



WP5 Open Science Support

Birger Larsen
Aalborg University, Copenhagen



Agenda

01 Introduction

WP5 on Open Science support

02 Open Science efforts

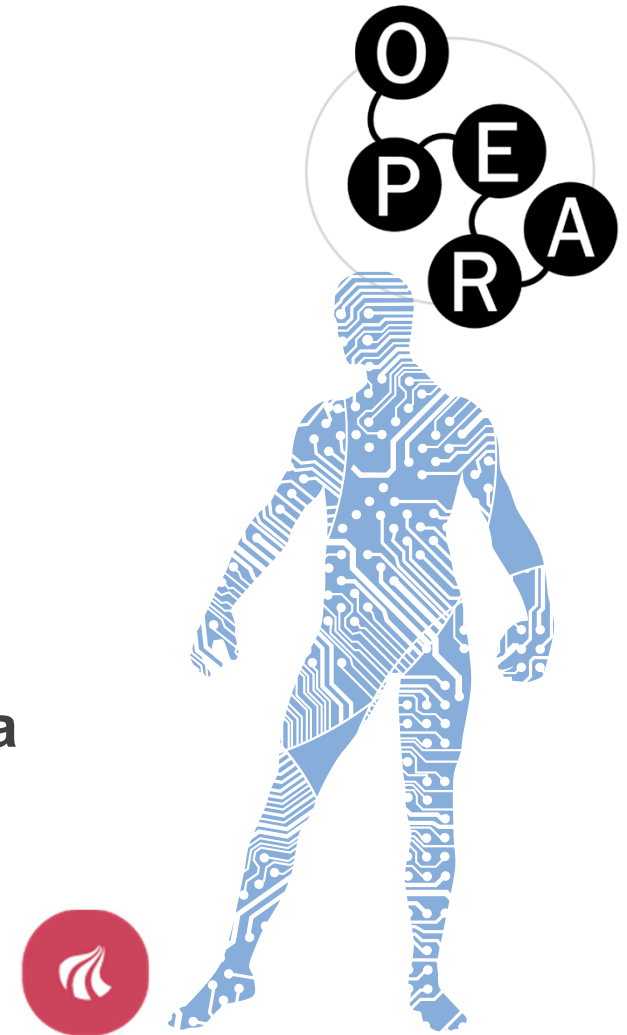
def. Open Science, Complexity of OS, Measuring OS?

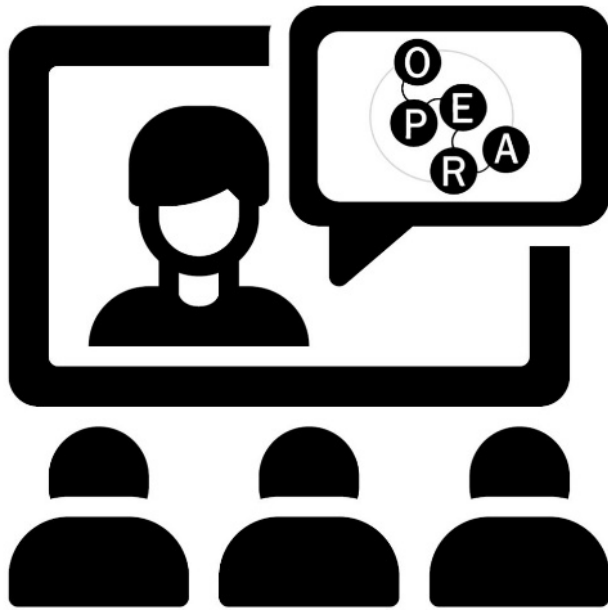
03 Data and indicators for Open Data

Initial WP5 results, examples of data and indicators

04 Concluding discussion

Summary of issues





Introduction

The OPERA project
WP5 on Open Science support



WP5 Open Science Support

WP5 aims at finding and evaluating ways Open Science efforts may form part of research analytics, metrics and evaluation
– and to prepare the inclusion of some of these approaches in analytics platforms like NORA
- and test them

Status:

- **review** of open science manifestos - in draft
- **review** of open data indicators - ready
- **usability test of NORA** with relevant stakeholders - planned

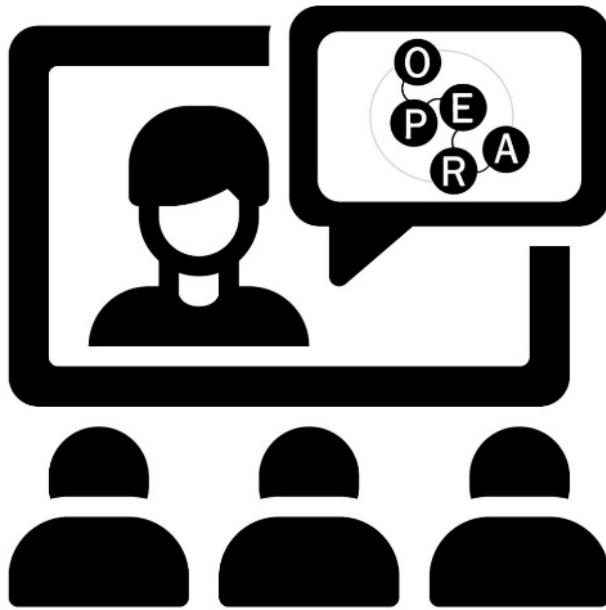


WP5 Open Science Support



Participants

- **Birger Larsen** (WP lead), Aalborg University
- **David Budtz** (deputy lead), Aalborg University
- **Pelle Annfeldt Israelsson**, Aalborg University
- **Mogens Sandfær**, Technical University of Denmark
- **Martin Collin**, Technical University of Denmark
- **Karen Hytteballe Ibanez**, Technical University of Denmark
- **Karsten Kryger Hansen**, Aalborg University Library
- **Nils Thideman**, Aalborg University Library
- ... and others



Open Science efforts

Defining Open Science
The complexity of Open Science
Measuring Open Science?



def. Open Science



"Open Science has the potential to increase the quality, impact and benefits of science and to accelerate advancement of knowledge by making it more reliable, more efficient and accurate, better understandable by society and responsive to societal challenges, and has the potential to enable growth and innovation through reuse of scientific results by all stakeholders at all levels of society, and ultimately contribute to growth and competitiveness of Europe."

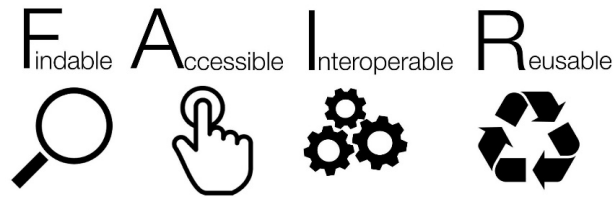
– Competitiveness Council, 2016

What?

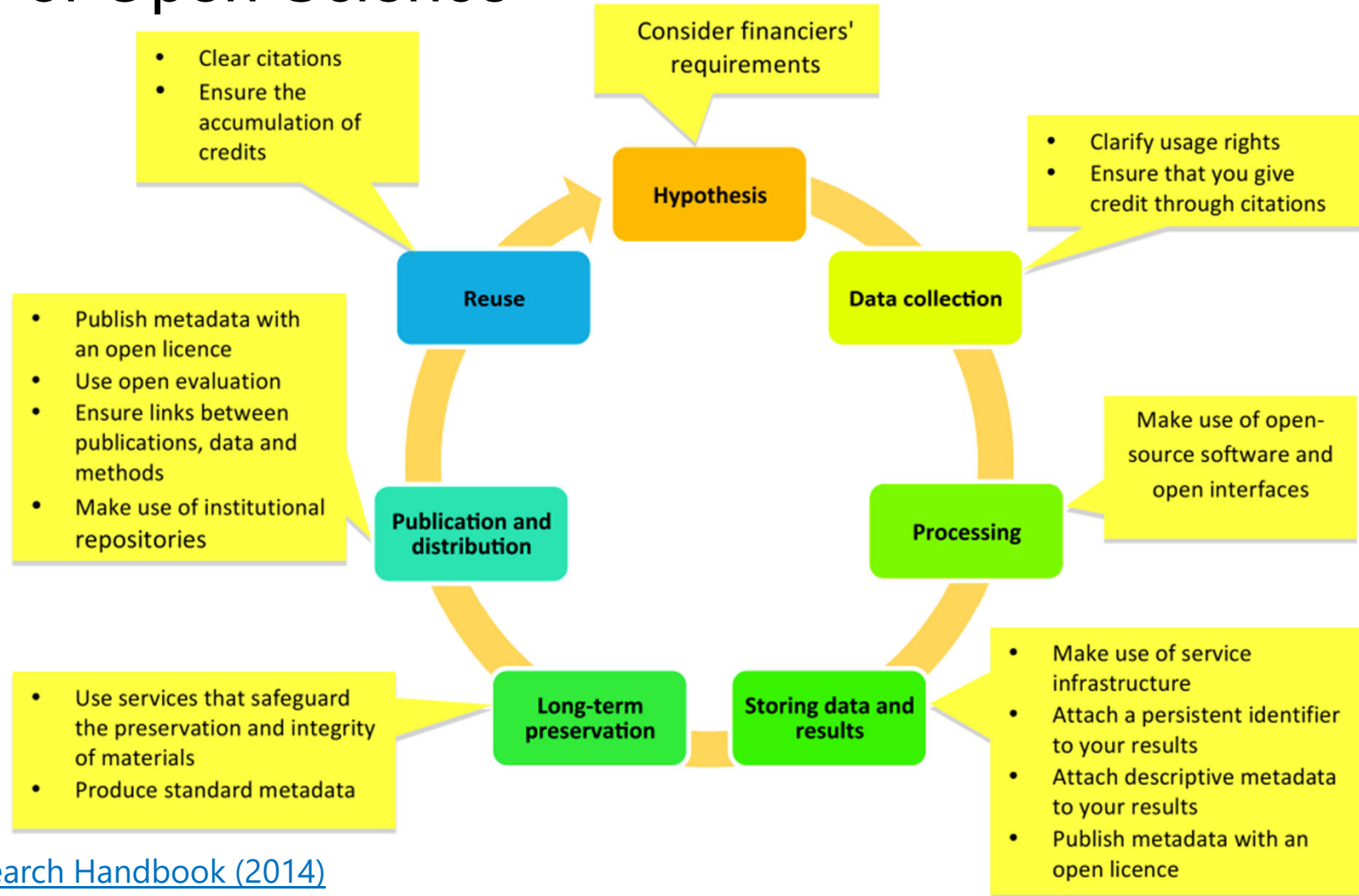
FOSTER defines Open Science (OS) as the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.

def. Open Science

- Open Science movement across scientific fields
- Manifestos and Principles
 - **Amsterdam Call for Action on Open Science** → “data sharing and stewardship” environment
 - **FAIR Guiding Principles for Open Data**
- Reproducibility crisis? → need for high quality **open** data



Complexity of Open Science



[Open Science and Research Handbook \(2014\)](#)

Complexity of Open Science

fteval JOURNAL

for Research and
Technology Policy Evaluation

September 2017, Vol. 44, pp. 50 -56
DOI: 10.22163/fteval.2017.276
© The Author(s) 2017



NEW INDICATORS FOR OPEN SCIENCE

POSSIBLE WAYS OF MEASURING THE UPTAKE AND IMPACT OF OPEN SCIENCE

DIETMAR LAMPERT, MARTINA LINDORFER, ERICH PREM, JÖRG IRRAN AND FERMÍN SERRANO SANZ

7 Open Science aspects (Lampert et al., 2017)

- **A. The scientific process**

1. Conceptualization and data gathering/creation
2. Analysis
3. Diffusion of results
4. Review and evaluation

- **B. The system level**

5. Reputation system, recognition of contributions, trust
6. Open science skills and awareness
7. Science with society

Quality of metadata	mean rating (0..10 max.)
quality of metadata (versioning, volume, data format, description of fields, etc.)	8.2
	PU R RFO

1. Conceptualization and data gathering/creation

Figure 1: Stakeholder groups - abbreviations and colour

R	researchers
RO	research (conducting) organisations
RFO	Research-funding organisations
PM	policy-makers
PU	publishers

Requirements from research funders	mean rating (0..10 max.)			
% of research funders that mandate the provision of the data / software code produced in the context of the funded activity AND who mandate the conformity to data (exchange) standards	7.9			
		RFO	PM	
Accessibility	mean rating (0..10 max.)			
accessibility of open data / code as % of all data / code produced by publicly (co-)funded projects	9.1			
		R	RO	RFO
Machine-readable	mean rating (0..10 max.)			
% of machine-readable data / metadata	7.9			
		PU	R	RFO
Availability of metadata	mean rating (0..10 max.)			
availability of explanatory metadata as % of all available data (resulting from publicly (co-)funded research)	7.5			
		PU	R	RFO
Quality of metadata	mean rating (0..10 max.)			
quality of metadata (versioning, volume, data format, description of fields, etc.)	8.2			
		PU	R	RFO
Simulation results	mean rating (0..10 max.)			
usability of simulation results (models, data, and code)	7.5			
		R	RFO	PU
Data services	mean rating (0..10 max.)			
(types of) open data services offered	8			
		PU	R	RO
Data compilation/publication costs incorporated	mean rating (0..10 max.)			
% of funded projects incorporating costs for data compilation / publication and maintenance (of the repository/data sets)	7.6			
		PM	RFO	RO
Long-term availability	mean rating (0..10 max.)			
is the (long-term) availability of the data guaranteed (availability of a sustainability plan (yes/no))	8.2			
		RFO	RO	PM
Sharing policies	mean rating (0..10 max.)			
# of sharing policies in research organisations (sharing of data, organisms, etc.)	7.6			
		RO		

Figure 2: Roles in the scientific process. Source: Liz Allen et al. (2014): Credit where credit is due; Amy Brand, Liz Allen, Micah Altman et al. (2015): Beyond authorship: attribution, contribution, collaboration, and credit.

Term	Definition
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims
Methodology	Development or design of methodology; creation of models
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components
Validation	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs
Formal Analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools
Data curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse
Writing – Original Draft	Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)
Writing – Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/data presentation
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team
Project Administration	Management and coordination responsibility for the research activity planning and execution
Funding acquisition	Acquisition of the financial support for the project leading to this publication.

6. Open science skills and awareness - e.g.
curating and maintaining large data sets

7. Science with society - promotion of the
engagement of citizens in science and research
e.g. OPERA WP2



Should we measure Open Science?

- Long awaited report...
- Very reluctant to propose concrete indicators; afraid of adverse affects
- Recommends to develop 'Indicator Frameworks' and 'Toolboxes' "...to **guide the responsible development, interpretation, and use of indicators by policy makers, research management and researchers...**"
- The **frameworks** enable the collective definition of the evaluative needs given the context of the research field and the epistemic culture in the relevant communities
- The **toolboxes**, on the other hand, are oriented towards more technical questions, and are based on the collective expertise of the relevant communities



Indicator Frameworks for Fostering Open Knowledge Practices in Science and Scholarship

Report of the Expert Group on Indicators for Researchers' Engagement with Open Science

Members and authors of the Report: *Paul Wouters (chair), Ismael Ràfols, Alis Oancea, Shina Caroline Lynn Kamerlin, J. Britt Holbrook and Merle Jacob*

Edited by Rene von Schomberg



An egocentric example...
... of the impact of Open Data
= personal motivation

Dimensions

[Sign In](#) | [Help](#)

[Search](#) | [Björger Larsen](#) | [X](#) | [Log Out](#)

FILTERS

- > GROUPS
- > PUBLICATION YEAR
- > RESEARCHER
- > FINDER
- > COUNTRY OF FUNDER
- > RESEARCH ORGANIZATION
- > LOCATION - RESEARCH ORGANIZATION
- > RESEARCH CATEGORIES
- > PUBLICATION TYPE
- > SOURCE TITLE
- > PUBLISHER
- > JOURNAL LIST
- > OPEN ACCESS

Björger Larsen
[Aalborg University · Aalborg, Denmark](#)
[View Profile](#)

Overview		Experience & Education				
Publications	Datasets	Grants	Patents	Clinical Trials		
100	0	0	0	0		

Citations
882

The information on this profile has been aggregated algorithmically from several different sources (including publication and public ORCID data).

PUBLICATIONS	DATASETS	GRANTS	PATENTS	CLINICAL TRIALS	POLICY DOCUMENTS
100	0	0	0	0	<small>selected filter not applicable</small>

Title, Author(s), Bibliographic reference - About the metrics ☐ Show abstract Sort by: Citations ▼

A review of the characteristics of 108 author-level bibliometric indicators

Lena Willgaard, Jesper W. Schneider, Björger Larsen
 2014, Scientometrics - Article

[\[Dataset\]](#) [\[12\]](#) [\[Metrics\]](#) [x](#) [View PDF](#) [% Add to Library](#) [In your ORCID record ▼](#)

Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: an empirical analysis of the potential

Gunnar Sivertsen, Björger Larsen
 2012, Scientometrics - Article

[\[Dataset\]](#) [\[6\]](#) [\[Metrics\]](#) [x](#) [Add to Library](#) [In your ORCID record ▼](#)

The publication-citation matrix and its derived quantities

Peter Ingversen, Björger Larsen, Ronald Rousseau, Jane Russell
 2001, Science Bulletin - Article

[\[Dataset\]](#) [\[4\]](#) [\[Metrics\]](#) [x](#) [Add to Library](#) [Add to ORCID](#)

The Interactive Track at INEX 2004

Anastasio Tombas, Björger Larsen, Saeeda Malik
 2005, Advances in XML Information Retrieval - Chapter

[\[Dataset\]](#) [\[42\]](#) [\[Metrics\]](#) [x](#) [Add to Library](#) [Add to ORCID](#)

FindCobies: A search engine for rare diseases

Rafel Gregoriou, Paula Petru, Christina Lomax, Björger Larsen, Henrik L. Jørgensen, Ingemar J. Cox, Lena Kai Hansen, Peter Ingversen,
 2013, International Journal of Medical Informatics - Article

[\[Dataset\]](#) [\[37\]](#) [\[Metrics\]](#) [x](#) [View PDF](#) [% Add to Library](#) [In your ORCID record ▼](#)

The polymersynthesis continuum in IR

Björger Larsen, Ingemar Jørgensen, Jørn Kjellstrand
 2006, Proceedings of the 1st international conference on Information interaction in context - IIX - Proceeding

[\[Dataset\]](#) [\[26\]](#) [\[Metrics\]](#) [x](#) [Add to Library](#) [In your ORCID record ▼](#)

The Scholarly Impact of CLEF (2000–2009)

Theodoris Tsalikis, Björger Larsen, Henning Müller, Stefan Schulzki, Erhard Duden
 2013, Information Access Evaluation, Multilinguality, Multimodality, and Virtualization - Chapter

[\[Dataset\]](#) [\[25\]](#) [\[Metrics\]](#) [x](#) [Add to Library](#) [In your ORCID record ▼](#)

A comparative study of first and all-author co-citation counting, and two different matrix reduction approaches applied for author-co-location analyses

Jesper W. Schneider, Björger Larsen, Peter Ingversen
 2009, Scientometrics - Article

[\[Dataset\]](#) [\[5\]](#) [\[Metrics\]](#) [x](#) [Add to Library](#) [In your ORCID record ▼](#)

Developing a Test Collection for the Evaluation of Integrated Search

Marianne Lykke, Björger Larsen, Håakon Lund, Peter Ingversen
 2010, Advances in Information Retrieval - Chapter

[\[Dataset\]](#) [\[34\]](#) [\[Metrics\]](#) [x](#) [Open Access](#) [% Add to Library](#) [In your ORCID record ▼](#)

→

→

About Dimensions · Likelihood · Twitter

Privacy policy · Cookie settings · Legal notices

© 2020 Digital Research & Research Solutions Ltd.

data →

data →

data →



Birger Larsen

Professor in Information Analysis and Information Retrieval, Aalborg University Copenhagen (Denmark)

Verified email at hum.aau.dk - [Homepage](#)

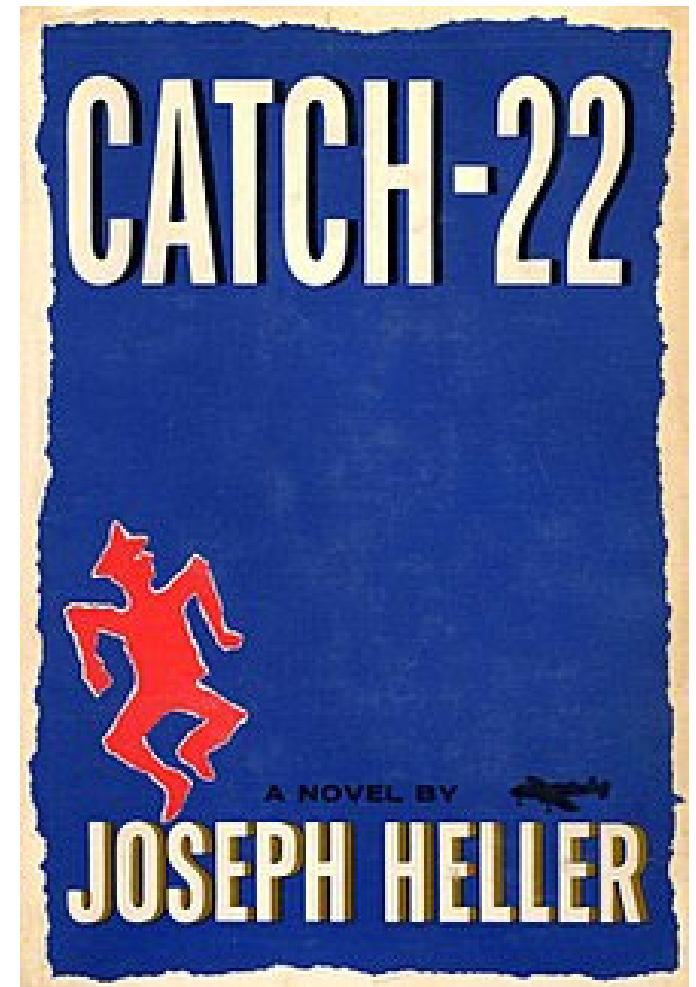
[Information Retrieval](#) [Digital Libraries](#) [Informetrics](#) [Bibliometrics](#)

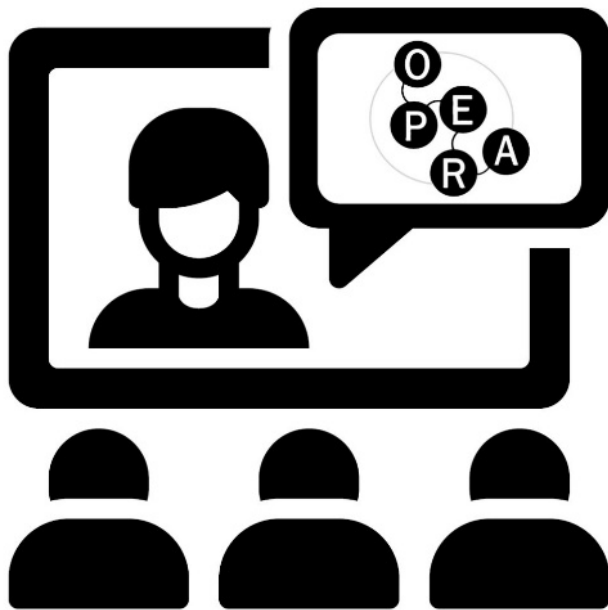
 FOLLOW

TITLE	CITED BY	YEAR
A review of the characteristics of 108 author-level bibliometric indicators L Wildgaard, JW Schneider, B Larsen Scientometrics 101 (1), 125-158	141	2014
Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential G Sivertsen, B Larsen Scientometrics 91 (2), 567-575	121	2012
The publication-citation matrix and its derived quantities P Ingwersen, B Larsen, R Rousseau, J Russell Chinese Science Bulletin 46 (6), 524-528	84	2001
Developing a test collection for the evaluation of integrated search M Lykke, B Larsen, H Lund, P Ingwersen European Conference on Information Retrieval, 627-630	72	2010
The interactive track at INEX 2004 A Tombros, B Larsen, S Malik International Workshop of the Initiative for the Evaluation of XML Retrieval ...	63	2004
FindZebra: a search engine for rare diseases R Dragusin, P Petcu, C Lioma, B Larsen, HL Jørgensen, IJ Cox, ... International Journal of Medical Informatics 82 (6), 528-538	59	2013
The polyrepresentation continuum in IR B Larsen, P Ingwersen, J Kekäläinen Proceedings of the 1st international conference on Information interaction ...	49	2006
Supporting polyrepresentation in a quantum-inspired geometrical retrieval framework I Frommholz, B Larsen, B Piwowarski, M Lailmas, P Ingwersen, ... Proceedings of the third symposium on Information interaction in context ...	48	2010
Applying diachronic citation analysis to ongoing research program evaluations P Ingwersen, B Larsen, I Wormell The Web of Knowledge: A Festschrift in Honor of Eugene Garfield, Information ...	48	2000
Influence of a performance indicator on Danish research production and citation impact 2000–12 P Ingwersen, B Larsen Scientometrics 101 (2), 1325-1344	41	2014
The interactive track at INEX 2005 B Larsen, S Malik, A Tombros International Workshop of the Initiative for the Evaluation of XML Retrieval ...	41	2005

Open Science summary

- Large movement, backed by central actors
- Hard not to agree to visions
- Huge potential impact for science and society
- Complex...
- Not for free through... Incentives?
- How to measure? Any reliable data available?





Data and indicators for Open Data

Initial WP5 results
Examples of data and indicators



OPERA WP5 Review of Existing and Proposed Indicators for Open Science Activities


- Review WP5B: examine existing and proposed indicators for Open Science activities with a focus on data sharing in fields that have a long tradition for Open Data. We aim to select the most relevant and promising indicators for inclusion in Research Analytics Platforms and Research Information Systems
- **Motivation:** A prerequisite for making data sharing visible is an understanding how agencies, organisations, platforms and repositories facilitate data sharing, either as part of the Open Sciences movement or as part of the traditions within their field
- We therefore examine central examples of how existing data portals operate and how data sharing and data citation is facilitated in them


Examples

- Physics, astronomy, space and environment research are all datacentric fields of research
 - **NASA** was chosen as a representative of how research data are shared between researchers in a multifaceted scientific community
 - The **Global Biodiversity Information Facility (GBIF)** was selected because it illustrates how data collected by researchers across the world are created and shared in order to understand nature, and as it is a good example of the needs for standardisation of datasets and data citation practices
- Several new initiatives are aiming to collect and mediate open data
 - **Mendeley data** is a new initiative from Elsevier creating a data repository connected to their existing publishing and library platform
 - **Google Dataset Search (beta)** utilises the Google search engine to identify datasets across the web and the different existing data depositories making these datasets accessible from a single-entry point



- GBIF - the Global Biodiversity Information Facility – was established in 2001 based on an OECD memorandum of understanding. GBIF is an international network and research infrastructure funded by the world's governments and aimed at **providing anyone, anywhere, open access to data about all types of life on Earth**
- The GBIF repository was created so that the knowledge for the natural world could expand and dissemination in a manner that avoids duplication of effort and expenditure. GBIF acts as coordinator and provides institutions with the common standards and open-source tools which enable participants to engage with the natural scientific community
- A typical dataset consists of counts of some species in certain locations. The current number of datasets can be seen in the GBIF search engine: at the time of writing a total of **52,434 datasets**, including 19,427 occurrence datasets, 31,237 checklist datasets, 1,457 sampling events and 303 metadata datasets
- GBIF itself is more interested in the **number of species included** its data – which cannot easily be counted as a single number but lies somewhere between 1 and 2.3 million. Also of interest is the **number of occurrences of species**, which is more than 1.5 billion in GBIF at present



[Get data](#)
[How-to](#)
[Tools](#)
[Community](#)
[About](#)


[Login](#)

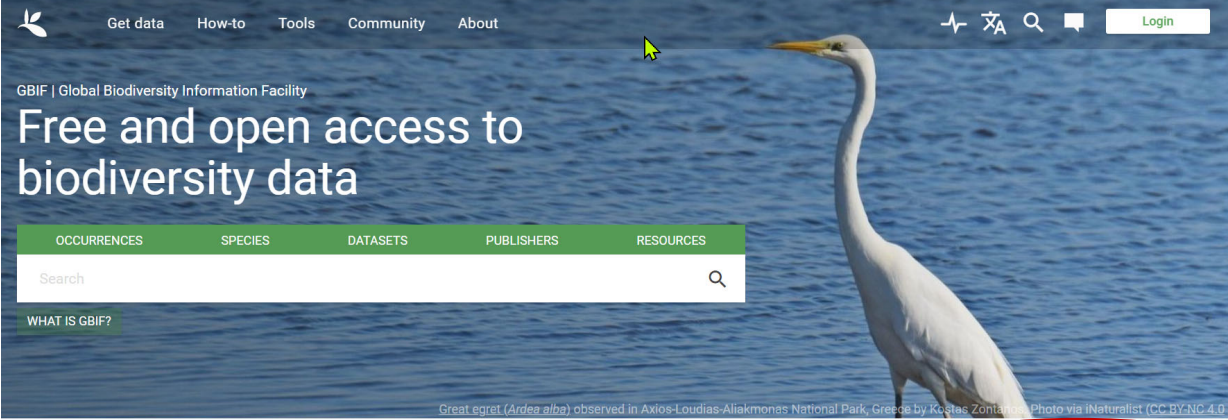
GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

[OCCURRENCES](#)
[SPECIES](#)
[DATASETS](#)
[PUBLISHERS](#)
[RESOURCES](#)



[WHAT IS GBIF?](#)




Great egret (*Ardea alba*) observed in Axios-Loudias-Aliakmonas National Park, Greece by Kostas Zontanos. Photo via iNaturalist (CC BY-NC 4.0)

Occurrence records
1,599,689,022

Datasets
54,224


Publishing institutions
1,639

**Peer-reviewed papers using data
4,682**




News

Volunteers complete Arabic, Russian, and Ukrainian translations of GBIF.org



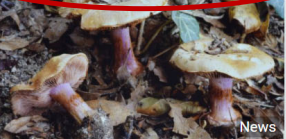
News

BID call for proposals: Sub-Saharan Africa 2020




News

Call for data papers from European Russia




News

New data-clustering feature aims to improve data quality and reveal cross-dataset connections




Data use

The evolution of cleaning behaviour in marine fishes




News

BIFA programme awards funding to nine new projects in Asia



News

Guide to publishing sequence-derived data opens for peer-review



Taxonomy

New species described: *Liodessus alto-peruensis*

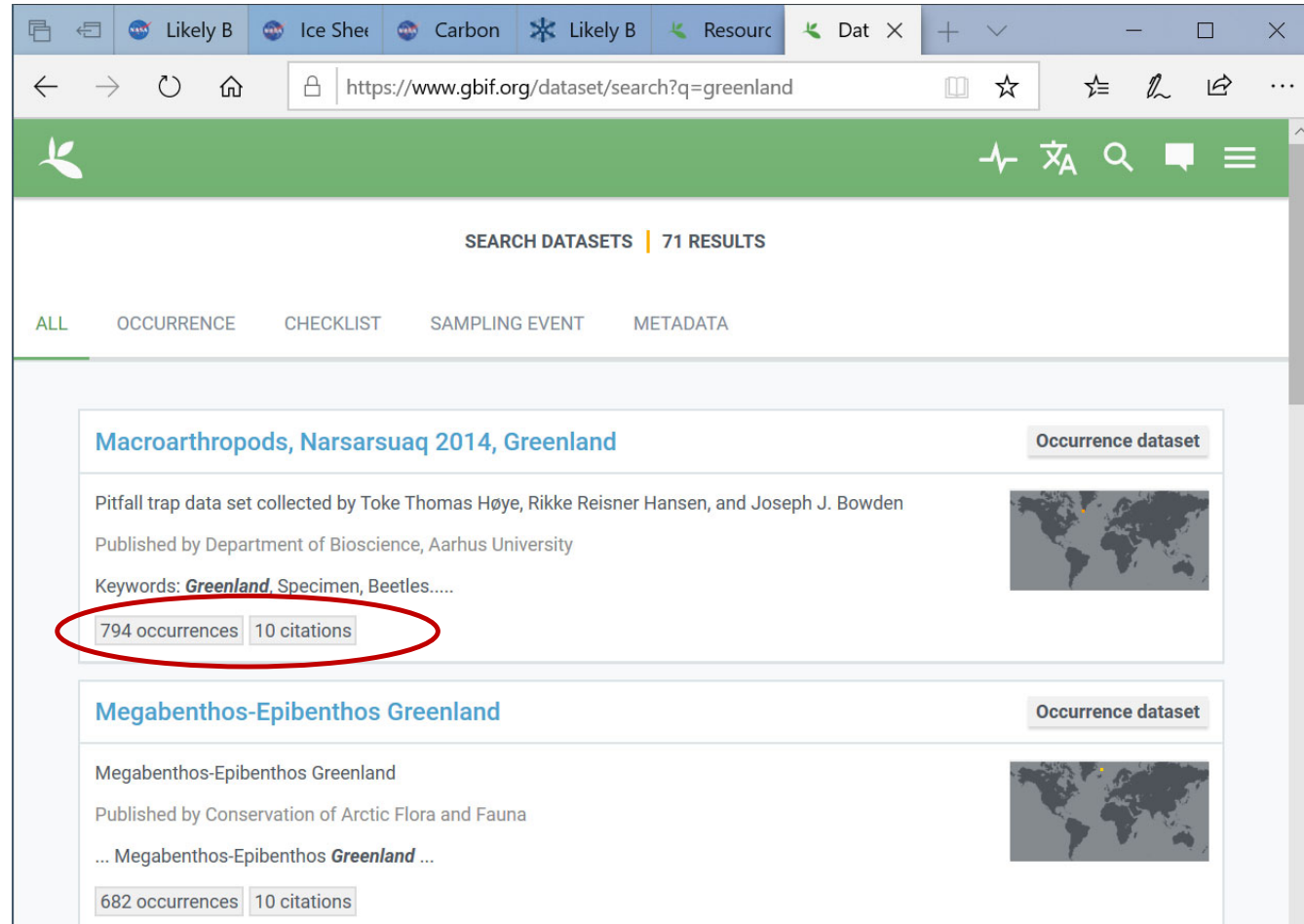
Two new species of *Liodessus* Guignot, 1939 diving beetles from Northern Peru (Coleoptera, Dytiscidae, Hydroporinae)

19 November 2020

- GBIF requires users who download individual datasets or search results and use them in research or policy to **cite them using a DOI**
- **Detailed citation guidelines are provided**, including instructions for how to cite downloads with multiple datasets, individual datasets, datasets accessed through third-party tools (such as python or R), as well as custom datasets exports
- **Users must be registered to download.** To aid users an email with dataset specific citation instructions is sent every time a dataset is downloaded, and a list of all downloaded datasets are listed in each user's profile to further aid correct citation
- Note that **downloads often consist of data selected from multiple datasets**, e.g. someone interested in bumblebees (*genus Bombus*) would get results for the over 250 species of bumblebee from datasets that include these. Such downloads with selected data from multiple datasets are assigned their own unique DOI

- GBIF also **actively searches for research uses and citations** of biodiversity information accessed through GBIF's global infrastructure
- **Daily searches** are carried out in Google Scholar, Scopus, Wiley Online Library, SpringerLink, NCBI Pubmed and bioRxiv, and the **results are curated** and added to a database from which **citation statistics** can be extracted
- These citing articles are shown on the main <http://gbif.org> search page when searching for datasets with details available on each dataset page and can also be searched directly

- *GBIF example dataset search results – including data set size and number of citing publications*



The screenshot shows a web browser window with the URL <https://www.gbif.org/dataset/search?q=greenland>. The page displays search results for datasets related to Greenland. The top navigation bar includes the GBIF logo and a search icon. The main heading is "SEARCH DATASETS | 71 RESULTS". Below this, there are tabs for "ALL", "OCCURRENCE", "CHECKLIST", "SAMPLING EVENT", and "METADATA". The "ALL" tab is selected. The first result is "Macroarthropods, Narsarsuaq 2014, Greenland", labeled as an "Occurrence dataset". It includes the description "Pitfall trap data set collected by Toke Thomas Høye, Rikke Reisner Hansen, and Joseph J. Bowden", the publisher "Published by Department of Bioscience, Aarhus University", and keywords "Greenland, Specimen, Beetles....". At the bottom of this result, a red circle highlights the statistics: "794 occurrences" and "10 citations". The second result is "Megabenthos-Epibenthos Greenland", also labeled as an "Occurrence dataset". It includes the description "Megabenthos-Epibenthos Greenland", the publisher "Published by Conservation of Arctic Flora and Fauna", and the text "... Megabenthos-Epibenthos **Greenland** ...". At the bottom of this result, the statistics are "682 occurrences" and "10 citations".

SEARCH DATASETS | 71 RESULTS

ALL OCCURRENCE CHECKLIST SAMPLING EVENT METADATA

Macroarthropods, Narsarsuaq 2014, Greenland Occurrence dataset

Pitfall trap data set collected by Toke Thomas Høye, Rikke Reisner Hansen, and Joseph J. Bowden

Published by Department of Bioscience, Aarhus University

Keywords: **Greenland**, Specimen, Beetles....

794 occurrences 10 citations

Megabenthos-Epibenthos Greenland Occurrence dataset

Megabenthos-Epibenthos Greenland

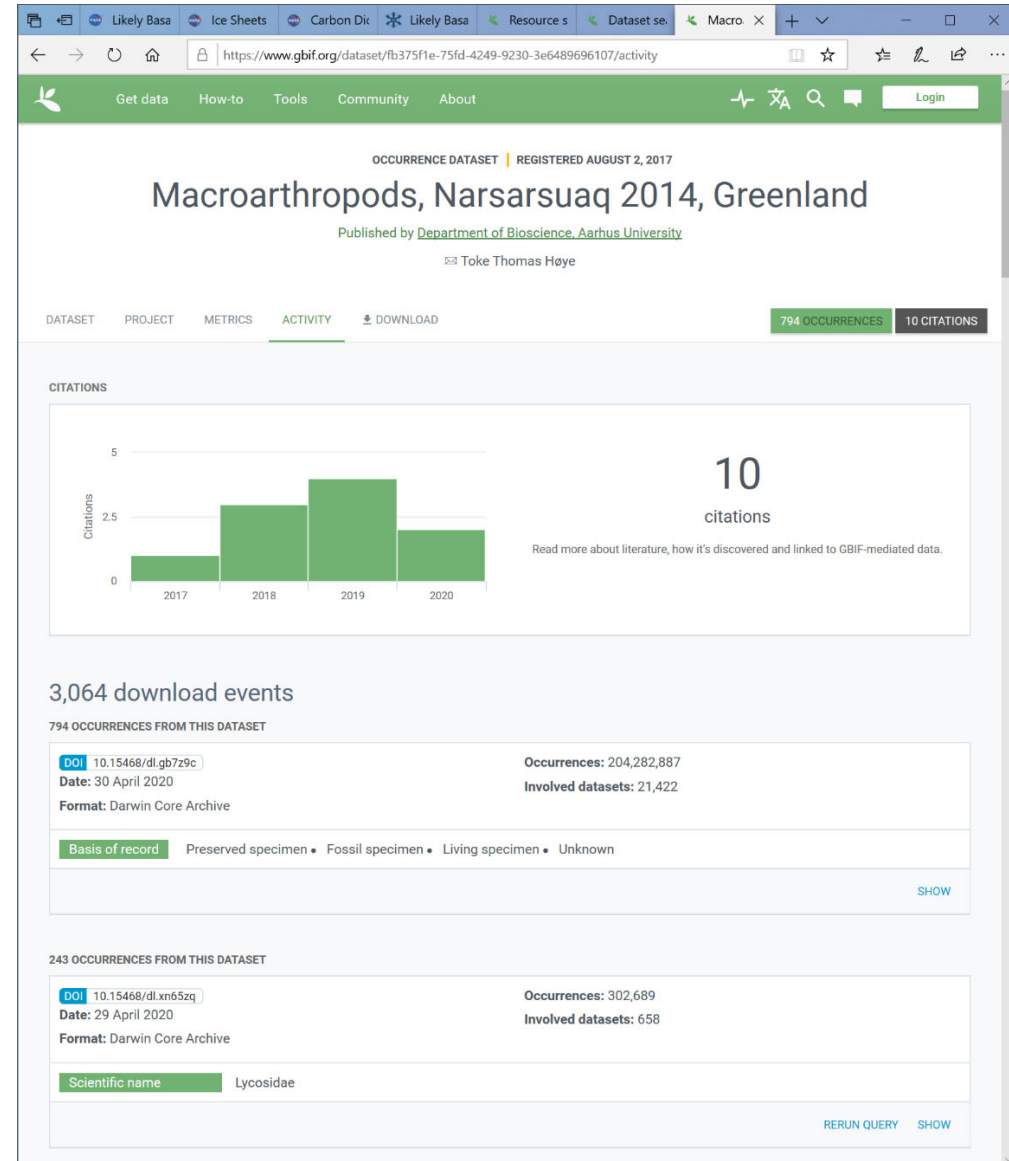
Published by Conservation of Arctic Flora and Fauna

... Megabenthos-Epibenthos **Greenland** ...

682 occurrences 10 citations

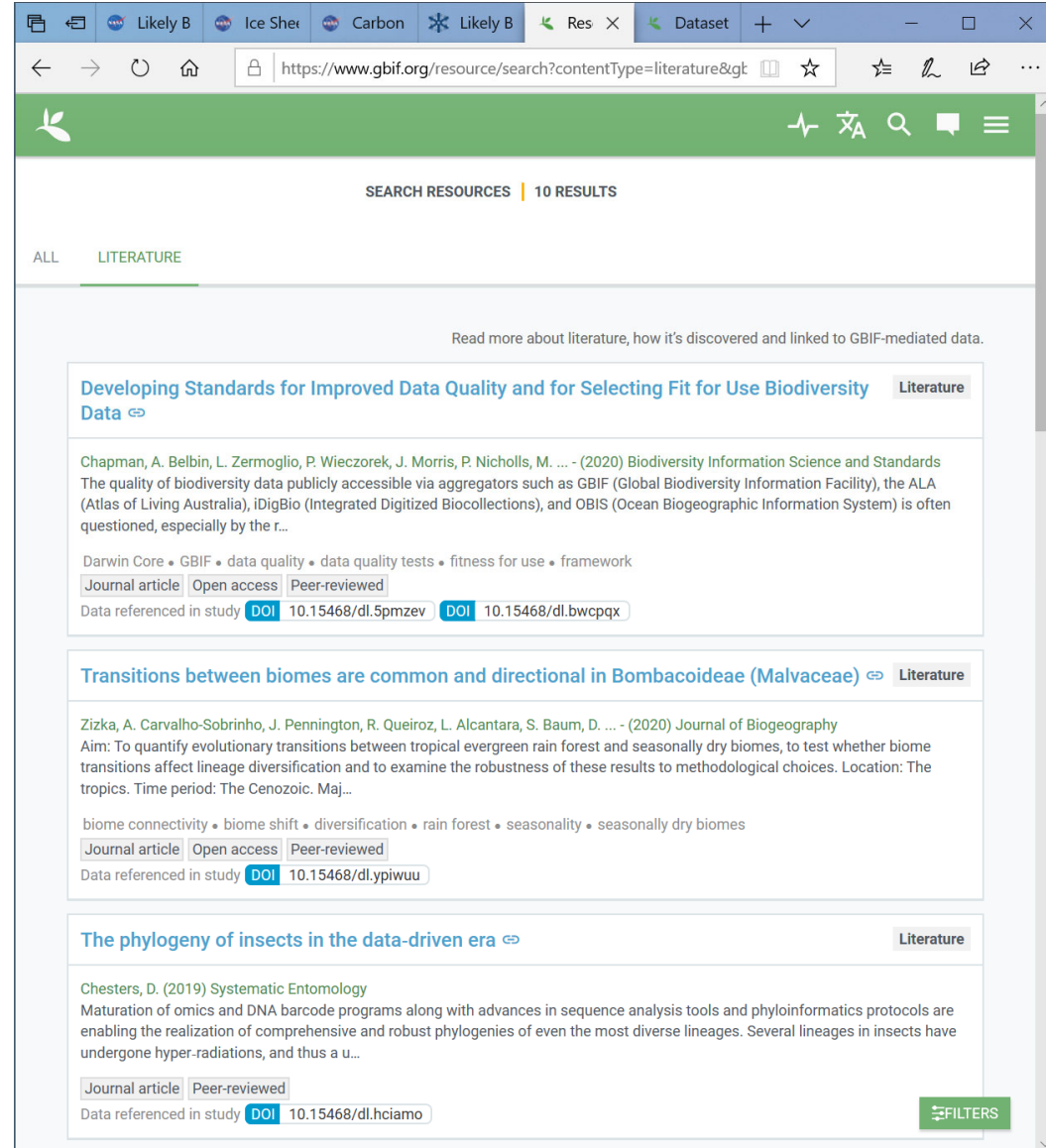
- *GBIF dataset example – with citation and download details*
- *The dataset has 794 occurrences – in some cases all were included in the 3,064 download events, in other cases only some of the occurrences*

19 November 2020



- *Example of metadata of publications citing a GBIF dataset*
- *Where possible publications are linked to external fulltexts*

19 November 2020



The screenshot shows a web browser window with the GBIF search results page. The address bar shows the URL: <https://www.gbif.org/resource/search?contentType=literature>>. The page title is "SEARCH RESOURCES | 10 RESULTS". The "LITERATURE" tab is selected. The page displays three search results, each with a title, authors, abstract, and links to the full text and data referenced in the study.

Developing Standards for Improved Data Quality and for Selecting Fit for Use Biodiversity Data Literature

Chapman, A. Belbin, L. Zermoglio, P. Wiecek, J. Morris, P. Nicholls, M. ... - (2020) Biodiversity Information Science and Standards
The quality of biodiversity data publicly accessible via aggregators such as GBIF (Global Biodiversity Information Facility), the ALA (Atlas of Living Australia), iDigBio (Integrated Digitized Biocollections), and OBIS (Ocean Biogeographic Information System) is often questioned, especially by the r...

Darwin Core • GBIF • data quality • data quality tests • fitness for use • framework
Journal article | Open access | Peer-reviewed
Data referenced in study [DOI 10.15468/dl.5pmzev](#) [DOI 10.15468/dl.bwcpqx](#)

Transitions between biomes are common and directional in Bombacoideae (Malvaceae) Literature

Zizka, A. Carvalho-Sobrinho, J. Pennington, R. Queiroz, L. Alcantara, S. Baum, D. ... - (2020) Journal of Biogeography
Aim: To quantify evolutionary transitions between tropical evergreen rain forest and seasonally dry biomes, to test whether biome transitions affect lineage diversification and to examine the robustness of these results to methodological choices. Location: The tropics. Time period: The Cenozoic. Maj...

biome connectivity • biome shift • diversification • rain forest • seasonality • seasonally dry biomes
Journal article | Open access | Peer-reviewed
Data referenced in study [DOI 10.15468/dl.ypiwuu](#)

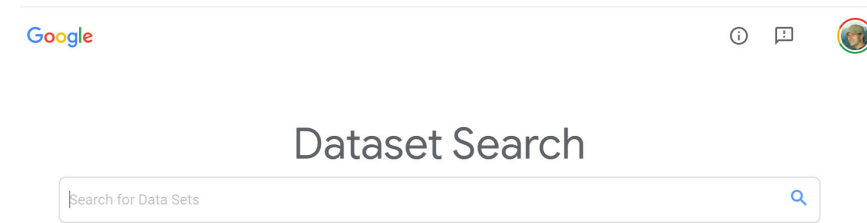
The phylogeny of insects in the data-driven era Literature

Chesters, D. (2019) Systematic Entomology
Maturation of omics and DNA barcode programs along with advances in sequence analysis tools and phyloinformatics protocols are enabling the realization of comprehensive and robust phylogenies of even the most diverse lineages. Several lineages in insects have undergone hyper-radiations, and thus a u...

Journal article | Peer-reviewed
Data referenced in study [DOI 10.15468/dl.hciamo](#)

[FILTERS](#)

Google Dataset Search



- Google Dataset Search is a new dataset search function, providing access to **datasets identified by Google on the open web**
- Datasets can be **included if they have assigned correct schema.org metadata**. Once metadata have been added, Google needs to be notified and the dataset metadata can be crawled
- **Google Dataset Search does not store the datasets themselves** but acts as a platform that links to data providers
- In case several providers provide access to the same dataset, **Google attempts to deduplicate** this and provides links from the dataset to all providers
- In addition, if the dataset is **cited in Google Scholar, the number of Google Scholar citations is shown - and links to an automatic Google Scholar search**

Google Dataset Search

- *Example dataset in Google Dataset Search – with links to data providers and to citing articles in Google Scholar*

Greenland geothermal heat flux distribution and estimated Curie Depths, links to gridded files

Explore at PANGAEA

Explore at pangaea.figshare.com

Explore at search.datacite.org

2 scholarly articles cite this data set ([View in Google Scholar](#))

19 November 2020

The screenshot shows the Google Dataset Search interface. The search bar at the top contains the word 'greenland'. Below the search bar, a list of datasets is displayed. The dataset 'Greenland geothermal heat flux distribution and estimated Curie Depths, links to gridded files' is highlighted. To the right of this dataset, a sidebar provides detailed information: it lists the unique identifier (DOI: 10.1594/PANGAEA.892973), the date the data set was updated (Aug 13, 2018), the provider (PANGAEA), the author (Yasmina M Martos), the license (Attribution 3.0 CC BY 3.0), the area covered, and a description of the data. The description mentions that the data is derived from spectral analysis of data from the World Digital Magnetic Anomaly Map2 and provides a corresponding geothermal heat flux map. Below the dataset list, there are three buttons: 'Explore at PANGAEA', 'Explore at pangaea.figshare.com', and 'Explore at search.datacite.org'. At the bottom of the dataset entry, it states '2 scholarly articles cite this data set (View in Google Scholar)'. The browser's address bar shows the search URL: datasetsearch.research.google.com/search?query=greenland&docid=wjJGT719zHZsu3WAAAAA...

Updated Apr 22, 2020

Greenland Internet Speed
tradingeconomics.com
Updated Oct 25, 2017

Emigration from Greenland
2019, by destination country
www.statista.com
Updated Apr 28, 2020

Iceland exports from Greenland
tradingeconomics.com
Updated Jun 7, 2017

Greenland geothermal heat flux
distribution and estimated...
doi.pangaea.de
pangaea.figshare.com
+1more
tsv, html
Updated Aug 13, 2018

Greenland Ice & Ocean Mask -
Greenland Mapping Project...
developers.google.com

2000 Greenland Mosaic -
Greenland Ice Mapping Project...
developers.google.com

Most popular Facebook pages
in Greenland 2020, by number...
www.statista.com
Updated Feb 28, 2020

Average number of employees
in Greenland 2018, by industry

Greenland geothermal heat flux distribution and
estimated Curie Depths, links to gridded files

Explore at PANGAEA Explore at pangaea.figshare.com
Explore at search.datacite.org

2 scholarly articles cite this data set ([View in Google Scholar](#))

tsv, html

Unique identifier
<https://doi.org/10.1594/PANGAEA.892973>

Data set updated Aug 13, 2018

Data set provided by
PANGAEA

Authors
Yasmina M Martos

Licence
[Attribution 3.0 \(CC BY 3.0\)](#)
Licence information was derived automatically

Area covered

Variables measured
File content, File format, File name, File size, Uniform resource locator/link to file

Description
Curie depths beneath Greenland are revealed by spectral analysis of data from the World Digital Magnetic Anomaly Map2. A thermal model of the lithosphere then provides a corresponding geothermal heat flux map. This new map exhibits significantly higher frequency but lower amplitude variation than earlier heat flux maps, and provides an important boundary condition for numerical ice-sheet models and interpretation of borehole temperature profiles. In addition, it reveals new geologically significant features. Notably, we identify a prominent quasi-linear elevated geothermal heat flux anomaly running northwest-southeast across Greenland. We interpret this feature to be the relic of the passage of the Iceland hotspot from 80 to 50 Ma. The expected partial melting of the lithosphere and magmatic underplating or intrusion into the lower crust is compatible with models of observed satellite gravity data and recent seismic observations. Our geological interpretation has potentially significant implications for the geodynamic evolution of Greenland.

Google Dataset Search

- Automated search in Google Scholar from Google Dataset Search (see previous)
- Note that number of citations in Google Dataset Search does not appear to be recently updated (2 vs. 4 citing articles)

Google Scholar

"10.1594 pangea 892973" OR "pangea de 10.1594 pangea 892973"

Articles 4 results (0,03 sec)

My profile My library

Any time

Since 2020

Since 2019

Since 2016

Custom range...

Sort by relevance

Sort by date

☒ include patents

☒ include citations

☐ Create alert

Geothermal heat flux reveals the Iceland hotspot track underneath Greenland [PDF] wiley.com

YM Martos, TA Jordan, M Catalán... - Geophysical ..., 2018 - Wiley Online Library

Abstract Curie depths beneath Greenland are revealed by spectral analysis of data from the World Digital Magnetic Anomaly Map 2. A thermal model of the lithosphere then provides a corresponding geo...

☆ Cited by 18 Related articles All 5 versions

Surface expression of basal and englacial features, properties, and processes of the Greenland Ice Sheet [PDF] wiley.com

MA Cooper, TM Jordan, MJ Siegert... - Geophysical Research ..., 2019 - Wiley Online Library

Abstract Radar-sounding surveys measuring ice thickness in Greenland have enabled an increasingly "complete" knowledge of basal topography and glaciological processes. Where such observations are s...

☆ Cited by 2 Related articles All 8 versions

Sensitivity of the Northeast Greenland Ice Stream to Geothermal Heat [PDF] wiley.com

S Smith-Johnsen, NJ Schlegel... - Journal of ..., 2020 - Wiley Online Library

Page 1. manuscript submitted to JGR: Earth Surface Sensitivity of the Northeast Greenland Ice Stream to 1 Geothermal Heat 2 S. Smith-Johnsen1, NJ. Schlegel2, B. de Fleurian1and KH Nisancioglu1,3 3 1Department of Earth ...

☆ Cited by 2 Related articles

A constraint upon the basal water distribution and thermal state of the Greenland Ice Sheet from radar bed echoes [PDF] whiterose.ac.uk

TM Jordan, CN Williams, DM Schroeder... - ..., 2018 - eprints.whiterose.ac.uk

Page 1. This is a repository copy of A constraint upon the basal water distribution and thermal state of the Greenland Ice Sheet from radar bed echoes. White Rose Research Online URL for this paper: http://eprints.whiterose.ac.uk/150981/ Version: Published Version ...

☆ Cited by 10 Related articles All 16 versions

Lessons learnt

- Overall, the analysis of the existing portals shows that there are several different initiatives that facilitate open data sharing – both field specific and generic, both commercial and sponsored by governments or research organisations
- Some of these
 - function as **aggregators of metadata** (and do not offer any archiving of data themselves)
 - some publish data from certain platforms or organisations
 - some **facilitate self-archiving of datasets**
- Most aggregators do a fairly good job of presenting consistent metadata, e.g. preserving titles, author information, and DOIs and pointing back to the original source
- However, different metadata levels and metadata specific to some sources can be a challenge – with some fields being empty in an aggregator, and some information from the original source that does not fit into the aggregator scheme

Lessons learnt

- Most of the examined examples attempt to give statistics on **the number of dataset views and dataset downloads**.
 - However, as the same dataset can be discovered in several aggregators the views downloads statistics are also distributed and are hard to aggregate and analyse. Thus getting an overview and correct total for these figures is difficult
- (This situation is of course not unlike that of citation counts for publications where the same article may have different citation counts in Web of Science, Scopus, Google Scholar and ResearchGate...)

Lessons learnt

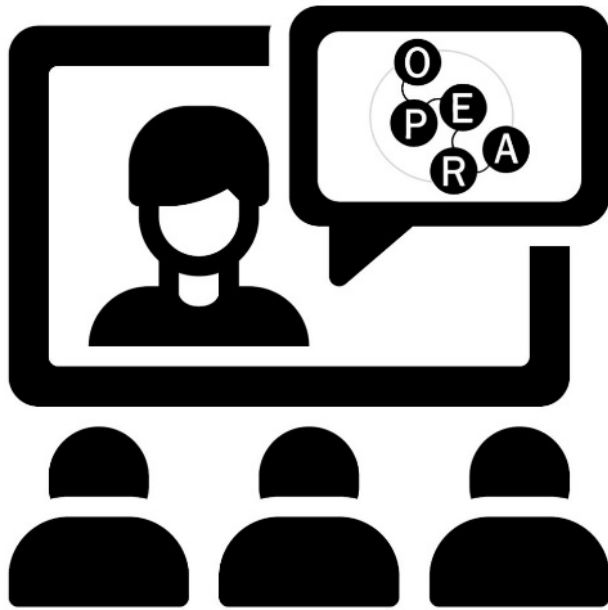
- In addition to views and downloads, actual usage of data that leads to a **dataset citation** in new publications is interesting and important to monitor
- Google Dataset Search reports the number of citations in Google Scholar - automatically identified via a search on DOIs and archive name
- GBIF does daily automated searches in a number of sources, and manually curates these

Lessons learnt

- Identifying dataset citations is made difficult because a data citation culture is still to be established in most fields
- Many citations to datasets may be missed because
 - 1) many different ways of citing datasets is being used with little consistency (e.g. referring to the dataset in the main text, vs. in a footnote or in the reference list)
 - 2) some may not be used to citing data, but cites the article describing the data instead or not at all

Lessons learnt

- To counter this, several aggregators and dataset repositories give **detailed instructions on how to cite the dataset**, e.g. by posting a reference that can be readily copied in a manuscript
- GBIF has the most advanced solution where not only each dataset can be cited, but also subsets and aggregates receive their own citable DOI. The disadvantage of this is that **the same data can be cited with several different DOIs**
- Even with such elaborate support in place example studies show that the **data citation culture is still weak** – see Kahn, Thellwall and Koucha (2019) for GBIF



Concluding discussion

Summary of issues



Data and Indicators summary

- Open Science and Open data are complex phenomena – we can propose a wide range of potential data and indicators that might be useful
- However, for many of these we have no or limited data – or data that is not very reliable...
- Even when we have data, e.g. on searches, views, downloads and data citations – we know little about what they actually *mean*
- Views, downloads and citations are often reported... these and more advanced indicators can be hard to interpret
- **= we are at a very early stage in relation to measuring Open Science efforts in a meaningful and productive way**

What to do then in OPERA??

- Very scarce data on Open Science are available → cannot be directly imported to all or even many records in the OPERA NORA
- Examples of data on Open Science Efforts can be found
 - E.g. open datasets, with size information, views, downloads and citations
 - **Add these in NORA as examples (even simulated) and study how relevant stakeholders react to and interpret these**
 - → final WP5 deliverable: **usability test of NORA**
 - Live NORA + mock-ups with sample Open Science indicators
 - Eyetrack+ deans, research managers, researchers, doctoral students etc. as they interact with NORA and the mock-ups, interview them about perceptions and usefulness





Thank You

Acknowledgments

- OPERA partners
- Pelle Annfeldt Israelsson
- Brian Kirkegaard Lunn
- Internal OPERA reviewers



References

- Ingwersen, P., & Chavan, V. (2011). **Indicators for the Data Usage Index (DUI): an incentive for publishing primary biodiversity data through global information infrastructure.** *BMC Bioinformatics*, 12 Suppl 1(Suppl 15). <https://doi.org/10.1186/1471-2105-12-S15-S3>
- Khan, N., Thelwall, M. and Kousha, K. (2019). **Data citation and reuse practice in biodiversity - Challenges of adopting a standard citation model.** In: Catalano, G., Daraio, C., Gregori, M., Moed, H. F. and Ruocco, G. (eds.) *17th International Conference on Scientometrics & Infometrics, ISSI2019: Proceedings, Volume I.* Italy: International Society for Scientometrics and Informetrics/Edizione Efesto, pp. 1220-1225. <https://wlv.openrepository.com/handle/2436/623005>
- Lampert, D., Lindorfer, M., Prem, E., Irran, J. & Sanz, F. S. (2017). **New indicators for open science - Possible ways of measuring the uptake and impact of open science.** *fteval Journal for Research and Technology Policy Evaluation*, 44. pp. 50-56. <https://repository.fteval.at/316/>
- Open Science and Research Initiative (2014). **The Open Science and Research Handbook.** <https://www.fosteropenscience.eu/sites/default/files/pdf/3986.pdf>
- Wouters, P., Ràfols, I., Oancea, A., Kamerlin, S.C.L., Holbrook, J.B. Jacob, M. Edited by Rene von Schomberg (2019). **Indicator Frameworks for Fostering Open Knowledge Practices in Science and Scholarship. Report of the Expert Group on Indicators for Researchers' Engagement with Open Science.** Directorate-General for Research and Innovation, Social Challenges 6 Programme - Horizon 2020
- **ReACT - Responsible Impact Project** - <https://www.communication.aau.dk/research/Research+Projects/react/>
- **NASA Open Data Portal** - <https://data.nasa.gov/>
- **Global Biodiversity Information Facility (GBIF) Repository** - <https://www.gbif.org/search>
- **Mendeley Data** - <https://data.mendeley.com/>
- **Google Dataset Search** - <https://datasetsearch.research.google.com/>