalogoues
	Press releases	Moved	into the reports category

Czech National Repository Typology

Parent category	*Document type*	*Status*	*Note*	*GL indication*
	books	New	includes NUSL Monographs marked GL and a part of Thematic collections	GL selectively
	book chapters	New		GL selectively
	articles	New	includes Preprints and a part of Synopses marked as preprints	GL selectively - preprint
	reviews	New		GL selectively
Conference materials	conference papers			GL selectively
	conference proceedings		includes a part of Thematic collections	GL selectively
	conference programmes			GL
	conference posters			GL
Theses (ETDs)	bachelor theses			GL
	master theses			
	rigorous theses			
	doctoral theses			
	post-doctoral theses			
Methodologies and procedures	certified methodologies	New	extracted from Methods	GL
	methodologies without certification		rest of Methods	
	preservation procedures	New		
	treatment procedures	New		
Reports	annual reports			GL
	research reports		includes Technical reports	
	project reports		merged Progress and Final reports, and Grant reports	
	statistical and status reports		merged types	
	conservation reports	New		
	field reports	New	includes Survey reports	
	business trip reports			
	press releases			
Other specialized materials	trade literature		category became a type; includes - Trade print, Product catalogues, Gazettes, Brochures and Flyers	GL
	studies and analyses		merged types	
	specialized maps	New		
	educational materials	New	includes a part of Synopses	
	exhibition catalogues and guides		merged types	

DIY Data Creation as Scholarly Communication*

Andrea Marshall,
Centre for Media and Celebrity Studies, North Brunswick, New Jersey, USA

Abstract:
Social media sites and zines function as grey literature that generate intersubjectivities between diverse disciplinary factions and layperson peer groups through discussions of shared ideas and debates concerning emergent research topics and trends. Social media sites and zines as critical data artifacts increase accessibility to much larger populations due to a lack of conventional barriers such as bibliographic databases, institutional repositories and textbooks. Zines and social media sites can also be classified as digital makerspace environments that support do it yourself (DIY) data creation as a revolutionary methodological apparatus that both increases awareness of evolving knowledge structures and expands participatory research access to lay persons.

Introduction

Scholarly communication within contemporary digital cultures often disrupt binary divisions between formal and informal modalities of distributed cognition. Infrastructural approaches to engaging with this quandary have included the evolution of social networking sites for academics (Hailu and Wu 2021 p.1). Many scholars have referred to these digital research terrains as "academic social networking sites" (Hailu and Wu 2021 p. 1); some of these began as bibliographic bookmarking sites such as Mendeley and Zotero (Hailu and Wu 2021 p. 1). Furthermore, some researchers such as Michael Wood (2020) posit that within the new digital modalities of scholarly communication, emergent scholarship could benefit if scholarly communities could "make the evaluation system more critical, multi-dimensional and responsive to the requirements of different audience groups, changing circumstances, and new ideas" (p.1). Digital environments with porous participatory architectures include zines and traditional social networking sites that include as Facebook, Twitter and Instagram. Zines and public social networking sites, with their emphases on grass roots communal interactions, diverse peer groups and audiences, and informal knowledge structures, function as grey literature. Dobrica Savic (2019) has observed that a "digital mindset requires a deep understanding that the power of technology can democratize" (p.7); contemporary scholarly communication in zines and public social networking sites have evolved into archived datasets where laypersons interact with and contribute to research discussions, disciplinary questions and become components of DIY (do it yourself) data creation.

In order to further investigate the nascent possibilities of transcending ASNs and traditional peer review protocols, my research presents zines and popular social networking sites as two digital environments that can expand one's understanding of emancipatory research that both informs a wider audience and demolishes conventional barriers to accessing scholarly outputs. Zines have the unique distinction of existing both within print (analog) and digital cultures. Zines were conceived as analog infrastructures that in turn served as epistemic objects; these objects functioned as artifacts that explicated and amplified the ideologies and beliefs of multiple communities of practice. These zine subcultures, while originally situated within arts, music, and activist contexts, soon evolved to include and reflect the values of diverse and sundry populations.

* First published in the GL2020 Conference Proceedings, February 2021.

One could also posit that Zines served as early prototypes for social networking sites with the emphasis on an DIY ethos, a collapsing of conventional amateur and expert hierarchies and the use of creative cognitive tools to articulate often subversive and postmodern principles and beliefs. Social media sites perform as both metanarratives and paratextual architectures in contemporary digital cultures. Scholarly dialogues that occur on social media sites often have collaborative and communal informalities that generate new epistemologies and ways of knowing. Zines and social media sites construct themselves as loci of scholarly communication that dismantle conventional power dynamics.

Zines

Zines are "self-published, do-it yourself (DIY) publications that come in many formats" (Du Laney et al. 2020 p.1); the term zine is a colloquial construct that evolved from the semantic synergy of the words fanzine and magazine (Du Laney et al. 2020 p. 1). Since Zines are often "small independent print publications that often fall outside of the mainstream collective memory" (Lymn 2013 p. 3) they are construed as ephemeral and additionally defy conventional indexing, classification and bibliographic infrastructures and practices. Zines are also digital artifacts though they began as analog creations; they are often (though not always) freely available to users on the internet. Visibility for zines has greatly increased due to the advent of the internet; furthermore they often reflect the values of subcultural groups such as "the feminist community, LGBTQ+ community, the punk community, fandoms etc." (The Feminist Bibliothecary). Open source materials resist familiar boundaries of paywalls, institutional constraints and sponsored publications.

Zines as open source objects are at liberty to publish content without the limitations imposed by institutions, publishing houses and corporate organizations. Zines as critical data artifacts deliberately confront conventional standards, traditional values, and popular cultural attitudes towards controversial topics. Hays (2020) observes that they "have value to scholars as articulations of identity….and as articulations of the self"(p.16). Zines as grey literature contribute to scholarly conversations about agency, power, citizen science, the custodianship of knowledge, and discussions of accessibility.

Social Networking Sites

Social media sites are digital environments that support the intraconnectedness of modern global denizens; data literacy is an increasingly crucial component in discussions of pedagogy, learning, and scholarly communication trends. Raffeghelli et al. (2020) explain that "the massive adoption of social media has crossed paths with learning management systems, creating new forms of data" (p. 2). These "new forms of data" (p.2) evolve into "data epistemologies" (Raffeghelli et al. 2020 p. 2); social media sites are environments wherein these knowledge structures are informally constructed. These informally constructed data artifacts and data epistemologies reflect the makerspace DIY ethos supported in online and analog Zine cultures and other digital environments such as wikis. Turkle and Papert discuss DIY learning patterns (1992) in the context of scientific research; their analysis of "bricolage" (p.6) or learning by doing can be applied to social media sites and informal knowledge production processes. They observe that "…bricoleurs have goals but set out to realize them in a spirit of a collaborative venture" (Turkle and Papert, 1992, p. 6); this "spirit of a collaborative venture" (p.6) is supported by social networking sites that promote and encourage diverse populations conduct interlocutions with scholars and other laypersons.

Social media networking sites are loci of progressive collaborations between scholars and their lay audiences. Discussions about data transparency in human subjects research, and cooperative and communal interactions and relationships with study participants remain pivotal elements of dialogues about ethics, power dynamics, and agency. Less formalized

processes and fewer barriers between laypersons and researchers generate circumstances wherein accessibility and diversity can be nonnegotiable ethically for future data epistemologies. Scholarly communication as a visible, flexible and adaptable construct allows for actualized social change in terms of the collapsing of binaries of expert and amateur, and the collapsing of boundaries between scholars and laypersons.

Future work

Scholars are just beginning to understand how zine cultures both analog and digital contribute to discussions of the historical records of various populations and societies; scholars generating their own data in zines and on social media sites permit laypersons to unpack the black box processes that lead to research protocols becoming knowledge structures. Those knowledge structures evolve into institutionalized caveats, systematically produced policies and sociopolitical infrastructures that have global impact. Therefore, social media sites and zines as grey literature occupy mitigating positions within scholarly dialogues regarding critical data literacies, collaborative knowledge production, and the extant values within scholarly communities of practice. Grey literature has the capacity, through social media sites and zines, to apply revolutionary methodological protocols to ontological, ideological and pedagogical approaches that directly impact what we learn, how we learn, and who determines what we learn. Privileges and power dynamics that exist within scholarly communities cannot be dismantled without diverse voices and perspectives to support new ways of knowing and alternate modalities of communication and information. Grey literature can contribute meaningfully to this purpose through bricolage; scholars-as-bricoleurs can support DIY data creation as not only scholarly communication, but critical information thinking that can change the world.

Works Cited

Du Laney, C., Maher, M., & Schindler, A. (2020). The Librarian's Guide to Zines for Classroom and Community.

Franklin, I. (2020). Grey matter/literature/area: Bucketfull of Brains, fanzine form and cultural formation. Punk & Post-Punk, 9(2), 247-265.

Fife, K. (2019). Not for you? Ethical implications of archiving zines. Punk & Post Punk, 8(2), 227-242.

Hailu, M., & Wu, J. (2020). The Use of Academic Social Networking Sites in Scholarly Communication: Scoping Review. Data and Information Management, 1(ahead-of-print).

Hays, A. (2020). A Citation Analysis about Scholarship on Zines. Journal of Librarianship and Scholarly Communication, 8(1), p.eP2341.

Lymn, J. (2013). The zine anthology as archive: archival genres and practices. Archives and Manuscripts, 41(1), 44-57.

Raffaghelli, J. E., Manca, S., Stewart, B., Prinsloo, P., & Sangrà, A. (2020). Supporting the development of critical data literacies in higher education: building blocks for fair data cultures in society.

Saqr, M., Al-Mohaimeed, A., & Rasheed, Z. (2020). Tear down the walls: Disseminating open access research for a global impact. International journal of health sciences, 14(5), 43.

Savić, D. (2020, January). Digital Transformation and Grey Literature Professionals. In CONFERENCE ON GREY LITERATURE AND REPOSITORIES (p. 14).

Smith, D. R. (2017). Confessions of a science blogaholic: Highs, lows, and increasing liabilities. Frontiers in Communication, 1, 15.

The Feminist Bibliothecary. (2018, March, 18). Where to Find Free Digital Zines. The Feminist Bibliothecary. https://thefeministbibliothecary.wordpress.com

Turkle, S., & Papert, S. (1992). Epistemological pluralism and the revaluation of the concrete. Journal of Mathematical Behavior, 11(1), 3-33.

Watson, A., & Bennett, A. (2020). The felt value of reading zines. American Journal of Cultural Sociology, 1-35.

Weiland, S. (2017). Social Scholarship? Academic Communications in the Digital Age.

Wood, M. (2020). Beyond Journals and Peer Review: Towards a More Flexible Ecosystem for Scholarly Communication (Pre Print).

Library, Information Science & Technology Abstracts™ with Full Text

Available via EBSCO*host*®

The definitive professional information resource designed for librarians and information specialists…

Library, Information Science & Technology Abstracts™ with Full Text is an indispensable tool for librarians looking to stay current in this rapidly evolving field.

Comprehensive content includes:

- Full text for more than 270 journals and nearly 20 monographs
- Indexing for more than 550 core journals, 50 priority journals and nearly 125 selective journals
- Includes books, research reports, proceedings and author profiles
- Access to 6,800 terms from reference thesauri
- Coverage extends back as far as the mid-1960s

Subject coverage includes:

- Bibliometrics
- Cataloging
- Classification
- Information Management
- Librarianship
- Online Information Retrieval
- And much more…

Contact EBSCO Publishing to learn more about *Library, Information Science & Technology Abstracts™ with Full Text*, or to request a free trial.

Phone: 800.653.2726
Email: request@ebscohost.com

www.ebscohost.com

The Grey Side of the Green Road : Empirical Assessment of Academic Publishing in the HAL Open Repository*

Joachim Schöpfel and Eric Kergosien, University of Lille, ULR 4073 - GERiiCO, France
Hélène Prost, CNRS and GERiiCO, France
Florence Thiault, PREFICS Laboratory, University of Rennes 2, France

Abstract
The paper provides empirical results on the part of grey literature in the French national open repository HAL. The results are based on a quantitative analysis of 973,968 deposits from 1,272 research laboratories affiliated the ten most important research universities in France (IDEX program). 35% of all deposits (documents and records) are grey literature. The main types of these items are conference papers (71%), PhD theses (11%) and preprints and working papers (8%). Their overall accessibility (degree of openness) is 37%, which is higher than commercial publications. Differences between document types, disciplines and institution are described. Further research is needed for a better understanding of how research laboratories are handling grey literature (licensing …). A follow-up study in 2021 will provide this insight.
Keywords: Open science, open access, open repositories, research laboratories, universities, grey literature

Introduction

HAL – "Hyperarticles en ligne" - is the national multidisciplinary open repository for documents and data from French scientists[1]. HAL was launched in 2001 by the Center for Direct Scientific Communication (CCSD)[2]. Funded by research organisations, the University of Lyon and the French Ministry of Higher Education, Research and Innovation, the CCSD is the host and operating agent for HAL. Actually (November 2020), HAL contains more than 2.4 million records and nearly 800,000 scientific documents in full text. The deposits are organized in institutional portals and collections from research units and projects. The following paper presents evidence on how grey literature is represented in the HAL collections of French research laboratories of ten major French research universities. Since 2016 and the vote of a "Digital Law", published results from public research can be made openly available after six months (for science, technology and medicine) or twelve months (for social sciences and humanities) (Nérisson, 2016). Some major research organisations and institutions adopted open access mandates, and a growing number of research structures and universities make use of HAL or of their own institutional repository for research assessment and monitoring.

Grey literature has been defined in various manner, especially by members and authors of the International Grey Literature Network (Farace & Schöpfel, 2010). We consider grey literature as "manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body" (Schöpfel, 2011). However, usually and especially in systematic reviews, grey literature is often described as "unpublished", "not peer reviewed" and "not in databases" and meaning most of the time all kinds of reports and conference papers (Schöpfel & Prost, 2020).

"Not in databases": the lack of bibliographic control, of evaluation and monitoring is the main reason why open archives and particularly institutional repositories have been seen from the

* First published in the GL2020 Conference Proceedings, February 2021.
[1] HAL https://hal.archives-ouvertes.fr/
[2] CCSD https://www.ccsd.cnrs.fr/en/

beginning on as a potential solution for grey literature, in order to increase its findability, accessibility and long-term preservation. With the words of Daniela Luzi, grey literature is "at home in open archives" (Luzi, 2010).

We identified five opportunities for the further development of grey literature in institutional repositories (Schöpfel et al., 2011):

1. Universities need a solution for the processing, disseminating and archiving of electronic theses and dissertations (ETD). Institutional repositories offer an interesting solution and may at least be an element in the global academic information system for ETD.
2. Institutions want control on research output and content, and this includes unpublished documents.
3. Institutions want to improve presence and impact on the web. Grey literature in repositories adds to both, due to broader dissemination and increased use of grey items, increasing prestige and visibility for the institution.
4. The open access initiative is not limited to published documents.
5. The evolution from "collection development" to "content recruitment" in academic libraries may act in favour of deposit of grey literature in institutional repositories.

Ten years later, how is grey literature at home in open archives? Which is its part? Which kind of grey literature? The paper presents original empirical results from a follow-up study to former research on open access strategies of French laboratories (Schöpfel et al. 2018, 2019), based on the assessment of HAL deposits of more than 1,200 research laboratories from the ten most distinguished French research universities (IDEX, *excellence initiative*). These laboratories cover the whole range of scientific disciplines, including medicine, law, economics and management. The ongoing research takes account of all document and data types while the following paper limits the focus on the grey part of the academic output, in particular conference papers, reports, working papers, theses and dissertations. Empirical results are provided on the percentage of grey items in HAL, on the part of open access, differences between grey document types, disciplinary specificities and institutional strategies. Our intention is to contribute to a better understanding of the relative impact and the "degree of openness" (Schöpfel & Prost, 2014) of grey literature in open repositories.

Methodology

The study was conducted as part of a research project on open access strategies of French research laboratories. The sample consists of 1,272 research laboratories from the ten universities part of the French excellence initiative (IDEX). The laboratories were identified and described by direct search on the universities' websites, with the French national directory of research structures[1] and with the public discovery tool for French innovation and research ScanR[2] (January-March 2020).

In a second step, we recovered the laboratories' identifiers in the HAL repository, and we used them for the API search of each laboratory's deposits (July-September 2020). The results were analysed based on the laboratory and deposit metadata (domain, university, deposit type, resource category). The API search retrieved 26 different HAL resource categories; we merged and described them in the following way in order to simplify the description of the results (table 1):

HAL category	Merged category	Resource type
Journal article	Articles	White
Book (author)	Books (author)	White
Book (editor)	Books (editor)	White
Book chapter	Book chapters	White
Other	Other (white)	White
Communication	Conference papers	Grey
Poster	Posters	Grey
PhD thesis	PhD theses	Grey
Report	Reports	Grey
Activity report	Reports	Grey
Short report	Reports	Grey
Report chapter	Reports	Grey
Other report	Reports	Grey
Undefined	Preprints, working papers…	Grey
Habilitation	Other (grey)	Grey
Lecture	Other (grey)	Grey
Master dissertation	Other (grey)	Grey
BA dissertation	Other (grey)	Grey
Lecture note	Other (grey)	Grey
Presentation	Other (grey)	Grey
Patent	Other (grey)	Grey
Image	Data	Data
Map	Data	Data
Software	Data	Data
Audio	Data	Data
Video	Data	Data

Table 1. Resource categories and types

The HAL category "other" consists of book reviews, encyclopaedia entries, translations etc., most of them reviewed publications. The HAL category "undefined" consists mainly of working papers, preprints and other, non-reviewed and unpublished documents. Some of the HAL resource categories are specific to a particular, institutional workflow (i.e., ingestion of records from institutional partnerships), such as BA dissertations, lecture notes and report chapters; they represent very small figures (<100) and have been merged (reports) or included in the "other" category.

Results

The part of grey literature

The API query retrieved 973,968 deposits which represent 30% of the total HAL content. These are the items which have been authored or co-authored by scientists affiliated to one of the 1,272 research laboratories of our sample. From all these items, 35% fall under the category of grey literature (figure 1).

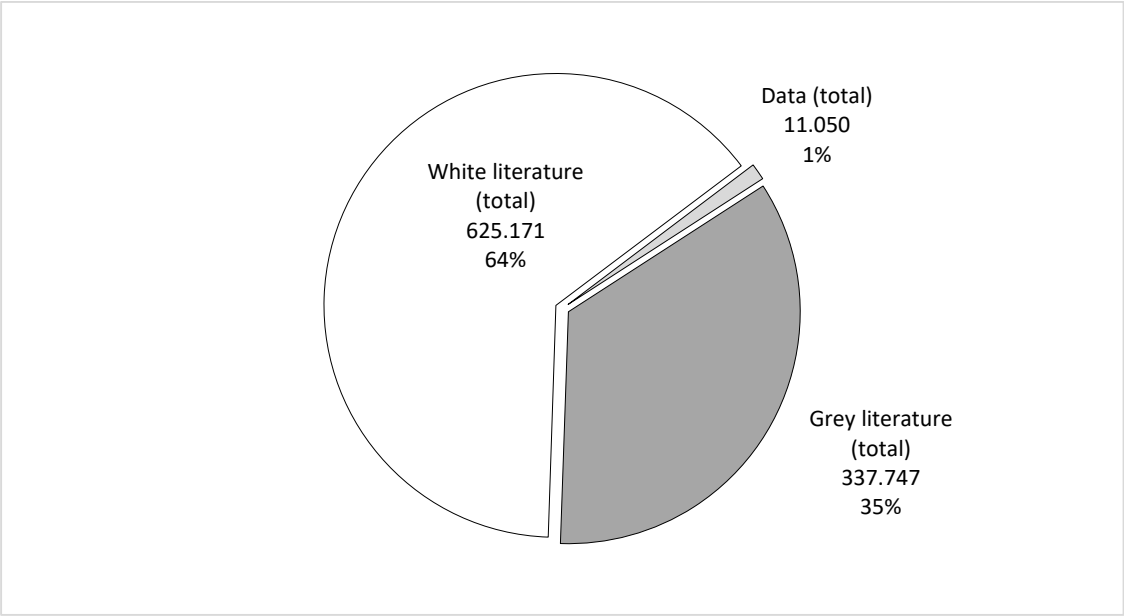


Figure 1. Part of grey literature (N=973,968 deposits)

Most of the grey literature consists of conference papers (71%), followed by PhD theses (11%) and working papers or preprints (8%). The different types of reports (project reports, activity or annual reports, short reports and report chapters) represent 4%, similar to posters (4%). Other resource types are less important, such as BA and Master dissertations, habilitation theses or lectures, totalling together 2% (figure 2).

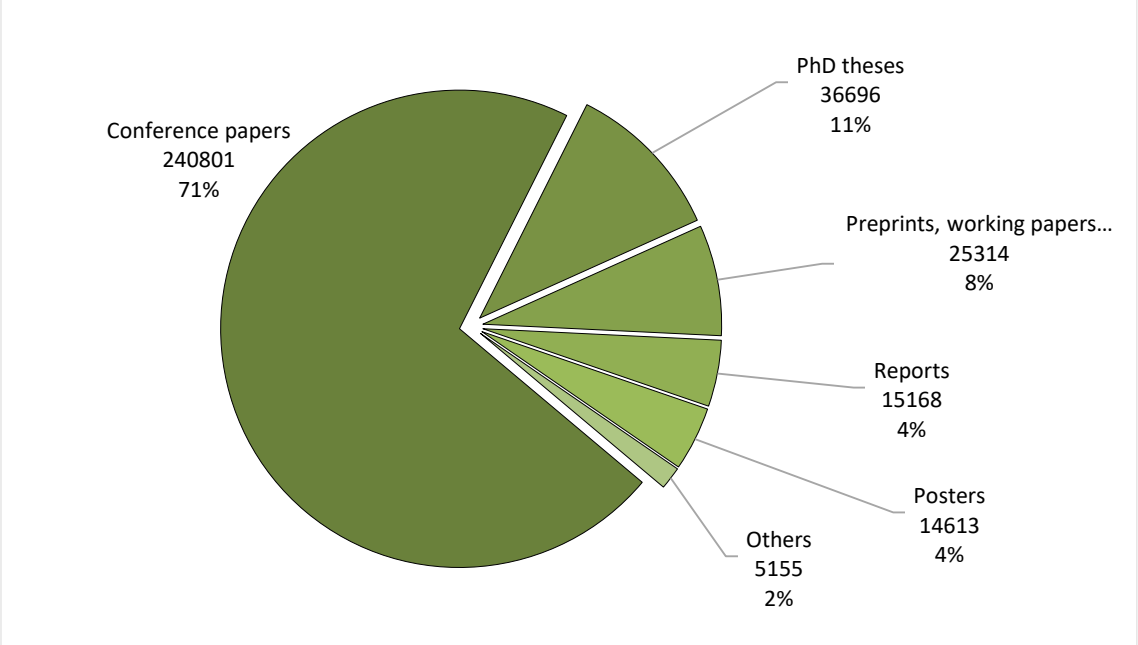


Figure 2. Types of grey literature (N=337,747 deposits)

Figure 1 shows that 1% of the deposits are neither white nor grey literature but datasets. HAL was designed as a document server, an open repository for academic publications and other scientific documents. So far, HAL has not been (re)defined as a data repository. However, since 2010, the MédiHAL portal of HAL allows to deposit visual and sound data (still images, videos and sounds), produced within the framework of scientific research. Also, there are some software deposits (codes) due to the partnership between HAL and the international Software Heritage project.

Degree of openness

The HAL repository contains deposits of full text and files as well as records, i.e. metadata without documents or files. The global part of deposits of full text is 24.3% for the whole HAL content. In our sample, this part is higher, 30.1% of the items have been deposited together with a text or data file. This part of freely and openly available research output can be interpreted as degree of openness. If all deposits would consist of metadata and data (documents), this degree of openness would be 100%.

Figure 3 shows that the part of deposits with document files is significantly higher for grey literature (37%) than for white literature (25%).

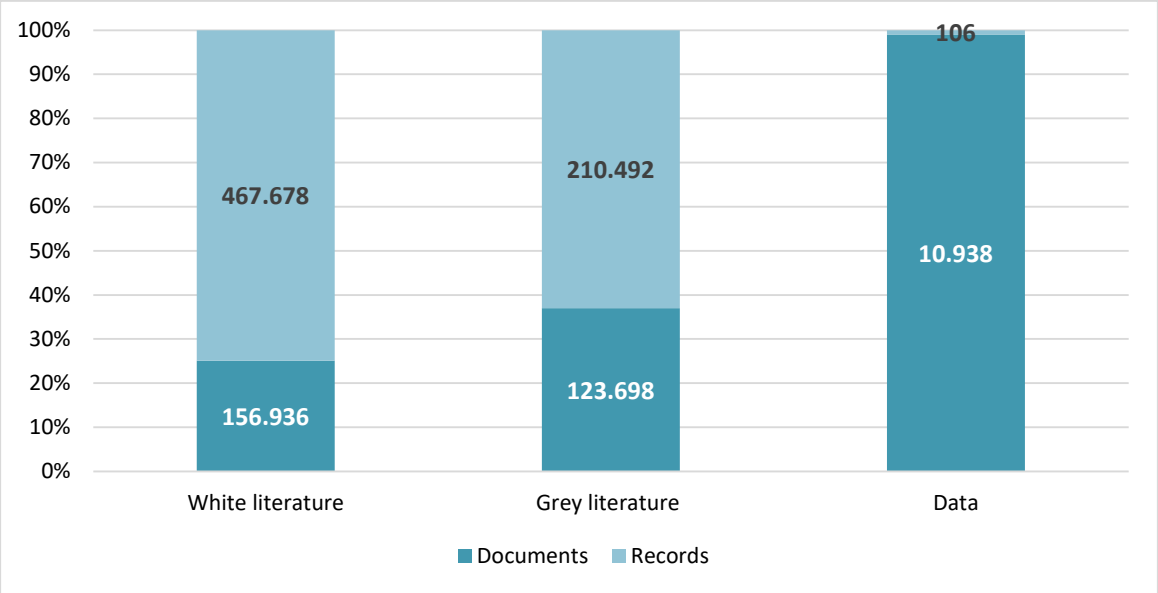


Figure 3. Deposits with and without document and data files (N=969,848 deposits)

The differences between the document types are important. Figure 4 shows the part of full text for the different types of grey literature.

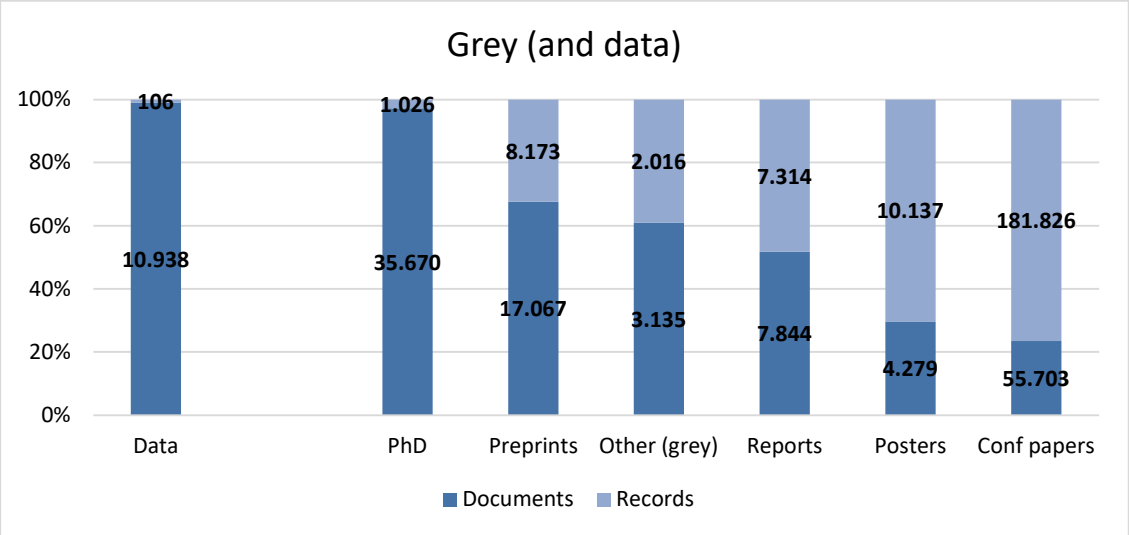


Figure 4. Grey literature deposits with and without document files (N=337,747 deposits)

The degree of openness of grey deposits is generally well above the average percentage, between 50% and near to 100%. The explanation of the exceptional part of openly available PhD theses (97%) is that the self-archiving of a PhD thesis on the HAL platform requires systematically the deposit of the text file.

The open part of the conference contributions, papers and posters, are lower, with 29% open posters and 23% open papers. This lower degree of openness is similar to the percentage of full text deposits of articles, books, chapters etc. (figure 5).

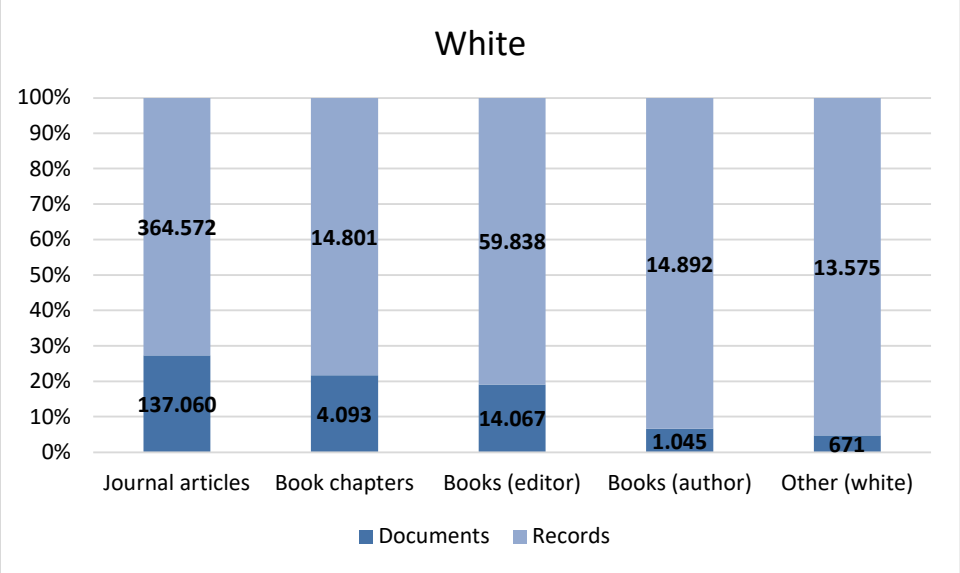


Figure 5. White literature deposits with and without document files (N=625,171 deposits)

27% of the journal articles are openly available on the HAL platform; the degree of openness of chapters, edited and authored books and other white resources is even lower.

Two other observations may be interesting. First, the part of grey literature of all deposits with full text is 44% which is significantly higher than the overall part of grey literature (37%). Second, figures 3 and 4 include the degree of openness of datasets which is exceptionally high (99%); nearly all datasets have been deposited with the data files. Again, the reason is rather simple: MédiHAL, the data portal of HAL, requires the deposit of the data files for each data deposit. Only some software records have been created without the code files.

Disciplinary differences

Each research laboratory has been indexed with a large scientific domain and with a more specific research discipline. In the following, we present the analysis of the HAL deposits regarding four large scientific domains, i.e., science and technology (SciTech), life and medical sciences (BioMed), social sciences and humanities (SSH), and law, economics and management (Law, Econ). Figure 6 shows that there are significant differences between the four domains.

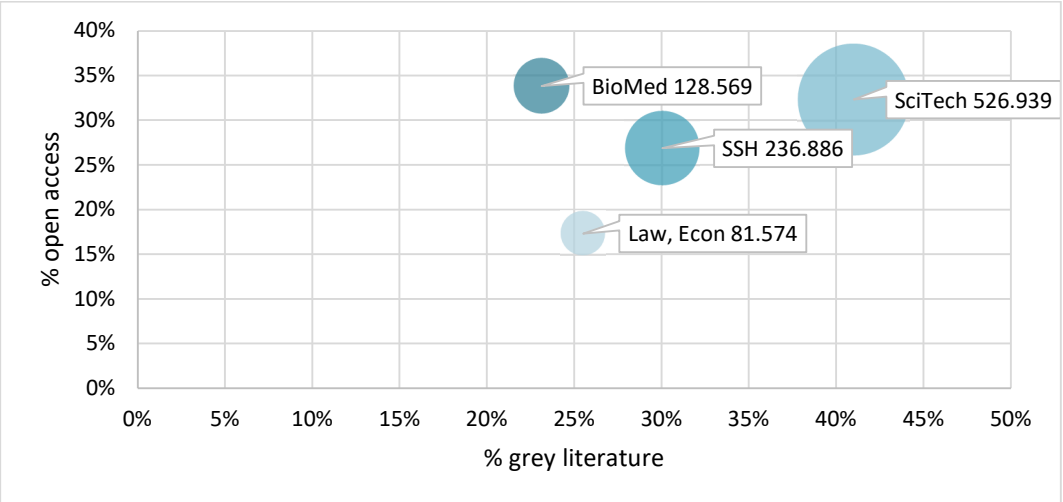


Figure 6. Degree of openness and part of grey literature in four scientific domains (N=973,968 deposits)

The deposits of the laboratories in life and medical sciences have the highest degree of openness (34%), followed by those in science and technology (32%). The same indicator is lower in social sciences and humanities (27%) and in law, economics and management (17%).
On the other hand, the laboratories in science and technology have the highest part of grey literature (41%), followed by those in social sciences and humanities (30%), law, economics and management (26%) and life and medical sciences (23%).

Institutional differences

As each laboratory is affiliated with at least (mostly) one of the ten French research universities, we finally analysed the laboratories' deposits also regarding their institutional affiliation. In other words, we determined the degree of openness and the part of grey literature for each university of our sample. Figure 7 reveals significant differences but also two clusters. Each bubble refers to a university, and the size of each bubble represents the number of deposits for this university.

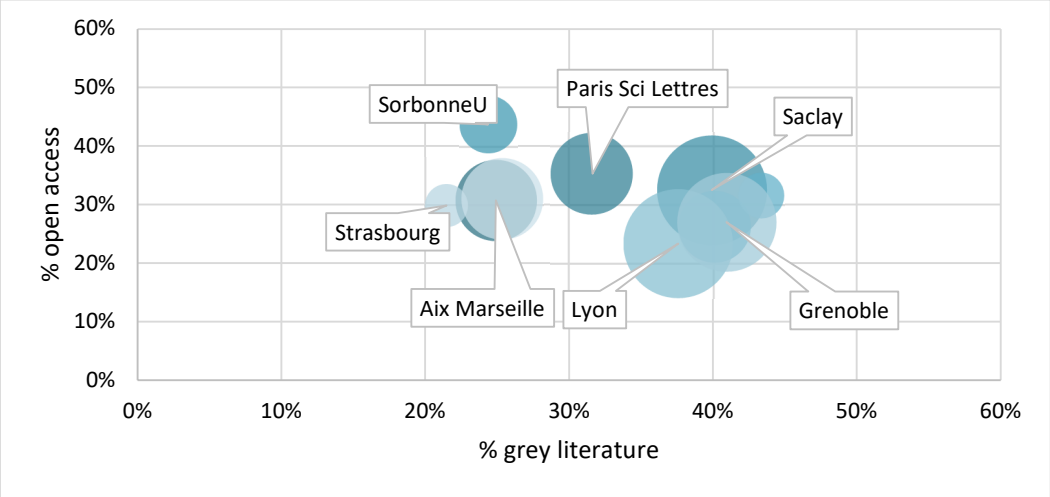


Figure 7. Degree of openness and part of grey literature per university (N=973,968 deposits)

Following figure 7, some universities appear "more open" and "greyer" than others. For instance, the degree of openness of the deposits from laboratories affiliated to Sorbonne Université (44%) and to the university of Paris Sciences Lettres (35%) than, for instance, Grenoble (27%), Bordeaux (26%) and Lyon (23%).
Regarding the part of grey literature, the universities of Nice (43%), Grenoble (41%), Bordeaux and Paris Saclay (both 40%) have deposited relatively more conference papers, PhD theses etc. than for instance Sorbonne Université (24%) or Strasbourg (21%).
Without a more detailed analysis, it is perhaps to early to speak of institutional clusters. Yet, figure 7 seems to indicate the existence of two clusters of rather similar universities (one with Strasbourg, Aix Marseille and Sorbonne Paris Cité, the other with Lyon, Grenoble, Saclay, Bordeaux and Nice) and that the main difference between these clusters is above all the different percentage of grey literature. But more research is needed for a more detailed evidence and a better understanding of these differences, which may be determined by institutional open access policies as well as by disciplinary differences.

Discussion

The impact of grey literature

Our results show that a significant part of the deposited scientific documents on HAL is grey literature - in our sample, these documents represent 35% of all deposited items, a percentage which is higher than in former surveys with French open repositories (less than 20% grey items, see Schöpfel & Stock, 2008 and Schöpfel & Prost, 2010), Spanish open repositories (23% grey items, see Melero et al., 2009) and European institutional research repositories (20% grey items, see Vernooy-Gerritsen et al., 2009). On the other hand, at the same time Luzi et al. (2009) already estimated the part of grey items eligible for the institutional repository of the Italian National Research Council at about one third of all documents. Obviously, this percentage appears to be a realistic estimation of the real impact of non-commercial and "unconventional" scientific literature, outside of and non-controlled by the academic journal and book publishing market.

Ten years ago, we observed that the number of grey documents in repositories was rapidly growing (Schöpfel & Prost, 2010). Compared to the 2010 figures, obviously not only the number but also the percentage of these grey documents increased in a significant way. This seems to confirm past appraisals that open repositories are the right place for grey literature, that it is "at home in open archives" (Luzi, 2010).

The French HAL repository provides a solution for at least three issues: it guarantees long-term conservation of deposited items through a partnership with the National Computing Center for Higher Education (CINES), it associates a unique identifier to each deposit, and it applies a generic metadata standard compliant with the Dublin Core, with some additional fields for specific document types, including theses and dissertations, conference papers and reports.

On the other side, a more detailed classification and indexing of unconventional document types is missing, which is regretful especially for potentially interesting items like working papers and preprints which can't be appropriately retrieved, but also for conference papers where a distinction between commercial publishing and grey items is more or less impossible. This is even more regretful as communications are the main part of the grey literature in our sample. A better classification would increase the findability of grey resources and contribute to the bibliodiversity, i.e., a large diversity of academic publishing.

Regarding the global part of grey literature, our results confirm domain-specific differences (which is not a surprise) but, more surprisingly, institutional differences – some large research universities have more deposits falling under the categories of grey literature than others. More analysis of our data is needed to assess whether this fact is correlated to disciplinary profiles or to the number of PhD students, or if we can identify other reasons.

The accessibility of grey literature

Our study shows that the part of accessible documents is higher for grey literature than for commercial publications. 37% of the grey literature is available in open access because deposited with the full text, while only 25% of the articles, books and book chapters are openly accessible. Furthermore, the results reveal differences between communications, theses, working papers, posters and reports which confirm former observations of different degrees of openness not only between repositories but also between document types (Schöpfel & Prost, 2015). Nearly all PhD theses and most of the preprints are in open access, while only slightly more than 20% of the conference papers are openly available on HAL.

One major reason for this difference is that HAL requires the deposit of the full text of PhD theses but not for conference papers. Another probable reason is that one part of the conference papers has been published in serials, journals or proceedings, published by commercial academic publishers. Here again, more insight is needed, for instance based on the DOI of the conference papers, to estimate the real grey, non-commercial part of conference papers.

However, there is no such explanation for the relatively low part of open access for reports and posters – except for the general observation that research units, similar to our own laboratory (see Schöpfel et al., 2018) increasingly use the HAL repository as a reservoir of metadata with the purpose of performance monitoring and assessment, through a massive deposit (creation or import) of records without the document. Open repositories are useful for research information management (see Schöpfel & Azeroual, 2021); however, if their primary function, i.e., sharing and preservation of academic production, declines or fails, their overall utility for the scientific communication may be at risk.

Another issue is the potential reusability of the openly accessible grey literature. All these documents are available in gratis open access, i.e., free of charge. Are they also in "libre" access, i.e., free for some kinds of further use and reuse (such as text mining), which presupposes some kind of open license permitting uses not permitted by default?[3] This seems unlikely because the by default condition is not licensing but intellectual property protection. Further analysis of the legal conditions of these deposits is needed to answer this question.

Our results reveal also differences between the large research domains, grey literature from science, technology and medicine being more open than in social sciences, humanities, law, economy and business. The reason may be the higher number of PhD theses but again, more research is required to determine the reasons for these differences, building on previous surveys like Fry et al. (2016).

Obviously, there are also institutional differences, some universities being more open (in terms of deposits of document files) than others. Here, further qualitative investigation will try to assess the reasons; we expect a multiple set of factors, like institutional open access policy, disciplinary differences, particular choices on the level of the individual research units and partnerships with research organizations.

Limitations

This paper provides empirical evidence from an ongoing research project. More detailed analysis is still to come, for instance regarding scientific disciplines, in order to answer questions about disciplinary particularities of HAL deposits and the correspondence between each laboratory's main field of research and the indexed disciplines of each of its deposits.

Moreover, based on the analysis of the HAL deposits, we will conduct in 2021 two surveys with the whole sample of 1,272 laboratories and with a smaller sample of some representative and particular laboratories in order to assess their strategies and practice regarding open science, open access and the HAL repository.

This second, qualitative part of the project may also fill a gap of this first study which limited the research field – open access strategies – to the deposits on HAL which is for sure the first and most important research repository in France but not the only way to publish in green open access. In particular, our first analyses exclude the self-archiving via preprint servers (like arXiv, medRxiv and bioRxiv) and via academic social networks (like Academia and ResearchGate).

This exclusion of preprint servers and social networks makes it impossible to assess the real open access part of all grey literature, except for PhD theses. In France, there is no directory for reports, working papers, Master dissertations, posters or conference papers. Also, there is no centralized or federated system for the French research information management, and no consolidated figures for the research output. Whenever a laboratory makes use of the HAL repository for an (quasi-) exhaustive reporting of its academic production, similar to a research information system, the part of grey literature will be representative if not complete. Hopefully, our future surveys will reveal this approach.

[3] See Peter Suber, The rise of libre open access. *SPARC Open Access Newsletter*, issue #164, June 2, 2012. http://legacy.earlham.edu/~peters/fos/newsletter/06-02-12.htm#libre

Concluding remarks

The purpose of our research project HAL/LO is to identify open access strategies of French research laboratories regarding the national open repository HAL, thus putting the focus on the "green road" of open access. From a sample of 1,272 laboratories of the ten universities that are part of the French excellence initiative IDEX, we analyse the HAL deposits in order to distinguish particular approaches to HAL which could be labelled as "green open access clusters".

The large and representative number of research laboratories and their deposits provides an opportunity for an analysis of the grey literature among these items. The limitations of this methodology have been discussed above, especially the issues with indexing and the mixture of commercial and unconventional conference papers. Nevertheless, the main result appears reliable enough: about one third of all deposits is grey literature, and their accessibility (degree of openness) is higher than of commercial academic publications.

Further analysis of our sample will contribute to a deeper understanding of the "place" of grey literature in open repositories. In particular, we will conduct three complementary studies:

- The evolution of the deposit of grey literature: is the part of one third stable over the time? How does the composition of grey literature change with the years? How did the 2016 "Digital Law" and following open access mandates impact the deposit of grey literature?
- The licensing of grey literature: what are the conditions of reuse of grey literature? Is open access to grey literature more "gratis" or more "libre"?
- The existence of DOI for conference papers: which is the part of commercial publishing of conference papers?

Further, qualitative research should also contribute to a better understanding of how research laboratories consider and handle non-commercial, unconventional academic documents, especially preprints, working papers, reports and conference papers. Of particular interest will be a follow-up study in 2021, insofar the major French research organisation CNRS decided in 2020 the mandatory use of HAL for the reporting and assessment of the performance of the CNRS research laboratories and individual researchers. For this reason, the analysis of the 2020 deposits will probably provide an exhaustive and reliable photography of the academic output and of the part of the grey literature.

Funding information

The study is part of the research project HAL/LO funded by GIS "Réseau Urfist" from 2019-2021 with 10,000 euros.

Acknowledgments

We would like to thank El Hadji Ibrahima Ndiaye from the University of Lille for his contribution to the project's database.

Conflict of interest

The authors declare no conflict of interest.

References

Farace, D., & Schöpfel, J. (Eds.). (2010). *Grey Literature in Library and Information Studies*. Berlin, De Gruyter Saur.

Fry, J., Spezi, V., Probets, S., & Creaser, C. (2016). Towards an understanding of the relationship between disciplinary research cultures and open access repository behaviors. *Journal of the Association for Information Science and Technology*, 67(11), 2710–2724. https://doi.org/10.1002/asi.23621

Luzi, D. (2010). Grey Documents in Open Archives. *The Grey Journal*, 6(3), 137–144.

Luzi, D., Di Cesare, R., Ruggieri, R., & Cerbara, L. (2009). Towards an Institutional Repository of the Italian National Research Council: A survey on Open Access experiences. *The Grey Journal*, 5(1), 35–47.

Melero, R., Abadal, E., Abad, F., & Rodríguez-Gairín, J. (2009). The situation of open access institutional repositories in Spain: 2009 report. *Information Research*, 14(4). http://informationr.net/ir/14-4/paper415.html

Nérisson, S. (2016). Le droit d'exploitation secondaire de l'auteur scientifique. *I2D - Information, Données & Documents*, 53(3), 27. https://doi.org/10.3917/i2d.163.0027

Schöpfel, J. (2011). Towards a Prague definition of grey literature. *Grey Journal*, 7(1), 5–18. https://hal.archives-ouvertes.fr/sic_00581570

Schöpfel, J., & Azeroual, O. (2021). Current Research Information Systems and Institutional Repositories: From Data Ingestion to Convergence and Merger. In D. Baker & L. Ellis (Eds.), *Future Directions in Digital Information. Predictions, Practice, Participation* (pp. 19–37). Oxford: Chandos Publishing.

Schöpfel, J., & Prost, H. (2010). *Développement et Usage des Archives Ouvertes en France. Rapport. 1e partie : Développement*. Villeneuve d'Ascq, Université de Lille 3. http://hal.archives-ouvertes.fr/sic_00497389/

Schöpfel, J., & Prost, H. (2014). Back to Grey. Disclosure and Concealment of Electronic Theses and Dissertations. *ITlib*, (2). http://itlib.cvtisr.sk/archiv/2014/2/back-to-grey-disclosure-and-concealment-of-electronic-theses-and-dissertations.html?page_id=2686

Schöpfel, J., & Prost, H. (2015). Degrees of openness: Grey literature in institutional repositories. *The Grey Journal*, 11(2), 103–112. https://hal.archives-ouvertes.fr/hal-01290339v1

Schöpfel, J., & Prost, H. (2020). How scientific papers mention grey literature: a scientometric study based on Scopus data. *Collection and Curation*, ahead-of-p(ahead-of-print). https://doi.org/10.1108/CC-12-2019-0044

Schöpfel, J., & Stock, C. (2008). Grey literature in French digital repositories: a survey. *GL10 Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society*, Amsterdam, 8-9 December 2008. https://archivesic.ccsd.cnrs.fr/sic_00379232/

Schöpfel, J., Le Bescond, I., & Prost, H. (2011). Open is not enough. A case study on grey literature in an OAI environment. *GL13 Thirteenth International Conference on Grey Literature: The Grey Circuit. From Social Networking to Wealth Creation*. Washington, DC, 5-6 December 2011, 75–86. https://hal.archives-ouvertes.fr/sic_00908862

Schöpfel, J., Prost, H., Fraisse, A., & Chaudiron, S. (2018). Valoriser les publications d'un laboratoire universitaire dans l'environnement de la science ouverte : Retour d'expérience de la collection GERiiCO sur HAL. In *ICOA 2018 3e colloque international sur le libre accès, 28-30 November 2018, Rabat, Morocco*. https://hal.archives-ouvertes.fr/hal-01940352

Schöpfel, J., Prost, H., & Ndiaye, E.H.I. (2019). Going Green. Publishing Academic Grey Literature in Laboratory Collections on HAL. In *GL21, Twenty-First International Conference on Grey Literature. Open Science Encompasses New Forms of Grey Literature, 22-23 October 2019, TIB, Hannover, Germany*. https://hal.archives-ouvertes.fr/hal-02300017

Vernooy-Gerritsen, M., Pronk, G., & van der Graaf, M. (2009). Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe. *Ariadne*, (59). http://www.ariadne.ac.uk/issue/59/vernooy-gerritsen-et-al/

[1] RNSR https://appliweb.dgri.education.fr/rnsr/

[2] ScanR https://scanr.enseignementsup-recherche.gouv.fr/

TIB
LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY

"AS AN INFORMATION CENTRE FOR THE DIGITISATION OF SCIENCE AND TECHNOLOGY, OUR OBJECTIVE IS TO SUPPORT RESEARCHERS AT ALL STAGES OF THEIR WORK BY PROVIDING THEM WITH OUR SERVICES."

Professor Dr. Sören Auer

WWW.TIB.EU

The Grey-side of Audio Archives*

Monica Monachini, Institute of Computational Linguistics "A. Zampolli" - CNR, Pisa, Italy
Maria Francesca Stamuli, Ministry of Culture and Tourism – Florence, Italy
Silvia Calamai, Human Sciences and Intercultural Communication, Siena University
Niccolò Pretto, Institute of Computational Linguistics "A. Zampolli" - CNR, Pisa, Italy
Silvia Bianchi, Human Sciences and Intercultural Communication, Siena University, Italy

Abstract
Archives often include documents that can hardly be considered publications or grey literature as such, yet they maintain their documentary value and play a role of primary sources for the specialists. These documents, indeed, can help archivists to reveal the sedimentation process of the archive itself and to preserve the authentic context of the documentary production. They also appear to be very useful for the community of researchers and scholars. This happens more frequently with oral archives which include 'non-conventional sources', thus bringing together audio documents, fieldworks notes, correspondence, slipcases, analogic compact cassettes or open reels. At the cross-road of two disciplines, Archival Science and Grey Literature, this paper aims to argue the applicability of the concept of grey literature to this wide range of documentary materials, by showing the experience of Archivio Vi.Vo, a regional project aiming at building a model for archiving, preserving, managing and disseminating audio documents.

1. Introduction

1.1. The background: audiovisual heritage

The audiovisual heritage of the twentieth century is at risk of being lost. Both analogue and digital documents are in danger: the former because of carrier degradation and playback device obsolescence, the latter because of the hardware and software obsolescence as well as the rapid evolution of standards and formats. On the other side, knowledge conveyed by oral culture may be as disseminated, reliable and accessible as written resources, counteracting the negative effect of technological obsolescence. Looking ahead and careful planning can minimize the repercussions on memory preservation.

1.2. Setting the problem: Archival Science, oral archives and Grey Literature

Why would Archival Science matter to Grey Literature? It appears that this question has a relevant place in the domain of digital archives, as we try to argue in the present paper. Building a digital archive entails the creation of different types of digital documents: i) documents of the archive itself; ii) documents about the archive. These are deliverables, technical papers, sketches and other different typologies of documents falling uncontroversially into different types of grey literature. They all appear to be extremely useful for supporting these projects as well as studying the evolution of the archive itself. In addition, they often play a role of primary sources for the specialists and prove to be useful in helping the archivists to reveal the sedimentation process of the archive itself and preserve the authentic context of the documentary production. Finally they also appear to be beneficial for the community of researchers and scholars in their studies.

The picture become even more complex in the realm of audio documents and oral archives, that preserve, organize and make available such documents. Oral archives appear to be rather complex and delicate, mainly because of the multifactor 'fragility' that characterizes audiovisual heritage (see Calamai, Casellato and Stamuli, forthcoming).

* First published in the GL2020 Conference Proceedings, February 2021.

In fact, the archival treatment of audio documents deals with 'oral sources'. Those kinds of sources are at the crossroads of several domains of knowledge and disciplines, each with its own jargon: oral history, linguistics, anthropology, ethnomusicology, digital humanities. In archival science, 'oral sources' are considered 'new sources' or 'not conventional sources' (Carucci, 1993; Mulè 2005). What makes audio documents different from prototypical archival and documentary records is mainly their hybrid nature, characterized by a dichotomy between the 'content' (the audio signals) and the carrier, that can potentially contain more than one audio recording, even different kinds of audiovisual records. The digitisation process of analog audio recordings for their preservation (Miliano, 1999) provides additional material that describes the operations undertaken. Doing so is essential for reporting the history of the document's transmission.

Given this background, the research question raised in this paper can be summarised as follows: "how much grey" is this wide range of documentary materials?

1.3. The Case Study: Archivio Vi.Vo.

Archivio Vi.Vo. is an Italian regional project supported by Regione Toscana, whose aim is to explore methods and services for long term preservation and secure access for audiovisuals (Calamai et al. 2020; Stamuli et al., in press). It constitutes a pilot study dealing with the description and cataloguing of Caterina Bueno's[1] audio recordings (digitized during the PAR-FAS project *Grammo-foni Le soffitte della voce*, 2010-13) composed of about 450 carriers (audio reels and compact cassette tapes corresponding to nearly 700 hours of recording). The services for long-term preservation and access are developed within the framework of CLARIN-IT, the Italian counterpart of the European Research Infrastructure for Social Sciences and Humanities, CLARIN-ERIC. The Italian node of the European data-center federation, ILC4CLARIN, offers long-term solutions and technology services for sustaining, archiving, deploying, connecting, and analyzing linguistic data. By means of *Archivio Vi.Vo*., Caterina Bueno's audio recordings and its metadata will be systematized under the CLARIN-IT domain, through an experimental and cross-disciplinary approach to preservation, management and access to audio-video data, aimed at adopting the model and the high-performance computing and archiving services of the GARR network infrastructure, built along the CLOUD paradigm. This model will be disseminated to the scientific community interested in managing audiovisual records and will make it possible for the general public to access the oral sources produced in the territory. The project is thus meant to support and advance activities regarding this topic in the CLARIN-IT consortium and the CLARIN infrastructure.

2. Research approach and first outcomes

One of the main objectives of *Archivio Vi.Vo* is the creation of a model for dealing with issues of preservation, archiving, management and dissemination of audio documents, adoptable by institutions, archives and researchers. In order to achieve this objective, grey literature assumptions have a prominent role.

Starting from the very beginning of the archive's creation, the developers should collect, analyze and make accessible to professional and non-professional audience (1) the overall set of different documents related to the production of the audio records, and (2) the documents concerning the design and development of the platform for the long-term preservation and access to the audio archive.

Concerning the former, in the arrangement phase of the archive itself, archival science prescribes to carry out a careful analysis and survey of the archival documentation. In fact, if

[1] Caterina Bueno (1943-2007) was an Italian ethnomusicologist and singer, highly appreciated for its cultural value. Her work allowed the collection of many Tuscanian and central Italy's folk songs that have been passed down orally from one generation to the next until the 20th century.

we conceive the archive not as a dead entity, but as an evolving reality we need to reveal the sedimentation process and give the documents their authentic context of production, shaped by the activities of the producer of the archive itself. This is particularly important for 'not conventional sources' as oral sources are. In Caterina Bueno's sound archive, everything reveals the archive sedimentation process and lends the appropriate and authentic context of production to the audio records: her fieldwork notes, her correspondence, the slipcases of compact cassettes and of the open-reels that she wrote, her diaries, the newspaper clippings she collected, the numbering labels that she used, etc.
Archivio Vi.Vo. makes all this data fully retrievable, thus making audio records fully understandable by adopting both bibliographic and archival recommended standards.

The second kind of grey literature documents considered here belongs to the well-known typology of "technical documents" (Pejšová, 2011). The production of these documents is strictly related to the software engineering field, and in particular, to the adopted software process model (sometimes called Software Development Life Cycle model). There are no universal software engineering methods that are suitable for all systems and all companies (Sommerville, 2016) and the software process model has to be chosen considering the kind of application will be developed. The complexity of the *Archivio Vi.Vo.* infrastructure prevents a development from scratch, but, as most of the complex business systems, it has to be developed configuring existing systems and integrating new functionalities. For this reason, within the project, we adopt the process model that Sommerville defines as "integration and configuration", in order to re-use and adapt existing software. Each task of this model has different outcomes: products or deliverables. The latter can be reports, models, sketch, UML schemas, system and user documentation (which are precious for the development, maintenance and possible extension of the software since the early stage of its development, as well as for other similar projects), *i.e.* a vast kind of material that will be published within the infrastructure itself and offered to the scientific community as grey literature.
From this respect, special attention is paid to the technical reports produced inside the project: that is, the description of Caterina Bueno's archive, its archival structure, the metadata scheme, the glossary (in order to have a common vocabulary inside the research project), and the legal documentation related to privacy and copyright. Conversely, certain members of the project staff are producing a rather different type of technical reports, which are useful inside the project in order to document the activities of every single post-doc researcher, for preparing the hand-over from one researcher to another, for communicating accurately through a cross disciplinary project and to express to others the path on which the project is moving on. According to GL, different levels of accessibility and different types of public can thus be envisaged: i) open, for the general public (e.g., the glossary, the structure of Caterina Bueno's archives), and ii) restricted, for internal use (e.g., the final reports of each post-doc researcher involved in the project).

3. Final Remarks

As Serini (2003) points out, among the archival documents there are numerous materials which, according to their publication status and the interest they can have among specialists, can be quite rightly grey literature. The cross-disciplinary approach adopted by the *Archivio Vi.Vo.* project will bring to the forefront the technical documentation and the archival materials as 'grey literature', as data distinguished by content and type, but united by the objectives of research, organization and dissemination. According to Serini, what makes a wide range of documents 'grey literature' is the combination of different, but interwoven, layers such as their production (characteristic of a particular research endeavor or to an organization) and the type of distribution.

Archivio Vi.Vo. aims at producing a digital ecosystem with audio recordings at the heart. The unconventional documentation produced in the realm of audio archives will be treated in the light of grey literature, conceived as a product in relation to a service, a vehicle of direct communication, participatory democracy, transparency and organisational clarity, and the recovery of memory (Sardelli 2000).

At the same time, the building itself of such an ecosystem produces a flood of different and diverse objects which might be undisputedly inserted inside the GL approved taxonomy. On the other hand, it cannot be excluded that the GL taxonomy might also benefit from the work carried out by *Archivio Vi.Vo.* given the cross-disciplinarity and multimediality of audio archives.

Bibliography

Bologna, M. (2015). *La sedimentazione storica della documentazione archivistica*. In: Giuva L and Guercio M. (Eds) *Archivistica. Teorie metodi, pratiche*, 212-235. Roma: Carocci.

Calamai, S., Stamuli M.F., Casellato, A., (in press), Collaborative best practices: An Italian *Vademecum* on the conservation, the description, the use and the re-use of oral sources. Sonorités. Sonorités. Bulletin del l'AFAS, Association fraınçaise des archives orales sonores et audiovisuelles. Questions de droit, d'éthiques et de terrain: qu'est-ce qui change dans les pratiques de recherche? Law, ethics and fieldwork: how are research practices changing?.

Calamai, S., Pretto, N., Monachini, M., Stamuli, M.F., Bianchi, S., Bonazzoli, P. (2020) *Building a Home for Italian Audio Archives*. In Navarretta, C. and Eskevich, M. (eds.) Proceedings of CLARIN Annual Conference 2020, p. 112-116. Virtual Edition, 2020. Retrieved from: https://office.clarin.eu/v/CE-2020-1738-CLARIN2020_ConferenceProceedings.pdf; last visited December 28, 2020.

Carucci, P. (1993) Prefazione. In Barrera G., Martini A. and Mulé A. (Eds.) *Fonti orali. Censimento degli istituti di conservazione,* 13-23. Roma: Pubblicazioni degli Archivi di Stato. (Quaderni della Rassegna degli Archivi di Stato, 71).

International Council on Archives ICA, Brothman B. (1999) *ISAD(G): General International Standard Archival Description Second Edition Adopted by the Committee on Descriptive Standards Stockholm, Sweden, 19-22 September 1999*. Ottawa: International Council on Archives. Retrieved from: https://www.ica.org/en/isadg-general-international-standard-archival-description-second-editio; last visited March 25, 2020.

Miliano, M. (Ed.). (1999). *IASA cataloguing rules (IASA 1999).* International Association of Sound and Audiovisual Archives IASA Editorial Group. Web edition. Retrieved from: https://www.iasa-web.org/cataloguing-rules; last visited April 6, 2020.

Mulè A. (2005) *Le fonti orali in archivio. Un approccio archivistico alle fonti orali*. In: Ghezzi A. G. (Ed.) L'Archivio: teoria, funzione, gestione e legislazione, 284-296. Milano: Pubblicazioni dell'ISU Università Cattolica.

Serini, P. (2003). *Attualità della letteratura grigia. Il ruolo delle biblioteche nella sua valorizzazione*, Biblioteche oggi - gennaio-febbraio 2003, Retrieved from: http://www.bibliotecheoggi.it/2003/20030106101.pdf; last visited March 26, 2020.

Stamuli, M.F., Calamai, S., Monachini, M. (*in press*) *L'Archivio Vi.Vo: una piattaforma per la conservazione e l'accesso alle fonti orali in Toscana*. In: Brunetti, D., Robotti, D., Salvalaggio, E. Documenti sonori. Voce, suono, musica in archivi e raccolte, Torino, Centro studi piemontesi (Archivi e Biblioteche in Piemonte, 5).

Pejšová, P. (2012). A linked-data vocabulary of grey literature document types: Version 1.0. Libraries In The Digital Age (LIDA) Proceedings, 12. Retrieved from http://ozk.unizd.hr/proceedings/index.php/lida/article/view/87/55;

Sardelli, A. (2000). *Il fascino discreto della letteratura grigia: ridefinizione e ambiti di applicazione*. La letteratura grigia: politica e pratica. 3° Convegno nazionale, atti a cura di V. Alberani e P. De Castro, Roma, ISS, p. 15-20.

Sommerville, I. (2016). *Software Engineering* (10 ed.). Pearson.

Data Papers provide an Innovative Tool for Information and Data Management: A Use Case*

Dominic Farace, GreyNet International, Netherlands
Joachim Schöpfel, University of Lille, France

Abstract:
In earlier work dealing with Data Papers, emphasis initially focused on its definition, the construction of a standardized template in compiling its component parts, its further publication, access and data preservation[1]. this was followed by a study seeking to engage the citation and referencing of data papers and the further sharing and reuse of the data described therein[2]. a more recent work outside GreyNet's immediate community of practice[3] addressed the diverse formats used to compile and publish data papers. that study also discussed the automatic generation of data papers, and the differing opinions as to whether they contribute to data, information and/or knowledge production. drawing on the results of the above-mentioned work, this case study seeks to demonstrate how the data paper provides an innovative tool for information and data management, as part of an "ecosystem" of conference proceedings, journal articles, research data and open repositories. it relies upon GreyNet's current collection of 41 published datasets and 11 data papers[4]. this study highlights the importance of the human contribution for the writing of data papers and the enrichment of their metadata. To this end, key shared components of GreyNet's collection of data papers are discussed, namely the stakeholders, linked metadata, open data archiving, preservation, and issues of quality and information rights. The study concludes from a user perspective by addressing the value of data papers drawn from available usage statistics. The results of this study are expected to move beyond a simple case study to a use case in which the key components of data papers can be implemented in other communities of practice dealing with non-conventional 'grey' literature.

1. Introduction – Data Paper Template And Key Components

The data paper template and its component parts provide an innovative tool for data management. A completed template is not a data paper but provides for the content needed to produce a published data paper. The template is standardized and consists of five sections in which to compile and order the information and data required for a data paper. These include an overview of the bibliographic linked metadata, the methods applied, a technical description of the dataset, the potential for reuse of the data including any limitations, and key linked references. Each section is aided by a note field in order to assist the researcher/author in completing the template and further writing the data paper. Emphasis is placed on the description of the data and not on its analysis. It is understood that the analysis of data allows for multiple outputs, while the description of the data by the researcher/author is a priori defined.

Another way of examining the performance of this tool is to look at the function of its component parts as they apply to information and data management. The creators of the data are assigned an ORCiD, a unique persistent identifier, establishing their professional identity, it shows that the data is housed in a trusted data archive carrying the CoreTrustSeal and that not only is the data published in an archive, but its accompanying data paper is published as a journal article, thus driving further awareness to the research data. The DOI, digital object identifier, guarantees the findability of the data even if it changes archives. The CC0 waived rights of the data allow for its open access, and together these components of the data paper

* First published in the GL2020 Conference Proceedings, February 2021.

are shown to adhere to the FAIR data principles (Findability, Accessibility, Interoperability, and Reuse) – an indication of the quality of the data.

2. Data Paper In Line With Other Grey Literature Document Types

Research data and its accompanying data paper stand not alone but in line with other related document types, which together form GreyNet's document trail. In 2012, GreyNet's first datasets within its Enhanced Publications Project were published in the DANS Easy Archive. The following year, it became part of GreyNet's information and data workflow. It was not until 2017 within GreyNet's Data Papers Project, that its first data paper was published. In that following year, the benefits of this tool for information and data management soon became evident and measurable. Not only was GreyNet's document trail extended with a further document type in both preprint and article format, but it opened (cross) links to other related documents, it allowed for further statistics on document use and downloads. One example is from the quarterly statistics provided by EBSCO on the number of downloads of GreyNet data papers published in The Grey Journal. These totalled 290 downloads over the period 2018-2019. It is also thought to increase the potential for data citation, referencing, and the reuse of datasets. However, it remains a fact that GreyNet's collection of published datasets linked to a data paper is relatively small. This is seen when compared with other of GreyNet's collections, such as slide presentations and conference papers. However, because the data paper prompted further development of its workflow as well as its capacity to manage information and data across collections, it rightly can be considered an innovative tool.

3. Data Papers Enrich Greynet's Published Datasets

In 2012, GreyNet entered together with DANS Easy Archive in an Enhanced Publications Project to begin housing its datasets linked to its collection of Conference Papers. Since then, the number of GreyNet's datasets entered in the DANS Archive averages circa 5 per year. In 2015, GreyNet undertook a project on leveraging its sustained information resources. The results of a stakeholder survey indicated that the use and awareness of its collection of its datasets were significantly lower than those of its other sustained information resources.

In 2017, GreyNet together with one of its associate members - the University of Florida; George A. Smathers Libraries – initiated a Data Papers Project, which undertook the construction of a standardized template in compiling its component parts, its further publication, open access and preservation. As part of this project, three workshops on data papers have since been organized and carried out together with three of its associate members in the United States, Italy, and the Czech Republic.

The results of the Data Papers Project on the use of GreyNet's published datasets can be summarized as follows: Between mid-October 2012 and early April 2020, GreyNet's collection of published datasets number 41 housed in the DANS Easy Archive. During this period, the 41 datasets were downloaded a total of 771 times, averaging 18.8 downloads per dataset. Of those 41 datasets, 11 have an accompanying Data Paper, which accounts for 331 (42.9%) of the 771 downloaded datasets. On average, they account for 30.1 downloads per dataset.

As of April 4, 2020 **GreyNet's Collection of Published Datasets in DANS Easy Archive**	**With Data Papers**	**Without Data Papers**
41 Datasets	11 (26.8%)	30 (73.2%)
771 Downloads	331 (42.9%)	440 (57.1%)
Average Downloads per Dataset	**30.1**	**18.8**

Table 1: Comparison of dataset downloads with and without an accompanying data paper

The table above indicates that on average GreyNet's datasets with an accompanying Data Paper show an increase of 11.3 (37.5%) downloads per dataset compared with those that do not include a Data Paper

4. Data Papers for Other Communities of Practice

Given the background, developments, and the results to date in GreyNet's Data Papers Project, the question now raised is whether data papers can serve other communities of practice within the field of grey literature. Here, we briefly discuss GreyNet's recent initiative to expand its project.

In February 2020, researchers/authors outside of GreyNet's community of practice were invited to submit their research data and write a Data Paper using the standardized template described above. The selection was made based on recent conference papers dealing with grey literature that contain statistical data. Now two and half months on, there have been 4 responses to the request – the phases of each to date are recorded in the table below.

Response Number:	Dataset(s) Submitted to GreyNet	Data Paper Drafted	Dataset(s) published in DANS	Data Paper Uploaded in DANS	Data Paper Preprint in RGL, GreyGuide	Data Paper Published in TGJ Vol.16
R1.	☑	☑	☑	☑	☑	☑
R2.	☑	☑	☑	☑	☑	☑
R3.	☑	☐	☐	☐	☐	☐
R4.	☑	☑	☐	☐	☐	☐

Table 2: Workflow from submission of the dataset to publication of the data paper as a journal article

In brief, two of the four researchers/authors (R1 and R2) responded to the request by submitting their datasets and accompanying data papers. These two data papers await formal publication in The Grey Journal[5], Volume 16, Number 2, Summer 2020. Another researcher/author (R3) submitted the dataset, however the data paper has yet to commence. And, (R4) entered later in the project, however has since submitted both the data and draft data paper.[6]

5. Concluding Remarks

While it may be premature to speculate on the implementation of Data Papers in the workflow of other grey literature communities, the benefits of this tool for GreyNet's management of data and information are clear. To a certain extent, GreyNet is prepared to enable the publication of datasets outside of its immediate catchment and further publish their related data papers. Nonetheless, its continuing role remains in education and training on Data Papers as a tool for information and data management.

References

[1] FARACE D., FRANTZEN J., and SMITH II P. L, (2018). Data Papers are Witness to Trusted Resources in Grey and Access to Grey Literature, Rome, Italy 23-24 October 2017 https://doi.org/10.26069/greynet-2018-000.002-gg

[2] FARACE D., FRANTZEN J., and SCHÖPFEL J, (2019). Open Data engages Citation and Reuse: A Follow-up Study on Enhanced Publication. In: Twentieth International Conference on Grey Literature, Loyola University New Orleans, USA December 3-4 2018 vol. 20 pp. 117-122. https://doi.org/10.26069/greynet-2019-000.070-gg

[3] SCHÖPFEL, J., FARACE, D., PROST, H., and ZANE, A. (2019). Data Papers as a New Form of Knowledge Organization in the Field of Research Data. Knowledge Organization, 46(8), 622–638. https://doi.org/10.5771/0943-7444-2019-8-622

[4] GREYNET – GREY LITERATURE NETWORK SERVICE; Amsterdam - The Netherlands; (2012): Thematic collection: GreyNet Enhanced Publications Project. https://doi.org/10.17026/dans-zp3-pd7e

[5] FARACE D. (ed.) 2020. The Grey Journal: An International Journal on Grey Literature / GreyNet, Grey Literature Network Service. - Amsterdam: TextRelease, Vol. 16, 2020. – e-ISSN 1574-180X (PDF) http://www.greynet.org/thegreyjournal.html

[6] **Note:** Since the completion of the text writing to submission of this conference paper for GL2020, all four data papers mentioned in Table 2 have since appeared published in The Grey Journal.

Data from "OpenGrey, System for Information on Grey Literature in Europe"

https://doi.org/10.17026/dans-xtf-47w5
URN: urn:nbn:nl:ui:13-i3-x6uc

Joachim Schöpfel, University of Lille, France
https://orcid.org/0000-0002-4000-807X
Project Leader / Database Reformat / Final Text

Dominic Farace, GreyNet International, Netherlands
https://orcid.org/0000-0003-2561-3631
Project Administration / Draft Text

Jerry Frantzen, GreyNet International, Netherlands
https://orcid.org/0000-0002-3405-7078
Technical Support

Abstract
OpenGrey, System for Information on Grey Literature in Europe, is your open access to 1 014 872 bibliographic references of grey literature produced in Europe and allows you to export records and locate the documents. Examples of grey (gray) literature include technical and research reports, doctoral dissertations, conference papers, official publications, and other types of grey literature. In 1980, this information resource first began as a bibliographic database under the name SIGLE. In 2006, the records migrated to OpenSIGLE - an open access database. In 2010 its name was changed to the OpenGrey Repository given that full-texts documents began to be ingested and preserved. And in 2020, Inist-CNRS (service provider) gave notice that OpenGrey will be discontinued. In order to guarantee open access to this resource, it is now archived as a database in DANS EASY – a data archive certified with the Nestor and CoreTrustSeal.

Keywords: Data Paper; SIGLE; OpenSIGLE; OpenGrey; EAGLE; Grey Literature; Gray Literature

Subject Area: Science, Technology, Biomedical Science, Economics, Social Science, Humanities

Methods Applied

● Steps

Each participating European organization used a standard record template (form) with the aid of manuals (e.g., Subject Category List and associated key terms as well as a Corporate Register with preferred Names/spellings). Each participating organization made a selection of their acquisitions for submission to the SIGLE database. If a document title was not in English, then an English translation was provided. As such, each SIGLE record contained informative titles in English and/or in the original language, the author's name, academic degree, and the research organization or educational institution, the document's date of publication and type, number of pages, report numbers, and language, as well as subject classifications. Many records include keywords and abstracts. Each national structure submitted records in their own language. A search through the entire database was made possible by providing an English translation of the title or English keywords. One of these fields was mandatory. Moreover, each record contained a clear mention of availability.

● Sampling strategy

The selection of records entered in the SIGLE database was made by several institutions representing the EAGLE member countries. The following list is not exhaustive.

Country	Institution
Belgium	Université Catholique de Louvain
Czech Republic	National Library of Technology
Germany	Fachinformationszentrum Karlsruhe
Germany	German National Library of Science and Technology (TIB)
Spain	Centro de Informaciòn y Documentaciòn Cientìfica
France	Institut de l'Information Scientifique et Technique-CNRS
United Kingdom	British Library Document Supply Service
Italy	Consiglio Nazionale delle Ricerche
Luxembourg	Bibliothèque Nationale de Luxembourg
Latvia	Latvian Academic Library
Portugal	Fundação para a Ciência e a Tecnologia
Russian Federation	VNTIC Scientific & Technical Information Centre of Russia
Slovakia	Slovak Centre of Scientific and Technical Information

The sampling strategy of the former SIGLE records was decided by the EAGLE association, coordinated by the SIGLE technical committee, and carried out by the member institutions following agreed and explicit rules. Later input in the OpenSIGLE/OpenGrey repository that originated from the International Conference Series on Grey Literature was provided by GreyNet International, the Grey Literature Network Service. Regarding the property of the records and the database, all usage rights of the SIGLE database lapsed upon the complete liquidation of the association while the copyright on input remained with each member organization that supplied the records. EAGLE's last General Assembly asked the operating agent, FIZ Karlsruhe, for the interim preservation of the SIGLE records in XML format beyond the liquidation of SIGLE. This was for the purpose of archiving and integration into a new European non-profit project i.e., OpenSIGLE hosted and managed by INIST-CNRS. Nearly all of the former EAGLE members signed a declaration of intention regarding the future use of their existing input in the SIGLE database. The complete liquidation of EAGLE was formally published by the Luxemburg Register of Commerce and Societies on August 23, 2006 (Schöpfel et al., 2006).

● Quality Control

The SIGLE descriptive cataloguing rules were based on those of the International Nuclear Information System (INIS), and the subject classification scheme was a modified version of that endorsed by the Committee on Scientific and Technical Information (COSATI) of the US Federal Council for Science and Technology. Each member institution was responsible for the quality of their national records. The operating agent (technical processing center) played a key role in the quality control, rejecting non-compliant input. The SIGLE technical committee supervised the compliance with the agreed rules. However, "in operating SIGLE, no attempt has been made to be rigid in formulating rules of membership (...) considerable flexibility has enabled a number of countries to participate to a greater or lesser extent depending on their capabilities" (Wood & Smith, 1993, p.21). Furthermore, the last operating SIGLE agent and the host of the OpenSIGLE/OpenGrey repository, FIZ Karlsruhe and INIST-CNRS, executed complementary quality controls at different moments during the transfer, launch, and update of the metadata records (Schöpfel et al., 2006).

Dataset Description

File name:	OpenGrey
Format:	MySQL
Size:	360,358,816 bytes
***Path*:**	mysql.sql.gz
Creation dates:	*from* 1980-01-01 *to* 2018-07-27
Language:	English
License:	CC0 Waiver - no rights reserved
Archive name:	DANS EASY Archive
Publication date:	2021-02-06
DOI:	10.17026/dans-xtf-47w5
URN:	urn:nbn:nl:ui:13-i3-x6uc

Potential Reuse of the Data

In 1980, this information resource first began as a bibliographic database under the name SIGLE. In 2006, the records migrated to OpenSIGLE - an open access database. In 2010 its name was changed to the OpenGrey Repository given that full-texts documents began to be ingested and preserved. And in 2020, Inist-CNRS (service provider) gave notice that OpenGrey will be discontinued. In order to guarantee open access to this resource, it is now archived as a database in DANS EASY – a data archive certified with the Nestor and CoreTrustSeal. The database file can be downloaded in XML and CSV format. It is accompanied by a ReadMe text file and a document on MySQL. The ReadMe file contains information about the eight tables of the database. The bibliographic metadata data can be useful to enrich scientific bibliographic databases or other academic discovery tools. The data can also be useful for scientometric studies on academic output in the European research area between 1980 and 2005. Also, because of the records' harmonized multidisciplinary and multilingual characteristics, the data can be used for research and development in the field of scientific terminology. Furthermore, the data are published on the DANS EASY platform under a Creative Commons public domain license.

Linked References:

Farace D., Frantzen J., Stock C., Henrot N., & Schöpfel J, (2009). OpenSIGLE, Home to GreyNet's Research Community and its Grey Literature Collections: Initial Results and a Project Proposal. In: Tenth International Conference on Grey Literature - Designing the Grey Grid for Information Society, Science Park Amsterdam, NL. December 8-9, 2008 vol. 10 (1), pp. 118-122. https://doi.org/10.26069/greynet-2020-000.239-gg

Farace D., Frantzen J., Stock C., Henrot N., & Schöpfel J, (2010). OpenSIGLE - Crossroads for Libraries, Research and Educational Institutions in the field of Grey Literature. In: The Grey Journal: An International Journal on Grey Literature, Library of Congress, Washington D.C., USA 14-15 December 2009 vol. 11 (1), pp. 122-128. https://doi.org/10.26069/greynet-2020-000.212-gg

Schöpfel, J. (2005). MetaGrey Europe, a Proposal in the aftermath of EAGLE-SIGLE. *GL7 Seventh International Conference on Grey Literature*, 5-6 December 2005, Nancy. http://archivesic.ccsd.cnrs.fr/sic_00001722/en/

Schöpfel, J., Stock, C., & Henrot, N. (2006). From SIGLE to OpenSIGLE and beyond: An in-depth look at Resource migration in European context. *GL8 Eighth International Conference on Grey Literature,* 4-5 December 2006, New Orleans. https://archivesic.ccsd.cnrs.fr/sic_00181592/

Wessels R. H, (1998). An Eagle's eye-view of grey literature research. In: Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information, Luxembourg 13-14 November 1997 vol. 3 (1), pp. 212-213. https://doi.org/10.26069/greynet-2021-000.398-gg

Wood, D. N., & Smith, A. W. (1993). SIGLE: A Model for International Co-operation. *Interlending & Document Supply*, 21(1), 18–22. https://doi.org/10.1108/02641619310154331

APPENDIX

History

The European database SIGLE (System for Information on Grey Literature) was created in 1980 in order to collect and to make available grey literature produced in the countries of the European Community by the intermediary of an online database. The main purpose was to provide access to documents, and incidentally to improve bibliographic coverage. SIGLE was launched with the support of the European Community. The base was produced from 1985 onwards by EAGLE (European Association for Grey Literature Exploitation), with an operating agent (KNAW, The British Library, FIZ Karlsruhe) and with input from several Member States of the European Community (Wood & Smith, 1993; Wessels, 1998). Each country was represented by one or more documentation centers or libraries of national importance, and each national center held the national grey collection or at least, guaranteed the document supply of distributed holdings. Until 2005, the SIGLE database was distributed by BLAISE, STN, EI and Ovid (Silverplatter) and in the early nineties by SUNIST in France. Records were also sold to organizations like NERAC for specific use (Schöpfel et al., 2006).

In 2005, the members of EAGLE decided to dissolve the association and to stop the input of new documents. The database SIGLE was taken from all hosts proposing the database online. In the context of the Open Access movement and with the consent and approval of the former EAGLE members, EAGLE's last operating agent (FIZ Karlsruhe) transferred the SIGLE records to the French INIST-CNRS (Institut de l'Information Scientifique et Technique) which then launched the OpenSIGLE project, i.e., all former SIGLE records, in a simplified XML format, were made freely available on a DSpace open repository (Schöpfel, 2005; Schöpfel et al., 2006; Farace et al., 2009, 2010).

In 2009-2010, GreyNet International joined OpenSIGLE, adding the preprints of the GL conference series to the repository. By providing full text access to the conference papers, it offered open access to international research on grey (gray) literature.

In 2011 OpenSIGLE changed its platform and its name and was relaunched as the open repository OpenGrey. OpenGrey provided new features and new content. Thus, all former SIGLE bibliographical records were searchable again, free of charge, with OpenGrey.

Contents

OpenGrey[*] is a multidisciplinary European database, covering science, technology, biomedical science, economics, social science and humanities. The records are in English. They include bibliographical data, the SIGLE classification codes with their wording, and possibly an English abstract (starting in 1997). Each record has either an English title and/or English keywords. Document types include technical or research reports, doctoral dissertations, conference papers, official publications, and other types of grey literature. The OpenDOAR directory entry on the OpenGrey Repository[†] indicates 1,014,842 metadata, giving access to the full text of 18,158 items (data provided by the Jisc CORE service).

[*] Former OpenGrey website: http://www.opengrey.eu/

[†] OpenDOAR https://v2.sherpa.ac.uk/id/repository/2181

Its content in March 2021 shown in figures:

Scientific domain	Records
Sciences	833,500
Social Science and Humanities	325,000
Biomedical Science	116,000
Science	170,000
Technology	222,500

Table 1. Scientific domain of the OpenGrey records (multiple entries)

Language	Records
English	568,494
French	263,215
German	157,833
Italian	15,966
Spanish	9,170
Czech	4,576
Latvian	3,238
Dutch	2,080

Table 2. Most important languages of the OpenGrey records

Document type	Records
Theses and dissertations	534,648
Reports	189,646
Conference papers	20,521
Miscellaneous	244,531

Table 3. Most important document types of the OpenGrey records

Country	Records
United Kingdom	451,896
France	283,689
Germany	199,322
Italy	31,350
Russia	19,536
Spain	11,286
Czech Republic	5,927
Belgium	5,252
Latvia	3,962
Portugal	1,782
Slovakia	851

Table 4. Geographical origin: most important countries of the OpenGrey records

Table	Rows	Count
handle2url	2	30
item	8	942,052
item2subject	2	1,501,299
modif	5	0
news	4	0
partner	5	15
subject	2	274
translation	3	85

Table 5. Information about the 8 database tables

Sample SIGLE Record

Accession Number

98:4462DE SIGLE

Title

Anaerobic co-treatment of biological waste materials and sewage sludge. Final report. Gemeinsame anaerobe Behandlung von Bioabfall und Klaerschlamm. Abschlussbericht.

Author

Risse, H.

Corporate Source

Technische Hochschule Aachen (DE). Lehrstuhl und Inst.fuer Siedlungswasserwirtschaft

Funding Organization

Oswald-Schulze-Stiftung, Gladbeck (DE)

Source

Jan 1998. 83 p.

Availability

Available from TIB Hannover: F98B1076+a.

Document Type

Numerical Data

Country

Germany, Federal Republic of

Language

German

Abstract

The main goals of co-fermentation are sewage sludge stabilisation, production of a cyclable product and minimisation of process emissions. This research project focused on anaerobic wet fermentation. Criteria of evaluation were the quantity and quality of the resulting fermented sludge, biogas, and process water. In view of the scarcity of data on optimum process conditions, the investigations were to provide information, e.g.: Optimum ratio of biological waste to sewage sludge, influence of process conditions (temperature, charge volume, flow rate) on biogas production and process water burden. (orig.)

Classification Code

10U Biomass energy

Terms

BIOLOGICAL WASTES; SEWAGE SLUDGE; ANAEROBIC DIGESTION; METHANE; CHEMICAL OXYGEN DEMAND; AMMONIA; WASTE WATER; PHOSPHORUS; METALS; TEMPERATURE DEPENDENCE; EVALUATION; EXPERIMENTAL DATA.

Figure 1. Sample of a former SIGLE record (source: Wikipedia)[‡]

[‡] Wikipedia https://en.wikipedia.org/wiki/System_for_Information_on_Grey_Literature_in_Europe

Twenty-Third International Conference on Grey Literature

Digital Transformation of Grey Literature: Exploring Next Generation Grey

OBA Forum - Amsterdam, Netherlands
December 6-7, 2021

Conference Announcement

Grey literature is once again at a crossroads in its known 80 years in the vernacular. It appears that every quarter or more century this field of information faces a change in direction.

In the mid-20th Century, it originated with the collection of government documents – namely war reports. In the years succeeding, it expanded to numerous other types of documents and collections produced by organizations in government as well as academics and business. This period was occupied with the acquisition and document delivery of these scientific and technical materials. It was also during this period that the many problems associated with these materials surfaced, namely their lack of indexing, translation, open access, preservation, and assessed value for science and society.

Since the final decade of the 20th Century up to the present, the field of grey literature has initiated and undertaken evidenced based research hand in hand with technological developments and sustained information management. The problems of the prior period became challenges and this led to programs of research and education in grey literature. This period coincides with the digital transformation of grey literature. While much has been accomplished in connecting the supply and demand sides of grey literature, the field once again finds itself in transition driven by the fact that the entire information landscape is itself in a period of unprecedented change and flux.

GL2021 offers the many and diverse communities of practice in grey literature a unique opportunity to collaborate in addressing and defining the next phase in the digital transformation of grey literature. Together this can be accomplished by unlocking the potential next generation grey holds for information science and society.

Conference Topics

- Cite Grey
- Citizen Science
- Communities of Practice
- Digital Grey
- Evidence based Grey
- Funding Grey
- Linking Grey Resources
- Optics of Grey
- Policy development
- Reuse of Data
- Vanished OA Journals
- Transformation of Preprints
- Other topics

GL2021 Dateline

March 25	April 8	April 15	April 20	Oct. 25	Nov. 15	Dec. 6-7
Close Call for Papers	Program Committee Meeting	Author Notification	Call for Posters Open	Close Early Conference Registration	Conference Papers and Posters due	GL2021 Conference

TextRelease
GL2021 Program and Conference Bureau

Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
www.textrelease.com • conference@textrelease.com
Tel.+31-20-331.2420

Twenty-Third International Conference on Grey Literature

Digital Transformation of Grey Literature: Exploring Next Generation Grey

OBA Forum - Amsterdam, Netherlands

December 6-7, 2021

Call for Papers

Title of Paper:	**Conference Topic(s):**
Author Name(s):	**Phone:**
Organization(s):	**Email:**
Postal Address:	**URL:**
Postal Code – City – Country:	

Guidelines for Abstracts

Participants who seek to present a paper dealing with grey literature are invited to submit an English language abstract between 300-400 words. The abstract should address the problem/goal, the research method/procedure, an indication of costs related to the project, as well as the anticipated results of the research. The abstract should likewise include the title of the proposed paper, conference topic(s) most suited to the paper, name(s) of the author(s), and full address information. Abstracts are the only tangible source that allows the Program Committee to guarantee the content and balance in the conference program. Every effort should be made to reflect the content of your work in the abstract submitted. Abstracts not in compliance with the guidelines will be returned to the author for revision.

Related Conference Topics		
❑ Cite Grey	❑ Evidence based Grey	❑ Policy development
❑ Citizen Science	❑ Funding Grey	❑ Reuse of Data
❑ Communities of Practice	❑ Linking Grey Resources	❑ Vanished OA Journals
❑ Digital Grey	❑ Optics of Grey	❑ Preprint Transformation
❑ Other Related Topic:		

Due Date and Format for Submission

Abstracts in MS Word must be emailed to conference@textrelease.com on or before **March 25, 2021**. The author will receive verification upon its receipt. By mid-April, shortly after the Program Committee meets, authors will be notified of their place on the conference program. This notice will be accompanied by further guidelines for submission of full text papers, biographical notes, accompanying research data, PowerPoint slides, and required Author Registration.

TextRelease
GL2021 Program and Conference Bureau

Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
www.textrelease.com • conference@textrelease.com
Tel.+31-20-331.2420

Author Information

Bianchi, Silvia **41**
Silvia Bianchi is a research fellow at the University of Siena - DSFUCI department; former official of the Archival an Bibliographic Superintendence of Tuscany. PhD in archivistic science, she has been working for 16 years as an archivist and archives manager at public and private institutions. She deals with documentary studies, analysis of different archival assets as well as their preservation, protection and safeguard, and also archival communication and archival teaching. She has written various articles and contribution on the sector of interest, moreover she is also a freelance journalist and member of archivists and historians associations.
ORCID_ID https://orcid.org/0000-0002-9879-0615
ROR_ID https://ror.org/01tevnk56

Calamai, Silvia **41**
Silvia Calamai is associate professor in Linguistics and Sociolinguistics at the University of Siena. She is member of the CLARIN Legal Issues Committee. At present, she is coordinating the project on social evaluation of non native accents (Fregati dall'accento!, Fondazione ALSOS 2018-2020); the project on the description and cataloguing of Caterina Bueno's oral archive ("Archivio Vivo", Regione Toscana 2019-2021); the project "Chinese Culture and Languages in Italy" (Wenzhou University & Siena University) and the scientific committee of the Historical Archive of the Arezzo psychiatric hospital (2017-). She is in the board of Italian Association of Speech Sciences and of Sonorités Bulletin de l'AFAS Association française des détenteurs de documents audiovisuels et sonores. Her main research interests are sociophonetics, oral archives and dialectology. Her work has been published in various peer-reviewed journals, refereed conference proceedings, and collective volumes. In addition to over 100 scholarly contributions and two manuals for university students, she has written four books based on phonetic and dialectological fieldwork.
ORCID_ID https://orcid.org/0000-0002-6585-2576
ROR_ID http://ror.org/01tevnk56
Email silvia.calamai@unisi.it

Černohlávková, Petra **13**
Petra graduated with a master's degree in Information science and librarianship at the Charles University in Prague. She is currently working in the National Library of Technology (NTK) in Prague. She is a data curator of the Czech National Repository of Grey Literature and of the Institutional Repository of NTK (both on Invenio SW). She used to work at the Europen University Institute with institutional repository (DSpace). Her research interests are grey literature, all about repositories (new tools, development, etc.), open science, research data, linked data (emphasised on research projects and their outputs) and free licences.
ROR_ID https://ror.org/028txef36
ORCID_ID https://orcid.org/0000-0001-8683-6156
Email petra.cernohlavkova@techlib.cz

Farace, Dominic **45, 49**
Dominic Farace is Head of GreyNet International and Director of TextRelease, an independent information bureau specializing in grey literature and networked information. He holds degrees in sociology from Creighton University (BA) and the University of New Orleans (MA). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked since 1976. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (SWIDOC/KNAW), Farace founded GreyNet, Grey Literature Network Service in 1992. He has since been responsible for the International Conference Series on Grey Literature (1993-2013). In this capacity, he also serves as Program and Conference Director as well as managing editor of the Conference Proceedings. He is editor of The Grey Journal and provides workshops and training in the field of grey literature. Email: info@greynet.org
ORCID_ID https://orcid.org/0000-0003-2561-3631

Frantzen, Jerry **49**
Jerry Frantzen graduated in 1999 from the Amsterdam University of Applied Sciences/Hogeschool van Amsterdam (HvA) in Library and Information Science. Frantzen is the technical editor of The Grey Journal (TGJ). And, since 1996, he is affiliated with GreyNet International, Grey Literature Network Service, as a freelance technical consultant.
ORCID_ID https://orcid.org/0000-0002-3405-7078

Kergosien, Eric **29**
Eric Kergosien, MCU since 2014, is working at the University Lille in the GERiiCO laboratory. His main area are Knowledge management and Text Mining applied to heterogeneous textual data with a special interest in semantic and geographic resources. One aspect of his work is Information Extraction with NLP and Text mining methods in several fields (scientific and technical data, libraries, land-use planning and medical). A second important aspect of his work is domain ontology engineering.
ORCID_ID https://orcid.org/0000-0002-2397-5519
ROR_ID https://ror.org/02kzqn938
Email: eric.kergosien@univ-lille.fr

Marshall, Andrea **25**
Andrea Marshall is a librarian and feminist researcher who has published work in human computer interaction, information science, feminist approaches to maker culture and STEM education and fan studies. Her current work includes gendered analyses of online Star Wars fanfiction communities and zines as grey literature.
ORCID_ID http://orcid.org/0000-0003-1229-4694
Email andrearieth marshall@gmail.com

Monachini, Monica **41**
Monica Monachini is a Senior Researcher at CNR-ILC. Field of expertise: computational linguistics, computational lexicography, semantics, lexical semantics, language resources, ontologies, lexicon, terminologies, metadata, validation, methods for retrieving information in different areas (biology, environment, civil protection, oceanography, social media, humanities and social sciences, ...), infrastructural issues related to language resources. Active in many standardisation activities for harmonising lexical information. Involved and responsible of the Pisa team in many international projects for language engineering. Over the last years, she has published articles in the field of lexical resources and information extraction in different areas. Currently, she focused her activities on digital humanities. Member of various Scientific Committees; UNI delegate for ISO/TC37/SC4. ORCID_ID https://orcid.org/0000-0003-3356-3988
ROR_ID https://ror.org/028g3pe33
Email Monica.Monachini@ilc.cnr.it

Author Information *CONTINUED*

Pretto, Niccolò **41**
Niccolò Pretto is a postdoctoral research fellow at the Institute for Computational Linguistics "A. Zampolli", CNR, Pisa, Italy. He completed his Ph.D. at the Department of Information Engineering, University of Padova. His research is primarily focused on sound and music computing, and preservation and access to historical audio documents and cultural heritage in general. More specifically, his work consists of research and development of innovative methodologies, and applications to preserve, analyze, and experience musical heritage. His work is highly multidisciplinary; during his research he has collaborated with engineers (computer science, telecommunication, civil, and the like), as well as musicologists, linguists, archaeologists, archivists, psychologists, architects, earth scientists, and physicians.
ORCID_ID https://orcid.org/0000-0002-3742-7150
Email niccolo.pretto@ilc.cnr.it

Prost, Hélène **29**
Hélène Prost is information professional at the Institute of Scientific and Technical Information (CNRS) and associate member of the GERiiCO research laboratory (University of Lille 3). She is interested in empirical library and information sciences and statistical data analysis. She participates in research projects on evaluation of collections, document delivery, usage analysis, grey literature and open access, and she is author of several publications.
ORCID iD: https://orcid.org/0000-0002-7982-2765
ROR_ID https://ror.org/02mn0vt57
Email: helene.prost@inist.fr

Schöpfel, Joachim **29, 45, 49**
Joachim Schöpfel is associate professor in information and communication sciences at the University of Lille, member of the GERiiCO laboratory and independant consultant at the Ourouk consulting office, Paris. He is interested in scientific information, academic publishing, open science and grey literature. He is a member of GreyNet, euroCRIS and NDLTD.
ROR_ID https://ror.org/02kzqn938
ORCID_ID https://orcid.org/0000-0002-4000-807X
Email joachim.schopfel@univ-lille.fr

Sheehan, Jerry **5**
As Deputy Director, NLM, Jerry Sheehan shares responsibility with the Director for overall program development, program evaluation, policy formulation, direction and coordination of all Library activities. In addition, he is responsible for the day-to-day operations of the Library and, in the absence of the Director, assumes full responsibility for all functions performed by the National Library of Medicine. Sheehan joined the National Library of Medicine as Assistant Director for Policy Development in September 2006. In this position, he was responsible for monitoring, evaluating and advising NLM officials on a broad range of science, technology, information and health policy issues that affect the Library. He also managed the trans-NIH BioMedical Informatics Coordinating Committee and served as a key link to a variety of communities in the public and private sectors that develop and implement such policies. From September 2015 to January 2017, Mr. Sheehan was on detail to the White House Office of Science and Technology Policy where he served as Assistant Director for Scientific Data and Information and helped advance open science policies across the Federal Government.
ROR_ID https://ror.org/0060t0j89
Email: sheehanjr@nlm.nih.gov

Stamuli, Maria Francesca **41**
Maria Francesca Stamuli, Ph.D. in Philology and Linguistics and MA Librarian, has been working from 2018 as state archivist for the regional office of archives and libraries of Tuscany (the "Soprintendenza Archivistica e Bibliografica della Toscana"). Before she worked for the Museum of the Royal Palace of Naples (1999-2009) and for the National Library of Naples (2010 - 2017). In the 2017, she was one of the young librarians involved in the re-organization of the Library of the Neapolitan Congregation of Girolamini.
Email: mariafrancesca.stamuli@beniculturali.it

Thiault, Florence **29**
Senior lecturer in Library and Information Science at University Rennes 2, laboratory PREFICS. Co-manager of URFIST (regional unit of training in scientific and technical information) of Rennes.
ORCID_ID https://orcid.org/0000-0001-7466-5855
ROR_ID https://ror.org/01m84wm78
Email florence.thiault@univ-rennes2.fr

Vyčítalová, Hana **13**
Hana graduated with a master's degree in Information science and librarianship at the Charles University in Prague. She works in the National Library of Technology in Prague (NTK). Currently, she is leader of the Center of repositories in the NTK and coordinator of the international Conference on Grey Literature and Repositories in the Czech Republic. She is interested in grey literature, digital repositories, open access, research data, enhanced publications and free licences.
ROR_ID https://ror.org/028txef36
Email hana.vycitalova@techlib.cz

Notes for Contributors

Non-Exclusive Rights Agreement

- I/We (the Author/s) hereby provide TextRelease (the Publisher) non-exclusive rights in print, digital, and electronic formats of the manuscript. In so doing,
- I/We allow TextRelease to act on my/our behalf to publish and distribute said work in whole or part provided all republications bear notice of its initial publication.
- I/We hereby state that this manuscript, including any tables, diagrams, or photographs does not infringe existing copyright agreements; and, thus indemnifies TextRelease against any such breach.
- I/We confer these rights without monetary compensation and with the understanding that TextRelease acts on behalf of the author/s.

Submission Requirements

Manuscripts should not exceed 15 double-spaced typed pages. The size of the page can be either A-4 or 8½x11 inches. Allow 4cm or 1½ inch from the top of each page. Provide the title, author(s) and affiliation(s) followed by your abstract, suggested keywords, and a brief biographical note.
A printout or PDF of the full text of your manuscript should be forwarded to the office of TextRelease. A corresponding MS Word file should either accompany the printed copy or be sent as an attachment by email. Both text and graphics are required in black and white.

REFERENCE GUIDELINES

General

i. All manuscripts should contain references
ii. Standardization should be maintained among the references provided
iii. The more complete and accurate a reference, the more guarantee of an article's content and subsequent review.

Specific

iv. Endnotes are preferred and should be numbered
v. Hyperlinks need the accompanying name of resource; a simple URL is not acceptable
vi. If the citation is to a corporate author, the acronym takes precedence
vii. If the document type is known, it should be stated at the close of a citation.
viii. If a citation is revised and refers to an edited and/or abridged work, the original source should also be mentioned.

Examples

Youngen, G.W. (1998), Citation patterns to traditional and electronic preprints in the published literature. - In: College & Research Libraries, 59 (5) Sep 1998, pp. 448-456. - ISSN 0010-0870

Crowe, J., G. Hodge, and D. Redmond (2010), Grey Literature Repositories: Tools for NGOs involved in public health activities in developing countries. – In: Grey Literature in Library and Information Studies, Chapter 13, pp. 199-214. – ISBN 978-3-598-11793-0

DCMI, Dublin Core Metadata Initiative Home Page http://purl.oclc.org/metadata/dublin_core/

Review Process

The Journal Editor first reviews each manuscript submitted. If the content is suited for publication and the submission requirements and guidelines complete, then the manuscript is sent to one or more Associate Editors for further review and comment. If the manuscript was previously published and there is no copyright infringement, then the Journal Editor could direct the manuscript straight away to the Technical Editor.

Journal Publication and Article Deposit

Once the journal article has completed the review process, it is scheduled for publication in The Grey Journal. If the Author indicated on the signed Rights Agreement that a preprint of the article be made available in GreyNet's Archive, then browsing and document delivery are immediately provided. Otherwise, this functionality is only available after the article's formal publication in the journal.

Contents

'EXPLORING THE GREY SIDE OF OPEN SCIENCE'