

ISSN 1574-1796

An International Journal on Grey Literature



Autumn 2019, Volume 15, Number 3

‘ADVANCING OPEN SCIENCE AND GREY LITERATURE’

GreyNet

www.textrelease.com

Grey Literature Network Service

www.greynet.org

The Grey Journal

An International Journal on Grey Literature

COLOPHON

Journal Editor:

Dr. Dominic Farace
GreyNet International,
Grey Literature Network Service
Netherlands
journal@greynet.org

Associate Editors:

Julia Gelfand
University of California, Irvine
United States

Dr. Debbie L. Rabina
School of Information and Library Science
Pratt Institute, United States

Dr. Joachim Schöpfel
Université Charles de Gaulle Lille 3
France

Prof. Gretta E. Siegel (Emeritus)
Portland State University
United States

Dr. Plato L. Smith
University of Florida
George A. Smathers Libraries
United States

Technical Editor:

Jerry Frantzen, TextRelease

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 15, Number 3, Autumn 2019. - 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.
ISSN 1574-1796 (Print)
ISSN 1574-180X (PDF)

Subscription Rates:

€240 institutional, including postage & handling

Contact Address:

Back Issues, Document Delivery, Advertising,
Inserts, Subscriptions:

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

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.



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.

© 2019 TextRelease

Copyright, all rights reserved. No part of this publication may be reproduced, stored in or introduced into a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without prior permission of the publisher.

Contents

'ADVANCING OPEN SCIENCE AND GREY LITERATURE'

OpenAIRE: Advancing Open Science	141
<i>Paolo Manghi, Michele Artini, Claudio Atzori, Miriam Baglioni, Alessia Bardi, Sandro La Bruzzo, and Michele De Bonis; Institute of Information Science and technologies, Italian National Research Council, Italy; Harry Dimitropoulos, Ioannis Foufoulas, Katerina Iatropoulou, Natalia Manola, and Stefania Martziou; Athena Research Center in Information, Communication and Knowledge Technologies, Greece; Pedro Principe, University of Minho, Portugal</i>	
Semantic Query Analysis from the Global Science Gateway	147
<i>Sara Goggi, Gabriella Pardelli, Roberto Bartolini, and Monica Monachini, ILC-CNR, Italy Stefania Biagioni and Carlo Carlesi, ISTI-CNR, Italy</i>	
The U.S. Government Publishing Office: Keeping America Informed in the 21st Century and Beyond ...	157
<i>Cynthia Etkin, U.S. Government Publishing Office, GPO, United States</i>	
Analysis of folk literature in grey literature from the National Library of China	163
<i>Cui Yue, National Library of China, NLC, China</i>	
Data from "Data Papers as a New Form of Knowledge Organization in the Field of Research Data"	170
<i>Joachim Schöpfel, University of Lille, GERiiCO laboratory, France; Dominic Farace, GreyNet International; Hélène Prost, CNRS, GERiiCO laboratory, France; Antonella Zane, University of Padova, Italy</i>	
Data from "AccessGrey, Securing Open Access to Grey Literature for Science and Society"	173
<i>Dominic Farace and Jerry Frantzen, GreyNet International, Netherlands Stefania Biagioni and Carlo Carlesi, Institute of Information Science and Technologies, Italy</i>	

Colophon.....	136
Editor's Note.....	139
Advertisements	
GL2020 Pre-Conference Announcement.....	138
OpenAIRE: Advancing Open Science.....	140
WorldWideScience.org: An International Partnership Supporting Open Science.....	156
GreyGuide: Research and Knowledge Sharing in the Field of Grey Grey Literature.....	175
NTK, National Library of Technology, Czech Republic	176
Journal Subscription Form 2020.....	177
Author Information.....	178
Notes for Contributors.....	181

Pre-Conference Announcement

Twenty-Second International Conference on Grey Literature

'Applications of Grey Literature for Science and Society'



Italian National Research Council
Rome, Italy • November 19-20, 2020



EDITOR'S NOTE

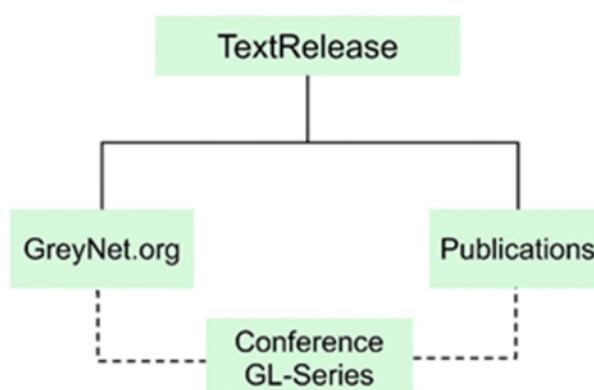
GreyNet International Welcomes Change for Growth in 2020

Since its relaunch in 2003, the Grey Literature Network Service has operated under the legal and fiscal entity of TextRelease. During these more than 15 years, GreyNet has experienced increased and sustained growth in the information resources it provides those both within and outside the field of grey literature. In order to accommodate further growth, enhance its existing resources, and rise to challenges required for innovation, GreyNet powered by TextRelease will outsource its International Conference Series on Grey Literature starting in 2020.

2003 – 2019



2020



GreyNet as an international organization and corporate author to the serial publications under TextRelease will continue with these activities. Likewise, all partnerships, licensed agreements, collaborations with service providers, and contracts undertaken by TextRelease on behalf of GreyNet will remain in effect. The work by GreyNet's Resource Policy Committee, Social Media Publishing Committee, and Library and Information Science Committee will likewise remain intact. GreyNet will further maintain representation on the Conference Program Committee outsourced to new management in 2020.

*Dominic Farace,
Journal Editor*

OpenAIRE

Advancing Open Science

A pillar of the European Open Science Cloud



The Pan-European Open Knowledge Infrastructure

- + OpenAIRE is a **participatory, community-led, and technology-savvy infrastructure** ensuring optimal convergence of Open Science efforts in Europe.
- + OpenAIRE is a **one-stop shop for scholarly communication services** addressing all stakeholders' needs to improve science through openness & transparency.
- + OpenAIRE is the **Europe-wide Research Information System** encompassing all research and enabling ready-made reporting, monitoring & analysis services.
- + OpenAIRE's **content acquisition policy** enables the harvesting of **all global research output** and enriches them with e.g. funding information.
- + OpenAIRE is Europe's **gateway to worldwide Open Science**, embedded in a global system of repository networks.
- + OpenAIRE **accelerates innovation in open scholarly communication** and serves as a **testbed for novel services & models**.
- + OpenAIRE is a **trusted open (research) data hub** for 3rd party providers to **build innovative value added services** for research and society.
- + OpenAIRE builds bridges to **Citizen Scientists** and the public, enabling access and reuse of research to non-professional researchers.

Ilaria Fava (<https://orcid.org/0000-0002-0298-0239>), Najla Rettberg (<https://orcid.org/0000-0003-1888-2294>)
Andreas Czerniak (<https://orcid.org/0000-0003-3883-4169>), Jochen Schirwagen (<https://orcid.org/0000-0002-0458-1004>)
Iryna Kuchma (<https://orcid.org/0000-0002-2064-3439>)

Visit the
OpenAIRE portal
www.openaire.eu

Follow us
on Twitter
[@openaire_eu](https://twitter.com/openaire_eu)

Connect
on Facebook
facebook.com/groups/openaire

Contact
Project Coordinator Mike Hatzopoulos
mike@di.uoa.gr

Göttingen State and University Library
Bielefeld University Library
Electronic Information for Libraries (EIL)

Connect on LinkedIn
linkedin.com/groups/OpenAIRE-3893548


OpenAIRE


European
Commission

OpenAIRE: Advancing Open Science*

Paolo Manghi, Michele Artini, Claudio Atzori, Miriam Baglioni, Alessia Bardi,
Sandro La Bruzzo, and Michele De Bonis, Institute of Information Science and technologies,
Italian National Research Council, Italy

Harry Dimitropoulos, Ioannis Foufoulas, Katerina Iatropoulou, Natalia Manola, Stefania Martziou, Athena
Research Center in Information, Communication and Knowledge Technologies, Greece

Pedro Principe, University of Minho, Portugal

Abstract

OpenAIRE¹, the point of reference for Open Access in Europe, is now addressing the problem of enabling the Open Science paradigm. To this aim it will provide services to: (i) overcome the limits of today's scientific communication landscape, by allowing research communities and the relative e-infrastructures to fully publish, interlink, package and reuse their research artefacts (e.g. literature, data, and software) and their funding grants within the European and global ecosystem as supported/promoted by OpenAIRE, (ii) enable end-users (e.g. researchers, funder officers) to search and consult a rich and up-to-date knowledge graph of research results and (iii) enable scientific and educational information repositories and publishers to subscribe and be notified of changes in the OpenAIRE knowledge graph. These combined actions will bring long-term and immediate benefits to research communities, research organisations, repository managers, and funders by affecting the way research results are disseminated and reused. On the one hand, publishing the interlinked and packaged research literature, data and software via OpenAIRE drives research communities to an Open Science transition in a consistent and interoperable fashion. On the other hand, the resulting infrastructure concretely enables the construction of Open Science oriented services, supporting practices such as machine-assisted research reproducibility and evaluation.

1. Introduction

Open Science is frequently defined as an umbrella term that involves various movements aiming to remove the barriers for sharing any kind of output, resources, methods or tools, at any stage of the research process 9. This is intended as a means for accelerating research by enhancing transparency and collaboration, and fostering innovation and reproducibility. Scientists and organizations see Open Science as a way to speed up, improve quality, and more effectively reward research activities, while funders and ministries see it as a means to optimize cost of science and leverage innovation. Open Science is an emerging vision, a way of thinking, whose challenges always gaze beyond its actual achievements. Today, the effective implementation of Open Science calls for a scientific communication ecosystem capable of enabling Open Science publishing principles. The ecosystem should allow research communities to share (for discovery and transparent evaluation) and re-use (for reproducibility 213) their scientific results by publishing all intermediary and final research artefacts, beyond scientific literature. Artefacts can be research data, software and research methods (e.g. workflows, protocols, algorithms, etc.), which should be deposited in repositories for scientific communication (e.g. institutional repositories, data archives, software repositories, CRIS systems), and should be published together with the semantic links between them. To complete the picture, such ecosystem should support publishing of packages of artefacts (e.g. research objects 214, enhanced publications 515, RMap 16) to allow discovery, evaluation, and reproducibility of science (e.g. workflows or experiments with input datasets).

Today's scientific communication landscape is far from supporting this vision, mainly due to its inability to:

1. *Support publishing of all kinds of research artefacts.* For example, research methods publishing workflows are generally not best practice, i.e. no research method repositories, no persistent identifiers for methods, no citation practices and, therefore, no scientific reward;
2. *Keep a complete and up-to-date record of research artefacts relationships.* For example, publication, data, software repositories and publishers do not keep bi-lateral links between each other's artefacts, and the links they keep are not in-sync with the updates of the artefacts (e.g. links to new versions of the data, obsolete links);
3. *Find agreements on how to share and publish packages of artefacts.* Solutions exists (e.g. research objects, enhanced publications, RMap) but are specific to rather small communities of scientists, implying that research packages, as well as research methods, are not regarded as first-class citizens in the scientific communication domain.

* First published in the GL19 Conference Proceedings, 2018.

¹ OpenAIRE www.openaire.eu

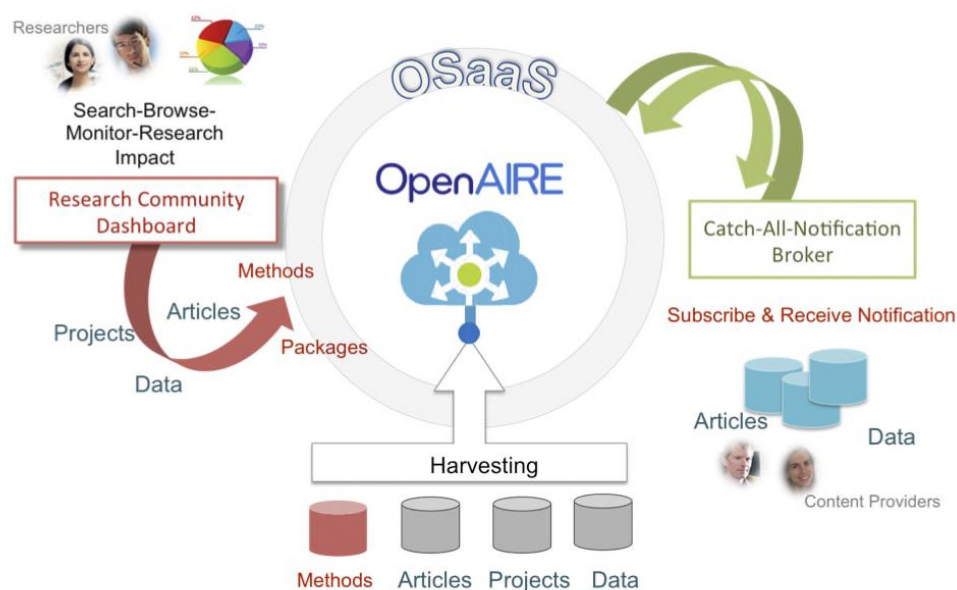
De facto, today's scientific communication ecosystem lacks tools and practices for engaging research communities at adopting the aforementioned novel Open Science publishing principles, even when researchers are already in the position of publishing interlinked artefacts and/or packages.

OpenAIRE fosters transparent evaluation of results and facilitates reproducibility of science for research communities by enabling a scientific communication ecosystem where artefacts, packages of artefacts, and links between them can be exchanged across communities and across content providers. To this aim, OpenAIRE, via the OpenAIRE-Connect project, introduces and implements the concept of Open Science as a Service (OSaaS) on top of the existing OpenAIRE infrastructure¹, by delivering services in support of Open Science. Following the NIST definition of *aaS* service models², the service model "Open Science as a Service" provides the consumers with the capability of accessing tools that implement Open Science principles, transparently with respect to the underlying technical infrastructure. Tools are accessible through either a thin client interface, such as a web browser, or an application program interface (API).

OpenAIRE-Connect³ will realize and operate two services for Open Science. The first, Research Community Dashboard, will serve research communities to publish research artefacts, packages, and links, and to monitor their research impact. The second, Catch-All Notification Broker Service, will engage and mobilize content providers, and serve them with services enabling notification-based exchange of research artefacts, to leverage their transition towards the Open Science paradigm. Both services will be served on-demand according to the OSaaS paradigm, hence be re-usable by different disciplines and providers, each with different practices and maturity levels, so as to favor a shift towards a uniform cross-community and cross-content provider scientific communication ecosystem.

The adoption of these services, eased by the OSaaS approach, aims at incepting Open Science publishing within the existing scholarly communication landscape. By introducing an OSaaS approach, OpenAIRE-Connect will deliver on-demand Open Science publishing services to research communities and content providers, aligning practices and mechanisms that address transparent evaluation and reproducibility (see Figure 1). By complementing the technological efforts with networking activities that will strengthen the emerging Open Science social environment, OpenAIRE-Connect will facilitate a cultural and technological shift towards common Open Science publishing practices.

To achieve its objectives, OpenAIRE-Connect involves key stakeholders of scientific communication: a pool of forward-looking research communities, today publishing or in the need of publishing research data and methods, international representatives of Open Access publishers (Jisc 3), publication repositories (COAR 9), and data archives (ICSU World Data Systems/WDS 11), willing to support and benefit from such a change.



Figure

1. Research Community Dashboard and Catch-All Notification Broker Service

² Peter Mell and Timothy Grance (September 2011). The NIST Definition of Cloud Computing (Technical report). National Institute of Standards and Technology: U.S. Department of Commerce. Special publication 800-145. <http://doi.org/10.6028/NIST.SP.800-145>.

³ OpenAIRE-Connect <https://www.openaire.eu/connect>

2. OpenAIRE services in support of Open Science

OpenAIRE-Connect extends the technological services today offered by the OpenAIRE infrastructure in order to foster the adoption of Open Science publishing practices and facilitate the emergence of shared solutions. Specifically, OpenAIRE-Connect introduces two classes of new services:

1. *Research community services* offering support for a uniform transition of research communities towards Open Science publishing via the Research Community Dashboard;
2. *Content provider services* leveraging the transition of content providers towards Open Science publishing via the Catch-All Notification Broker Service.

Continuing and building on OpenAIRE's openness and sharing of content, services and practices, OpenAIRE-Connect will develop a *uniform, common strategy* for approaching and engaging with research communities (especially Research Infrastructures targeting ESFRI 8 programmes), which will be a major outcome of the project. This strategy will be used in due course by the wider OpenAIRE constituency, i.e. the National Open Access Desks (NOADs), in its strategic synergies within the emerging European Open Science Cloud (EOSC 12) ecosystem to increase awareness on the Open Science topics and to promote the adoption and uptake of the new services.

In synergy with EOSC, towards the widely discussed need of the gluing social aspect (support and human infrastructure), OpenAIRE-Connect aims to design and deliver a *targeted support and training programme for research communities* and relevant stakeholders. This will inform them about the benefits and use of the OpenAIRE-Connect services, to pass on best practices and to lower the barriers of participation in the Open Science ecosystem and, particularly, in the OpenAIRE infrastructure.

2.1. Research Community Dashboard

OpenAIRE-Connect will support the evolution of today's fragmented scientific communication landscape by providing researchers of specific communities with services giving access to facilities for collaboratively maintain an up-to-date knowledge graph of their interlinked or packaged research artefacts, e.g. literature, data, software and methods. This research community graph will be built as an extension of the broader knowledge graph today populated by the OpenAIRE infrastructure by adding software and other research artefacts (all the artefacts that are different from literature, dataset, and software). It will therefore integrate and inherit links to funders, projects, literature and datasets as inferred from article full-texts (today more than 4 millions) or harvested by OpenAIRE from content providers (today more than 1000). These facilities will be provided by a new infrastructural service, the OpenAIRE Research Community Dashboard, that each community will be able to request and configure according to its specific needs. Each dashboard will serve two types of community users (researchers and research community operators, who have an "administrative role" for the community and can configure the dashboard) with a suite of value added functionalities:

- *Connect and Link*: researchers authoritatively provide and curate links between artefacts related to their scientific community, a process moderated by the research community operators.
- *Deposit*: researchers who have not a repository of reference can deposit in OpenAIRE's Zenodo 67 files and metadata relative to their research literature, data, software, methods, and packages, and obtain a DOI.
- *Enrich Content* (configurable inference): research community operators authoritatively tune the configuration of OpenAIRE text mining algorithms with community specific rules to identify artefacts relevant for the research community;
- *Learn your Impact*: the research community can reliably monitor and report the research impact of their scientific production w.r.t. several European (and beyond) funders, visualize trends, classifications resulting from OpenAIRE's analytics services; the knowledge graph inclusive of all community artefacts, with links between them and relative attribution of work, enables transparent evaluation of science.
- *Discovery and reuse*: researchers can restrict their search, browse, and navigation focus to the subpart of the OpenAIRE information space associated to the community; discovering and accessing packages of artefacts fosters reproducibility of science.

OpenAIRE-Connect will also develop, in collaboration with other international initiatives, interoperability guidelines to enable the exchange of research artefacts and packages and will offer APIs enabling third-party services to bulk-feed research artefacts into the OpenAIRE knowledge graph according to the established guidelines.

OpenAIRE-Connect involves its onset researchers from a wide range of communities looking into pragmatic solutions to research data and methods publishing in Open Science settings. With different levels of maturity, touching upon interdisciplinary, they will deliver end-user requirements for the realization of the *Dashboard* service, and engage in pilots to test, assess, and adopt the services:

- *Earth and Environmental Sciences (UniHB)*: the ATLAS⁴ community relies on thematic data archives (Pangaea) whose datasets are packages of datasets and related literature, aiming to link to different scientific domains.
- *Neuroinformatics (CNRS)*: the France Life Imaging national infrastructure⁵ produces data images, links them with methods (software and services), and produces packages.
- *Fisheries and aquaculture management (CNR, IDR)*: the BlueBridge⁶ and MARBEC⁷ infrastructures are moving towards collaborative editing of “dynamic publications”, looking for Open Science solutions.
- *Humanities and Cultural Heritage (PIN)*: PARTHENOS⁸, a cluster of research infrastructures from Linguistics, Humanities, Cultural Heritage, History, Archaeology, with different types of data to interconnect.
- *Environment & Economics (ICRE8)*: the national/EU node of the United Nations Sustainable Development Solutions Network⁹ sets out to build an infrastructure to gather all publications and data available in repositories and in Public Sector Information portals, and link them to European and national funding.

2.2. Catch-All Notification Broker Service

Research artefacts repositories, a.k.a. content providers, (e.g. institutional and thematic repositories, aggregators, data archives) are key in serving research communities towards their Open Science goals. To ease the adoption of Open Science principles among researchers it is important to lower the barriers to Open Science publishing.

The “publish/deposit once” practice is a decisive means in the overall Open Science roadmap, and will only be achieved when content providers seamlessly connect to the wider open scientific communication ecosystem. This will allow them to pro-actively exchange information about research artefacts and links of value to all interested communities or stakeholders, without the researchers having to worry *where*, *when*, and *how* to publish in order to fulfill the numerous mandates.

As part of the OSaaS portfolio, OpenAIRE-Connect will develop and deploy a Catch-All Notification Broker Service that connects all types of content providers (institutional repositories, publishers, data repositories, and CRIS systems). The Catch-All Notification Broker will extend OpenAIRE’s notification brokering service 4, which serves literature repositories, and will broaden the content provider base with the ones that serve specific research communities.

Thanks to its functionality, providers can be notified of metadata records relative to artefacts that are “of interest to them” (i.e. metadata records that should be in the content provider’s data base), or “linked to them” (i.e. a scholarly link exists between one of the provider’s artefact and the identified artefact).

Notifications are sent only to subscribed providers, following a subscription and notification pattern, and can be delivered by mail, OAI-PMH interfaces, or, currently under investigation, via push APIs (e.g. SWORD protocol), FTP and ResourceSync.

This will effectively allow content providers to complete or enrich their collection of artefacts with up-to-date information from the wider OpenAIRE ecosystem, and research communities or infrastructures to have a direct communication line with content providers via OpenAIRE.

The idea behind the service is to disseminate and advocate the principle that scholarly communication data sources are not a passive component of the scholarly communication ecosystem, but rather active and interactive part of it. They should not consider themselves as thematic silos of products, but rather as hubs of products semantically interlinked with any kinds of research artefacts and, more broadly, up-to-date with the evolving research ecosystem.

OpenAIRE-Connect brings on board leading representatives of institutional repositories (e.g. COAR), data repositories (UniHB/Pangaea, UniHB/WDS), publishers (via Jisc) that are already moving towards Open Science-oriented publishing and are committed to provide requirements and engage in the experimentation of the brokering services. Beyond the ones on board to the project, i.e. Pangaea (UniHB) and Zenodo (CERN), a number of content providers have already indicated their interest in the Catch-All Notification Broker Service and are ready to engage in a number of pilots for testing and adopting the service: the German GESIS Datorium repository for datasets and scripts in the Social

⁴ ATLAS: <https://www.eu-atlas.org/>

⁵ France Life Imaging: <https://www.francelifeimaging.fr/>

⁶ BlueBridge: <http://www.bluebridge-vres.eu/>

⁷ MARBEC: <http://www.umr-marbec.fr/en/?lang=en>

⁸ PARTHENOS: <http://www.parthenos-project.eu/>

⁹ United Nations Sustainable Development Solutions Network: <http://unsdsn.org/>

Sciences, the FCT-FCCN network of Portuguese institutional repositories, Australian ANDS data archive, and Open Access publishers, such as eLife, Frontiers, EuropePMC.

In addition, OpenAIRE-Connect will implement a pilot for exchanging subscriptions and notifications between the Catch-All Broker Service and the Jisc UK Notification Router, in order to serve each other's "customers" with a wider range of subscriptions and notifications (see Figure 2). This cooperation goes in the direction of a scientific communication ecosystem where brokers conforming to common requirements for the exchange of subscriptions and notifications can interoperate to collaboratively (by granting or delegating subscriptions to the network) channel the information they collect from producers of events (OpenAIRE for the Catch-All Broker Service and publishers for the Jisc Notification Router) to interested consumers (e.g. any content provider in OpenAIRE and UK repositories for Jisc Notification Router).

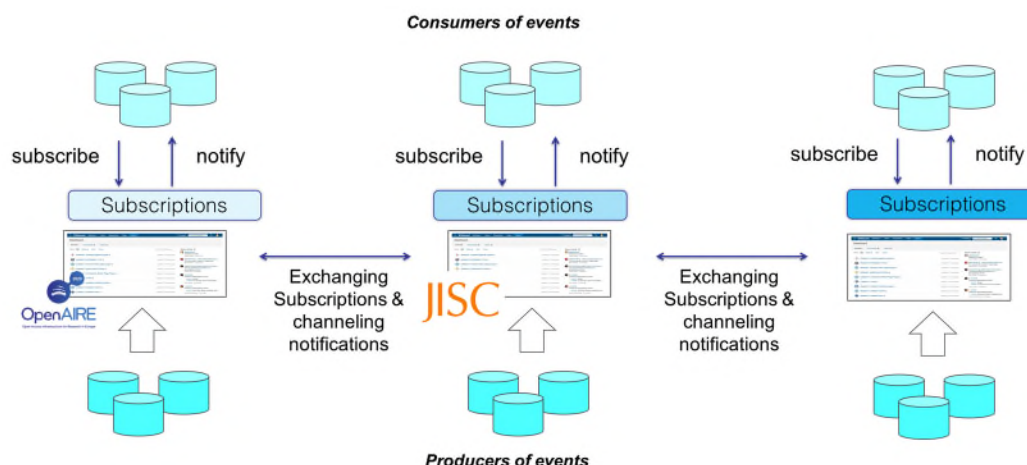


Figure 2. Interoperability between the Broker services

3. Conclusions

The effective implementation of Open Science calls for a scientific communication ecosystem capable of enabling the "Open Science publishing principles" of transparency and reproducibility. Such ecosystem should provide tools, policies, and trust needed by scientists for sharing and interlinking (for "discovery" and "transparent evaluation") and re-using (for "reproducibility") all research artefacts produced during the scientific process, e.g. literature, research data, methods, software, workflows, protocols.

OpenAIRE fosters Open Science by advocating its publishing principles across Europe and research communities and by offering technical services in support of Open Access monitoring, research impact monitoring, and Open Science publishing. Its aim is to provide Research Infrastructures (RIs) with the services required to bridge the research life-cycle they support (where scientists produce research artefacts) with the scholarly communication infrastructure (where scientists publish research artefacts) in such a way science is reusable, reproducible, and transparently assessable. OpenAIRE is fostering the establishment of reliable, trusted, and long lasting RIs by compensating the lack of Open Science publishing solutions and by providing the support required by RIs to upgrade existing solutions to meet Open Science publishing needs (e.g. technical guidelines, best practices, Open Access mandates). To this aim, OpenAIRE is working closely with existing RIs to extend its portfolio by implementing the concept of "Open Science as a Service" (OSaaS) and offer two new services: the Research Community Dashboard and the Catch-All Notification Broker Service.

The first beta release of the services is planned on March 2018. A set of testing sessions will be conducted by five research communities (for the Research Community Dashboard) and a number of content providers (for the Catch-All Notification Broker Service) before the first production public release, expected in June 2019.

Acknowledgments

This research was supported by EU funded project OpenAIRE-Connect (grant agreement: 731011; Call: H2020-EINFRA-2016-1). We thank our colleague Stefania Biagioni for her help during the writing of this paper.

References

1. Manghi, P., Bolikowski, L., Manold, N., Schirrwagen, J., & Smith, T. (2012). OpenairePlus: the european scholarly communication data infrastructure. *D-Lib Magazine*, 18(9), 1. <http://doi.org/10.1045/september2012-manghi>
2. Bechhofer, Sean, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch et al. "Why linked data is not enough for scientists." *Future Generation Computer Systems* 29, no. 2 (2013): 599-611. <https://doi.org/10.1016/j.future.2011.08.004>
3. Jisc Publication Router <http://www.jisc.ac.uk>
4. Artini, Michele, Claudio Atzori, Alessia Bardi, Sandro La Bruzzo, Paolo Manghi, and Andrea Mannocci. "The OpenAIRE literature broker service for institutional repositories." *D-Lib Magazine* 21, no. 11/12 (2015). <http://doi.org/10.1045/november2015-artini>
5. Hoogerwerf, Maarten, Mathias Lösch, Jochen Schirrwagen, Sarah Callaghan, Paolo Manghi, Katerina Iatropoulou, Dimitra Keramida, and Najla Rettberg. "Linking data and publications: towards a cross-disciplinary approach." *International Journal of Digital Curation* 8, no. 1 (2013). <http://doi.org/10.2218/ijdc.v8i1.257>
6. Zenodo. <https://zenodo.org>
7. Potter, M., & Nielsen, L. H. (2015, September 1). Zenodo Information Architecture and Usability. <http://doi.org/10.5281/zenodo.31870>
8. ESFRI https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
9. COAR repository <https://www.coar-repositories.org>
10. FOSTER <https://www.fosteropenscience.eu/node/1420>
11. WDS <https://www.icsu-wds.org>
12. EOSC <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>
13. Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean?. *Science translational medicine*, 8(341). <http://doi.org/10.1126/scitranslmed.aaf5027>
14. Research Object <http://www.researchobject.org>
15. Bardi, A., & Manghi, P. (2014). Enhanced Publications: Data Models and Information Systems. *LIBER Quarterly*, 23(4), 240–273. <http://doi.org/10.18352/lq.8445>
16. RMap. <http://rmap-project.info>

Semantic Query Analysis from the Global Science Gateway*

Sara Goggi, Gabriella Pardelli, Roberto Bartolini, and Monica Monachini, ILC-CNR, Italy
Stefania Biagioni and Carlo Carlesi, ISTI-CNR, Italy

1. Introduction

Nowadays web portals play an essential role in searching and retrieving information in the several fields of knowledge: they are ever more technologically advanced and designed for supporting the storage of a huge amount of information in natural language originating from the queries launched by users worldwide.

Given this scenario, we focused on building a corpus constituted by the query logs registered by the GreyGuide: *Repository and Portal to Good Practices and Resources in Grey Literature*¹ and received by the *WorldWideScience.org*² (*The Global Science Gateway*) portal: the aim is to retrieve information related to social media which as of today represent a considerable source of data more and more widely used for research ends.

The following quotation by Bronson gives a good description of the *WorldWideScience* search engine:

The database is available at <<http://worldwidescience.org/>>. It is based on a similar gateway, Science.gov, which is the major path to U.S. government science information, as it pulls together Web-based resources from various agencies. The information in the database is intended to be of high quality and authority, as well as the most current available from the participating countries in the Alliance, so users will find that the results will be more refined than those from a general search of Google. It covers the fields of medicine, agriculture, the environment, and energy, as well as basic sciences. Most of the information may be obtained free of charge (the database itself may be used free of charge) and is considered "open domain." As of this writing, there are about 60 countries participating in WorldWideScience.org, providing access to 50+databases and information portals. Not all content is in English. (Bronson, 2009)

While the World Wide Web keeps on growing, the development of ever more sophisticated search tools within the universe of public and private infrastructures allows to optimize the users' approach to technology: a new generation of web users – such as the so-called "millennials" – is exponentially connected for getting access and share information on social networks.

In 2005 Bettelle states: "Most searchers –and linguists may be no exception – are instead incredibly lazy, generally typing in a few words and expecting the engine to bring back perfect results, ignoring that it is only the act of offering more data in the query that often dramatically improves the results" (Battelle 2005: 23-25). Nowadays it is possible to retrieve knowledge from the web also by means of query logs, linguistic elements which are useful for monitoring a wide range of information.

This Corpus – called GSGCorpus (Global Science Gateway Corpus) – has been processed with Natural Language Processing (NLP) tools: it talks the *web language*, made up of terms originating from the most various domains and styles. The analysis mainly concentrates on the semantics of the queries received from the portal clients: it is a process of information retrieval from a rich digital catalogue whose language is dynamic, is evolving and follows – as well as reflects – the cultural changes of our modern society.

* First published in the GL20 Conference Proceedings, February 2019.

¹<http://greyguide.isti.cnr.it/> GreyGuide is the online forum and repository of good practices and resources in Grey Literature. It was created - and is now edited - by GreyNet International (content provider) and ISTI-CNR, Pisa Italy (service provider): its launch was in December 2013 and since then GreyGuide provides a unique resource in the field of grey literature, which was long awaited and responds to the information needs of a diverse, international grey literature community. GreyNet International is one of the WorldWideScience Associate Members <https://worldwidescience.org/alliancemembers.html>.

² <https://worldwidescience.org/>

It is a global science gateway comprised of national and international scientific databases and portals. *WorldWideScience.org* accelerates scientific discovery and progress by providing one-stop searching of databases from around the world. WorldWideScience.org is maintained by the U.S. Department of Energy's Office of Scientific and Technical Information as the Operating Agent for the WorldWideScience Alliance.

2. Methods and Tools

This project includes eight months of query logs³ registered between July 2017 and February 2018 for a total of 445,827 queries.

The preliminary phase has essentially dealt with the huge amount of non-relevant information, the so-called *noise* which had to be filtered and eliminated.

Therefore, in order to analyze the available information a considerable pre-processing on four levels has been carried out:

- at the first level, the set of queries has been cleaned: duplicates, alphanumeric strings, strange graphical forms, IP addresses, etc. have been eliminated;
- at the second level, filters have been added and alphabetical order inserted for having a first picture of the contents of these queries;
- the third step consisted of several trials for choosing the focus;
- lastly, natural language processing (NLP) tools have been applied for processing the information and building the sample.

Since the corpus is made up of queries collected in only eight months and the cleaning process reduced them consistently, as a result the final is relatively small. In addition, only the queries in English have been registered while those in other languages have been eliminated (there are a few in French, Spanish, Italian, Portuguese, Polish, Albanese, Galician, Corsican, and so on).

Coming to the NLP analysis, the software team has decided to follow these two steps:

1. free information extraction: it measured the frequency of all the words contained in the corpus. This preliminary investigation provided us with the whole scenario of the lexical variety of the queries and allowed us to focus on a set of terms from which we built a micro-ontology with meaningful terms relating to the queries launched on the portal;
2. ontology-based extraction: the extraction has been performed again using this micro-ontology which has been essentially used for enriching the domain. In this way, the search engine retrieved each single occurrence of those terms (monograms, bigrams, trigrams) which can be found starting from the ontology.

In the following paragraph, it is described the process of information retrieval from a rich digital catalogue of queries.

2.1 NLP Analysis

The free information extraction from the GSGCorpus measures the frequency of the words contained in the corpus; examines the lexical variety of the queries and finally focuses on a set of terms to build a micro-ontology. This extraction was preceded by the cleaning process already described above and of which some examples are listed here:

- Graphical variants:
 - (*micro-fluidic* "micro fluidic*" microfluidic*)
 - (*trypodes basapyrazilique pdf, trypodes Base pyrazilique PDF*)
 - (*Adam Smith, Adams, SM.*)
- Queries in languages other than English:
 - (*alimentos proteicos, Alteracion proteica, Entrainement isometrique, 1. Peste-des-petits-ruminants virus fusion protein F) gene, complete cds, Trypanosoma cruzi: contribution Ã l'identification de substances chimiques et naturelles ayant une activitÃ© trypanocide, alfa1 fetoproteina, hrabanus maurus lehrer, abt und bischof, anabolismo proteina, AND Tanztheater Pina Bausch: Spiegel Gesellschaft, poliì tico, Espanìfa*)
- Duplicates:
 - (*spinal injury) electrical nerve electrical stimulation*)
 - (*spinal injury) electrical nerve electrical stimulation*)
- Alphanumeric strings and IP addresses:
 - (*AND colegio MÃ©xico hazaÃ±a*)
 - (*TÃ©cnicas*)
 - (*http://www.repositoriodgb.buap.mx:2095/*)
 - (*http://www.scielo.org.mx/pdf/tca/v6n3/v6n3a2.pdf*)

³ The General Query Log is the record of each SQL statement received from clients, in addition to their connection and disconnection time.

- <http://www.sciencedirect.com/science>
 - <http://www.sciencedirect.com/science/article/pii/S0019850199001133>
 - <http://www.sciencedirect.com/science/article/pii/S0048712002732908>
 - <http://www.sciencedirect.com/science/article/pii/S0176161712001848>
 - <http://www.sciencedirect.com/science/article/pii/S0211563811001684>
 - <http://www.sciencedirect.com/science/article/pii/S0308814614003628>
 - <http://www.sciencedirect.com/science/article/pii/S1632347500719722>
 - http://www.xvideos.com/video27283249/jynx_maze_anal_banged_on_bangbros_chongas_in_1080p_ch13211_
 - <https://AsPredicted.org/wsxx7.pdf>
 - <https://cirworld.com/index.php/jssr/article/view/3380>
 - <https://doaj.org/article/c4a86ed60b7a4961baf52e2b45951d65>
- Disambiguation:
- AND *terapeuta errores Jaquelin cortez*
AND *terapeuta errores jaquelin Fortes*
 - *sandra massoini*
sandra massoni
 - *bio pelicula [spanish]*
bio peliculua [termine inesistente]
 - *social media reslut [termine inesistente]*
 - *facebook adverstis [termine inesistente]*
 - *social nedia marketing [termine inesistente]*
- “Empty” words like articles and prepositions.

As for the most frequent words, the extraction from the GSGCorpus has provided:

- I. the decreasing frequency of monograms like nouns, adjectives and adverbs, that is the number of occurrences of each word;
- II. the decreasing frequency of bigrams and trigrams and the number of occurrences of each of them in the corpus.

On the basis of their frequency, monograms have been divided in three areas depending on their frequency: high, medium and low. In the highest there are a very few words, while in the lowest there are many but with an irrelevant number of occurrences and the presence of *hapax legomena*⁴. Table 1 and Figure 1 show the twelve most recurrent monograms and how the adjective <social> ranks first with 2412 occurrences.

TERMS	# OCCURRENCES
SOCIAL	2412
MARKETING	1714
MANAGEMENT	1682
EFFECT	1578
MEXICO	1160
SCIENCE	1135
EDUCATION	1122
BUSINESS	1054
CHILD	994
PSYCHOLOGY	916
HEALTH	910
RESEARCH	906

Table 1– Twelve most recurrent monograms

As shown in Figure 1, the most relevant nouns in the corpus are: <business>, <child>, <education>, <effect>, <health>, <management>, <marketing>, <Mexico>, <psychology>, <research>, <science>. The only adjective retrieved is <social>.

⁴ In *corpus linguistics*, a *hapax legomenon* (*/ˈhæpəks lɪˈɡɒmɪnɒn/* also */ˈhæpəks/* or */ˈheɪpəks/*; [1][2] pl. *hapax legomena*; sometimes abbreviated to *hapax*) is a [word](#) that occurs only once within a context, either in the written record of an entire language, in the works of an author, or in a single text.

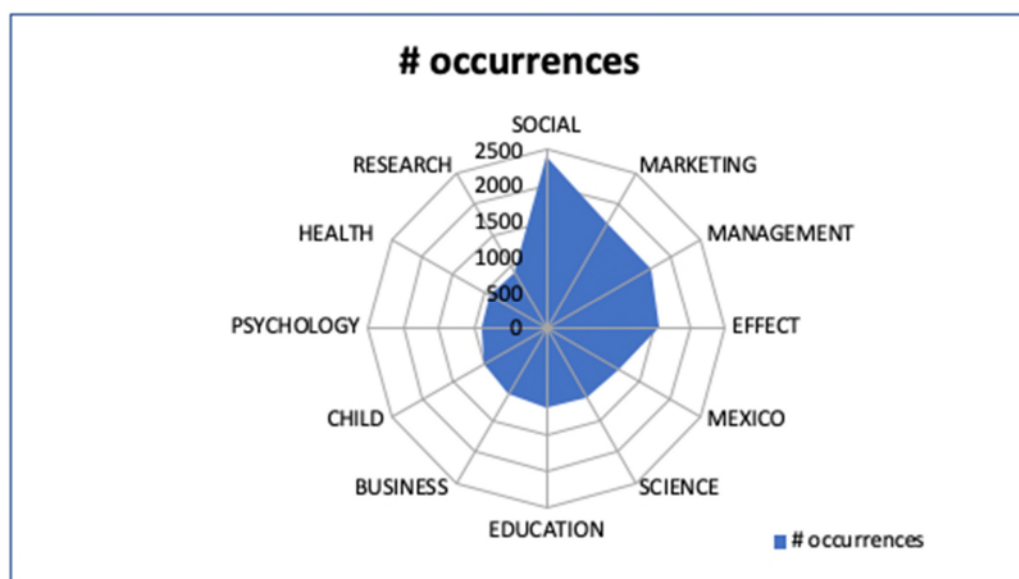


Figure 1 – Most frequent words

Figure 2 shows the bigrams with the higher frequency in the Corpus.

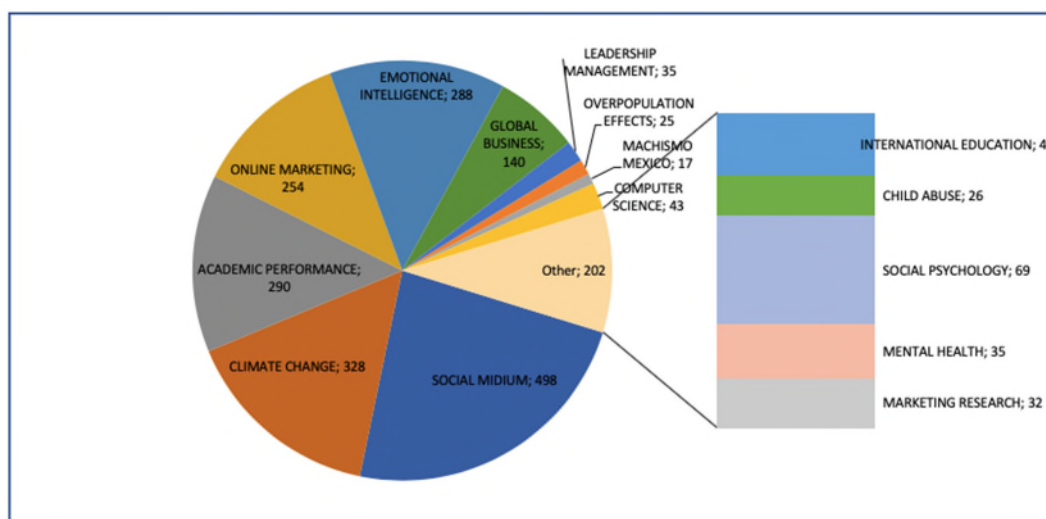


Figure 2 – Most frequent bigrams

3. A case study

At the end of this pre-processing phase, we chose to focus on a flow of queries launched on the *WorldWideScience* platform concerning only the bigram *social media*. A case study has been carried out involving medicine, psychiatry and 'social media'.

Why social media?

- ✓ nowadays social media are obviously a very effective means of communication but can even vehiculate knowledge as their various types (eg.: blogs, YouTube, Facebook, Twitter, etc.) are by now often quoted in bibliographical references amongst the more traditional categories (books, journals and so on);
- ✓ the subject involves document types pertaining to Grey Literature.

NLP analysis allowed to browse the corpus through the most and less queried terms: once *social* has been identified as the most frequent one, the analysis was channeled into '*social media*' and the pertinent contexts.

Ontology-based extraction: enrich the domain; retrieve each occurrence of those terms contained in the ontology by using a search engine.

In particular:

- ✓ Some low-frequency terms (hapax) carry a negative connotation⁴, similarity and diversity. in relation to the use of 'social media';
- ✓ An analysis of negative connotations in connection with child/children, is further investigated.

3.1 Results

The topics concerning social media selected from the corpus of queries (Figure 3):

- ❖ The terms of the queries which reveal who are the subjects involved:
 - *The celebrities* [The term "celebrities" connote people according to the fame and public attention accorded by the mass media]
 - *The influencers* [The "influencers" can be defined as internet users who have established a relevant number of virtual relationships]
 - *The Millennials* [The Millennials - or generation Y - are those born in full digital revolution]
 - *The students*
 - *The teenagers*
- ❖ The terms of the queries which follow the expression <social media **cause**> with a negative polarity:
 - *anxiety*
 - *depression*
 - *dietary diseases*
 - *disadvantages*
 - *distraction*
 - *insecurity*
 - *sleep deprivation*
 - *teenager becomes antisocial*
 - *teenager neglect real world interaction*
- ❖ The terms of the queries which follow the expression <social media **impact on**>:
 - *anxiety*
 - *democracy*
 - *families*
 - *hospitality industry*
 - *maintaining relationships with others*
 - *sales*
 - *society*
 - *teens' lives*
- ❖ The terms of the queries which follow the expression <social media **Social media influence**>:
 - *brand loyalty*
 - *consumers purchase decision*
 - *criminality*
 - *fashion trends*
 - *teenagers' body image*
- ❖ The terms of the queries which follow the expression <social media **help**> with a positive polarity:
 - *business growth*
 - *maintain relationships with people*
 - *young people stay connected to distant people*
- ❖ The terms of the queries involving the words<child/children>: bigrams and trigrams with a negative polarity, see Figure 3:
 - *child abuse*
 - *child labor Indonesia*
 - *Child Marriage*
 - *Child psychiatry*
 - *depression child*

⁴ "In [linguistics](#), a **polarity** is a [lexical item](#) that can appear only in environments associated with a particular [grammatical polarity](#) – affirmative or negative. A polarity item that appears in [affirmative](#) (positive) contexts is called a **positive polarity item** (PPI), and one that appears in negative contexts is a **negative polarity item** (NPI)", Wikipedia.

- domestic violence children
- domestic violence children behavior
- internet safety children
- Obesity child
- obesity children
- Psychopatic assassin children
- punishment children
- The Evil Child Marriage
- violent video games child

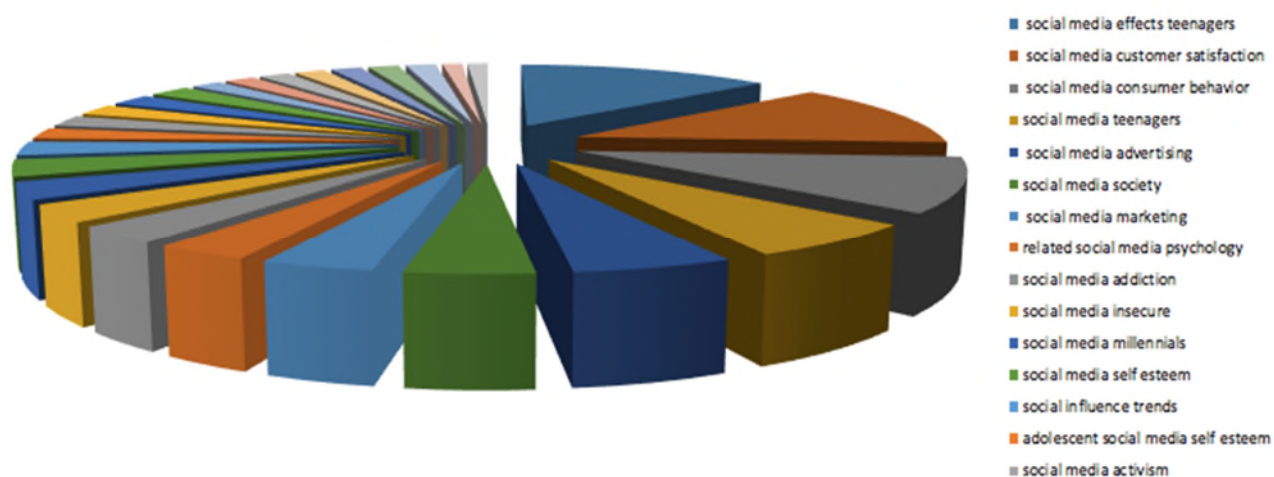


Figure 3 'Social media' occurrences

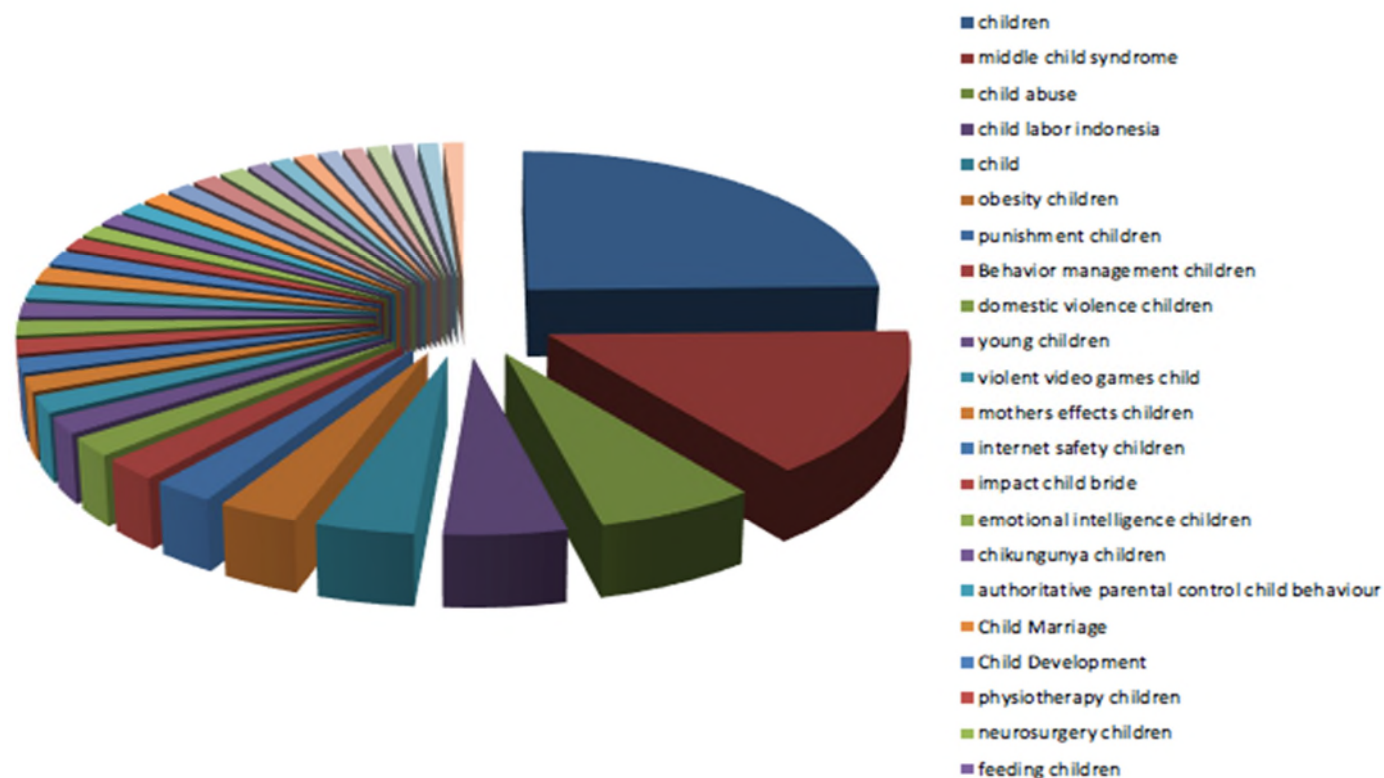


Figure 4 – Child/children context queries

4. Conclusion and final remarks

This work, which does not have any sociological purpose, provides two possible keys for interpreting the sample data of the GSGCorpus.

- I. use of terms resulting from the recently exploded digital technology, in particular the contexts of the bigram <social media>⁵ with either positive, negative or neutral polarity (see Figures 3, 5).
- II. use of terms with a mostly negative polarity which have been retrieved from the contexts of the words <child/children>: examples are given by this string of words which have been extracted from the corpus <prevalence child marriage Burkina Faso>⁶ as well as by the search of the book "The Evil Child Marriage" by Radhika Kapu⁷.

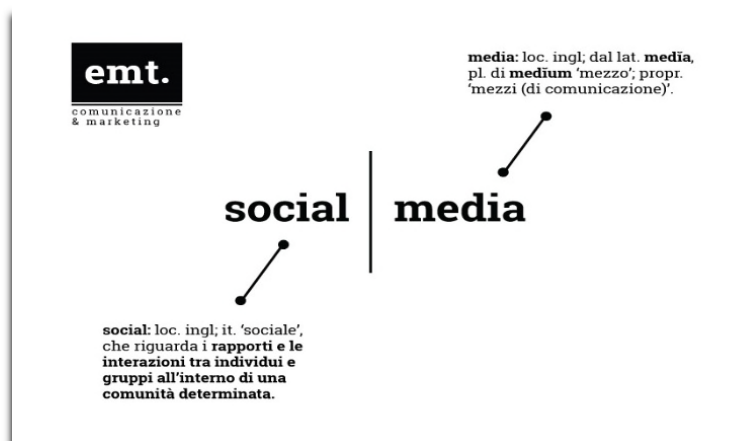


Figure 5. Meaning of the two terms

a) From the sample of queries where the bigram <social media> appears, polar verbs (affirmative and negative) and bipolar verbs (without a clear-cut tendency) have been extracted. In agreement with Manfred Klenner and Stefanos Petrakis, "we observed verbs with a relatively clear positive or negative polarity preference, as well as cases of verbs where positive and negative polarity preference is balanced (we call these bipolar-preference verbs)".

Just to make an example, a verb like 'to cause' has a negative polarity while 'to help' has a positive polarity; 'to impact' and 'to influence' have a neutral polarity (please see par. 3.1 for a detailed description of the terms with a negative polarity used in conjunction with <social media>).

The main results show that the use of social media

- can cause physical harm and widespread diseases;
- can help strengthening virtual relationships;
- can sustain e-commerce and therefore business grown.

b) From the sample of queries associated to the words <child/children>, only a negative polarity of the lexical forms emerges. The use of the following verbs, nouns and adjectives clearly certifies this assumption: 'to abuse', 'to labor', 'marriage', 'psychiatry', 'depression', '(domestic) violence', 'obesity', 'assassin', 'punishment'; 'domestic', 'psychopathic', 'violent'.

Terms extracted from the corpus of queries are largely referring to topics pertaining to the major problems of today's society, eg. *alcoholism, depression, obesity, pornography, drugs, violence, etc.*

Some critical issues can be identified in the following points: a diachronic analysis of the terms was not possible given the short temporal window taken into account; queries in different languages and many spelling/grammatical errors made our task more complicated by weighing the cleaning process down.

⁵ <https://enricomtomassi.com/cosa-social-media/>

⁶ Burkina Faso has a child marriage prevalence rate of 52%. On average, almost one out of two girls in Burkina will be married before the age of 18. The rates of child marriage vary from one region to another, and are as high as 86% in the Sahel region and 76% in the East region.

< <https://www.girlsnotbrides.org/child-marriage/burkina-faso/> >

⁷ https://www.researchgate.net/publication/323771530_Child_Marriage_The_Social_Evil

Essential Bibliography

- Battelle, J. 2005. *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. Nicholas Brealey Publishing.
- Bronson Fitzpatrick R., (2009) WorldWideScience.org: The Global Science Gateway, *Medical Reference Services Quarterly*, 28:4. {363-370, DOI:10.1080/02763860903256121}.
- Chiari, I. 2007-2008. Liste di frequenza. *Analisi del testo letterario* 1 -, 2007-2008
- Hendry David G., Jenkins J. R., McCarthy Joseph F. (2006). Collaborative Bibliography, *Inf. Process. Manage.*, May 2006, volume 42,3. Pergamon Press Inc. Pages 805-825. {<http://dx.doi.org/10.1016/j.ipm.2005.05.007>}.
- Kalgariff A., Grefenstette G. *Introduction to the Special Issue on the Web as Corpus*. ACL {<https://www.mitpressjournals.org/doi/pdf/10.1162/089120103322711569>}.
- Klenner M., Petrakis S. (2012). Polarity Preference of Verbs: What Could Verbs Reveal about the Polarity of Their Objects? In G. Bouma et al. (Eds.): *NLDB 2012, LNCS 7337*, pp. 35–46, 2012. Springer-Verlag Berlin Heidelberg 2012
- Smith A., Anderson M. (2018). *Social Media Use in 2018*, Report, March 1, 2018. Pew Research Center Internet & Technology {<http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/>}.
- Trevisan M., Dini L. Barbu E., Barsanti I., Lagos Ni., Segond, F., Ruhlmann M., Vald, Ed (2012). Query Log Analysis with GALATEAS LangLog, in *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics, EACL '12, ACL, Stroudsburg*. Pages 87-91. {<http://dl.acm.org/citation.cfm?id=2380921.2380939>}.

Web Search Queries as a Corpus

<https://www.celi.it/blog/2016/06/web-corpora/>

<<http://www.fondazionemilano.eu/blogpress/weaver/2014/03/22/181/>>

<https://www.uniba.it/docenti/gatto-maristella/attivita-didattica/materiale-didattico/Gatto_light.pdf>

<<https://worldwidescience.org/>>

<<http://greyguide.isti.cnr.it/>>

https://www.researchgate.net/publication/323771530_Child_Marriage_The_Social_Evillarity_item

Appendix

SEMANTIC EXTRACTION SCHEME

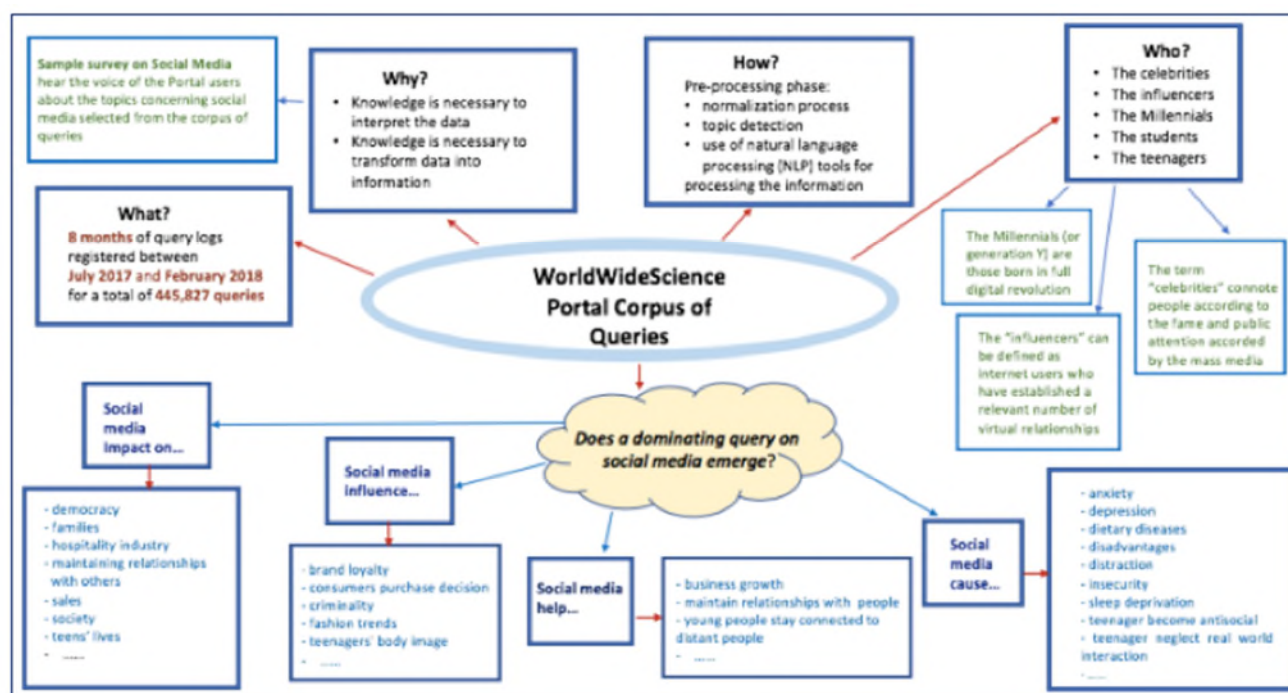


Fig. A1

THE GLOBAL SCIENCE GATEWAY

<<https://worldwidescience.org/index.html>>



Fig. A2

REPOSITORY AND PORTAL TO GOOD PRACTICES AND RESOURCES IN GREY LITERATURE

<<http://greyguide.isti.cnr.it/>>

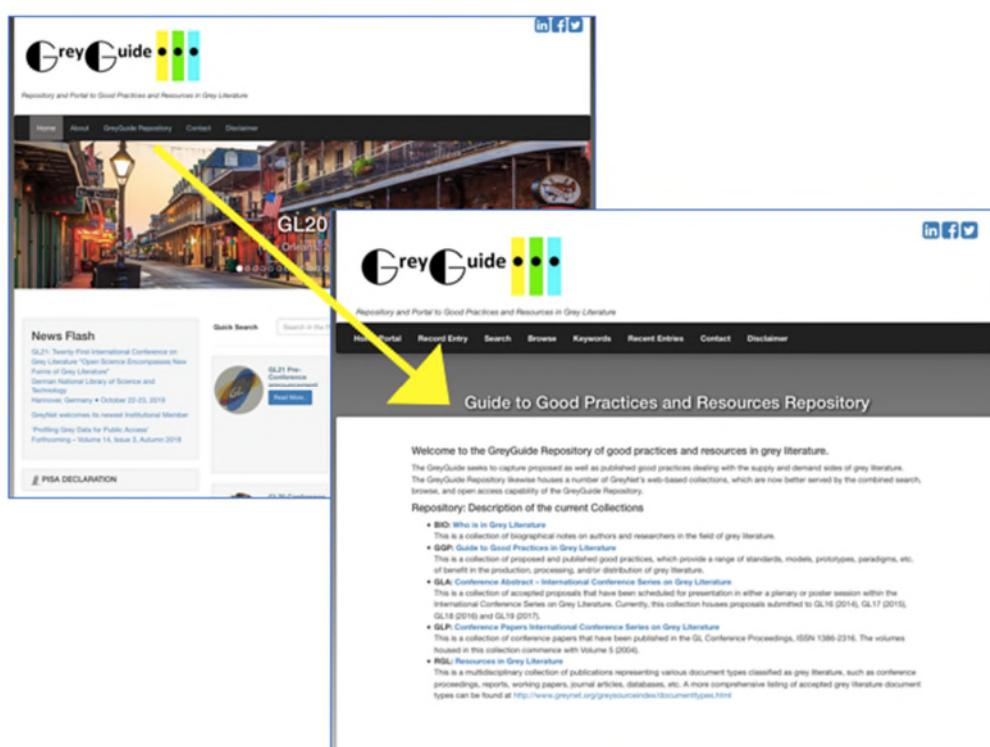


Fig. A3



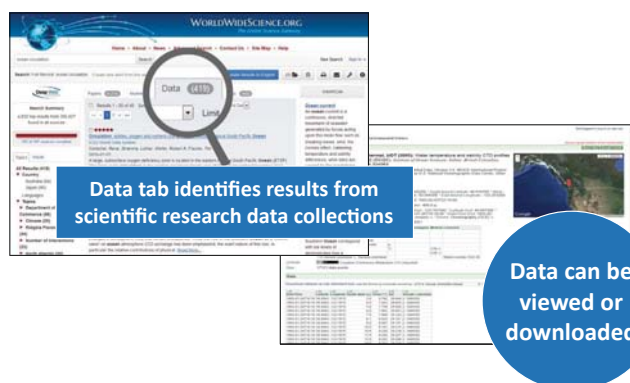
WorldWideScience.org

An International Partnership Supporting Open Science

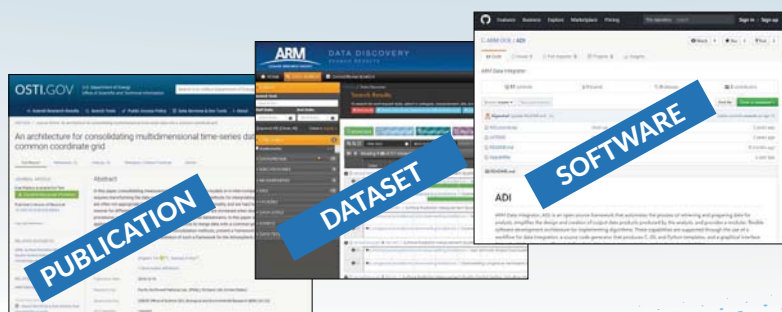


- ▶ Simultaneously explore over 100 national and international scientific databases and portals from 75 countries
- ▶ Search information in textual, multimedia, software, and scientific data formats
- ▶ Eliminate language barriers through multilingual translations across ten languages

Search and Access Scientific Research Data



Modern Science Demands Reproducibility and Open Access to Publications, Datasets, and Software



Operating Agent:

OSTI.GOV

WorldWideScience.org Operating Agent
Point of Contact: [Lorrie Johnson, JohnsonL@osti.gov](mailto:Lorrie.Johnson@osti.gov)

The U.S. Government Publishing Office: Keeping America Informed in the 21st Century and Beyond*

Cynthia Etkin, Senior Program Planning Specialist, GPO

Office of the Superintendent of Documents
U.S. Government Publishing Office, United States

Abstract

On June 23, 1860, the 36th Congress of the United States approved a joint resolution that created the Government Printing Office (GPO), and directed the Superintendent of Public Printing to have executed the printing and binding of documents approved by the Senate and House of Representatives, and the executive and judicial branch departments. GPO began operation with 350 employees on the day that Abraham Lincoln was inaugurated as the 16th President of the United States, March 4, 1861. The GPO has a long history and it is a remarkable story.

The Printing Act of 1895 overhauled existing printing laws and created a number of new authorities for GPO. Among them was to disseminate Government public documents through three channels: deposit in designated libraries; distribution to Government entities; and sales. Another new authority was to organize the Government's information, with the provision of a cataloging and indexing program. From this landmark legislation came two important public information programs, the Cataloging and Indexing Program (C&I) and the Federal Depository Library Program (FDLP). Both of these programs became integral to GPO's mission of Keeping America Informed.

More recently, with the increase in digital communication and expanding publishing technologies, GPO has streamlined and transformed from a print-only operation to an integrated publishing organization. Recognizing this, Congress re-designated the agency the U.S. Government Publishing Office on December 17, 2014.

The GPO has continually transformed itself throughout its history by adapting to changing technologies. In the ink-on-paper era this meant moving from hand-set to machine type-setting to digital type setting; from slower manual fed presses to high speed presses; and from hand to automated bookbinding.

New strategic priorities for GPO and a National Plan for Access to U.S. Government Information for GPO's public information programs are guiding another transformation for Keeping America Informed in the 21st century and beyond.

**Remarks before the Twentieth International Conference on Grey Literature
Loyola University, New Orleans, LA
December 3, 2018**

"Good morning! And let me add my welcome to New Orleans. There are other names you may hear by which the city is known: The Crescent City, The Big Easy, and The Birthplace of Jazz. I call it a special place, and I am happy to be here.

It is an honor to be here this morning representing the U.S. Government Publishing Office, the GPO, and delivering the Opening Address at this the Twentieth International Conference on Grey Literature. I have been at the GPO for more than 20 years now and I suspect, like many of you, like to talk about the place I work. I am proud of the agency, its mission, its history, and its future. This morning I will share a little bit about the history of GPO and then I will describe the path we are taking into the future.

In June 1860, the 36th Congress of the United States approved Joint Resolution 25, *Joint Resolution in Relation to Public Printing*, which created the Government Printing Office (GPO), and directed the Superintendent of Public Printing to execute the printing and binding of documents approved by Congress and by the departments of the executive and judicial branches of the Government. GPO began operation as an agency of the Legislative Branch of Government nearly a year later with 26 presses, the steam engines to run them, and 350 employees. It was March 4, 1861, the same day Abraham Lincoln was inaugurated as the 16th President of the United States.

The Printing Act of 1895 overhauled existing printing laws and created a number of new authorities for GPO. Among them was to disseminate Government public documents through three channels: deposit in designated libraries; distribution to Government entities; and sales. Another new authority was to organize the Government's information, with the provision of a cataloging

* First published in the GL20 Conference Proceedings, February 2019.

and indexing program. From this landmark legislation came two important public information programs, the Cataloging and Indexing Program (C&I) and the Federal Depository Library Program (FDLP). Both of these programs became, and remain, integral to GPO's mission of *Keeping America Informed*. More about these important programs in a minute.

As one might well imagine, GPO's business grew. And around the beginning of the twentieth century — and for several decades through the 1970s, GPO was known as "The Largest Print Shop in the World." The sheer number of presses put GPO in a league of its own. The 1926 inventory shows GPO having close to 200 presses, and it is said that a good sized commercial printer might have half as many presses. But it wasn't just the numbers, the variety of presses also made GPO unique. GPO's inventory contained every type of press found in commerce, allowing an almost unlimited scope of GPO products. "If a Government customer wanted a particular kind of printed document, GPO could produce it."

GPO's growth peaked in the late 1970s, early 1980s. At the time, in addition to the plant in Washington, D.C., there were six regional printing plants and 24 regional bookstores, and almost 9,000 employees. By the end of the 80s the number of presses had declined, but they were replaced with more productive and efficient presses that allowed production of billions of publications. In the 1990s the regional printing plants were closed with the exception of the Denver plant, which closed in 2003. In 2001 the bookstores located throughout the nation began to close.

GPO entered the online world in 1994. Our game-changing moment occurred in 1993 when, with the strong support of the library community, Congress enacted the *GPO Electronic Information Access Enhancement Act* (Public Law 103-40), which directed the GPO to create a publicly accessible system of online access to make available the *Congressional Record*, the *Federal Register*, and other publications as determined by the Superintendent of Documents.¹ GPO was given one year to implement this, and we met the challenge. This landmark legislation gave GPO the necessary authority to assume an essential role in the provision of information services and dissemination of electronic information. Access to this system, which became known as GPO Access, was free for Federal depository libraries and in 1995 it was made free to anyone. As a result of this action, the face of Government information as provided by the GPO began to shift away from tangible print toward a digital presence and in the process improved our operational efficiency and expanded our information dissemination capabilities. Fifteen years later *GPO Access* was replaced with *GPO's Federal Digital System* (FDsys), a significantly reengineered system that is OAIS-compliant (open archival information system). FDsys is a content management system, a preservation repository, and an advanced search engine. There are more than 2.2 million titles available through FDsys and it averages 45 million retrievals per month. Next week on 14 December FDsys will sunset and the third generation of online system of record will be **govinfo**.² **govinfo** was launched in early 2016 in beta.

The impact on the GPO business model and infrastructure has been dramatic. In 1993, before this act, we printed 25,000 copies of the Congressional Record every day. Today, we print 2,500 copies per day. We used to print 35,000 copies of the Federal Register each day. Today, we print fewer than 2,000. Although we print fewer copies of everything, our reach has significantly expanded because of the digital presence.

Today headquarters is at the same location on North Capitol St, NW where it first opened its doors in 1861. Now, however, there are 4 buildings that together have 33 acres or more than 13 hectares of floor space with 23 presses; a bookstore; 1,740 employees; and a strong online presence.

¹ Public Law 103-40 is codified in Title 44, *United States Code*, Chapter 41— Access To Federal Electronic Information. <https://www.govinfo.gov/content/pkg/USCODE-2017-title44/pdf/USCODE-2017-title44-chap41.pdf>.

² An update since the GL20 International Conference on Grey Literature, **govinfo** is now GPO's official system of record. Find **govinfo** at: <https://www.govinfo.gov/>.

FEDERAL DEPOSITORY LIBRARY AND CATALOGING AND INDEXING PROGRAMS

Now I want to touch on the Federal Depository Library Program (FDLP) and the Cataloging and Indexing Program. I mentioned these earlier when describing new authorities that came to GPO with the Printing Act of 1895. Both of these programs are under the purview of the Superintendent of Documents' Public Information Programs.

The Founding Fathers and early legislators recognized the importance of the free flow of information in a democratic society. They thought it essential that the citizenry be informed about its government and its workings so as to allow effective participation in the democratic process. It is this thinking that underlies the establishment of the Federal Depository Library Program (FDLP) as a network of libraries located throughout the country and territories with collections of Government publications of public interest and educational purposes that are freely accessible to the general public.

The library program traces its roots to 1813 when Congress passed a joint resolution for the printing and distribution of the journals of Congress to the executive branches and the legislative branches of every state and territorial legislature, one copy to each university and college in each state, and one copy to the Historical Society incorporated in each state.

Today we have 1,132 libraries — libraries of all types and sizes (we have some of the largest and some of the smallest libraries in the world) — participating in the depository library program. And I am pleased to say that there are ten depository libraries in New Orleans, and I want to recognize that our conference host, Loyola University of New Orleans, has two libraries in our program, the main campus library (since 1942) and the college of law library (since 1978).

GPO provides depository libraries with information products in digital format — and there's still some printed material (and microfiche) going to libraries, in return the libraries must provide free access to the public to these resources, and assistance to those who use them. Depository libraries are a critical link between "We the People" and the Government's information, and GPO's primary information dissemination program.

We now have a number of libraries that have joined the FDLP as "digital only" libraries, though some of our long-standing member libraries are moving in this direction by making available digital content and weeding their tangible collections of Government publications.

The Cataloging and Indexing Program is statutorily mandated to acquire Federal agency information products and develop a comprehensive and authoritative national bibliography of U.S. Government publications. Beginning with 1895, the bibliography was issued as a printed monthly catalog. It was replaced in 2004 with an online public access catalog, the *Catalog of U.S. Government Publications* (CGP).³ The initial record load included cataloging from 1976, which is when GPO first had machine readable records. GPO continues to work on retrospective conversion of records for inclusion in the catalog, while exploring new technologies and methods to enhance discovery of Government information.

In fiscal year 2017 GPO distributed 4,200 tangible titles to depository libraries, and created nearly 11,000 PURLs (persistent uniform resource locator) to digital content accessible through the bibliographic records in the *Catalog of U.S. Government Publications*. There were nearly 33.2 million searches of the catalog from around the world.

NATIONAL ACADEMY OF PUBLIC ADMINISTRATION REPORT

In 2012 at the request of Congress, the National Academy of Public Administration (NAPA) conducted an operational review of GPO, a 21st century readiness study. NAPA's final report, *Rebooting the Government Printing Office: Keeping America Informed in the Digital Age*,⁴ was released January 2013. Of great importance, NAPA determined that "GPO's core mission of authenticating, preserving, and distributing federal information remains critically important to American democracy," while at the same time they made recommendations to strengthen GPO's business model for the future. GPO is working to implement the recommendations of the NAPA panel.

³ *Catalog of U.S. Government Publications* is available from: <https://catalog.gpo.gov>. The statutory authority for the Cataloging and Indexing Program is in Title 44 *United States Code* §§1710-1711. <https://www.govinfo.gov/content/pkg/USCODE-2017-title44/pdf/USCODE-2017-title44-chap17-sec1710.pdf>.

⁴ *Rebooting the Government Printing Office: Keeping America Informed in the Digital Age* can be viewed and downloaded from GPO's website at: https://www.gpo.gov/docs/default-source/congressional-relations-pdf-files/gpo_napa_report_final.pdf?sfvrsn=2.

Among the recommendations of the NAPA Panel that greatly affects the Public Information Programs of the Superintendent of Documents is one related to preservation, safeguarding the historical documents of our democracy for future generations. It has wide-sweeping outcomes: “GPO should work with depository libraries and other library groups to develop a comprehensive plan for preserving the print collection of government documents.” Further they stated that the plan should include cataloging, digitizing, and preserving tangible copies of government publications, as well as a process for ingesting digitized copies into GPO’s system of record for online access. This is a huge undertaking, but the Superintendent of Documents was pleased to see this recommendation, for it gets to the heart of the Public Information Programs — permanent public access to Government information.

MOVING FORWARD

GPO continues to transform and move forward with two documents as guides: *GPO Strategic Plan*, a five-year rolling strategic plan; and *National Plan for Access to U.S. Government Information*, a framework for a user-centric service approach to permanent public access.

GPO STRATEGIC PLAN

The FY 18-22 Strategic Plan⁵ states GPO’s vision as “an informed nation that has convenient and reliable access to their government’s information through GPO’s products and services.” The plan has five strategic goals:

1. Exceed our stakeholders’ expectations
2. Enhance access to Federal Government information
3. Strengthen our position as the government-wide authority on publishing
4. Promote collaboration and innovation within government
5. Engage employees and enhance internal operations

I am not going to talk about all of these, but I do want to say a few things about enhancing access to Federal Government information and promoting collaboration and innovation within government.

Strategies for achieving the second goal, enhanced access to Government information, are to increase the amount of U.S. Government information available for free to the public and enhance access to information to meet evolving user needs; and support access and discoverability through the Federal Depository Library Program and the Cataloging and Indexing Program. This means incorporating missing records and enhancing existing bibliographic records for historical materials. We aim to make the *Catalog of U.S. Government Publications* comprehensive index of every document issued or published by a department, bureau, agency, or office that is not confidential in character. This also means that GPO will continue to add to the collections in **govinfo**, with the goal of offering complete and historic content.

Strategies for achieving the fourth goal, promoting collaboration and innovation within government, include developing partnerships with Federal agencies that will improve identification of and increase access to their information. This also includes the establishment of the Federal Publishing Council, an advisory group made up of Federal employees involved in all facets of the Federal printing and publishing community.

NATIONAL PLAN FOR ACCESS TO U.S. GOVERNMENT INFORMATION

The *National Plan for Access to U.S. Government Information*⁶ sets the strategic direction for the Federal Depository Library and Cataloging and Indexing programs, with the vision “to provide Government information when and where it is needed.” It is guided by GPO’s mission of *Keeping America Informed*, and it was informed by the NAPA study of GPO, external influences, and most particularly by the results of the FDLP Forecast Study. The Forecast Study, already underway when NAPA came to GPO for their study, was a data-gathering effort to obtain information that would allow GPO to better understand the issues and challenges facing depository libraries, document their needs, and their view of the ideal depository library program of the future.

⁵ GPO’s FY18-22 Strategic Plan is available for viewing and downloading from: <https://www.gpo.gov/who-we-are/our-agency/mission-vision-and-goals>.

⁶ *National Plan for Access to U.S. Government Information* is available for viewing and downloading from: <https://www.fdpl.gov/superintendent-of-documents-public-policies>.

The three strategic priorities set forth in the National Plan are:

1. Establish Library Services and Content Management processes and procedures that apply lifecycle management best practices for all formats and ensure permanent public access to Government information dissemination products in the digital age.
2. Provide a governance process and sustainable network structure that ensures coordination across the Federal Depository Library Program and allows the most flexible and effective management of depository libraries and their resources.
3. Deliver dynamic, innovative, strategic services and mechanisms to support the needs of Federal depository libraries in providing accurate Government information to the public at large in a timely manner.

The first strategic priority looks internally at the organization and process improvement, while the second looks at how GPO can better administer the FDLP to meet the changing needs of the libraries. The third considers what new or enhanced tools and services GPO can provide libraries that will better support them. Some of the outcomes are achievable in the near term, while others can only be achieved over the long-term.

The introduction of online digital content into the FDLP was a major paradigm shift for GPO. GPO has sent Government publications to libraries for more than 200 years. Materials sent to depository libraries remain the property of the U.S. Government; libraries are the stewards of those collections. Collections have always been in the depository libraries and now GPO has a digital collection to curate and a distributed tangible national collection to preserve.

A lot of activities taking place revolve around preservation, because you can't have permanent public access to content if it isn't preserved. We are building a sustainable preservation program within our Library Services and Content Management business unit. We're expanding the number of staff in this area, and we are getting ready to pilot how GPO can provide various preservation services to depository libraries. We have 40 partnerships⁷ with depository libraries and Federal agencies that are providing digital content for ingest into **govinfo**, preserving tangible content in their collections, sharing cataloging records, or providing access to or preserving digital content in their repositories. GPO recognizes that to be successful in preserving Government information, collaboration is necessary. We call our collaboration the Federal Information Preservation Network, or FIPNet.

In the area of digital preservation, GPO is currently undergoing an audit for **govinfo** to be certified as a trustworthy digital repository under ISO 16363: 2012, Space data and information transfer systems -- Audit and certification of trustworthy digital repositories. **govinfo** will be assessed against 109 criteria related to organization infrastructure, digital object management, and infrastructure and security risk management. The latest update on the audit? The auditing team will be making their site visit to GPO later this week.⁸

During the Senate confirmation hearing for a former Public Printer of the United States, the questions was asked, "Do we still need the depository library program when people can access Government information from their living room?" The response was, without hesitation, a resounding, "YES! But GPO needs to better support them, and help them manage in the digital age." And that's what GPO is doing, with the GPO Strategic Plan and the National Plan as guides.

A NEW MIDDLE NAME

On December 16, 2014 Congress redesignated the Government Printing Office to the Government Publishing Office. "This is a historic day for GPO. Publishing defines a broad range of services that includes print, digital, and future technological advancements. The name Government Publishing Office better reflects the services that GPO currently provides and will provide in the future," said Davita Vance-Cooks, who was then Director of the Government Publishing Office, the agency's chief executive officer.⁹ Modern publishing operations provide a suite of services to ensure access

⁷ For more information about partnerships see: <https://www.fdlp.gov/about-the-fdlp/partnerships>.

⁸ I am pleased to provide a further update on the ISO 16363 audit. After the onsite visit, the auditing team recommended to the certification committee that GPO be awarded certification as a trustworthy digital repository under the international standard; the committee concurred. For more information on GPO and the Trustworthy Digital Repository certification process see: <https://www.fdlp.gov/preservation/trusted-digital-repository-iso-16363-2012-audit-and-certification>.

⁹ GPO News Release No. 14-27, *GPO IS NOW THE GOVERNMENT PUBLISHING OFFICE*. December 17, 2014. <https://www.gpo.gov/docs/default-source/news-content-pdf-files/2014/14news27.pdf>.

to the information they disseminate, from conventional printing to eBooks, digital publishing, mobile access, social media, and other strategies. All of these services are provided by GPO today.

CONCLUSION

The importance of an informed public to the effective conduct of self-government prompted James Madison to write in August 1822:

A popular Government without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy, or perhaps both. Knowledge will forever govern ignorance, and a people who mean to be their own Governors, must arm themselves with the power which knowledge gives¹⁰.

Since its founding, GPO has performed that function by producing and distributing documents created by our government, serving as the “means of acquiring” information for “a people who mean to be their own Governors” through Federal depository libraries and now also through **govinfo** and the *Catalog of U.S. Government Publications*.

GPO is in a good place. Today, GPO is fundamentally different from what it was as recently as a generation ago: smaller, leaner, and equipped with digital production capabilities that are the bedrock of the information systems relied upon daily by Congress, Federal agencies, depository libraries, and the public to ensure open and transparent Government in the digital era. As it was with its history, GPO will continue to strategize, plan, and transform. And GPO will be *Keeping America Informed* for future generations, beyond the 21st century. Thank you very much.

¹⁰ United States. Government Publishing Office. *Keeping America Informed, the U.S. Government Publishing Office: A Legacy of Service to the Nation, 1861-2016*, Washington, DC: United States Government Publishing Office, 2016. Available from the GPO website at: <https://www.govinfo.gov/content/pkg/GPO-KEEPINGAMERICAINFORMED-2016/pdf/GPO-KEEPINGAMERICAINFORMED-2016.pdf>.

Analysis of folk literature in grey literature from the National Library of China*

Cui Yue, National Library of China, Beijing, China

Abstract:

As a nationwide library, national bibliographic center and document information center, the National Library of China (NLC) is not only comprehensively collecting official Chinese publications but as also, long before, attached importance to the acquisition of grey literature. The NLC's acquisition of grey literature has changed from a basic, scattered collection to a comprehensive collection and finally to a key collection. The institution was established through, in succession, the academic dissertation library, local chronicles and the genealogical documents collection.

At the same time, the NLC was exploring the construction of characteristic Chinese grey literature. Through the analysis of the present grey literature collections in NLC, the amount of folk literature, conference proceedings, research reports and document assembly has already taken shape.

Therefore, a new grey literature system has been formed, among which the most significant increase is in folk literature with distinctive regional characteristics and historical value.

Broadly speaking, folk literature refers to all literature bearing historical and cultural information and kept in folkloric. It contains the literature produced by the people and the official literature lost in folkloric.

This literature covers historical, cultural, artistic and other fields, such as folk poetry, intangible cultural heritage, memoirs, revolutionary historical materials, etc.

At present, literature collections and sorting institutions, such as libraries are not collecting and researching of folk literature to a great enough extent. As a result, researchers cannot easily find comprehensive literature. As a library that bears a mission of preserving cultural heritage, the collection and utilization of folk literature should be emphasized.

Taking folk literature in the grey literature collected by NLC over the past 10 years as a sample, this paper analyzes the content, composition, quantity, type, geographical distribution, value and significance of this kind of literature. To solve the problem of the construction of folk literature in grey literature, this paper innovates acquisition methods and the collection scope of folk literature within grey literature. Some ideas and suggestions are put forward on how to establish folk literature collection institutions and rationally utilize and develop existing resources.

1 Background

As early as January 1912, the Library had begun to collect books. In June 1921, the graduation thesis and dissertation submitted by foreign graduates were accepted in collection[1]. In April 1956, the National Library of China publicly collected scientific research materials through newspapers. In July 1964, the National Library of China, systemically collect nationwide academic conference materials[2].

By 80s, the process of the collection in the National Library was basically in a scattered and chaotic state. Fortunately, it was not until 1985 that the National Library had established a special force, "Domestic Data Collecting Department (DDCD)" which specialized in the collection of GL, and began to investigate, collect and arrange all the grey literature out there. In late 80s, the whole number of copies was up to 100,000[3]. It was unexpected such mass of materials were gathered in a short time. However, due to limited human resources, there were some staggering difficulties in terms of document processing. Therefore, according to the situation mentioned above, in early 90s, NLC adjusted its strategy, primarily focused on thesis, local history and genealogy instead of comprehensive collection. Meanwhile, certain institutions such as dissertation library, local chronicles and genealogy literature center were established. Afterward, on the other hand, the DDGD was abolished. At the beginning of 2003, the newly revised "Regulations on the Selection of National Library Documents" re-listed the collection of grey literature, formulated the reference range for collection[4]. In general, the National Library's collection of grey literature has roughly gone through 4 stages, "the fragmentation of the 1980s" - "the full entry, 1980s to the 1990s" - "the selective entry, early 1990s - mid-1990s" - and the "Special Collections, late 1990s - Present".

* First published in the GL20 Conference Proceedings, February 2019.

2 Status of grey literature

2.1 Overview

The "Chinese Information Group" established in 2008 by National Library of China, for providing valuable and professional services for some major national-level projects such as legislative decision-making and scientific research projects, have collected various and numerous GLs from a wide range of channels, like government, college, institute, etc. And as of now, it really has done an excellent job. According to the statistics from 2008 to 2017, a total of up to 150,000 books were included in the Chinese Literature Group (CLG) in the past 10 years. With the experience of the previous literature collection, major collection scope set by CLG included: academic conference literature, research reports, folk literature, data compilation, and other types of literature.

In a decade of collecting, it's observed that the folk literature accounted for around one-third of the total number of grey literature, ranking first. Followed by data compilation and research report literature, just as what you can see from the Fig. 1. It is not difficult to realize that the folk literature in the grey literature database from the National Library currently occupied a large proportion, which has prominent features in terms of both document characteristics and document value. As shown in Fig. 1:

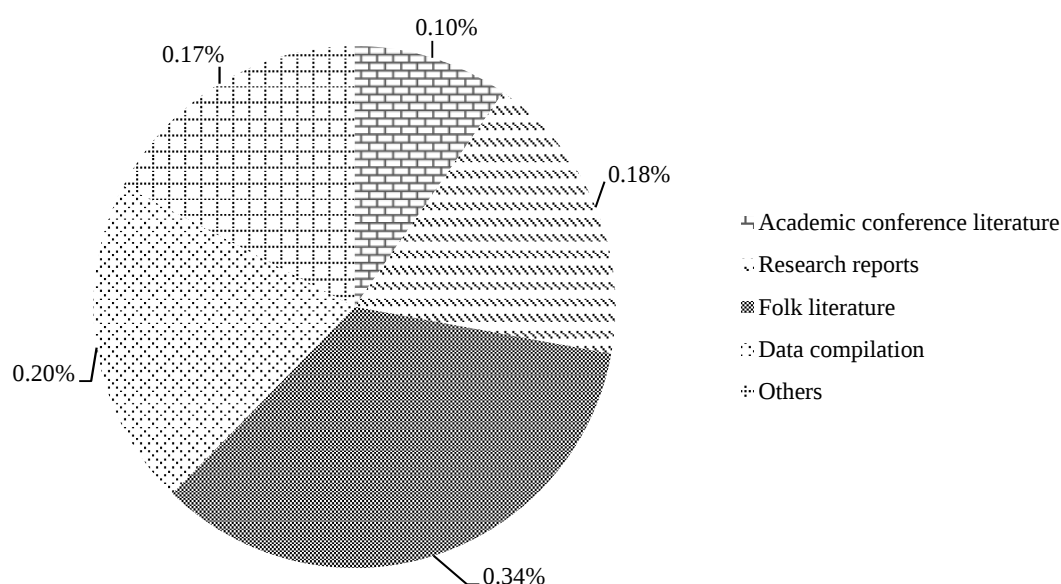


Fig. 1 The proportion of each type of grey literature during 2008-2017

2.2 Source

At present, the grey literature in our library, for the most part, is purchased from booksellers or accepted through donation projects. The purchase should be preceded by lists of book title for our selection supplied by booksellers. Acceptance of donations can be through individuals or groups. The first one is more dispersed but varied, while donations from group organizations are more fixed and systematic, and more continuous. According to Fig. 2, the donation frequency was higher in east China and north China, and their willingness to donate was more positive. On the other hand, it also reflects the uneven distribution of literature collection, should be strengthened to collect literature from other regions. As shown in Fig. 2:

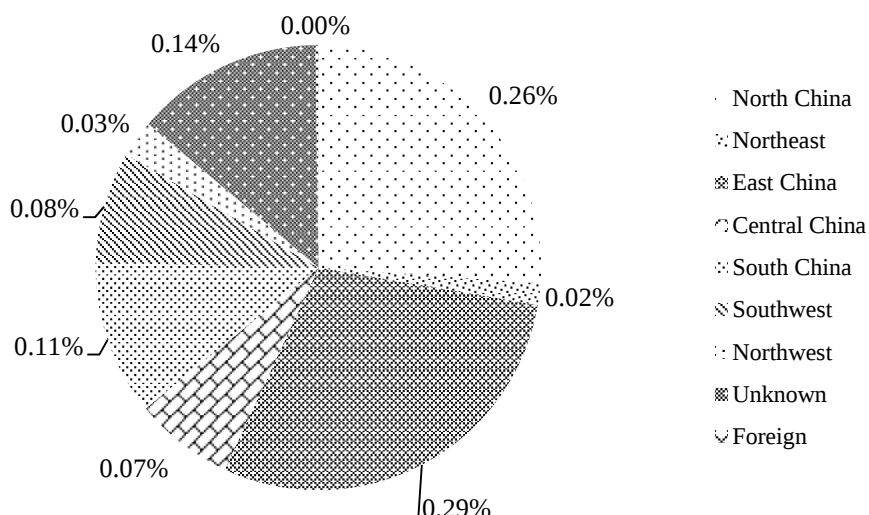


Fig.2 Donation distribution of grey literature in China

After ten years collecting of grey literature, the national library has a mature collection method and literature system of grey folk literature, the collection of literature is concentrated and the source of collection is stable, and has formed the relatively outstanding folk literature specialized collection.

3 Folk literature

3.1 Definition

Academically, it is clearly defined by Zhengzhenman: "Literatures formed and used in daily life of the people". The folk literature is produced by and spreads among people. So, to some degree, it could embody all aspects of people's life in the long historical process. Considering its historical, national and artistic nature, it could demonstrate kaleidoscope of society and history from multi-dimensional perspective for future generations and plays a key role of cultural understanding.

3.2 Classification

The grey folk literature in the National Library collection could generally be divided into three categories: folk customs, folk literary work, folk history materials. There are more than 20,000 kinds and 50,000 volumes of grey folk literature are collected in NLC. Among which folk literary work accounts for 67%, ranking first. Followed by folk history materials accounts for 28%. Specifically, the folk customs include folklores, parables, art and crafts, and folk culture materials. The most distinguished type among them could be ratified as intangible cultural heritage literature. Folk literary work can be represented by poetry and anthology. The poetry kind dominates folk literary work, accounting for 52%. The folk history materials mainly consist of revolutionary historical materials, college chronicle, memoir historical materials, etc. While the number of revolutionary historical materials has prominent position in the folk history materials collection, accounting for 79%.

3.3 Collection

To sum up, there are mainly two ways of obtaining grey literature in national library: institution and individual. Among them, the collection of personal literature is relatively scattered, irregular and discontinuous, which is difficult to trace this part of literature in the literature source. However, donations from group organizations are relatively fixed and systematic. According to statistics, the collection agencies of folk literature are mainly distributed in poetry groups, societies, associations, universities, education institutions, literary and artistic institutions, research institutions and so on.

4 Folk literature collection construction

The types of folk literature are varied and characteristic, so the resources construction of folk literature cannot be generalized. It should be collected, developed and utilized according to their own characteristics. The following is an elaboration on the construction of the typical folk literature.

4.1 The construction of intangible cultural heritage materials

The intangible cultural heritage documents record and preserve the original appearance of the intangible cultural heritage, providing effective help for the protection of the endangered intangible cultural heritage and research. Intangible cultural heritage documents are of historical, cultural and artistic value. Collecting such literature can inject new vitality into the library, and it is also an important supplement of literature resources.

According to statistics, there are 547 kinds of grey intangible cultural heritage documents in the national library. It can be divided into: Materials of intangible cultural heritage census, records of intangible cultural heritage, intangible cultural heritage research, laws and regulations related to intangible cultural heritage protection, intangible cultural heritage activities[5]. Materials of intangible cultural heritage census refer to a series of intangible cultural heritage census information generated in the work of intangible cultural heritage census; Materials of records of intangible cultural heritage refer to record the origin, development and inheritance of intangible cultural heritage; Materials of research on intangible cultural heritage refer to general or special research on intangible cultural heritage; Materials of intangible cultural heritage activities refers to the materials of lectures, exhibitions and publicity activities organized around intangible cultural heritage and protection. According to the statistics, the materials of intangible cultural heritage census information accounts for 62%[6], ranking first. On the one hand, it shows that the intangible cultural heritage census has achieved remarkable results in recent years; On the other hand, it also shows that the intangible cultural heritage census is the most basic and most direct way to protect the intangible cultural heritage. The documentation of records of intangible cultural heritage accounts for 23%, ranks the second. Indicating that under the influence of social development and environmental change, the endangered intangible cultural heritage can be recorded and preserved in the form of pictures and texts, which is also an important part of the protection of intangible cultural heritage.

Systematic solicitation. Literature of intangible cultural heritage should be collected in a targeted and systematic way according to its characteristics. Based on the above analysis, the literature of intangible cultural heritage census has the characteristics of large amount of systematic literature. Based on this, collection of intangible cultural heritage census materials according to literature sources, release time, release batches, etc. Secondly, according to the regional characteristics of intangible cultural heritage, systematic solicitation can be conducted in different regions. Thirdly, systematic solicitation can be carried out according to the grade of intangible cultural heritage, such as the collection of intangible cultural heritage system divided into national, provincial, municipal and county levels. Finally, a systematic collection of intangible cultural heritage batches can be conducted according to the intangible cultural heritage list.

Make special catalogue of intangible cultural heritage. Making a special catalogue can better understand the literature collection, timely find missing documents to supplement, and ensure the integrity of the literature. At the same time, it is conducive to literature collection and collation. The catalogue can be classified according to literature type, literature source, literature content and literature grade.

4.2 The construction of folk poetry materials

Chinese poetry originated in the pre-Qin Dynasty and flourished in the Tang Dynasty. The Chinese iambic verse originated in the Sui and Tang Dynasties and gain its popularity in the Song Dynasty. The poetry comes from the folk and belongs to the real grassroots literature. The poets of the Tang Dynasty, such as Li Bai, Du Fu, BaiJuyi are all well-known in the oriental culture. At present, the National Library's poetry and literature documents have a more comprehensive interview system, the interview direction is clear, the interviewees are very active, and we can acquire continuous relevant literature. which is the main reason for the significant growth in poetry literature in NLC database.

According to statistics, there are more than 5,000 kinds of grey folk poetry literature in the collection, among which contemporary poetry accounts for 98%. Showing poetry in China is greatly favored by the public and have a potential to develop further. On the other hand, it also shows that contemporary poetry is a kind of closer acquisition, easier to collect.

A total of 166 poetry associations, societies and other poetry-related groups in the country regularly donate poetry documents to the National Library. Donation groups are divided by region, as shown in Fig. 3. it can be easily seen that there are more poetry groups in Central South part of China, and their willingness to donate is more positive. As shown in Fig. 3 :

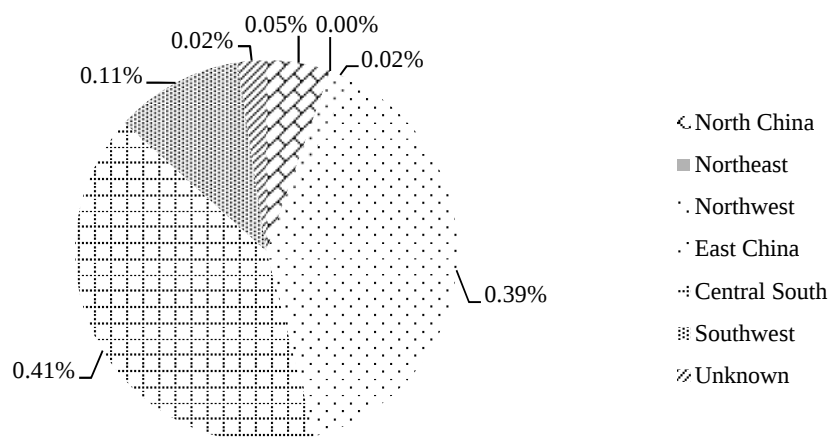


Fig. 3 Poetry donation distribution in China

According to the statistics, the folk poetry donation has some fixed poetry association and social groups. First, maintain the donation relationship with the institute, college, and certain poetry group, through which we could obtain stable and continuous poetry documents. And the Internet can be helpful too, like Chinese poetry net, and any other sites involving Chinese poetry, more selection of poetry and literature can be actively selected. Second, through the classification of poetry depending on the content of literature, we could set special catalog for better research and development.

4.3 The construction of revolutionary historical materials

The historical materials of revolution not only record the history of revolution, but also provide materials for the research of the history of revolution at a certain stage and provide help for the propaganda of traditional revolution education. At the same time, the historical data of revolution in grey literature recorded the history of revolution from a macro perspective, and vividly and succinctly described the history of revolution from the individual perspective. It complements the revolutionary history in many ways.

At present, the revolutionary history materials in the National Library Collection is expected to be more than 2,000 copies, including: History of revolutionary war; Memoir in revolutionary; Record of heroes and other historical materials. Among which History of revolutionary accounts for 38%, ranking first. Followed by Memoir in revolutionary and Record of heroes. As shown in Fig. 4 :

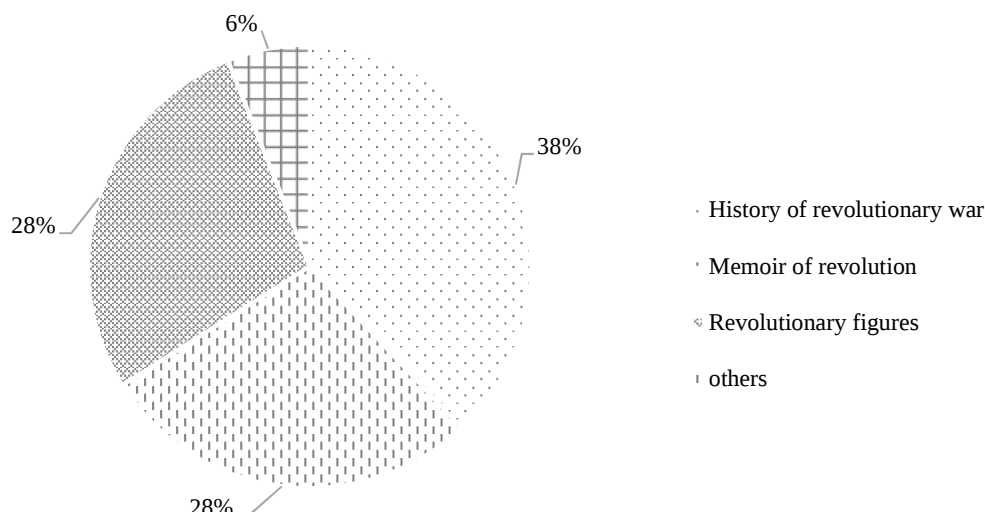


Fig. 4 Classification of revolutionary historical materials

The history of revolutionary war materials mainly records relevant materials about the war in the warfare process. The memoir in revolutionary age materials mainly refers to oral and written account of revolutionary events. The record of heroes materials mainly refers to the life story, biography and revolutionary of hero and heroine, etc.

Tracking the literature according to the source of the literature. it could ensure the validity, continuity and stability. As it statistic, revolutionary historical materials mainly come from individuals and institutions, and personal publishing part possess a large proportion, taking up 73.2%. while the rest materials come from institution such as Civil affairs bureau, Seminar, Retirement center for veteran cadres, Archives bureau, etc.

Secondly, Classifying on the basis of the types of revolutionary historical materials, for further development and utilization. In fact, our library has begun to compile the "Revolutionary History" bibliography as early as the early 1980s. By compiling certain bibliography, we could better understand the contents of the collection and conduct more efficient and reasonable classification management. We could according to the different publishing institutions to classify revolutionary historical materials, such as the History Research Office catalog, the Revolutionary Research Society institution catalog, the Archives bureau catalog etc. From the perspective of the literature form, it can be divided into historical materials catalog, memory catalog, research catalog, publicity and education catalog, etc.

5 Conclusion

At present, the collection of grey literature in China is still in the stage of continuous exploration. There are still many improvements in the collection and utilization of the folk literature. The folk literature is diversified, the literature sources are relatively scattered, and it is not easy to track and collect. In addition, the public's cognition of grey literature donation is simple, and the folk groups know little about the rules of grey literature entry. So, the publicity and promotion of grey literature collection should be strengthened to enhance the awareness of folk groups on grey literature and enhance their awareness of donation. Without active solicitation, the grey folk literature may be lost. At the same time, the collection of grey literature also needs the support of the government to actively promote the introduction of the grey literature legal deposit system. It provides a better guarantee for the collection of folk literature. Then, build resource sharing platform and open access resource. By building a resource sharing platform, the folk literature of different units and different regions are comprehensively integrated. (1) The collection information of folk literature is provided by culture-related units, libraries of various types, professional non-governmental organizations and individuals, etc. so as to achieve the largest collection of the substantive resources of folk literature. (2) The most extensive retrieval and utilization of the data is realized through the folk literature bibliographic data Shared by the libraries. (3) The public will gain a deeper understanding of the grey folk literature by publishing the solicitation information, collection process and collection system on the Shared platform. Finally, the digitization of folk grey

literature. At present, social development has achieved digitalization, networking and is now entering the direction of artificial intelligence. There are some limitations in the acquisition of physical books and the quantity of literatures. Some folk literatures only have one volume, so it is difficult to collect literatures. The valuable grey literature can be digitized and saved permanently, and the digital literature can be organized and analyzed. And then by using advanced technologies such as holographic images, valuable folk history and endangered intangible cultural heritage techniques will be reproduced and restored through three-dimensional images to provide more service space for the cultural field and the technical field.

References

- [1] Beijing library business research committee. collection of historical materials of Beijing library (1909-1949)[M]. Bibliography and Document press,1992
- [2] Compilation of historical materials of Beijing library (2) editorial committee. Compilation of historical materials of Beijing library (ii):1949-1966[M]. Beijing library press,1997
- [3] Yao rong. Development and utilization of Chinese materials[J].journals of Library Science,2004(4)
- [4] Wang dongbo, Zhao xiaohong. Improve the policy of document collection and selection and build the national general library -- an overview of the amendment of the regulations on document collection and selection of national library[J]. Journal of the National Library of China, 2004(1)
- [5] Cui yue. Some Observations of Improving the Document Catalog of Intangible Cultural Heritage—From the Perspective of Full Disclosure of the Literature Content[J].Library Development,2018(1)
- [6] Cui yue. Research on Intangible Cultural Heritage Literature Collection Development[J].Journal of the National Library of China.2018(3)

Data from:

“Data Papers as a New Form of Knowledge Organization in the Field of Research Data”

<https://doi.org/10.17026/dans-zk3-jkyb>

URN: [urn:nbn:nl:ui:13-iy-02u8](https://nbn-resolving.org/urn:nbn:nl:ui:13-iy-02u8)

Joachim Schöpfel, University of Lille

<https://orcid.org/0000-0002-4000-807X>

Project leader

Hélène Prost, CNRS, GERiCO laboratory

<https://orcid.org/0000-0002-7982-2765>

Dataset collector

Antonella Zane, Antonio Vallisneri Bio-medical Library

<https://orcid.org/0000-0001-7218-6068>

Dataset collector

Abstract

Data papers have been defined as scholarly journal publications whose primary purpose is to describe research data. Our survey provides more insights about the environment of data papers, i.e. disciplines, publishers and business models, and about their structure, length, formats, metadata and licensing. Data papers are a product of the emerging ecosystem of data-driven open science. They contribute to the FAIR principles for research data management. However, the boundaries with other categories of academic publishing are partly blurred. Data papers are (can be) generated automatically and are potentially machine-readable. Data papers are essentially information, i.e. description of data, but also partly contribute to the generation of knowledge and data on its own. Part of the new ecosystem of open and data-driven science, data papers and data journals are an interesting and relevant object for the assessment and understanding of the transition of the former system of academic publishing.

Keywords: Data papers, research data, knowledge organization, open science, data journals, FAIR principles, academic publishing

Subject Area: Library & information Science

Methods Applied

● Steps

In order to analyze specific features of data papers, we established a representative sample of data journals, based on lists from the European FOSTER Plus project, the German wiki forschungsdaten.org hosted by the University of Konstanz and two French research organizations. The complete list consists of 82 data journals, i.e. journals which publish data papers. They represent less than 0,5% of academic and scholarly journals. For each of these 82 data journals, we gathered information about the discipline, the global business model, the publisher, peer reviewing etc. The analysis is partly based on data from ProQuest's Ulrichsweb database, enriched and completed by information available on the journals' home pages.

Results

One part of the data journals is presented as “pure” data journals *stricto sensu*, i.e. journals which publish exclusively or mainly data papers. We identified 28 journals of this category (34%). For each journal, we assessed through direct search on the journals' homepages (information about the journal, author's guidelines etc.) the use of identifiers and metadata, the mode of selection and the business model, and we assessed different parameters of the data papers themselves, such as length, structure, linking etc. The results of this analysis are compared with other research journals (“mixed” data journals) which publish data papers along with regular research articles, in order to identify possible differences between both

journal categories, on the level of data papers as well as on the level of the regular research papers. Moreover, the results are discussed against concepts of knowledge organization.

● Sampling strategy

The sample was selected based on the data journal lists from the European FOSTER Plus project, the German forschungsdaten.org wiki hosted by the University of Konstanz and from two French research organizations. We compared the lists and selected pure and other data journals, trying to be exhaustive for the pure data journals and representative for the other journals with data papers.

● Quality Control

We checked and analyzed the information on the website of each data journal and compared with information from the cited lists and the Ulrichsweb database. In case of a problem with the data quality, the assessment was repeated and if necessary, corrected, by one or two other co-authors.

Dataset Description

<i>File name:</i>	ISKO 2019 all data journals.ods	ISKO 2019 pure data journals.ods
<i>Format name and version:</i>	ODS	
<i>Creation dates:</i>	From 2019-04 to 2019-05	
<i>Language:</i>	English	
<i>License:</i>	CC0 Waiver – no rights reserved	
<i>Archive name:</i>	DANS EASY Archive	
<i>Publication date:</i>	2019-06-03	

Potential Reuse of the Data

The data set can be reused by scientists in library and information science and academic librarians interested in the development of data journals and data papers and, more generally, in the field of academic publishing, new forms of scientific communication and research data management.

Linked References

BELTER Christofer W. (2014). Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets. *PLoS One*, March 26, 2014. Disponible sur : <https://doi.org/10.1371/journal.pone.0092590> (Consulté le 26/07/2019)

BORDELON Dominic, GROTHKOPF Uta, MEAKINS Sylvia, STERZIK Michael (2016). Trends and developments in VLT data papers as seen through telbib. *Proc. SPIE 9910, Observatory Operations: Strategies, Processes, and Systems VI*, 99102B (15 July 2016). Disponible sur : <https://www.eso.org/sci/libraries/SPIE2016/9910-89.pdf> (Consulté le 26/07/2019)

CALLAGHAN Sarah, DONEGAN Steve, PEPLER Sam et al. (2012). Making data a first-class scientific output: Data citation and publication by NERC's environmental data centres. *International Journal of Digital Curation*, vol. 7, no. 1, pp. 107-113. Disponible sur : <https://doi.org/10.2218/ijdc.v7i1.218> (Consulté le 26/07/2019)

CANDELA Leonardo, CASTELLI Donatella, MANGHI Paolo, TANI Alice (2015). Data Journals: A Survey. *JASIST*, vol. 66, no. 9, pp. 1747-1762. Disponible sur : <https://doi.org/10.1002/asi.23358> (Consulté le 26/07/2019)

CHAVAN Wishwas, PENEV Lyubomir (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, vol. 12, suppl. 15, S2. Disponible sur : <http://www.biomedcentral.com/1471-2105/12/S15/S2> (Consulté le 26/07/2019)

COSTELLO Mark J., MICHENER William K., GAHEGAN Mark, ZHANG Zhi-Quiang, BOURNE Philipp E. (2013). Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution*, vol. 28, no. 8, pp. 454-461. Disponible sur : <https://doi.org/10.1016/j.tree.2013.05.002> (Consulté le 26/07/2019)

DAVIS Grace H., PAYNE Eric, SIH Andrew (2015). Commentary: Four ways in which data-free papers on animal personality fail to be impactful. *Frontiers in Ecology and Evolution*, vol. 3, no. 102, pp. 1-3. Disponible sur : <https://doi.org/10.3389/fevo.2015.00102> (Consulté le 26/07/2019)

FARACE Dominic J., FRANTZEN Jerry, SMITH Plato L. (2018). Data Papers are Witness to Trusted Resources in Grey Literature: A Project Use Case. *The Grey Journal*, vol. 14, no. 1, pp. 31–36.

FRIEDMAN Rachel, PSAKI Stéphanie, BINGENHEIMER Jeffrey B. (2017). Announcing a New Journal Section: Data Papers. *Studies in Family Planning*, vol. 48, no. 3, pp. 291-292. Disponible sur : <https://doi.org/10.1111/sifp.12032> (Consulté le 26/07/2019)

GARCIA-GARCIA Alicia, LOPEZ BORRUL Alexandre, PESET Fernanda (2015). Data journals: eclosión de nuevas revistas especializadas en datos. *El profesional de la información*, vol. 24, no. 6, pp. 845-854. Disponible sur : <https://doi.org/10.3145/epi.2015.nov.17> (Consulté le 26/07/2019)

HUANG Xiaolei, HAWKINS Bradford A., QIAO Gexia (2013). Biodiversity Data Sharing: Will Peer-Reviewed Data Papers Work? *BioScience*, vol. 63, no 1, pp. 5-6. Disponible sur : <https://doi.org/10.1525/bio.2013.63.1.2> (Consulté le 26/07/2019)

LE DEUFF Olivier (2018). Une nouvelle rubrique pour la RFSIC : Le Data Paper. *Revue française des sciences de l'information et de la communication*, no. 15. Disponible sur : <http://journals.openedition.org/rfsic/5275> (Consulté le 26/07/2019)

LI Kai, GREENBERG Jane, DUNIC Jullian (2019). Data objects and documenting scientific processes: An analysis of data events in biodiversity data papers. Preprint accepted by *JASIST*. Disponible sur : <https://arxiv.org/abs/1903.06215> (Consulté le 26/07/2019)

MESRI (2018). *National Plan for Open Science*. Paris, Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. Disponible sur : https://libereurope.eu/wp-content/uploads/2018/07/SO_A4_2018_05-EN_print.pdf (Consulté le 26/07/2019)

MICHENER William K., BRUNT James W., HELLY John J., KIRCHNER Thomas B., STAFFORD Susan G. (1997). Nongeospatial metadata for the ecological sciences. *Ecological Archives*, vol. 7, no. 1, pp. 330-342. Disponible sur : <https://doi.org/10.2307/2269427> (Consulté le 26/07/2019)

Data from:

“AccessGrey, Securing Open Access to Grey Literature for Science and Society”

<https://doi.org/10.17026/dans-zzf-cje3>

Dominic Farace, GreyNet International, Netherlands

<https://orcid.org/0000-0003-2561-3631>

Project Administration/ Dataset Creator / Draft Text

Jerry Frantzen, GreyNet International, Netherlands

<https://orcid.org/0000-0002-3405-7078>

Visualization

Stefania Biagioni, Institute of Information Science and Technologies, Italy

<https://orcid.org/0000-0001-9518-0267>

Reviewer

Carlo Carlesi, Institute of Information Science and Technologies, Italy

<https://orcid.org/0000-0001-9808-6268>

Reviewer

Abstract

The data collected is based on the answers of 56 respondents to an online questionnaire carried out among a diverse population of stakeholders within GreyNet International, the Grey Literature Network Service¹. The dataset consists of responses to ten questions, the last of which is open ended and all of which allow for additional comments. The first nine questions deal with the uses, applications, and opinions related to persistent identifiers². Eight of the questions deal with the DOI³ and one pertains to the ORCID⁴ or other type of researcher ID. The data was collected in the second quarter of 2019 via SurveyMonkey, where it remains stored along with a copy housed in the DANS Easy Archive⁵.

Keywords: Data Paper; Persistent Identifiers; DOI; ORCID

Subject Area: Information Science; Grey Literature

Methods Applied

● Steps

In formulating the questions that would comprise the instrument used in this study that is geared to an understanding of persistent identifiers in the field of grey literature, a search was carried out in GLP, Collection of Conference Papers on Grey Literature housed in the GreyGuide Repository⁶. The search terms used were ‘persistent identifier’ and ‘DOI’. The search retrieved 16 full-text metadata records and enabled drafting the questions for this survey. The final edited version of the online questionnaire was entered in SurveyMonkey⁷ and the link generated was emailed to a defined population.

● Sampling strategy

The population of this stakeholder survey was drawn from GreyNet’s Distribution List spanning entries from January 2014 to April 2019 i.e. when the link to the online survey became operational. Only personal names with both surname and first name or initial were selected. The total population of survey recipients was 509. Reminder emails were sent out during the 5-week period in which the link to the survey was online accessible.

Survey Population	Survey Respondents	Survey Results %
2014 - (April 2019)		
509	56	11,0%

● Quality Control

There was no specific control carried out on the data acquired from the survey. None of the categories of responses were grouped or otherwise normalized. The number of questions where responses were skipped was negligible. Over half of the respondents provided their contact details recorded in the final question of the stakeholder survey. And, there is no cause to question the validity of the responses given that the respondents belonged to a defined survey population.

Dataset Description

File name:	Survey Results (Anonymous)
Format name and version:	PDF Standard
Creation dates:	from 2019-04-30 to 2019-06-04
Language:	English
License:	CC0 Waiver - no rights reserved
Archive name:	DANS EASY Archive
Publication date:	2019-08-12

Potential Reuse of the Data

The data collected from this stakeholder survey allows for potential reuse and further analysis by way of a follow-up study. The survey data may also be of interest to other communities of practice pertaining to persistent identifiers both within the field of grey literature as well as other fields of information science. On a more critical note, the data is limited by the number of respondents to the survey – 56 in total. While the percentage of responses is within the acceptable range of 10%-15%, it was a stakeholder survey and the total response of 11% remains on the lower end. In fact, the number of respondents does not formally allow for the expression of results in percentages. The data however remains preserved in a national archive, which carries the CoreTrustSeal⁸ and by way of this data paper demonstrates compliance with the FAIR principles⁹.

Linked References

¹ GreyNet International <http://www.greynet.org/>

² https://en.wikipedia.org/wiki/Persistent_identifier

³ https://en.wikipedia.org/wiki/Digital_object_identifier

⁴ <https://en.wikipedia.org/wiki/Wikipedia:ORCID>

⁵ DANS EASY Archive <https://easy.dans.knaw.nl/ui/home>

⁶ <http://greyguiderep.isti.cnr.it/listtitoli.php?authority=GreyGuide&collection=GLP&langver=en&RighePag=100>

⁷ SurveyMonkey, A Survey Platform <https://www.surveymonkey.com/home/>

⁸ <https://www.coretrustseal.org/wp-content/uploads/2018/04/DANS-Electronic-Archiving-SYstem-EASY-.pdf>

⁹ <https://www.force11.org/group/fairgroup/fairprinciples>



Repository and Portal to Good Practices and Resources in Grey Literature

Research and Knowledge Sharing in the Field of Grey Literature 2013 - 2019

Stefania Biagioni and Carlo Carlesi, CNR-ISTI-Pisa Italy
Dominic Farace, GreyNet International, Amsterdam, NL

Welcome to the GreyGuide: point of access to Grey Literature and Open Access Resources

Scenario

The GreyGuide - Repository and Portal to Good Practices and Resources in Grey Literature was launched in 2013 as a collaborative effort between GreyNet International and CNR-ISTI, Pisa, Italy.

Objective

To manage Open Source Repositories and provide a unique resource in the field of grey literature that is long awaited and which responds to the information needs of a diverse, international grey literature community adhering to Open Science guiding principles.

Ongoing Activity

- **Grey Guide Portal** managing and upgrading;
- **GreyForum Series** enabling access to material produced by the speakers;
- **DOI metadata field** for GL-Conference Papers and RGL documents in the GreyGuide Repository;
- **AccessGrey Project**.

2013-2017

- Repository
- Portal
- Document share

<http://greypguide.isti.cnr.it/>

2018-2019

- **GreyForum Series**
- **DOI Minting Service**
- **AccessGrey Project**



GreyGuide is GreyNet's Web Access Portal and Repository

With the recent addition of the RGL Collection (Resources in Grey Literature), the acquisition of generic, multidisciplinary, and diverse grey literature documents types is underway. Each new metadata, full-text record is assigned a DOI and system generated citation.

RGL Resources in Grey Literature:

- ✓ *Revise RGL online Template, Include DOI metadata field, Include System Generated Citation, Assign DOIs to existing records in RGL*
- ✓ *Merge GGP records with RGL Collection, Assign DOIs to the merged records*

Your Platform for Grey Literature

07 October 2019

GREY LITERATURE PUBLICATIONS

Your Platform for Grey Literature

1 COLLECTION

Multidisciplinary Document Types

• RGL Resources in Grey Literature

The GreyGuide Repository now welcomes your full range of document types such as *reports, theses, working papers, etc.* GreyNet as content provider now seeks to bring together work by the international grey literature community that has been produced both within and outside the GL-Conference Series.

Each new entry will receive a minted DOI as well as a system generated citation!

Just 3-EASY STEPS:

- Click <http://greypguide.isti.cnr.it/userarea.php?lang=en>
- Select RGL: Resources in Grey Literature
- Complete the online template and don't forget to press the submit button

Thank You!

AccessGrey Project

Securing Open Access to Grey Literature for Science and Society

Persistent Identifiers and Grey Literature

The Stakeholder Survey seeks to understand the influence that persistent identifiers have within grey literature communities.

Goal

- To learn the opinions, uses, and applications of persistent identifiers within the grey literature community
- To explore the use of persistent identifiers, namely the DOI, in the acquisition of grey literature

GreyForum Series Portal

GreyForum is a series of courses, seminars, and workshops where grey literature provides common ground for information professionals in the process of knowledge.

The GreyForum Series is now accessible at <http://greyforum.isti.cnr.it>



Share Research and Knowledge in the field of Grey Literature via GreyGuide Portal

New Entry

A Directory of Grey Literature Research Guides, Online Courses and Webinars is now online accessible via the GreyGuide Portal (Web Resources).

PARTNERS

GreyNet

<http://www.grey.net.org/>



CNR ISTI
<http://www.isti.cnr.it>

CONTENT PROVIDER

Dominic Farace
GreyNet
International
Netherlands

SYSTEM MANAGER

Stefania Biagioni
CNR-ISTI
NeMIS Lab
Italy

TECHNICAL DEVELOPER

Carlo Carlesi
CNR-ISTI
NeMIS Lab
Italy

COLLECTIONS

BIO: Who is in
Grey Literature

GLA: Conference
Proposals

GLP: Conference
Papers

RGL: Resource
in Grey
Literature



Remember
to endorse
The Pisa Declaration

National Repository of Grey Literature (NRGL)



NRGL is

digital
repository
for grey
literature

Free

online
access

Features

Provider:

National Library of Technology
Prague, Czech Republic

Records:

over 400,000 records

Collection provenance:

Czech Republic

Partners:

over 130 organizations (Academy of Science,
Public Research Institutions, Universities, State
Offices, Libraries, NGOs etc.)

International Cooperation:

OpenGrey, OpenAire, ROAR, OpenDOAR, BASE

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

What else?

Conference on Grey Literature
and Repositories

<http://nrgl.techlib.cz/conference/>

Informative Web pages

<http://nrgl.techlib.cz>

www.nusl.cz

NTK
4x 2,5x 4,5x
Národní technická knihovna
National Library of Technology

NUŠL
národní
úložiště
šedé
literatury



International Journal on Grey Literature

Subscription Order Form

For *GreyNet* Members free or reduced rates, see <http://www.greynet.org/membership.html>

TGJ Volume 16, 2020	Type of Subscription:	Amount in Euros	Total
THE GREY JOURNAL - PDF/ Email / PWP ISSN 1574-180X Annual Subscription, including Electronic Handling	☐ Institutional	€ 240	€

Customer information

Name:	
Organisation:	
Postal Address:	
City/Code/Country:	
E-mail Address:	

Check one of the boxes below for your Method of Payment:

- ☐ Direct transfer to TextRelease, Rabobank Amsterdam, Netherlands
BIC: RABONL2U IBAN: NL70 RABO 0313 5853 42, with reference to 'TGJ/The Grey Journal'
- ☐ MasterCard/Eurocard ☐ Visa card ☐ American Express
- Card No. _____ Expiration Date: _____
- Print the name that appears on the credit card, here _____
- Signature: _____ CVC II code: _____ (Last 3 digits on signature side of card)
- Place: _____ Date: _____

Note: Credit Card transactions will be authorized via Ingenico | Ogone, a designated payment service.

Correspondence Address:

TextRelease

Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
Tel +31-(0) 20-331.2420 • info@textrelease.com • www.textrelease.com

Author Information

Artini, Michele

Michele Artini is a research fellow at the Networked Multimedia Information Systems laboratory of the "Istituto di Scienza e Tecnologie dell'Informazione", Consiglio Nazionale delle Ricerche, Pisa, Italy. Since 2005 he is involved in EC funded projects for the realisation of aggregative data infrastructures like DRIVER, DRIVER II, BELIEF, HOPE, EFG, EFG1914, OpenAIRE, OpenAIREPlus and Openaire2020. He is interested in digital libraries, service-oriented infrastructures, database systems and workflow management systems.

Email: michele.artini@isti.cnr.it

Atzori, Claudio

Claudio Atzori received his MSc in "Information Technology" in 2009 at the University of Cagliari and completed his PhD at the Information Engineering at the Engineering School "Leonardo da Vinci" of the University of Pisa in 2016. He works as research fellow in the InfraScience research group, part of the Multimedia Networked Information System Laboratory (NeMIS), at the "Istituto di Scienza e Tecnologie dell'Informazione" (ISTI), National Research Council (CNR), in Pisa, Italy. He works on the realisation of aggregative data infrastructures for the Open Science and Scholarly Communication. He has also participated to the EC funded R&D projects: DRIVER-II, EFG, EFG1914, HOPE, EAGLE, OpenAIRE, OpenAIRE-Plus, OpenAIRE2020, OpenAIRE-Connect.

Email: claudio.atzori@isti.cnr.it

Baglioni, Miriam

Miriam Baglioni works as a research fellow at Networked Multimedia Information Systems (NeMIS) Laboratory of the Italian National Research Council - Institute of Information Science and Technologies (CNR-ISTI). She graduated in Computer Science with honors, in 2001, from the University of Pisa and received her PhD in Computer Science, in July 2005 from the University of Pisa. She was involved in the EU funded projects BRITE, MUSING, and GeoPKDD, and in the projects ClickWorld, Dantex, TetraModel and BiNet. She is currently collaborating at the EU funded projects EFG1914, OpenAIRE2020 and OpenAire-connect. She has worked on Data Mining, Knowledge Discovery, ontologies, social networks and bioinformatics. Her current research interests include data e-infrastructure for science, and science reproducibility. Email: miriam.baglioni@isti.cnr.it

Bardi, Alessia

Alessia Bardi is a researcher at Networked Multimedia Information Systems (NeMIS) Laboratory of the Italian National Research Council - Institute of Information Science and Technologies (CNR-ISTI). She graduated in Computer Science in 2009 at University

of Pisa and completed a PhD in Information Engineering at the Engineering Ph.D. School "Leonardo da Vinci" of the University of Pisa in 2016. She is involved in EC funded projects for the realisation aggregative data infrastructures. Her research interests include service-oriented architectures and data infrastructures for e-science and scholarly communication.

Email: alessia.bardi@isti.cnr.it

Bartolini, Roberto

Roberto Bartolini - Expertise on design and development of compilers of finite state grammars for functional analysis (macro-textual and syntactic) of Italian texts. Expertise on design and implementation of compilers of finite state grammars for analysis of natural language texts producing not recursive syntactic constituents (chunking) with specialization for Italian and English languages. Skills on acquiring and extracting domain terminology from unstructured text. Skills on semi-automatic acquisition of ontologies from texts to support advanced document management for the dynamic creation of ontologies starting from the linguistic analysis of documents.

Email: roberto.bartolini@ilc.cnr.it

Biagioni, Stefania

Stefania Biagioni graduated in Italian Language and Literature at the University of Pisa and specialized in Data Processing and DBMS. She is currently an associate member of the research staff at the Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (ISTI), an institute of the Italian National Research Council (CNR) located in Pisa. She is currently involved in the activities of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She has been head librarian of the Multidisciplinary Library of the CNR Campus in Pisa till August 2017. She was the responsible of ERCIM Technical Reference Digital Library (ETRD) Project and currently is the coordinator of the PUMA (PUBlication Management) & MetaPub, a service oriented and user focused infrastructure for institutional and thematic Open Access repositories looking at the DRIVER/OpenAIRE vision, <http://puma.isti.cnr.it>. She has coauthored a number of publications dealing with digital libraries and grey literature. Her research interest are focused on digital libraries, knowledge sharing and transfer in scientific area, scholarly communication infrastructures, Open Access and Open Science. She has been dealing with grey literature since 90's. Since 2013 she is involved on the GreyGuide Project.

Email: stefania.biagioni@isti.cnr.it

ORCID iD <https://orcid.org/0000-0001-9518-0267>

Author Information *CONTINUED*

Carlesi, Carlo

Carlo Carlesi, graduated in Computer Science, worked since 1970 at the IEI (now ISTI) of the CNR in Pisa. He is currently a Research Associate of the Institute ISTI and he is involved in the following projects: PUMA - Publication Management. The Digital Library service allows public access (when permitted) through Internet to the published documents produced by CNR Organizations. And GreyGuide, portal and repository of good practice and resources in the field of grey literature.

Email: carlo.carlesi@isti.cnr.it

ORCID iD <https://orcid.org/0000-0001-9808-6268>

De Bonis, Michele

Michele De Bonis is a research fellow at Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. He received his MSc in Computer Engineering in the year 2017 at the University of Pisa, Italy. He is currently researching in the areas of clustering methods. He is currently working for the developing of digital libraries and deduplication of the authors for the OpenAIRE project. Email: michele.debonis@isti.cnr.it

Etkin, Cynthia

Cynthia Etkin is the Senior Program Planning Specialist in the Office of the Superintendent of Documents (SOD), U.S. Government Publishing Office (GPO). She brings to this position 21 years of experience at GPO and almost 15 years of experience managing Federal depository and law library operations in academic libraries. Cynthia's job focuses on policy and planning for the public information programs of the SOD directly supporting GPO's mission of "Keeping America Informed" and the public's right to freely access its government's information. Recent efforts include preparing for GPO's trustworthy digital repository audit, developing the National Plan for Access to U.S. Government Information, and implementing an eLearning platform. She continues to work on implementing the National Plan while anxiously following Congressional efforts to modernize the statutory authority of the Depository Library Program. Email: cetkin@gpo.gov

Farace, Dominic

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 (1993-2018). 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

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, Grey Literature Network Service, as a freelance technical consultant.

Email: info@greynet.org

ORCID iD <https://orcid.org/0000-0002-3405-7078>

Goggi, Sara

Sara Goggi is a technologist at the Institute of Computational Linguistics "Antonio Zampolli" of the Italian National Research Council (CNR-ILC) in Pisa. She started working at ILC in 1996 working on the EC project LE-PAROLE for creating the Italian reference corpus; afterwards she began dealing with the management of several European projects and nowadays she is involved with organisational and managerial activities mainly concerning international relationships and dissemination as well as organization of events (e.g. LREC conference series). Currently one of her preminent activities is the editorial work for the international ISI Journal Language Resources and Evaluation, being its Assistant Editor. Since many years (from 2004) she also carries on research on terminology and since 2011 - her first publication at GL13 - she is working on topics related with Grey Literature. Email: sara.goggi@ilc.cnr.it

La Bruzzo, Sandro

Sandro La Bruzzo is a research fellow at Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. He received his MSc in Information Technologies in the year 2010 at the University of Pisa, Italy. Today he is a member of the InfraScience research group, part of the Multimedia Networked Information System Laboratory (NeMIS). His current research interests are in the areas of Service-Oriented Infrastructures for Digital Libraries, protocols for metadata exchanging, Database, Index. He is currently working for the development of the Digital Library and Data

Author Information

infrastructures for the European Commission projects OpenAIRE, OpenAIREplus, OpenAIRE2020, EFG1914, HOPE, and EAGLE. Email: sandro.labruzzo@isti.cnr.it

Manghi, Paolo

Paolo Manghi is a (PhD) Researcher in computer science at Istituto di Scienza e Tecnologie dell'Informazione (ISTI) of Consiglio Nazionale delle Ricerche (CNR), in Pisa, Italy. He is the technical director of the OpenAIRE infrastructure, technical manager and researcher for the EU-H2020 infrastructure projects OpenAIRE2020, SoBigData.eu, PARTHENOS, EOSC, and RDA Europe, and he is the scientific coordinator of the OpenAIRE-Connect project. He is active member of a number of Data Citation and Data Publishing Working groups of the Research Data Alliance; and invited member of the advisory boards of EC project and the Research Object initiative. His research areas of interest are today data e-infrastructures for science and scholarly communication infrastructures, with a focus on technologies supporting open science publishing, i.e. computational reproducibility and transparent evaluation of science. Email: paolo.manghi@isti.cnr.it

Molino, Anna

Anna Molino graduated in Linguistics at the University of Pisa in 2010. Since 2013, she works at CNR - ISTI ("Istituto di Scienza e Tecnologie dell'Informazione - A. Faedo") as member of the Networked Multimedia Information Systems Lab. (NeMIS). She has worked as project assistant and financial manager in various EU funded and national research projects for the Digital Libraries group of the NeMIS lab. She contributes in the language revision and translation of scientific papers. Email: anna.molino@isti.cnr.it

Monachini, Monica

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. Email: Monica.Monachini@ilc.cnr.it

Pardelli, Gabriella

Gabriella Pardelli was born at Pisa, graduated in Arts in 1980 at the Pisa University, submitting a thesis on the History of Science. Since 1984, researcher at the National Research Council, Institute of Computational Linguistics "Antonio Zampolli" ILC, in Pisa. Head of the Library of the ILC Institute since 1990. Her interests and activity range from studies in grey literature and terminology, with particular regard to the Computational Linguistics and its related disciplines, to the creation of documentary resources for digital libraries in the humanities. She has participated in many national projects. Member of board at Institute for Computational Linguistics. She is author and co-author a number of publications dealing with Computational Linguistics, Computational Terminology and Grey Literature. Email: gabriella.pardelli@ilc.cnr.it

Prost, Hélène

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. Email: helene.prost@inist.fr

Schöpfel, Joachim

Joachim Schöpfel is senior lecturer at the Department of Information and Library Sciences at the Charles de Gaulle University of Lille 3 and Researcher at the GERiICO laboratory. He is interested in scientific information, academic publishing, open access, grey literature and eScience. He is a member of GreyNet and euroCRIS. He is also the Director of the National Digitization Centre for PhD Theses (ANRT) in Lille, France.

Email: joachim.schopfel@univ-lille3.fr

ORCID iD <https://orcid.org/0000-0002-4000-807X>

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

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

Specific

- Endnotes are preferred and should be numbered
- Hyperlinks need the accompanying name of resource; a simple URL is not acceptable
- If the citation is to a corporate author, the acronym takes precedence
- If the document type is known, it should be stated at the close of a citation.
- 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.



International Journal on Grey Literature

'ADVANCING OPEN SCIENCE AND GREY LITERATURE'

OpenAIRE: Advancing Open Science	141
<i>Paolo Manghi, Michele Artini, Claudio Atzori, Miriam Baglioni, Alessia Bardi, Sandro La Bruzzo, and Michele De Bonis; Institute of Information Science and technologies, Italian National Research Council, Italy; Harry Dimitropoulos, Ioannis Foufoulas, Katerina Iatropoulou, Natalia Manola, and Stefania Martziou; Athena Research Center in Information, Communication and Knowledge Technologies, Greece; Pedro Principe, University of Minho, Portugal</i>	
Semantic Query Analysis from the Global Science Gateway	147
<i>Sara Goggi, Gabriella Pardelli, Roberto Bartolini, and Monica Monachini, ILC-CNR, Italy; Stefania Biagioni and Carlo Carlesi, ISTI-CNR, Italy</i>	
The U.S. Government Publishing Office: Keeping America Informed in the 21st Century and Beyond ...	157
<i>Cynthia Etkin, U.S. Government Publishing Office, GPO, United States</i>	
Analysis of folk literature in grey literature from the National Library of China	163
<i>Cui Yue, National Library of China, NLC, China</i>	
Data from "Data Papers as a New Form of Knowledge Organization in the Field of Research Data"	170
<i>Joachim Schöpfel, University of Lille, GERiCO laboratory, France; Dominic Farace, GreyNet International; Hélène Prost, CNRS, GERiCO laboratory, France; Antonella Zane, University of Padova, Italy</i>	
Data from "AccessGrey, Securing Open Access to Grey Literature for Science and Society"	173
<i>Dominic Farace and Jerry Frantzen, GreyNet International, Netherlands; Stefania Biagioni and Carlo Carlesi, Institute of Information Science and Technologies, Italy</i>	

Colophon.....	136
Editor's Note.....	139
Advertisements	
GL2020 Pre-Conference Announcement.....	138
OpenAIRE: Advancing Open Science.....	140
WorldWideScience.org: An International Partnership Supporting Open Science.....	156
GreyGuide: Research and Knowledge Sharing in the Field of Grey Grey Literature.....	175
NTK, National Library of Technology, Czech Republic	176
Journal Subscription Form 2020.....	177
Author Information.....	178
Notes for Contributors.....	181