RESEARCH DATA – MANAGEMENT, INFRASTRUCTURES, AND APPLICATIONS

Digital Workshop at the Ilia State University, International Doctoral School, Tbilisi, Georgia

Dr. Yves Vincent Grossmann, Max Planck Digital Library, Munich, Germany

გამარჯობა

SPEAKER

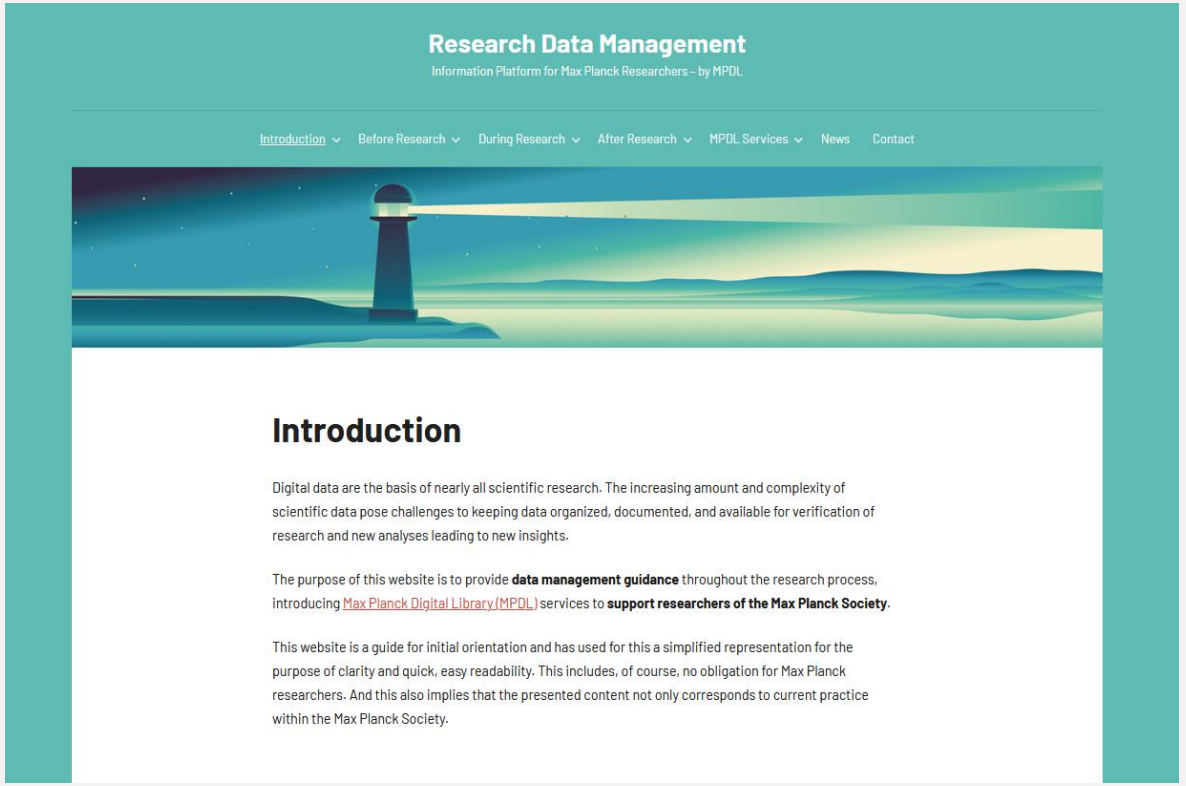
- Yves Vincent Grossmann
- https://orcid.org/0000-0002-2880-8947
- Referent for Research Data Management since October 2020 at the Max Planck Digital Library
- PhD thesis about the social history of industrial designers in (West-)Germany 1959-1990
- Contact: grossmann@mpdl.mpg.de

MAX PLANCK DIGITAL LIBRARY

- Information Services since 2007
- located in Munich, but no part of GV
- about 80 heads
- Developers, librarians, purchasers, administration
- www.mpdl.mpg.de

the only bookcase in the MPDL

RDM SUPPORT FOR MPG RESEARCHER



https://rdm.mpdl.mpg.de

WORKSHOP SCHEDULE

27th October 2021, 6-8 pm (UTC +4)

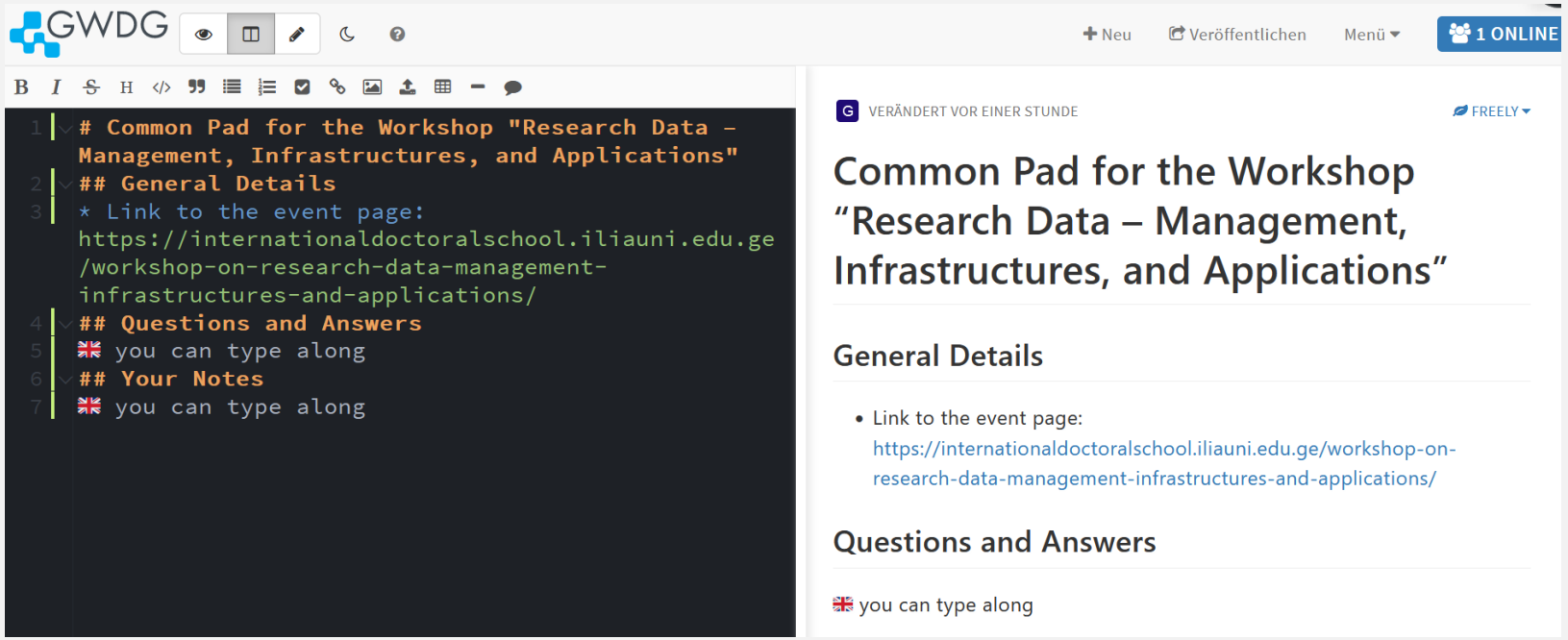
3rd November, 6-8 pm (UTC +5)

10th November, 6-8 pm (UTC +5)

17th November, 6-8 pm (UTC +5)

COMMON PAD FOR THE WORKSHOP

Pad for notes together: https://pad.gwdg.de/s/0Vwp7H1Cz



ORGANISATIONAL

- Questions and Remark? Ask immediately, wait for the end of a section, use the common pad
- Questions afterwards: grossmann@mpdl.mpg.de
- Slides will be available afterwards
- Live in spring/summer 2022 in Tbilisi

THANKS TO

- Professor **Marika Kapanadze** (Faculty of Business, Technology and Education, Ilia State University) for the inviation
- **Vakhtang Pataridze** (International Doctoral School, Ilia State University) for the organizational support
- you, for the participation

SHORT SURVEY:
YOUR AIMS AND BACKGROUNDS

STRUCTURE OF THE SESSION

Part 1: Research Data

Part 2: Infrastructure

Part 3: Research Data Management

Part 4: Applications

Part 5: Open Science

STRUCTURE OF PART 1

1. Examples of research data
2. Definition attempts
3. Classifications
4. Concret Examples of Research Data Publications
5. Research data cycles

STRUCTURE OF PART 2

1. Infrastructure
2. Research Data Repositores
3. Metadata
4. Persistent Identifier
5. Licenses

STRUCTURE OF PART 3

1. Why Research Data Management?
2. Data Management Plans
3. Research Data Policies
4. FAIR Princinples
5. Data Visualization
6. Jupyter
7. Problems

STRUCTURE OF PART 4

1. NFDI
2. EOSC
3. GOSC & GAIA-X
4. Open Research Data
5. Research Software
6. Electronic Laboratory Notebooks
7. Open Science
8. Pre-Registration
9. Open Peer Review
10. Open Educational Resources
11. Open Access

STRUCTURE OF PART 5

1. Open Science
2. Open Research Data
3. Linked Open Data
4. Citizen Science
5. Pre-Registration
6. Open Peer Review
7. Open Educational Resources
8. Open Access

AIMS OF THE COURSE

- Understanding research data
- Goals and applications of research data management
- Current developments
- Getting started
- Platform for discussion

SPECIAL TOPICS FOR NEXT SESSIONS BY YOU?

PART I

Research Data

STRUCTURE OF PART 1

1. Examples of research data
2. Definition attempts
3. Classifications
4. Concret Examples of Research Data Publications
5. Research data cycles

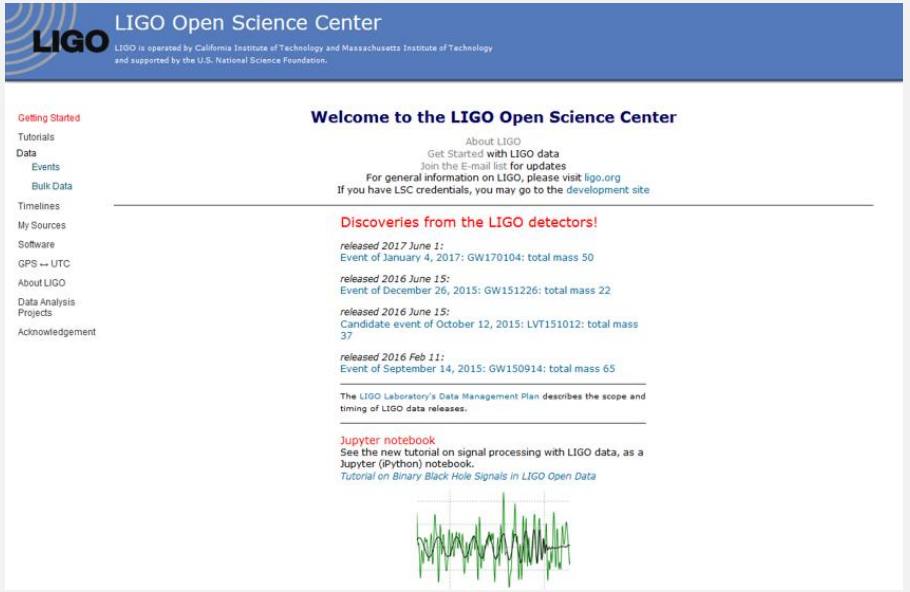
EXAMPLES OF RESEARCH DATA

RESEARCH DATA DEFINITION

- no straightforward definition applicable to all scientific fields
- the broadest one would be "anything needed to underpin scientific research", e.g. to validate and reproduce research findings

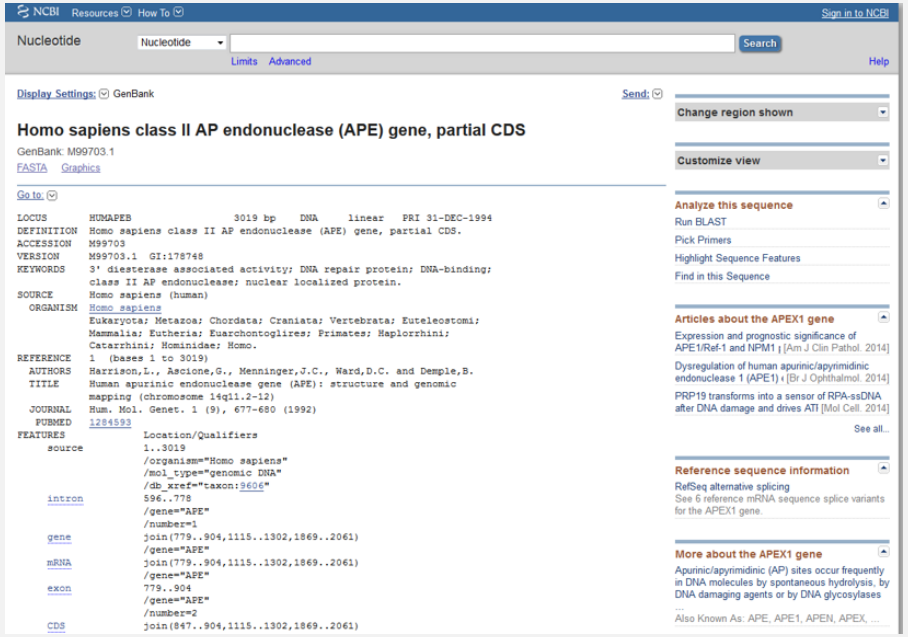
EXAMPLES FOR RESEARCH DATA

- Gravitational waves
- LIGO (https://ligo.org)



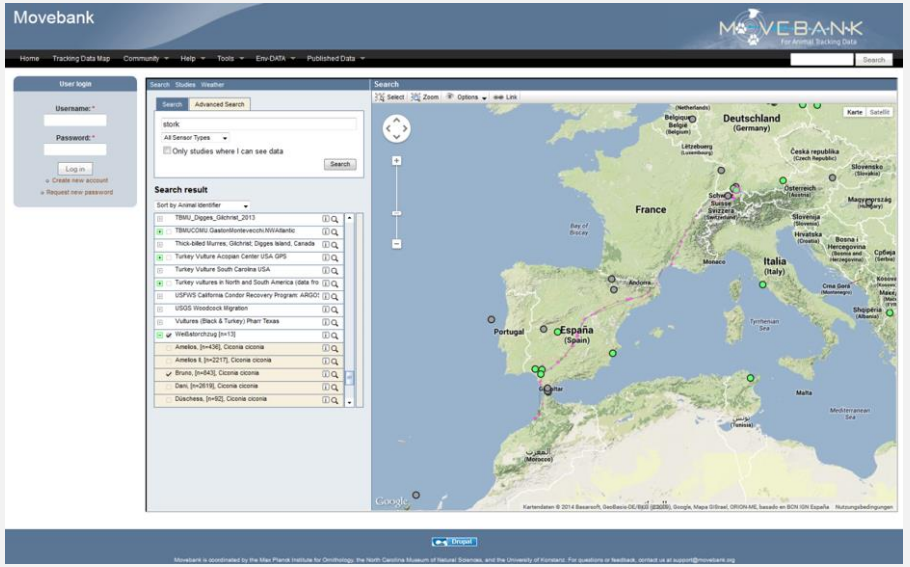
EXAMPLES FOR RESEARCH DATA

- Gen sequenzes
- GenBank (https://www.ncbi.nlm.nih.gov/genbank/)



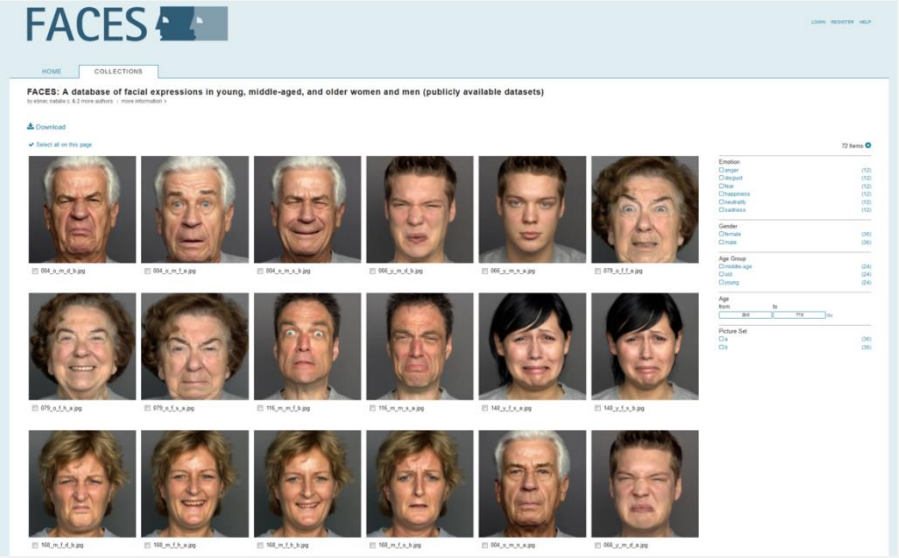
EXAMPLES FOR RESEARCH DATA

- Animal movments
- Movebank (https://www.movebank.org)



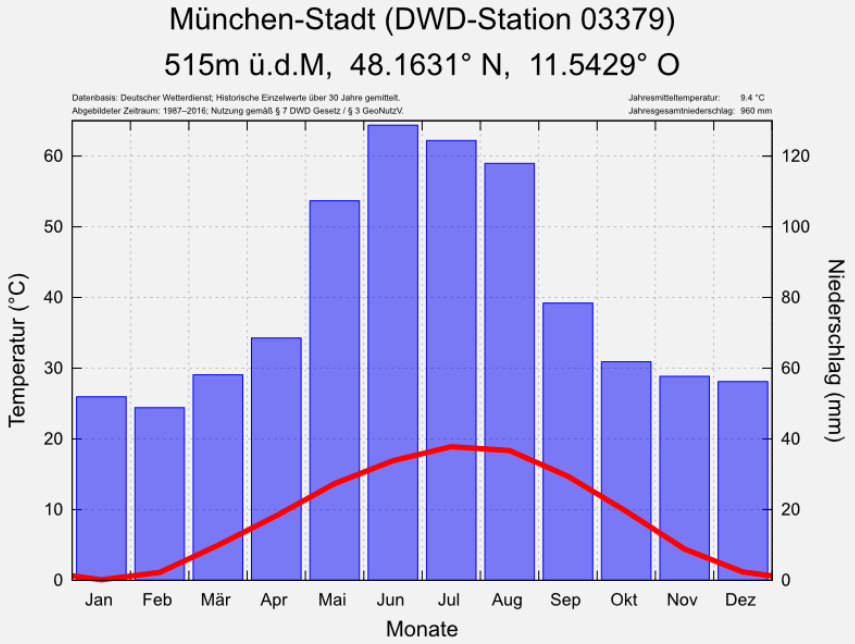
EXAMPLES FOR RESEARCH DATA

- Faces and mimiks
- FACES (https://faces.mpdl.mpg.de)



EXAMPLES FOR RESEARCH DATA

- Climate Data
- DKRZ World Data Centre for Climate: https://cera-www.dkrz.de/WDCC/ui/cerasearch/



Der-blaue-elefant, Klimadiagramm München-Stadt DWD 03379 Jahre 1987-2016, CC BY SA 4.0.

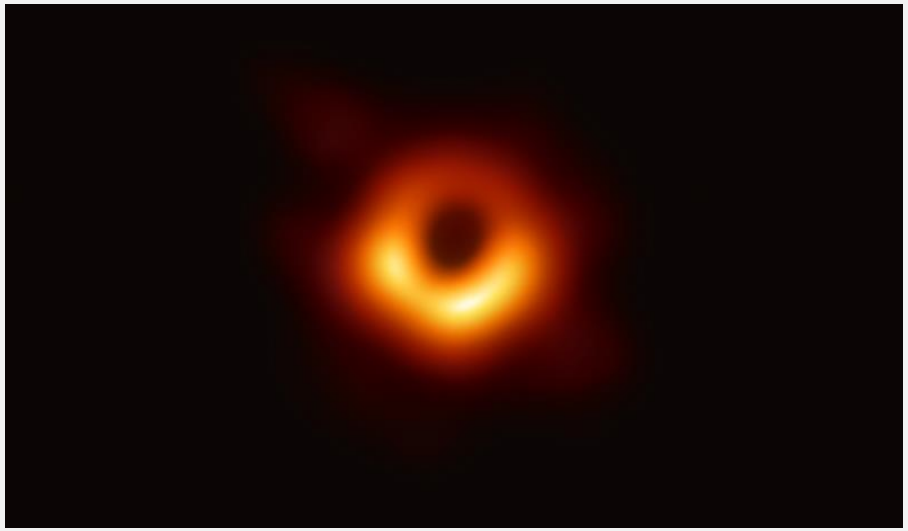
EXAMPLES FOR RESEARCH DATA

- Educational Study Data
- Max Planck Institute for Human Development, Berlin
- Berlin Age Study (BASE): www.base-berlin.mpg.de
- also VerbundFDB (German Network on Educational Reserach Data): www.forschungsdaten-bildung.de

Bundesarchiv, B 145 Bild-F081096-0027, Ulrich Wienke, Bad Godesberg, Altenbegegnungsstätte, CC BY SA 3.0

EXAMPLES FOR RESEARCH DATA

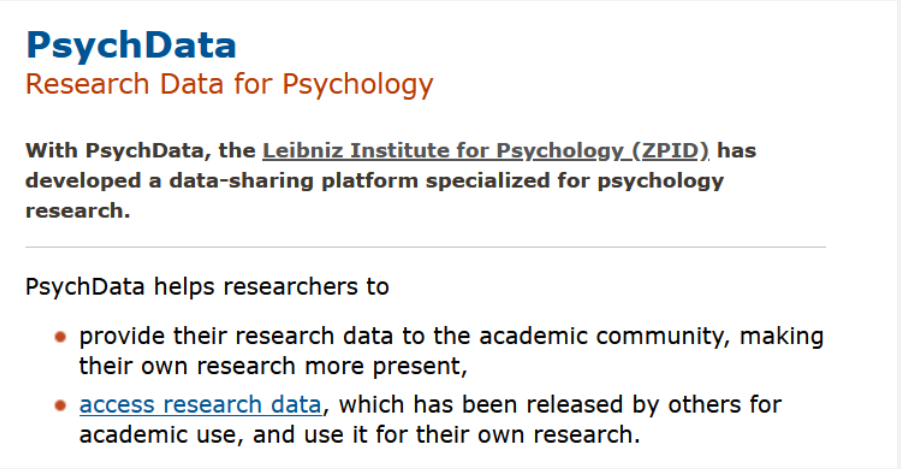
- Black holes



ESO, First Image of a Black Hole, 2019, CC BY 4.0, https://www.eso.org/public/images/eso1907a/

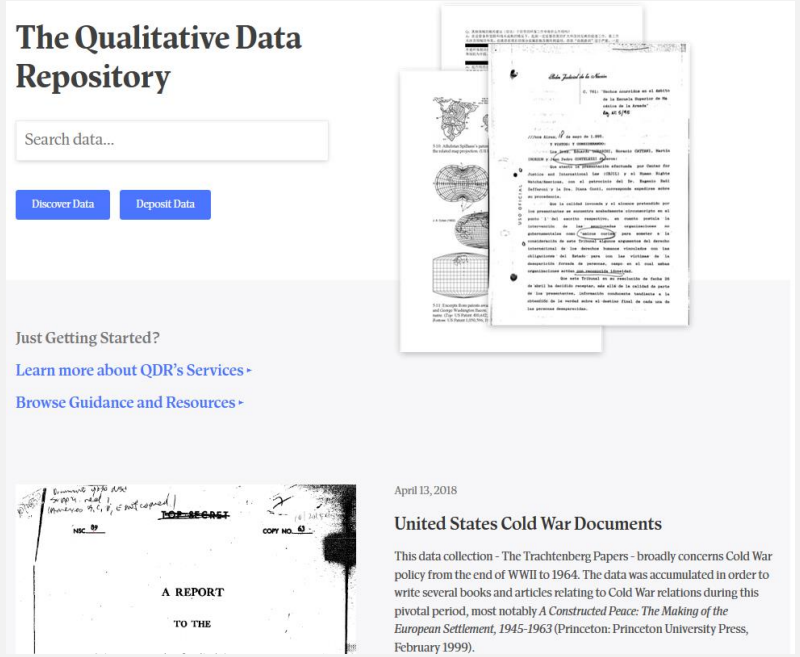
EXAMPLES FOR RESEARCH DATA

- Data from psychology
- PsychData (www.psychdata.de)



EXAMPLES FOR RESEARCH DATA

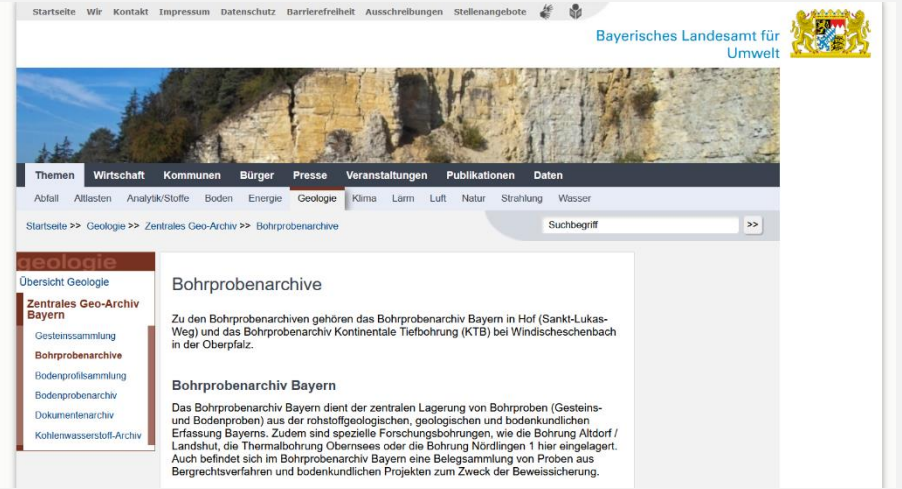
- Qualitative data
- United States Cold War Documents (https://qdr.syr.edu)



WHAT IS IT NOT ABOUT

- Physical Data
- Scientific publications
- Data about science

- i.e. [Drilling sample archives of the Bavarian State Office for the Environment]

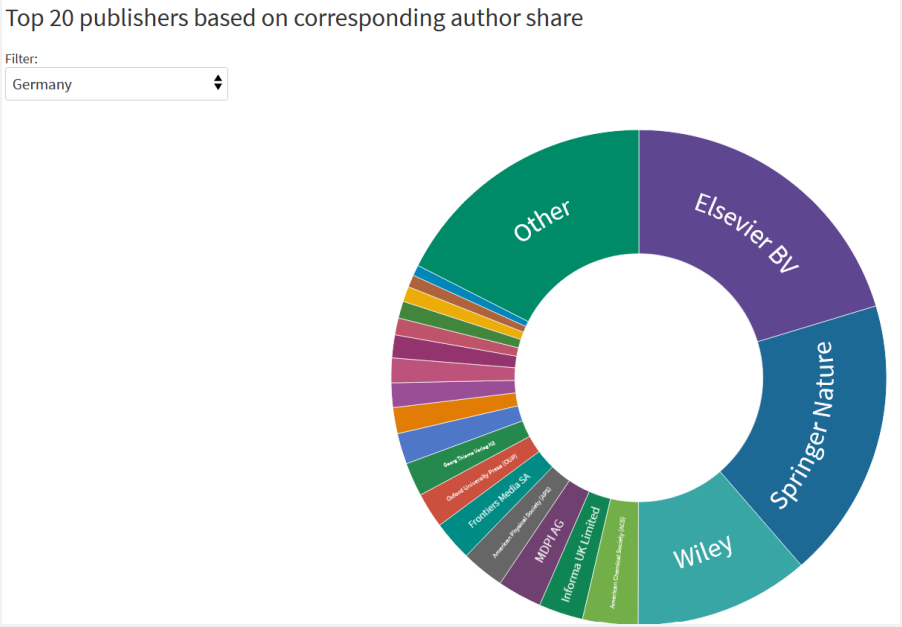


WHAT IS IT NOT ABOUT

- Physical Data
- Scientific publications
- Data about science

WHAT IS IT NOT ABOUT

- Physical Data
- Scientific publications
- Data about science



https://esac-initiative.org/market-watch/

DEFINITION ATTEMPS

DEFINITION ATTEMPT: 1

"*In very general terms, research data are data that are generated, collected or compiled in the research process and on the basis of which scientific hypotheses, models or theories are formed.*"

Competence Centre for Research Data at Heidelberg University, https://data.uni-heidelberg.de/faq.html

DEFINITION ATTEMPT: 2

"*Research data include measurement data, laboratory values, audiovisual information, texts, survey data or observational data, methodological test procedures and questionnaires.*"

DFG questionnaire, 10th June 2021, p. 1, https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf

DEFINITION ATTEMPT: 3

Research data are "*data generated in the course of scientific projects, e.g. through digitisation, source research, experiments, measurements, surveys or interviews.*"

Allianz AG "Forschungsdaten", Forschungsdatenmanagement - Eine Handreichung, 2018, p. 4, https://gfzpublic.gfz-potsdam.de/rest/items/item_3055893_5/component/file_3055894/content

DEFINITION ATTEMPT: 4

Research data refer "*to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation.*"

European Commission, H2020 Programme – Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, v.3.2, 2017, p. 4, https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

DEFINITION ATTEMPT: 5

"*Research data comprise all data generated in the scientific work process and processed in digital form.*"

Guideline on handling research data in the Leibniz Association, 2018, p. 1, https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/Forschung/Open_Science/Leitlinie_Forschungsdaten_2018.pdf

CLASSIFICATION OF RESEARCH DATA

CLASSIFICATION OF RESEARCH DATA

According to the character of the data

Qualitative data

- Texts (sources, transcripts)
- Images (scans, photos, microscope data, telescope data, satellite data)
- Multimedia (audio data, video data, 3D data, 4D data)

Quantitative data

- Number series (survey data, experimental data, sensor measurement series, gene sequences, geodata)
- Born-digitals (simulation data, algorithms, websites)

CLASSIFICATION OF RESEARCH DATA

According to the format of the data

Text (TXT, ODT, PDF, …)

Glossed text

See p. 12-13 in:

van den Berg, Margot C. & Smith, Norval S.H. 2013. Early Sranan. In: Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) *The Survey of Pidgin and Creole Languages. Vol. I: English-based and Dutch-based languages*, 3-14. Oxford: Oxford University Press.

The following lines, taken from the play in van Dyk's (c1765) language manual, present a detailed account of a plantation funeral that is generally considered to be true to life (van Trier-Guicherit 1991).

Zwarte Officier: *Tappe kisti meki wi go masra negere*
close coffin make 1PL go master black
Black Overseer: 'Close the coffin, let's go. Gentlemen,

aenoe sikkisi somma opo da bomen
2PL six person lift DET.SG woman
you six persons, lift up the woman.

hoe ply dem homan langa negere alle
Q place DET.PL woman with black all
Where is the woman with all the blacks?

aenoe wakke na bippi kry man mose
2PL walk LOC heap wail man must
You must walk as a group, the wailers must

wakke na fessi aenno no vergiti wan
walk LOC face 2PL NEG forget one
walk at the front. Don't forget a

santi kongo-kongo.
thing come-go
thing! Come, come!'

Zangers: *Da so wi jarri didde somma go mi jan do,*
that so 1PL carry dead person go 1SG play
Singers: 'That's how we carry the dead away. I celebrate. We

wi sarri fo joe alle da joe go lasi so, na
1PL regret for 2PL all that 2SG go expire so LOC
are very sorry for you. It is you who died like this.

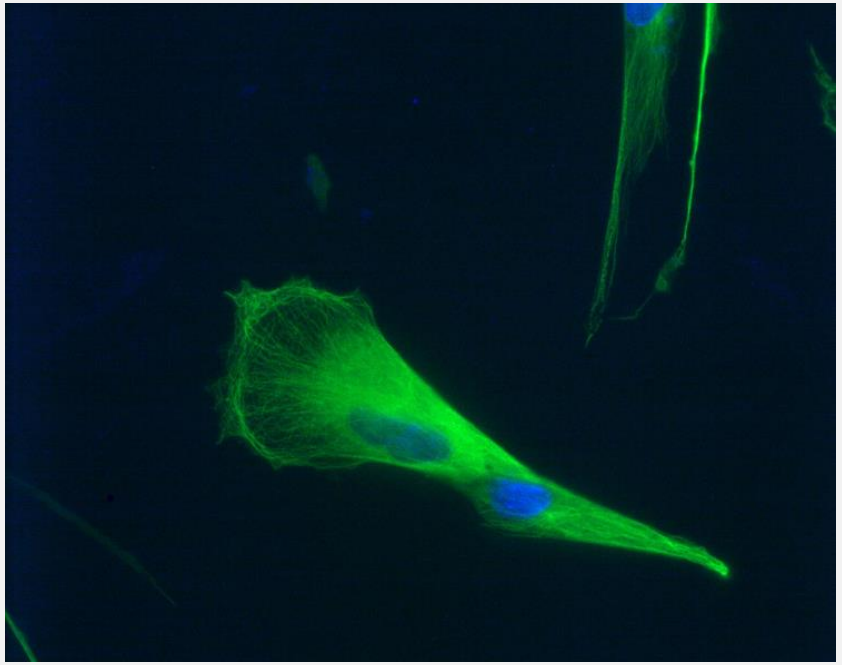
APiCS Consortium (2014): APiCS supplemental files, CC BY 4.0,
https://edmond.mpdl.mpg.de/imeji/collection/4WkY1hHhw8iEuNQ4

CLASSIFICATION OF RESEARCH DATA

According to the format of the data

Text (TXT, ODT, PDF, …)

Images (TIFF, JPEG, …)



Ries, Albert (Max Planck Institute for Biochemistry), IMR90 fibroblasts, 2014, CC BY 4.0, http://edmond.mpdl.mpg.de/imeji/collection/KvTf6TbHX0Yz49NC

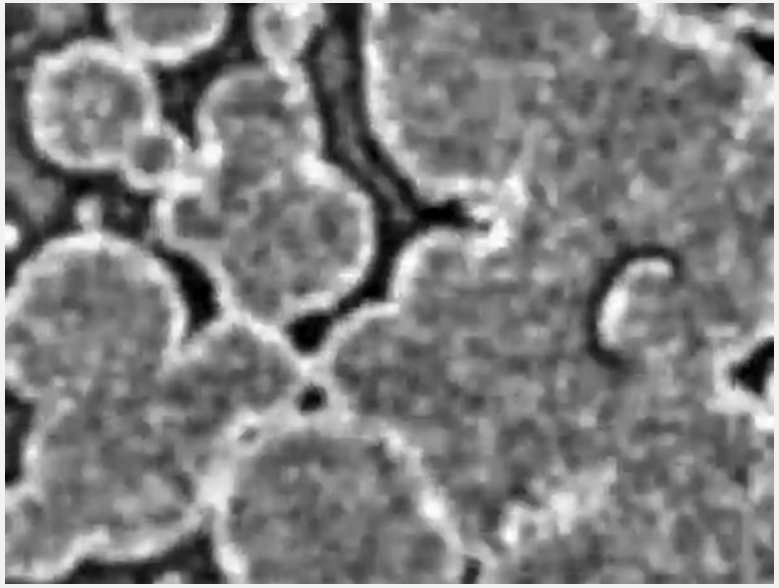
CLASSIFICATION OF RESEARCH DATA

According to the format of the data

Text (TXT, ODT, PDF, …)

Images (TIFF, JPEG, …)

Multimedia (MP4, Dicom, …)



Kaumudi Prabhakara (Max Planck Institute for Dynamics and Self-Organization), Spiralwaves, 2015, CC BY 4.0, https://www.youtube.com/watch?v=DmRZn073Uus

CLASSIFICATION OF RESEARCH DATA

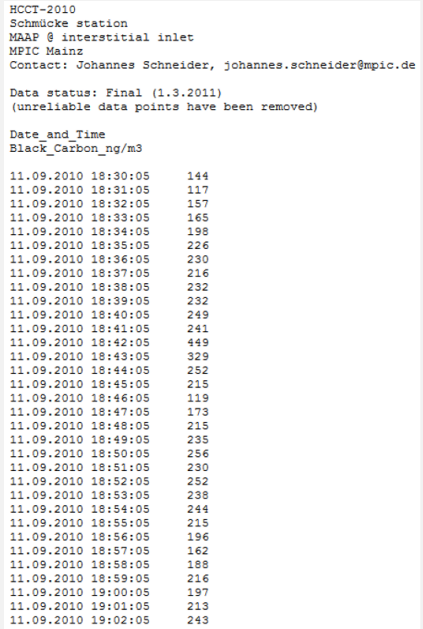
According to the format of the data

Text (TXT, ODT, PDF, …)

Images (TIFF, JPEG, …)

Multimedia (MP4, Dicom, …)

Number series (CSV, XLSX, …)



```
HCCT-2010
Schmücke station
MAAP @ interstitial inlet
MPIC Mainz
Contact: Johannes Schneider, johannes.schneider@mpic.de

Data status: Final (1.3.2011)
(unreliable data points have been removed)

Date_and_Time
Black_Carbon_ng/m3

11.09.2010 18:30:05     144
11.09.2010 18:31:05     117
11.09.2010 18:32:05     157
11.09.2010 18:33:05     165
11.09.2010 18:34:05     198
11.09.2010 18:35:05     226
11.09.2010 18:36:05     230
11.09.2010 18:37:05     216
11.09.2010 18:38:05     232
11.09.2010 18:39:05     232
11.09.2010 18:40:05     249
11.09.2010 18:41:05     241
11.09.2010 18:42:05     449
11.09.2010 18:43:05     329
11.09.2010 18:44:05     252
11.09.2010 18:45:05     215
11.09.2010 18:46:05     119
11.09.2010 18:47:05     173
11.09.2010 18:48:05     215
11.09.2010 18:49:05     235
11.09.2010 18:50:05     256
11.09.2010 18:51:05     230
11.09.2010 18:52:05     252
11.09.2010 18:53:05     238
11.09.2010 18:54:05     244
11.09.2010 18:55:05     215
11.09.2010 18:56:05     196
11.09.2010 18:57:05     162
11.09.2010 18:58:05     188
11.09.2010 18:59:05     216
11.09.2010 19:00:05     197
11.09.2010 19:01:05     213
11.09.2010 19:02:05     243
```

Schneider, Johannes (2017): HCCT2010, CC BY 4.0, https://dx.doi.org/10.17617/3.i

CLASSIFICATION OF RESEARCH DATA

According to the format of the data

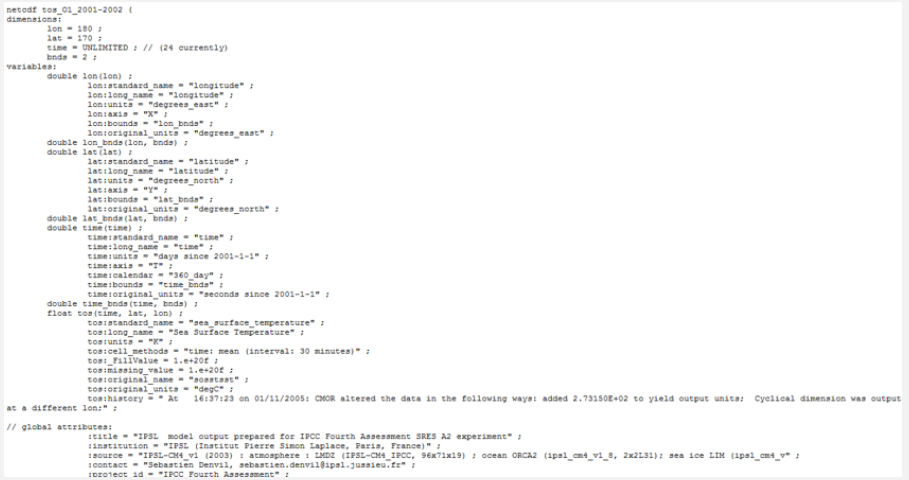
Text (TXT, ODT, PDF, …)

Images (TIFF, JPEG, …)

Multimedia (MP4, Dicom, …)

Number series (CSV, XLSX, …)

Born-digitals (netCDF, grib, HTML, …)



```
netcdf tos_O1_2001-2002 {
dimensions:
        lon = 180 ;
        lat = 170 ;
        time = UNLIMITED ; // (24 currently)
        bnds = 2 ;
variables:
        double lon(lon) ;
                lon:standard_name = "longitude" ;
                lon:long_name = "longitude" ;
                lon:units = "degrees_east" ;
                lon:axis = "X" ;
                lon:bounds = "lon_bnds" ;
                lon:original_units = "degrees_east" ;
        double lon_bnds(lon, bnds) ;
        double lat(lat) ;
                lat:standard_name = "latitude" ;
                lat:long_name = "latitude" ;
                lat:units = "degrees_north" ;
                lat:axis = "Y" ;
                lat:bounds = "lat_bnds" ;
                lat:original_units = "degrees_north" ;
        double lat_bnds(lat, bnds) ;
        double time(time) ;
                time:standard_name = "time" ;
                time:long_name = "time" ;
                time:units = "days since 2001-1-1" ;
                time:axis = "T" ;
                time:calendar = "360_day" ;
                time:bounds = "time_bnds" ;
                time:original_units = "seconds since 2001-1-1" ;
        double time_bnds(time, bnds) ;
        float tos(time, lat, lon) ;
                tos:standard_name = "sea_surface_temperature" ;
                tos:long_name = "Sea Surface Temperature" ;
                tos:units = "K" ;
                tos:cell_methods = "time: mean (interval: 30 minutes)" ;
                tos:_FillValue = 1.e+20f ;
                tos:missing_value = 1.e+20f ;
                tos:original_name = "sosstsst" ;
                tos:original_units = "degC" ;
                tos:history = " At   16:37:23 on 01/11/2005: CMOR altered the data in the following ways: added 2.73150E+02 to yield output units;  Cyclical dimension was output
at a different lon;" ;

// global attributes:
                :title = "IPSL  model output prepared for IPCC Fourth Assessment SRES A2 experiment" ;
                :institution = "IPSL (Institut Pierre Simon Laplace, Paris, France)" ;
                :source = "IPSL-CM4_v1 (2003) : atmosphere : LMDZ (IPSL-CM4_IPCC, 96x71x19) ; ocean ORCA2 (ipsl_cm4_v1_8, 2x2L31); sea ice LIM (ipsl_cm4_v" ;
                :contact = "Sebastien Denvil, sebastien.denvil@ipsl.jussieu.fr" ;
                :project_id = "IPCC Fourth Assessment" ;
```

Sea surface temperatures collected by PCMDI for use by the IPCC, 2015, https://www.unidata.ucar.edu/software/netcdf/examples/files.html

CLASSIFICATION OF RESEARCH DATA

According to the source of the data or the type of research

Observational: data captured in real-time, usually irreplaceable. For example, sensor data, survey data, sample data, neuroimages.

Experimental: Data from lab equipment, often reproducible, but can be expensive. For example, gene sequences, chromatograms, toroid magnetic field data.

Simulation: data generated from test models where model and metadata are more important than output data. For example, climate models, economic models.

Derived or compiled: data is reproducible but expensive. For example, text and data mining, compiled database, 3D models.

Reference or canonical: a (static or organic) conglomeration or collection of smaller (peer-reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals.

University of Edingburgh, Research Data Definitions, 2012, S. 2, https://www2.le.ac.uk/services/research-data/old-2019-12-11/documents/UoL_ReserchDataDefinitions_20120904.pdf

CLASSIFICATION OF RESEARCH DATA

According to the degree of aggregation or compression

Raw data:
Raw data should be defined as the total amount of all data available to empirical science as a basis for research.

Primary data:
Primary data are those data that are actually used for research as a subset of the raw data.

Secondary data:
Secondary data are data that have been obtained from primary data in process steps.

BREAK (5 MINUTES)

Search and count:
https://www.zooniverse.org/projects/penguintom79/penguin-watch

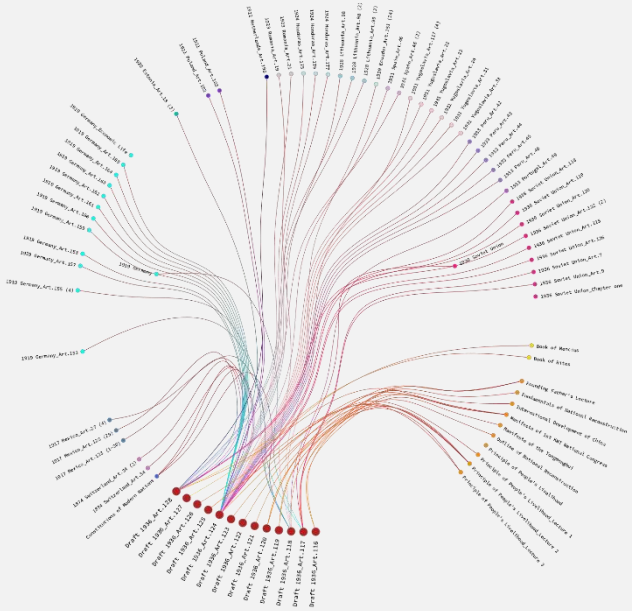
QUESTIONS?

CONCRET EXAMPLES OF RESEARCH DATA PUBLICATIONS

EXAMPLE FORM THE HUMANITIES

Wagner, Andreas, & Li, Fupeng. (2020): rg-mpg-de/fupeng: Citation Network Visualisation of two Drafts (1923 and 1936) of the Chinese Constitution (v1.0). Zenodo. https://doi.org/10.5281/zenodo.3699154.

https://github.com/rg-mpg-de/fupeng/tree/v1.0



EXAMPLE FOR THE BIBLIOMETRICS

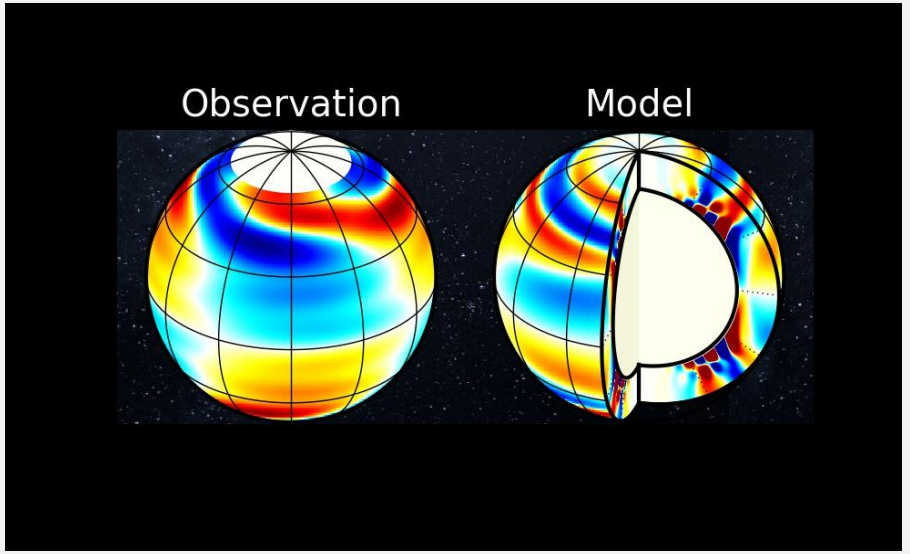
Bornmann, Lutz, Haunschild, Robin, Mutz, Rüdiger.(2021): Growth of Science. Max Planck Society, https://dx.doi.org/10.17617/3.7o.

Data from the study regarding growth rates of modern science: https://arxiv.org/abs/2012.07675

ear	npub_DIM	totpub_DI	growth_DI
1665	259	259	
1666	82	341	0,316602
1667	152	493	0,445748
1668	118	611	0,239351
1669	94	705	0,153846
1670	315	1020	0,446809
1671	111	1131	0,108824
1672	114	1245	0,100796
1673	201	1446	0,161446
1674	72	1518	0,049793
1675	90	1608	0,059289
1676	82	1690	0,050995
1677	96	1786	0,056805
1678	96	1882	0,053751
1679	42	1924	0,022317
1680	27	1951	0,014033
1681	156	2107	0,079959
1682	186	2293	0,088277
1683	123	2416	0,053642
1684	173	2589	0,071606
1685	166	2755	0,064117

EXAMPLES FORM ASTROPHYSICS

Gizon (2021). Solar inertial modes -- A&A Letter. Max Planck Society. https://edmond.mpdl.mpg.de/imeji/collection/oKXKmII3VbdVYFok.



Gizon, CC BY 4.0, https://edmond.mpdl.mpg.de/imeji/collection/oKXKmII3VbdVYFok/item/9Rvo3gaxSSCcA4gs?q=&fq=&filter=&pos=8#pageTitle

EXAMPLE FROM BIOLOGY

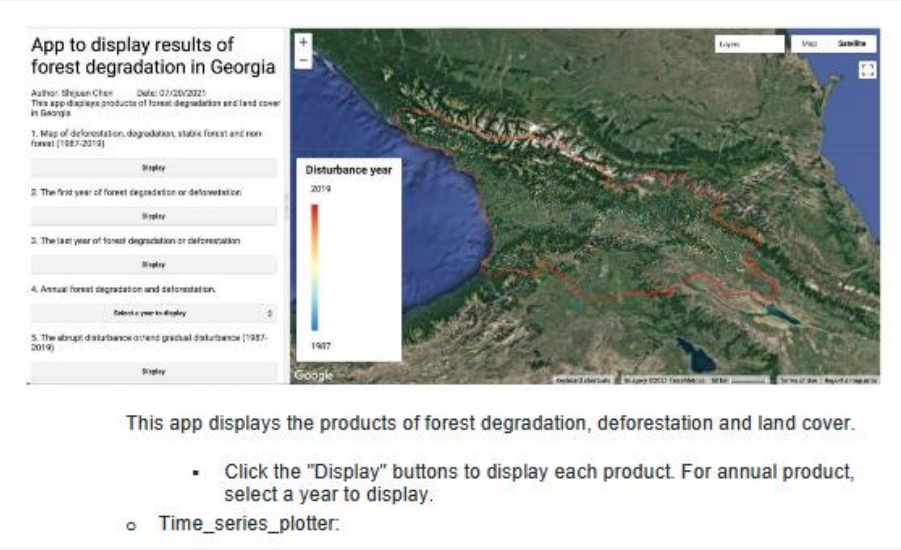
Epitashvili, Giorgi, Geiger, Matthias, Astrin, Jonas, Herder, Fabian, Japoshvili, Bella, & Mumladze, Levan. (2020). Supplementary material 1 from: Epitashvili G, Geiger M, Astrin JJ, Herder F, Japoshvili B, Mumladze L (2020) Towards retrieving the Promethean treasure: a first molecular assessment of the freshwater fish diversity of Georgia. Biodiversity Data Journal 8: e57862. https://doi.org/10.3897/BDJ.8.e57862 [Data set]. https://doi.org/10.3897/BDJ.8.e57862.suppl1

BOLD Sample ID	Identification	BIN	Collection Date	Institution Storing	Museum ID	River Drainage
WH02H10	*Acanthobrama microlepis*	BOLD:ADL5967	04. Aug 06	ZFMK	ZFMK-ICH-TIS-NB164	Kura
WH02H8	*Acanthobrama microlepis*	BOLD:ADL5967	04. Aug 06	ZFMK	ZFMK-ICH-TIS-NB164	Kura
KX189530	*Alburnoides eichwaldii*	BOLD:ADK3170	30. Sep 15	GenBank	NCBI:txid1201212	Kura
KX189532	*Alburnoides eichwaldii*	BOLD:ADK3170	30. Sep 15	GenBank	NCBI:txid1201212	Kura
KX189534	*Alburnoides eichwaldii*	BOLD:ADK3170	30. Sep 15	GenBank	NCBI:txid1201212	Kura
8005838	*Alburnoides eichwaldii*	BOLD:ADK3170	08-Jul-2019	ZFMK	ZFMK-ICH-106216	Kura
8007867	*Alburnoides eichwaldii*	BOLD:ADK3170	05-Jul-2019	ZFMK	ZFMK-ICH-106168	Kura
8007868	*Alburnoides eichwaldii*	BOLD:ADK3170	05-Jul-2019	ZFMK	ZFMK-ICH-106169	Kura
8007869	*Alburnoides eichwaldii*	BOLD:ADK3170	05-Jul-2019	ZFMK	ZFMK-ICH-106170	Kura
Ex24C12	*Alburnoides eichwaldii*	BOLD:ADK3170	02. Aug 06	ZFMK	ZFMK-ICH-TIS-NB137 & ZFMK ICH -103210-103241	Kura
Ex24D1	*Alburnoides eichwaldii*	BOLD:ADK3170	02. Aug 06	ZFMK	ZFMK-ICH-TIS-NB137 & ZFMK ICH -103210-103241	Kura
Ex24D4	*Alburnoides eichwaldii*	BOLD:ADK3170	02. Aug 06	ZFMK	ZFMK-ICH-TIS-NB143 & ZFMK ICH -103352-103353	Kura
Ex24F8	*Alburnoides eichwaldii*	BOLD:ADK3170	05. Aug 06	ZFMK	ZFMK-ICH-TIS-NB172 & ZFMK ICH -103565-103624	Kura
WH02H4	*Alburnoides eichwaldii*	BOLD:ADK3170	02. Aug 06	ZFMK	ZFMK-ICH-TIS-NB140	Kura
KX189557	*Alburnoides fasciatus*	BOLD:ADJ9477	08. Aug 15	GenBank	NCBI:txid870488	Bzyb
KX189558	*Alburnoides fasciatus*	BOLD:ADJ9477	08. Aug 15	GenBank	NCBI:txid870488	Bzyb
KX189559	*Alburnoides fasciatus*	BOLD:ADJ9477	08. Aug 15	GenBank	NCBI:txid870488	Bzyb
KX189560	*Alburnoides fasciatus*	BOLD:ADJ9477	08. Aug 15	GenBank	NCBI:txid870488	Bzyb
KX189561	*Alburnoides fasciatus*	BOLD:ADJ9477	08. Aug 15	GenBank	NCBI:txid870488	Bzyb
KX189562	*Alburnoides fasciatus*	BOLD:ADJ9477	08. Aug 15	GenBank	NCBI:txid870488	Bzyb
KX189563	*Alburnoides fasciatus*	BOLD:ADJ9477	09. Aug 15	GenBank	NCBI:txid870488	Kodor
KX189564	*Alburnoides fasciatus*	BOLD:ADJ9477	09. Aug 15	GenBank	NCBI:txid870488	Kodor
KX189565	*Alburnoides fasciatus*	BOLD:ADJ9477	09. Aug 15	GenBank	NCBI:txid870488	Kodor
KX189566	*Alburnoides fasciatus*	BOLD:ADJ9477	09. Aug 15	GenBank	NCBI:txid870488	Kodor
KX189567	*Alburnoides fasciatus*	BOLD:ADJ9477	09. Aug 15	GenBank	NCBI:txid870488	Kodor

EXAMPLE FOR SOFTWARE

Shijuan Chen et al. (2021): Monitoring Temperate Forest Degradation on Google Earth Engine Using Landsat Time Series Analysis, https://doi.org/10.5281/zenodo.5384552

The app displays the products of forest degradation, deforestation and land cover in Georgia.



EXAMPLE FROM GEOGRAPHY

Vasey, Dylan et al. (2020): Evolution of the Greater Caucasus basement and formation of the Main Caucasus Thrust, Georgia, Dryad, Dataset, https://doi.org/10.25338/B8D61N.

Table S4: Results of $^{40}Ar/^{39}Ar$ analyses

Step	$^{40}Ar/^{39}Ar^{1}$	± 1 s.d.	$^{38}Ar/^{39}Ar^{1}$	± 1 s.d.	$^{37}Ar/^{39}Ar^{1}$	± 1 s.d.	$^{36}Ar/^{39}Ar^{1}$	± 1 s.d.	^{39}Ar (mol)^{2}	^{39}Ar (%)	$^{40}Ar^{*3}$ (%)
SVANETI											
N6 (C16019) Muscovite											
1	185,73	2,41E+00	4,59E-01	1,77E-02	4,95E-03	4,95E-05	6,12E-01	1,85E-02	1,48E-17	0,09	1,5
2	188,16	2,41E+00	5,34E-01	1,62E-02	4,95E-03	4,95E-05	5,88E-01	1,53E-02	2,06E-17	0,16	6,4
3	43,489	3,53E-02	8,44E-02	2,32E-04	4,95E-03	4,95E-05	3,40E-02	1,26E-04	4,35E-15	15,4	76,5
4	45,570	3,53E-02	5,11E-02	1,65E-04	4,95E-03	4,95E-05	5,56E-03	5,62E-05	5,65E-15	35,2	96,3
5	46,641	4,75E-02	5,14E-02	2,57E-04	4,95E-03	4,95E-05	6,30E-03	1,02E-04	2,82E-15	45,1	95,9
6	45,878	3,79E-02	5,46E-02	1,57E-04	4,95E-03	4,95E-05	1,03E-02	7,90E-05	4,03E-15	59,2	93,2
7	45,700	4,62E-02	5,28E-02	1,58E-04	4,95E-03	4,95E-05	7,72E-03	6,16E-05	5,18E-15	77,4	94,9
8	45,902	4,80E-02	5,12E-02	2,73E-04	4,95E-03	4,95E-05	5,21E-03	9,83E-05	2,37E-15	85,7	96,5
9	47,477	5,30E-02	5,04E-02	2,26E-04	4,95E-03	4,95E-05	5,42E-03	9,55E-05	2,03E-15	92,9	96,4
10	49,382	6,76E-02	5,19E-02	2,50E-04	4,95E-03	4,95E-05	6,09E-03	9,53E-05	2,04E-15	100,0	96,2
KAZBEGI											
K1 (V16046D) Biotite											
1	13,023	4,95E-02	2,22E-02	6,53E-04	4,95E-03	4,95E-05	1,53E-02	9,88E-04	3,09E-16	1,2	62,3
2	17,233	3,77E-02	1,67E-02	3,64E-04	4,95E-03	4,95E-05	4,27E-03	4,75E-04	5,91E-16	3,4	91,1
3	16,164	3,35E-02	1,38E-02	2,29E-04	4,95E-03	4,95E-05	1,71E-03	2,00E-04	9,97E-16	7,2	95,8
4	17,662	3,34E-02	1,30E-02	1,79E-04	4,95E-03	4,95E-05	7,89E-04	2,04E-04	9,75E-16	10,8	97,7
5	19,127	3,35E-02	1,31E-02	1,09E-04	4,95E-03	4,95E-05	4,23E-04	1,10E-04	1,75E-15	17,4	98,8
6	21,409	2,86E-02	1,27E-02	1,37E-04	4,95E-03	4,95E-05	4,62E-04	1,02E-04	1,98E-15	24,9	98,9
7	24,361	3,31E-02	1,27E-02	1,15E-04	4,95E-03	4,95E-05	4,18E-04	7,69E-05	2,48E-15	34,3	99,2
8	28,616	3,13E-02	1,30E-02	8,50E-05	4,95E-03	4,95E-05	3,89E-04	5,28E-05	3,55E-15	47,6	99,4
9	33,840	3,21E-02	1,30E-02	8,10E-05	4,95E-03	4,95E-05	4,89E-04	3,01E-05	6,28E-15	71,3	99,5
10	40,105	3,19E-02	1,32E-02	7,32E-05	4,95E-03	4,95E-05	5,87E-04	2,95E-05	6,93E-15	97,4	99,5
11	44,053	8,61E-02	1,42E-02	2,57E-04	4,95E-03	4,95E-05	2,11E-03	2,94E-04	6,78E-16	100,0	98,0
K2 (V16052A) Muscovite											
1	42,636	4,30E-01	4,75E-02	3,75E-03	4,95E-03	4,95E-05	4,99E-02	6,00E-03	4,15E-17	0,5	60,4
2	44,555	2,24E-01	2,01E-02	1,24E-03	4,95E-03	4,95E-05	1,49E-02	3,21E-03	8,34E-17	1,4	86,9
3	49,892	2,87E-01	1,31E-02	1,45E-03	4,95E-03	4,95E-05	1,41E-02	1,79E-03	1,60E-16	3,3	90,0
4	52,938	2,40E-01	1,37E-02	5,72E-04	4,95E-03	4,95E-05	9,12E-03	1,02E-03	2,24E-16	5,9	93,8
5	56,872	2,07E-01	1,34E-02	5,11E-04	4,95E-03	4,95E-05	4,96E-03	7,95E-04	2,83E-16	9,1	96,6
6	62,278	1,83E-01	1,29E-02	3,12E-04	4,95E-03	4,95E-05	5,92E-03	5,97E-04	3,76E-16	13,5	96,6
7	74,045	1,11E-01	1,36E-02	1,44E-04	4,95E-03	4,95E-05	4,27E-03	1,45E-04	1,67E-15	32,8	98,2
8	68,062	1,14E-01	1,29E-02	1,32E-04	4,95E-03	4,95E-05	2,18E-03	1,18E-04	1,83E-15	53,9	98,9

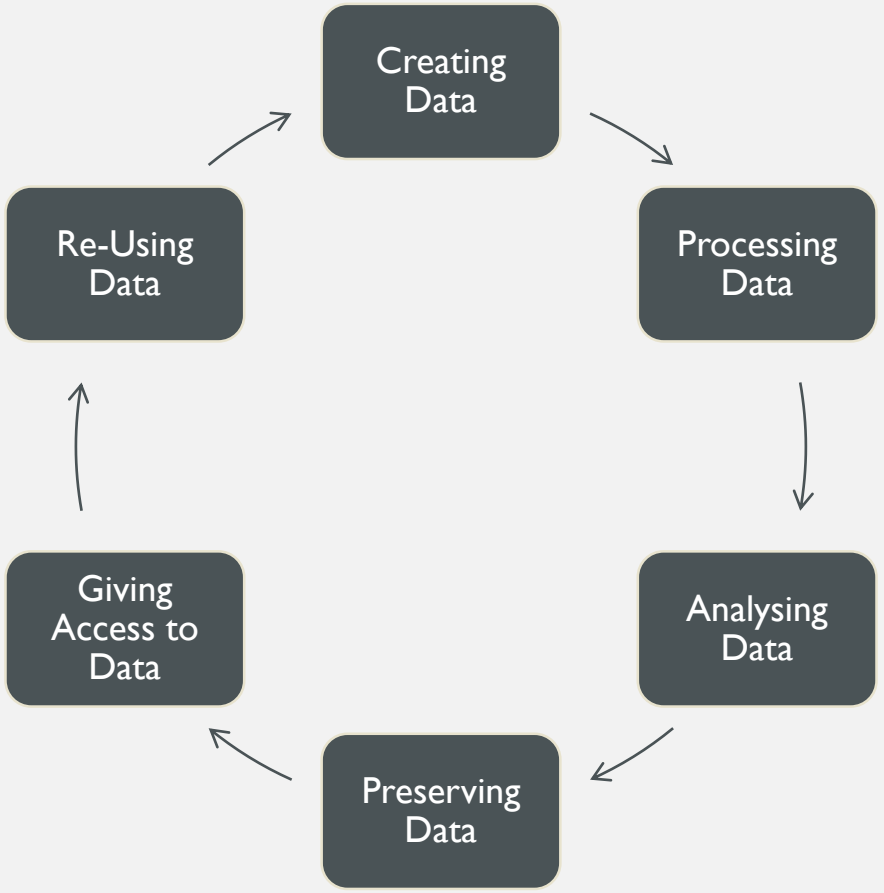
EXAMPLES FOR RESEARCH DATA BY YOU?

SIZES OF RESEARCH DATA SETS

- from kilobyte to zettabyte
- question of quantitative vs. qualitative

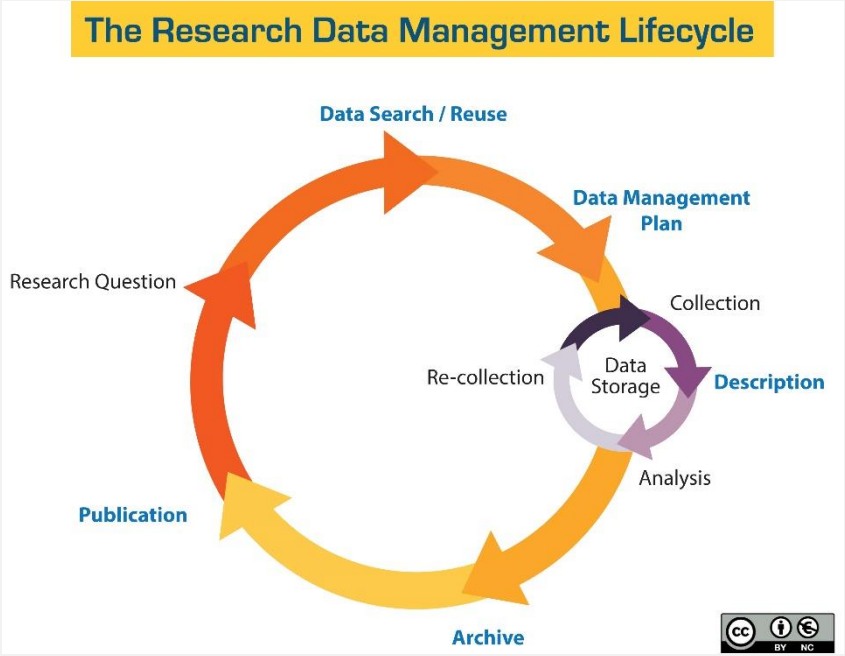
RESEARCH DATA CYCLE

RESEARCH DATA LIFECYCLE



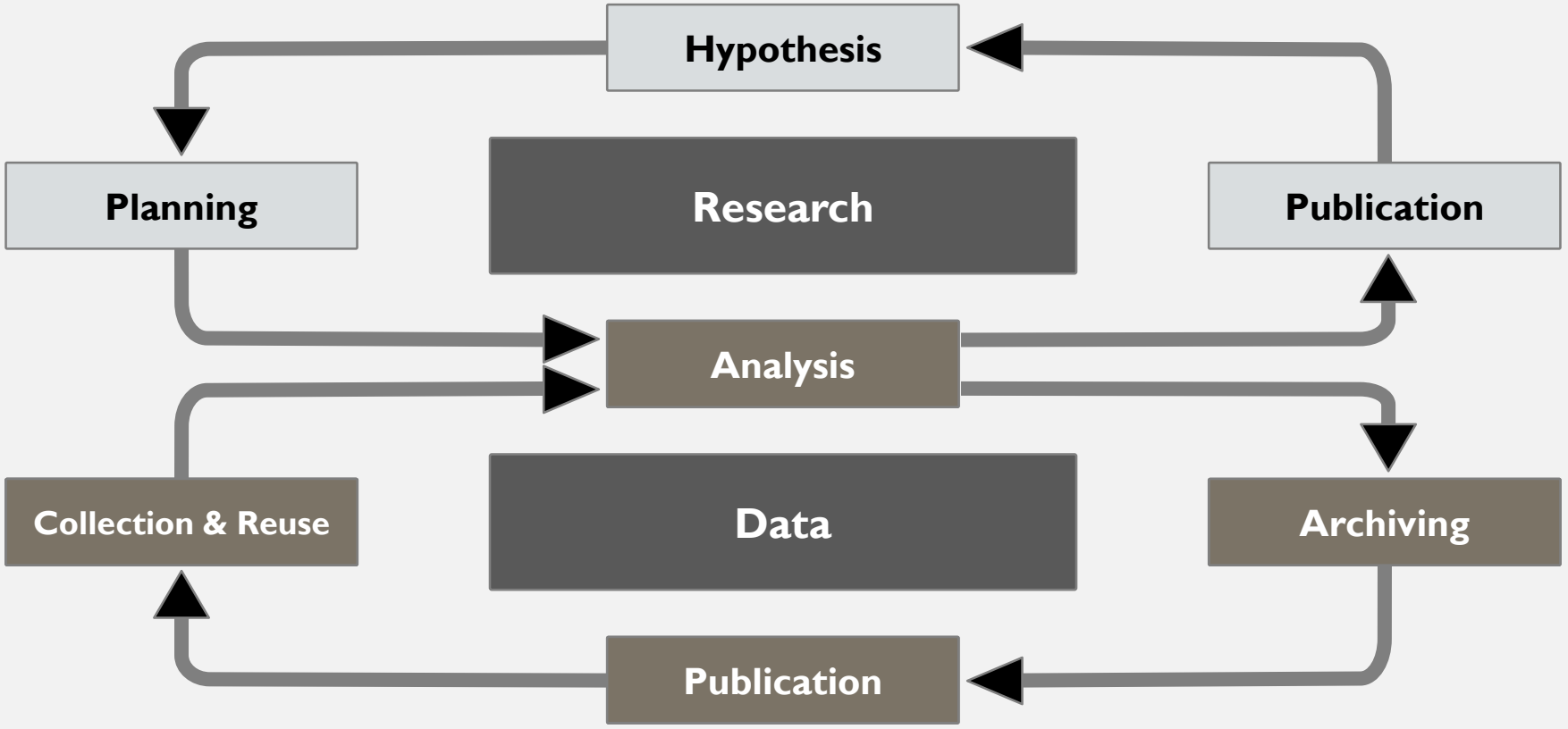
UK Data Service,
https://www.ukdataservice.ac.uk/manage-data/lifecycle

RESEARCH DATA LIFECYCLE



University Library, University of California, UC Santa Cruz, https://guides.library.ucsc.edu/datamanagement/

RESEARCH AND DATA LIFECYCLE



QUESTIONS?

PART 2

Infrastructure

STRUCTURE OF PART 2

1. Infrastructure
2. Research Data Repositores
3. Metadata
4. Persistent Identifier
5. Licenses

INFRASTRUCTURE?

e.g. transport infrastructure

- Roads, bridges
- Railways, waterways
- Filling station network, traffic radio, GPS
- Road traffic regulations, road patrols
- Road clearance service, toll system, vehicle tax
- Car industry, Ministry of Transport, public transport

Bundesarchiv, Bild 183-F0307-0001-046, Straßenverkehr in Leipzig, 1967, CC BY SA 3.0.

RESEARCH DATA INFRASTRUCTURE

- Scientific (large-scale) facilities
- Metadata, standards, identifiers
- Repositories, registries, data journals
- Legal bases(data centres, libraries, users...)
- National Research Data Infrastructure
- European Open Science Cloud

RESEARCH DATA REPOSITORIES

Image: Bundesarchiv, Bild 183-S90458 / CC-BY-SA 3.0.

RESEARCH DATA REPOSITORIES

Institutionell

- Open Data LMU
- Edinburgh DataShare
- Edmond

Subject

- PsychData
- Pangaea
- G-Node

General

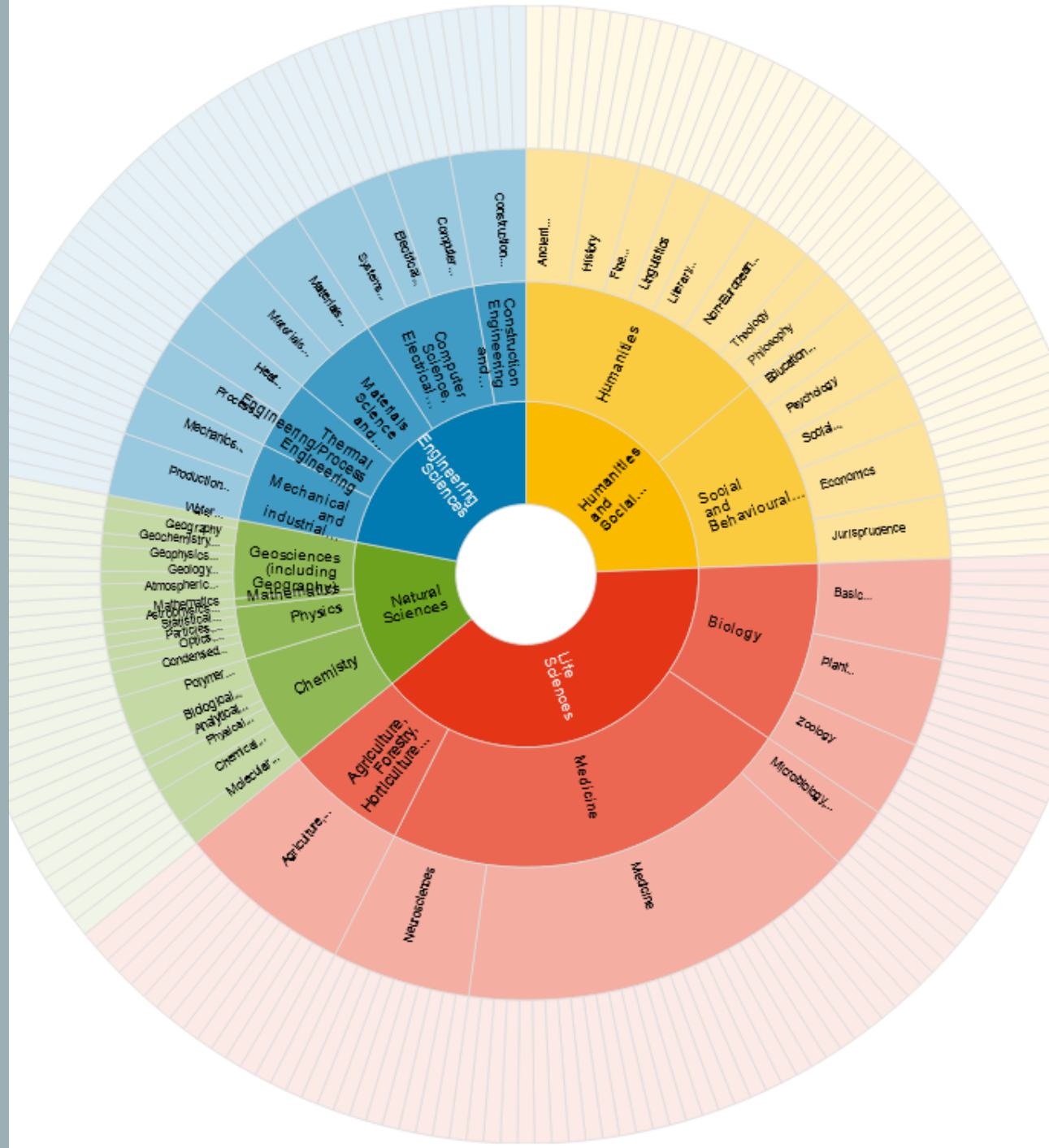
- Zenodo
- Dryad
- Radar
- Figshare
- Mendeley Data

RESEARCH DATA REPOSITORIES

How to find research data repositories:

Re3data (https://www.re3data.org)

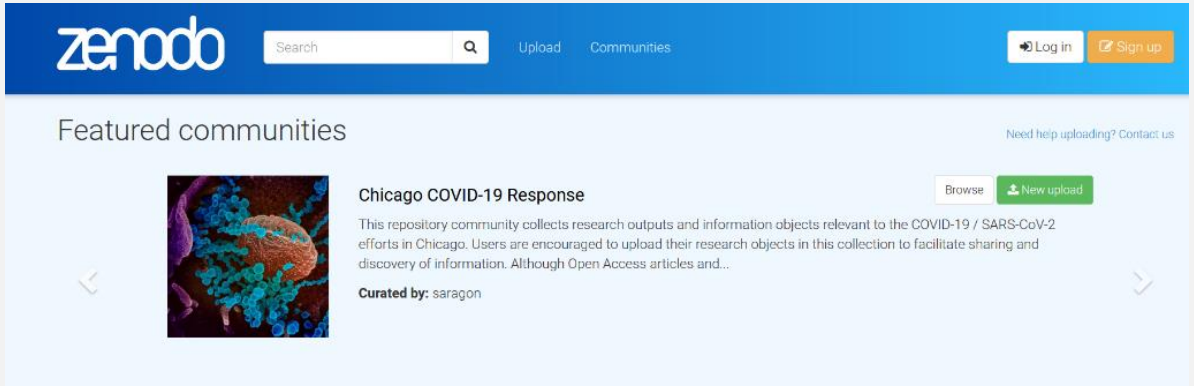
Images: Search by Subject, CC BY 4.0



REPOSITORIES: EXAMPLE 1

Zenodo

- https://zenodo.org
- by CERN



CERN Data Centre & Invenio, CC BY 4.0, https://zenodo.org.

REPOSITORIES: EXAMPLE 2

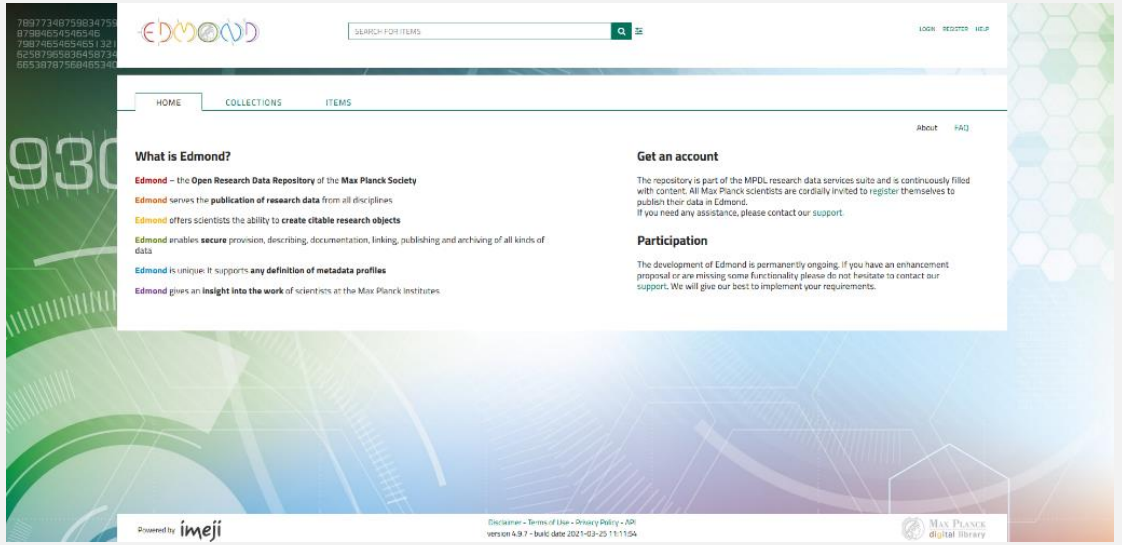
Discuss.Data

- https://discuss-data.net
- repository for Post-Soviet region
- run by University of Bremen and Göttingen State and University Library

REPOSITORIES: EXAMPLE 3

Edmond

- https://edmond.mpdl.mpg.de
- by Max Planck Digital Library for Max Planck Society



DATA REPOSITORY SOFTWARE

Dataverse

- Website: https://dataverse.org
- Example:
  - Harvard Dataverse: https://data.harvard.edu/dataverse
  - DataverseNO: https://dataverse.no

DATA REPOSITORY SOFTWARE

Invenio

- Website: https://inveniosoftware.org
- Example:
  - Zenodo by CERN: https://zenodo.org

DATA REPOSITORY SOFTWARE

Fedora

- Website: https://docs.fedoraproject.org
- Example:
  - Bibliothèque nationale de France: http://www.bnf.fr

DATA REPOSITORY SOFTWARE

DSpace

- Website: https://duraspace.org/dspace/
- Example:
  - The National Parliamentary Library of Georgia: https://dspace.nplg.gov.ge

RESEARCH DATA DISCOVERY

- Research Data Australia
- Portage Network (Kanada)
- CESSDA (Europa)

- Data Citation Index (Clarivate)

REPOSITORY CERTIFICATION

- CoreTrustSeal (https://www.coretrustseal.org)
- CLARIN certificate B (https://www.clarin.eu)
- DIN 31644 (https://www.beuth.de/en/standard/din-31644/147058907)
- DINI Certificate (https://dini.de/dienste-projekte/dini-zertifikat/)
- RatSWD (https://www.konsortswd.de/ratswd/)

DATA JOURNALS

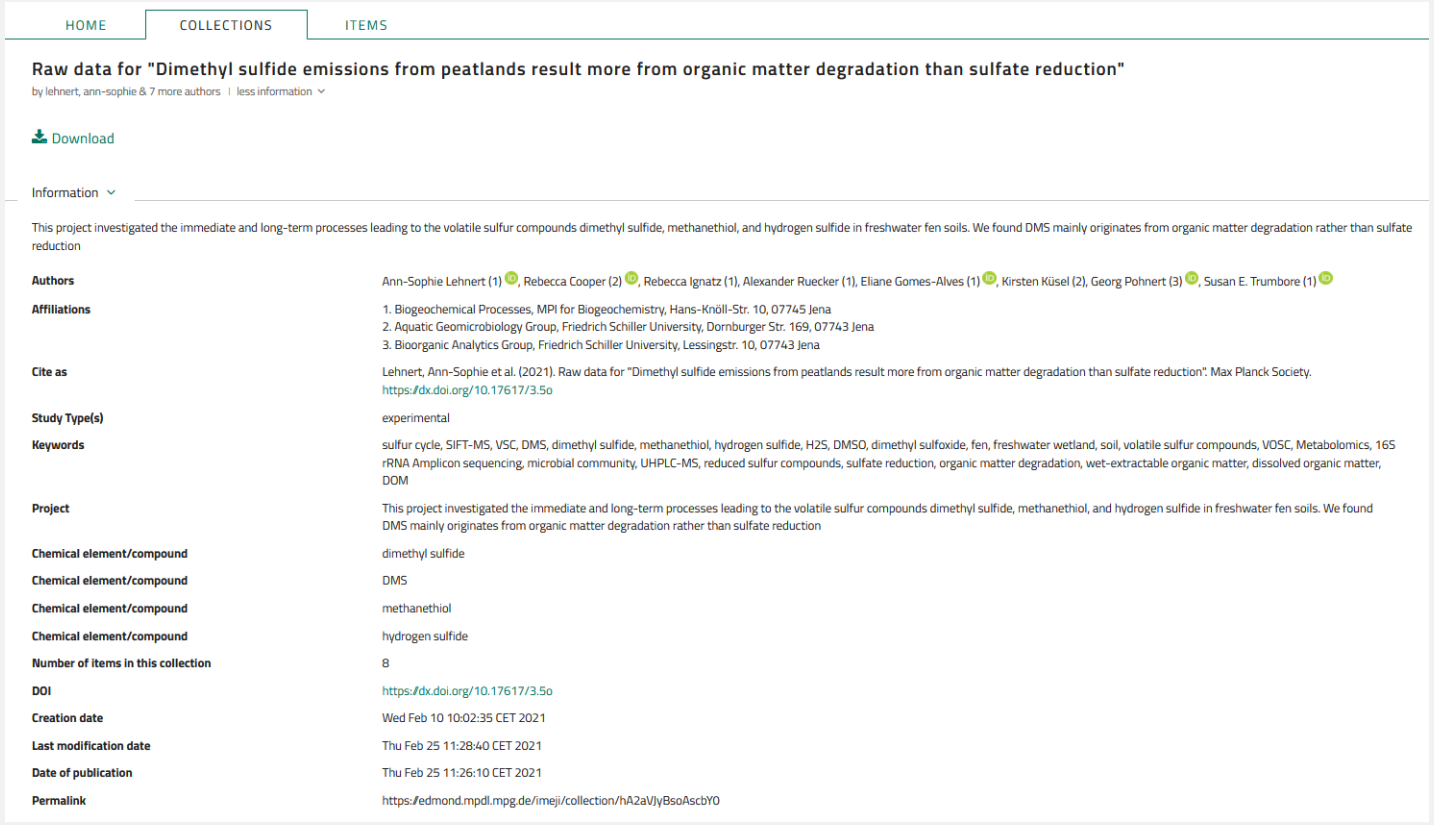
- Life Sciences
  - Gigascience
  - Biodiversity Data Journal
  - Open Health Data
- Earth Sciences
  - Earth System Science Data
- Natural sciences in general
  - Scientific Data

BREAK (5 MINUTES)

https://quickdraw.withgoogle.com

METADATA

METADATA



HOME | COLLECTIONS | ITEMS

Raw data for "Dimethyl sulfide emissions from peatlands result more from organic matter degradation than sulfate reduction"

by lehnert, ann-sophie & 7 more authors | less information

Download

Information

This project investigated the immediate and long-term processes leading to the volatile sulfur compounds dimethyl sulfide, methanethiol, and hydrogen sulfide in freshwater fen soils. We found DMS mainly originates from organic matter degradation rather than sulfate reduction

Authors	Ann-Sophie Lehnert (1), Rebecca Cooper (2), Rebecca Ignatz (1), Alexander Ruecker (1), Eliane Gomes-Alves (1), Kirsten Küsel (2), Georg Pohnert (3), Susan E. Trumbore (1)
Affiliations	1. Biogeochemical Processes, MPI for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena 2. Aquatic Geomicrobiology Group, Friedrich Schiller University, Dornburger Str. 169, 07743 Jena 3. Bioorganic Analytics Group, Friedrich Schiller University, Lessingstr. 10, 07743 Jena
Cite as	Lehnert, Ann-Sophie et al. (2021). Raw data for "Dimethyl sulfide emissions from peatlands result more from organic matter degradation than sulfate reduction". Max Planck Society. https://dx.doi.org/10.17617/3.5o
Study Type(s)	experimental
Keywords	sulfur cycle, SIFT-MS, VSC, DMS, dimethyl sulfide, methanethiol, hydrogen sulfide, H2S, DMSO, dimethyl sulfoxide, fen, freshwater wetland, soil, volatile sulfur compounds, VOSC, Metabolomics, 16S rRNA Amplicon sequencing, microbial community, UHPLC-MS, reduced sulfur compounds, sulfate reduction, organic matter degradation, wet-extractable organic matter, dissolved organic matter, DOM
Project	This project investigated the immediate and long-term processes leading to the volatile sulfur compounds dimethyl sulfide, methanethiol, and hydrogen sulfide in freshwater fen soils. We found DMS mainly originates from organic matter degradation rather than sulfate reduction
Chemical element/compound	dimethyl sulfide
Chemical element/compound	DMS
Chemical element/compound	methanethiol
Chemical element/compound	hydrogen sulfide
Number of items in this collection	8
DOI	https://dx.doi.org/10.17617/3.5o
Creation date	Wed Feb 10 10:02:35 CET 2021
Last modification date	Thu Feb 25 11:28:40 CET 2021
Date of publication	Thu Feb 25 11:26:10 CET 2021
Permalink	https://edmond.mpdl.mpg.de/imeji/collection/hA2aVJyBsoAscbY0

https://dx.doi.org/10.17617/3.5o

METADATA

Metadata Examples

- bibliographic information (title, author, …)
- contextual information (subject, geographical location, temporal coverage, keywords, …)
- administrative aspects (date of creation, file type, access rights, …)
- technical metadata (file properties and file sizes, …)

METADATA

- Bibliographic metadata
  - DC – Dublin Core Metadata Initiative
  - SKOS – Simple Knowledge Organization System
- Subject metadata
  - RDF – Resource Description Framework
  - Metadaten-Standards
  - (Metadaten-Registries)
- Technical metadata, licensing metadata, provision metadata

METADATA

Dublin Core

- widely used generic metadata standard
- https://dublincore.org

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description>
    <dc:title>Homage to Catalonia,</dc:title>
    <dc:creator>Orwell, George,1903-1950.</dc:creator>
    <dc:type>text</dc:type>
    <dc:publisher>London, Secker and Warburg</dc:publisher>
    <dc:date>[1938]</dc:date>
    <dc:language>eng</dc:language>
  </rdf:Description>
</rdf:RDF>
```

Library of Congress, Example of Dublin Core, CC0.

METADATA

DataCite Metadata Schema

- more specialized for research data
- https://schema.datacite.org

METADATA

Overview on Standards

- Data Curation Centre: http://www.dcc.ac.uk/resources/metadata-standards/list
- Research Data Alliance: http://rd-alliance.github.io/metadata-directory/

PARADATA

"Paradata is a term used to describe data generated as a by-product of the data collection process."

United States Census Bureau,
https://www.census.gov/topics/research/paradata.html

For example:

- Speech rate
- Time of interview
- Number of contacts

QUESTIONS?

PERSISTENT IDENTIFIER

PERISTENT IDENTIFIERS

- URL/URI: http://zuse.zib.de/collection/wI3XoEDHO8v0ImCa/item/VqNgKUacPrlhqPKu
- URN: urn:nbn:de:bsz:25-opus-14124
- Handle: hdl:11858/00-001M-0000-0019-D20F-6
- DOI: doi:10.1088/0004-637X/715/2/1453

- ORCID: 0000-0002-2880-8947

- ROR: 0061msm67

DIGITAL OBJECT IDENTIFIER

- for digital objects
- https://www.doi.org

Public domain

OPEN RESEARCHER AND CONTRIBUTOR ID

- for persons
- https://orcid.org

Public domain, CC0

RESEARCH ORGANIZATION REGISTRY

- for organizations
- https://ror.org

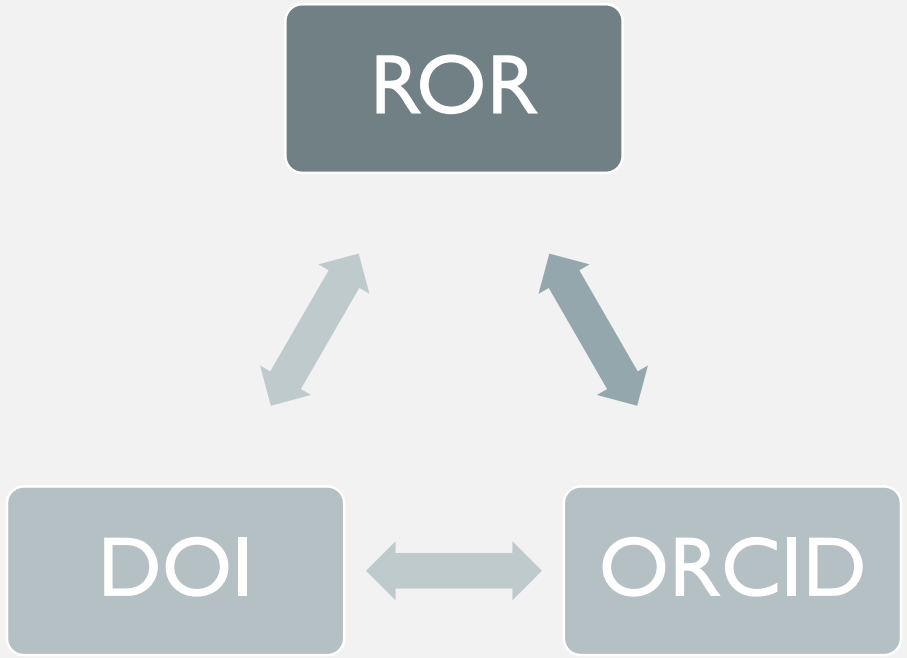
Research Organization Registry, CC BY 4.0,
https://doi.org/10.5281/zenodo.4701802

LIVE DEMONSTRATION

Ilia State University:

- https://ror.org/051qn8h41
- https://commons.datacite.org/ror.org/051qn8h41

IDENTIFIER NETWORK



LICENSES

DATA LICENSES

- Creative Commons (https://creativecommons.org)
  - CC 0
  - CC BY
- Open Data Commons (https://opendatacommons.org)
  - ODC-By
  - PDDL
- Datenlizenz Deutschland

CREATIVE COMMONS

- https://creativecommons.org
- initially designed for the licensing cultural material, but the latest version 4.0 is also suitable for scientific data.

OPEN DATA COMMONS

- https://opendatacommons.org
- explicitly designed for data

RECOMMENDATION FOR RESEARCH DATA

- CC0 or CC BY

QUESTIONS?

PART 3

Research Data Management

STRUCTURE OF PART 3

1. Why Research Data Management?
2. Data Management Plans
3. Research Data Policies
4. FAIR Principles
5. Jupyter

WHY RESEARCH DATA MANAGEMENT?

WHY RESEARCH DATA MANAGEMENT?

You **should** think about it, because:

- Re-use data, at least oneself, but also other
- Scientific reputation
- Acceptance of data as a separate publication is likely to increase in the future
- Because "fits somehow" has never won a Nobel Prize

WHY RESEARCH DATA MANAGEMENT?

You **must** think about it, because:

- Good Scientific Practice
- Code of Conduct
- Funding Agencies, i.e. by DFG and Horizon Europe, Wellcome Trust

DFG'S CODE OF CONDUCT "SAFEGUARDING GOOD RESEARCH PRACTICE"

Guideline 13: Providing public access to research results

[..] "If it has been decided to make results available in the public domain, researchers describe them clearly and in full. Where possible and reasonable, this includes making the research data, materials and information on which the results are based, as well as the methods and software used, available and fully explaining the work processes." (p. 17)

Guideline 17: Archiving

"Researchers back up research data and results made publicly available, as well as the central materials on which they are based and the research software used, by adequate means according to the standards of the relevant subject area, and retain them for an appropriate period of time." (p. 20)

http://doi.org/10.5281/zenodo.3923602

MAX PLANCK SOCIETY: "GOOD SCIENTIFIC PRACTICE"

„**a) General regulations governing scientific practice:** [..]

– reliable securing and storage of primary data for 10 years; clear and comprehensible documentation of the methods employed (e.g. lab book) and all important results," (p. 2)

„**4. Securing and storing primary data**

Primary data as a basis for publications must, as far as possible, be stored for at least ten years on durable, secure carriers in the institutes or research establishments in which they arose. Either the institute or the central organization must ensure that data remains readable for at least this length of time. Access to the data has to be granted for persons with a justifiable interest." (p. 4)

https://www.mpg.de/197494/rulesScientificPractice.pdf

POLICY FOR SCIENTIFIC INFRASTRUCTURE

Research ice breaker Polarstern (https://www.awi.de/expedition/schiffe/polarstern.html)

Data Flow Framework: https://www.awi.de/en/about-us/service/computing-centre/data-flow-framework.html

Mosaik Expedition 2019:

- Data Policy: https://mosaic-expedition.org/wp-content/uploads/2020/12/mosaic_datapolicy.pdf
- Science Plan: https://mosaic-expedition.org/wp-content/uploads/2020/12/mosaic_scienceplan.pdf

Janek Uin: MOSAIC-Expedition, Polarwinter 2019/2020, CC BY 4.0.

QUESTION: CODE OF CONDUCT OR GOOD SCIENTIFIC PRACTICE BY THE ILIA STATE UNIVERSITY?

EUROPEAN COMMISSION

ERC Work Programme 2021:

„Finally, as from 2021 it is no longer possible for applicants to opt out of the submission of Research Data Management plans." (p. 4)

European Commission Decision C(2021) 930, 22/02/2021

Marie Skłodowska-Curie Actions:

"[..] data management plan submitted at mid-term and an update towards the end of the project if needed" (p. 82)

European Commission Decision C(2021) 4200, 15/06/2021

EUROPEAN COMMISSION

Horizon Europe

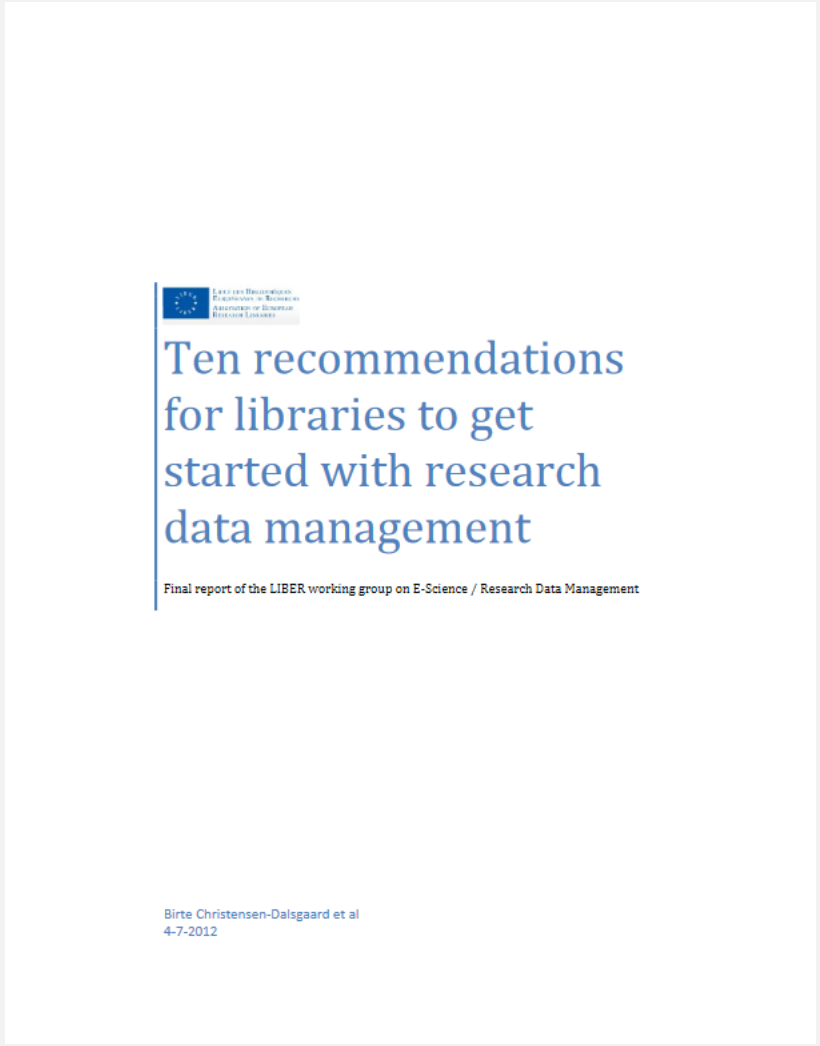
”Open science: research data management
The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- **establish a data management plan** ('DMP') (and regularly update it)
- as soon as possible and within the deadlines set out in the DMP, **deposit the data in a trusted repository** [..]
- as soon as possible and within the deadlines set out in the DMP, **ensure open access** — via the repository – to the deposited **data** [plus CC0, CC BY or equivalent]
- provide **information** via the repository about any research output or any other tools and instruments **needed to re-use or validate the data**.”

Horizon Europe and Euratom: General Model Grant Agreement, Version 1.0, 1. June 2021, p. 109, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-euratom_en.pdf

LIBER RECOMMENDATIONS FOR LIBRARIES

Final report of the LIBER working group on E-Science / Research Data Management, 2012, https://libereurope.eu/wp-content/uploads/2020/11/The-research-data-group-2012-v7-final.pdf.


Ten recommendations for libraries to get started with research data management

Final report of the LIBER working group on E-Science / Research Data Management

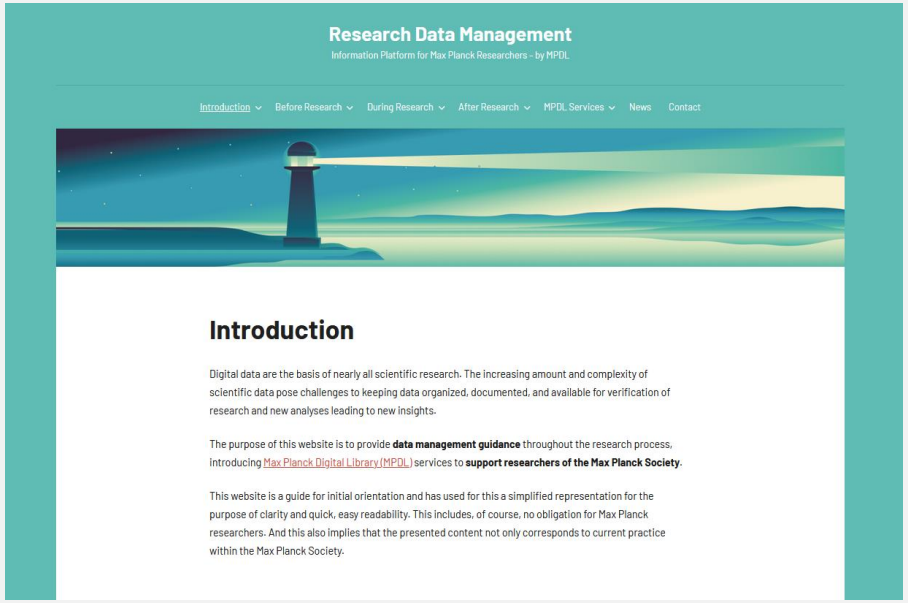
Birte Christensen-Dalsgaard et al
4-7-2012

DIFFERENT ASPECTS OF RDM

MPDL RDM INFORMATION PLATFORM

Sections:

- Before
- During
- After



https://rdm.mpdl.mpg.de

DATA MANAGEMENT PLANS

WHY WRITING DMPS?

– increasingly required by funding agencies

– costs for data management (software, hardware, technical expertise) are often eligible for project grants

WHO TO WRITE A DMP

1. Brainstorming yourself
2. Use a Template
3. Use a Tool

ASPECTS WITHIN A DMP

DMPs usually cover the following aspects:

1. Data description and collection
2. Documentation and metadata
3. Storage and backup
4. Legal and ethical requirements
5. Data sharing and long-term preservation
6. Responsibilities and resources

QUESTIONS WTHIN A DMP

The following questions can be answered in a DMP:

- What data is collected/used?
- How is the data processed?
- Which standards are used?
- How is the data documented?
- Where is the data stored (openly) and when?
- What, if any, speaks against publication?
- What expenses arise from research data management?

Checkliste by ETH Zurich (https://www.dlcm.ch/download_file/force/66/371)

HOW TO DMP

Digitale Tools:

- Argos (https://argos.openaire.eu)
- DMPTool (https://dmptool.org)
- DMPOnline (https://dmponline.dcc.ac.uk)
- RDMO (https://rdmo.aip.de)

Helbig, K., Paul-Stüve, T., & Rex, J. (2021): DMP-Toolguide, doi:10.5281/zenodo.4632308.

LIVE DEMONSTRATION OF RDMO

https://rdmo.forschungsdaten.info

Sign-up: https://rdmo.forschungsdaten.info/account/signup/

EXAMPLES OF DMPS

- Digital Curation Centre Example DMPs and Guidance: http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples
- LIBER Europe DMP Catalogue: https://libereurope.eu/dmpcatalogue/
- Examples for Horizon 2020 DMPs by the University of Vienna: https://phaidra.univie.ac.at/search#?page=1&pagesize=10&collection=o:1140797

COSTS OF RESEARCH DATA MANAGEMENT

- OpenAIRE: https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs
- Data Management Costing Tool by the UK Data Archives: https://www.ukdataservice.ac.uk/manage-data/plan/costing

FURTHER READING ON DMPS

- Michener, W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol 11(10): e1004525. https://doi.org/10.1371/journal.pcbi.1004525.
- Science Europe (2021) Practical Guide to the International Alignment of Research Data Management. https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf.

QUESTIONS?

LEGAL AND ETHICAL ASPECTS

MPDL RDM INFORMATION PLATFORM

https://rdm.mpdl.mpg.de/before-research/legal-and-ethical-aspects/

CASTELLUM

- A Privacy-Compliant Subject Management for Scientific Research
- Software Solution by the Max Planck Institute for Human Development
- https://www.mpib-berlin.mpg.de/research-data/castellum and https://castellum.mpib.berlin/
- Code: https://git.mpib-berlin.mpg.de/castellum

FILE HANDLING

MPDL RDM INFORMATION PLATFORM

https://rdm.mpdl.mpg.de/before-research/file-handling/

VERSION CONTROL WITH GIT

- https://github.com/topics/research-data
- Further Reading: https://rdm.mpdl.mpg.de/2021/06/01/using-git-for-research-data-management/
- More on Git systems in part 4

EXAMPLE FOR GIT VERSIONING 1

- GitHub Repository on education research data: https://github.com/Cghlewis/data-wrangling-functions
- Versioning research data: https://github.com/Cghlewis/data-wrangling-functions/commit/890ad14374bacc9124333ccdd7f2511afc4bae97

EXAMPLE FOR GIT VERSIONING 2

- GitHub Repository on education research data: https://github.com/CSSEGISandData/COVID-19
- Making research data available: https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports

FILE FORMATS

MPDL RDM INFORMATION PLATFORM

https://rdm.mpdl.mpg.de/before-research/file-formats/

DATA STORAGE AND BACKUP

MPDL RDM INFORMATION PLATFORM

https://rdm.mpdl.mpg.de/during-research/data-backup-and-storage/

SOME BACKUP STRATEGIES

3 – 2 – 1 rule

- 3 copies of your data
- on 2 different media
- 1 copy off side

10 steps by CESSDA

https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/4.-Store/Backup

RDM POLICIES

GUIDELINES (POLICY) FOR RDM

- Who owns the data?
- What Requirements are Imposed By Others?
- Which Data Should Be Retained?
- For How Long Should Data Be Maintained?
- How Should Digital Data Be Preserved?
- Are there Ethical Considerations?
- How are Data Accessed?
- How Open Should the Data Be?
- How Will Costs Be Managed?
- What are the Alternatives to Local Data Management?

Ricky Erway: Starting the Conversation: University-wide Research Data Management Policy (2013), OCLC Research

RDM POLICIES

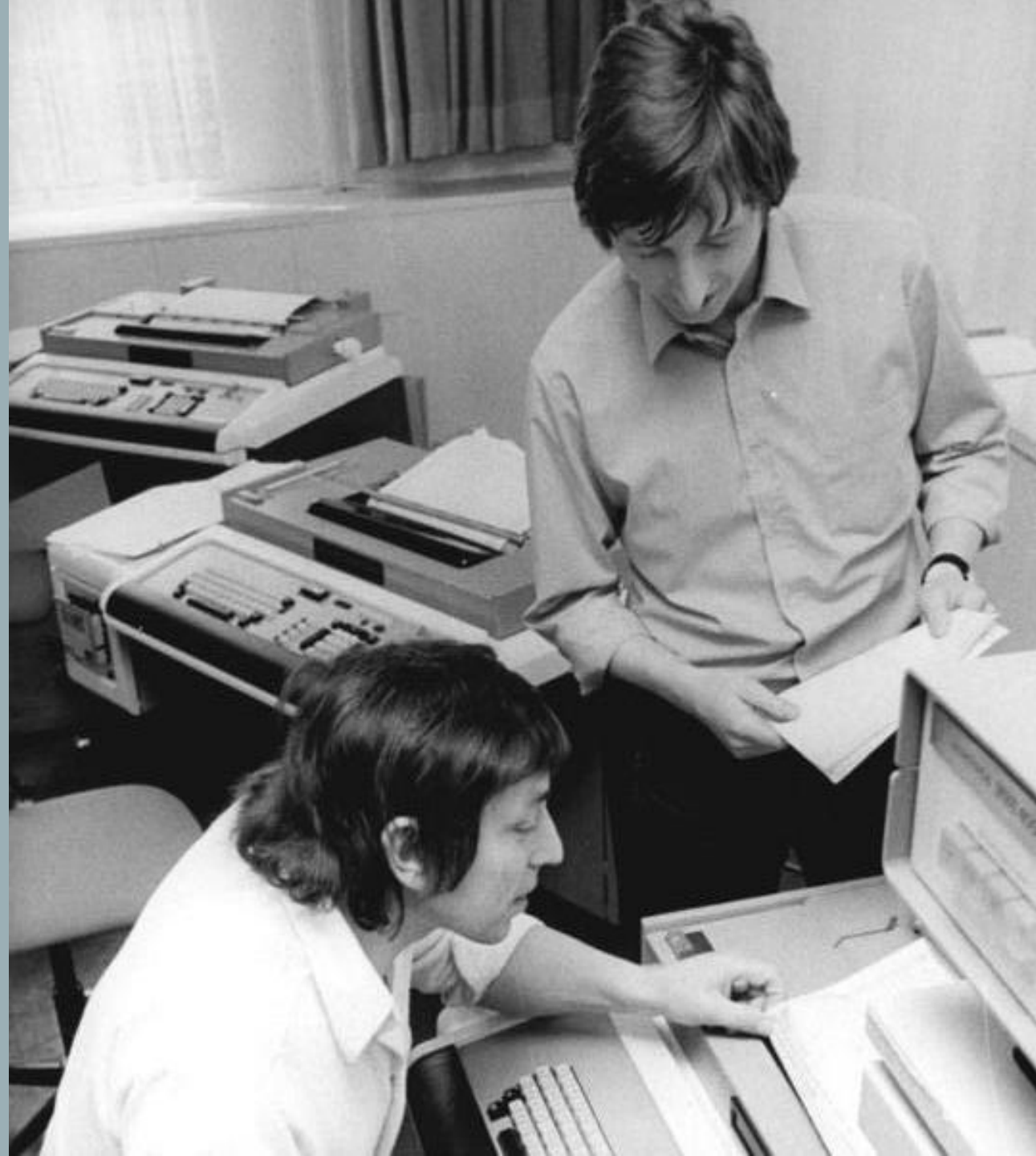
Framework for handling research data at (German) institution:

- Well-known German example: DFG "Guidelines for Safeguarding Good Scientific Practice“, https://wissenschaftliche-integritaet.de/en/research-integrity-by-the-dfg/
- Research data policy of Freie Universität Berlin from 2021, http://dx.doi.org/10.17169/refubium-30560
- Extensive listing: https://www.forschungsdaten.org/index.php/Data_Policies

BREAK (5 MINUTES)

https://jacksonpollock.org

FAIR PRINCIPLES



FAIR PRINCIPLES

Findable

Accessible

Interoperable

Reusable

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18.

SORBONNE DECLARATION

- Sorbonne Delcaration on Research Data Rights (https://www.leru.org/files/Sorbonne-declaration.pdf)

DIE ZEIT vom 30.01.2020

DIE ZEIT

DIE POSITION

Teilt euren Datenschatz!

Forschung lebt vom Austausch. Ihn zu stärken, versprechen Spitzenuniversitäten in der Sorbonne-Erklärung

Von Georg Krausch und Jan Wöpking

Daten sind das neue Öl, heißt es oft. Doch anders als Öl sind Daten erneuerbar, und ihr Umfang wächst täglich. Prognosen zufolge verdoppelt sich das weltweite Datenvolumen sogar alle 18 Monate. Ein riesiger Schatz, dessen Potenzial bisher kaum genutzt wird. Das muss sich ändern. Denn ob bei künstlicher Intelligenz, personalisierter Medizin, Elektromobilität oder industrieller Fertigung: Künftig geht nichts mehr ohne kluge Datenpolitik. Auch Bundesregierung und EU haben das erkannt und arbeiten an entsprechenden Grundsätzen.

Für die Innovationsfähigkeit unserer Gesellschaft sind Forschungsdaten von überragender Bedeutung. Sie bilden die Grundlage für bahnbrechende Entdeckungen in Medizin, Klima- und Materialforschung und eröffnen neue Forschungsfelder im Bereich der Gesundheits-, Sozial- und Kulturwissenschaften. Allerdings nur, wenn die Datenberge für Forschungszwecke zugänglich und nutzbar gemacht werden. Eine solche Öffnung würde die Verbindung von bisher getrennten Wissensfeldern erlauben und gerade in der Zweit- oder Drittverwendung von Daten enormes Innovationspotenzial freisetzen.

Um den Datenaustausch zu befördern, haben Anfang der Woche in Paris neun Verbände international führender Universitäten die »Sorbonne-Erklärung für offene Forschungsdaten« verabschiedet. Zusammen repräsentieren sie mehr als 160 der weltweit stärksten Forschungsuniversitäten, von Oxford bis Kapstadt, von Paris bis Tokio, von Berlin bis Toronto. Zu ihnen gehört auch der deutsche Verband forschungsstarker Universitäten, die German U15.

Die Sorbonne-Erklärung ist ein starkes Plädoyer für eine globale Kultur frei zugänglicher Forschungsdaten: Diese »sollen, soweit es nur geht, offen geteilt und wiederverwendet werden«. Zugleich hält die Erklärung fest, dass private, personenbezogene Daten geschützt und Urheberrechte respektiert werden müssen.

Was aber braucht es, um eine neue Kultur des Austauschs von Forschungsdaten umfassend und nachhaltig zu etablieren?

1. Die wissenschaftliche Gemeinschaft muss den komplexen Wandel in Richtung offener Forschungsdaten vorantreiben und gestalten. Sie muss auf hoher Datenqualität bestehen. Und sie muss bei der Gestaltung künftiger Clouds und Apps zum Teilen der Daten nicht nur auf technische Machbarkeit, sondern vor allem auf die Nutzerfreundlichkeit achten. Damit sie auch tatsächlich Akzeptanz unter den Wissenschaftlerinnen und Wissenschaftlern finden.

2. Freie Daten gibt es nicht umsonst. Es braucht Investitionen in bestehende und neue Infrastruktur. Vorhaben wie die Nationale Forschungsdateninfrastruktur und die European Open Science Cloud sind vielversprechende erste Ansätze. Zugleich müssen Politik und Fördermittelgeber aber auch Geld für das Forschungsdatenmanagement bereitstellen.

3. Das Teilen von Forschungsdaten muss politisch durch kluge Regeln und Gesetze unterstützt werden. Dabei ist es besonders wichtig zu verhindern, dass Abhängigkeitsverhältnisse zwischen kommerziellen Anbietern oder Plattformen und Forschungseinrichtungen entstehen.

4. Am wichtigsten ist jedoch: Eine offene Datenkultur braucht das Vertrauen der beteiligten Akteure und einen kooperativen Wettbewerb. Das sind die größten Herausforderungen, nicht die technischen Aspekte. Wenn es um das Teilen von Daten geht, herrscht oft eine Pattsituation, die an das spieltheoretische Gefangenendilemma erinnert: Obwohl alle Akteure von einer Öffnung profitieren würden, überwiegen Skepsis und Misstrauen. Die Sorbonne-Erklärung will auch hier ein Zeichen setzen, indem internationale Spitzenuniversitäten gemeinsam für eine offene Forschungskultur eintreten.

Nach der Berliner Erklärung von 2003, die heute als Meilenstein der Open-Access-Bewegung gilt, ist die Öffnung der Forschungsdaten der nächste Schritt.

FINDABLE

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

https://www.go-fair.org/fair-principles/

ACCESSIBLE

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

https://www.go-fair.org/fair-principles/

INTEROPERABLE

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

https://www.go-fair.org/fair-principles/

REUSABLE

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

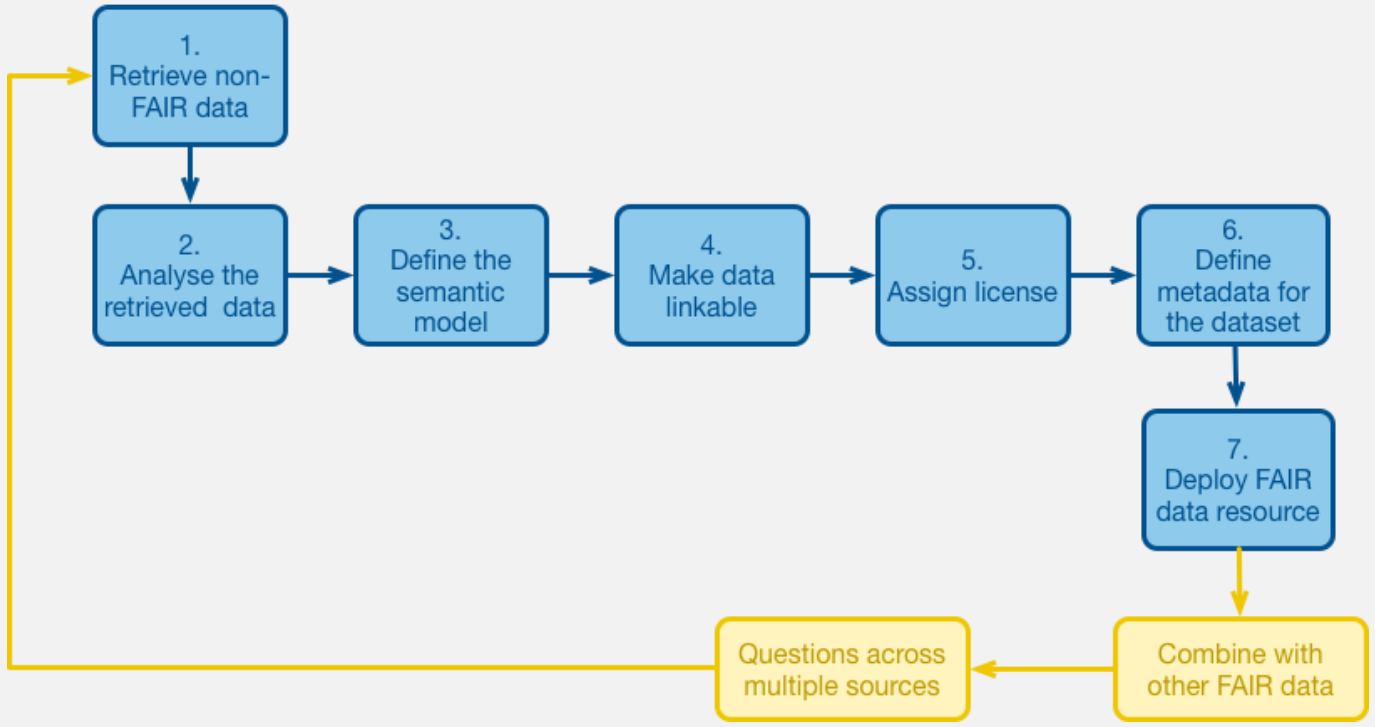
R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

https://www.go-fair.org/fair-principles/

FAIRIFICATION PROCESS



LIVE DEMONSTRATION: F-UJI

https://www.f-uji.net

i.e. https://dspace.nplg.gov.ge/handle/1234/19143

WEBINAR INFORMATION

User Experience with FAIR evaluation tools and services

https://www.cessda.eu/Training/Event-Calendar/Webinar-on-User-Experience-with-FAIR-evaluation-tools-and-services-for-example-the-FAIR-Evaluation-Services-test-informally-known-as-the-Wilkinson-s-FAIR-Test

Tomorrow 4th November 10:00 CET by GESIS – Leibniz-Institute for the Social Sciences in English

WEBINAR INFORMATION

Making Social Science Research Transparent

https://www.cessda.eu/Training/Event-Calendar/Making-Social-Science-Research-Transparent

11th November 13:30 CET by GESIS – Leibniz-Institute for the Social Sciences in English

BEST PRACTICE: RESEARCH DATA AND CODE FOR PUBLICATION

http://hdl.handle.net/21.11116/0000-0009-3F2C-1

QUESTIONS?

QUESTION FROM SESSION 1:

HOW COMPLEX IS IT TO TRANSFER DATA FROM ONE REPOSITORY TO ANOTHER?

TSAMMALEX DATASET

2014: https://edmond.mpdl.mpg.de/imeji/collection/d2JGQRxO19XTOEXG

vs.

2015: http://dx.doi.org/10.5281/zenodo.17571

vs.

2018: https://github.com/tsammalex/tsammalex/

vs.

https://tsammalex.clld.org

JUPYTER NOTEBOOK

JU PYT E R

- name Jupyter refers to the three programming languages Julia, Python and R. The file name extension is .ipynb.
- open source web application

JUPYTER NOTEBOOK

- allows you to create, work and share documents for live code, equations, visualisations and narrative text
- easy-to-use platform for data analyses
- Jupyter document is a .json document with a versioned schema consisting of a list of input and output cells and markdown text

JUPYTER NOTEBOOK

Advantages:

- enables you as researcher to work with text editors, terminals and custom components in a flexible, integrated and extensible way
- easy to use
- does not need to be installed, which is very practical in teaching

MARKDOWN + PYTHON = REPRODUCIBILITY

5/4/21	60563	131327	122999	13316	27284	1232	3047417	217407	29865	624

469 rows × 192 columns

Exploring the data: Plotting and filtering

Plotting

Appending `.plot()` on a dataset, plots all column values vs. the index column.

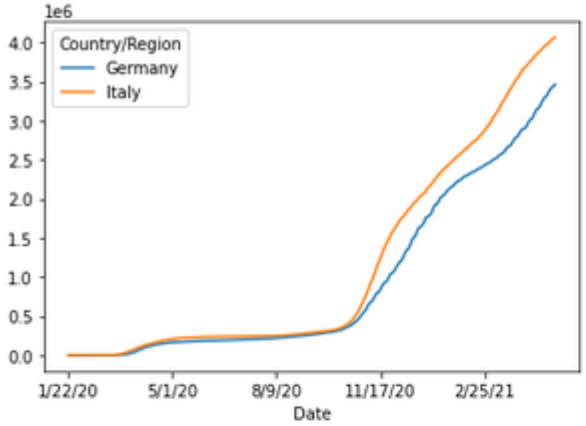
[97]:
```
cases.plot()
```

• • •

plot only selected countries

[64]:
```
cases.loc[:,["Germany", "Italy"]].plot()
```

[64]:
```
<AxesSubplot:xlabel='Date'>
```



Filtering

Selecting rows and columns from a dataset, based on index or column headers is facilitated through the `.loc[]` slice

Filter on dates (index).

[65]:
```
# Get cases on April 1 2020 (Row filter)
cases.loc["4/1/20"]
```

[65]:
```
Country/Region
Afghanistan      197
Albania          259
Algeria          847
```

GOOGLE COLAB

- Go futher and start immediately with Google CoLab: https://colab.research.google.com

SHARING JUPYTER

- GitHub (https://github.blog/2015-05-07-github-jupyter-notebooks-3/)
- NBViewer (https://nbviewer.org)

More than only sharing:

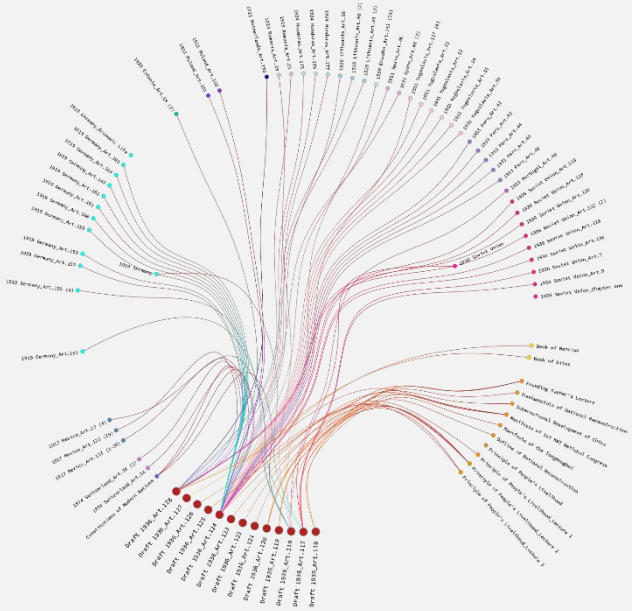
- https://mybinder.org/

MORE NOTEBOOKS

- RStudio (https://www.rstudio.com)
- Apache Zeppelin (https://zeppelin.apache.org)
- Spark Notebook (https://spark-notebook.io)

JUPYTER AND RESEARCH DATA IN THE HUMANITIES

Wagner, Andreas, & Li, Fupeng. (2020): rg-mpg-de/fupeng: Citation Network Visualisation of two Drafts (1923 and 1936) of the Chinese Constitution (v1.0). Zenodo. https://doi.org/10.5281/zenodo.3699154.



LIVE DEMONSTRATION JUPYTER NOTEBOOK

Covid19 Monitor

https://gitlab.gwdg.de/mpievolbio-scicomp/fdm2021/-/tree/master

https://mybinder.org/v2/git/https%3A%2F%2Fgitlab.gwdg.de%2Fmpievolbio-scicomp%2Ffdm2021.git/HEAD?urlpath=lab

JOURNALS & PUBLICATION PLATFORMS

Journals

- Copernicus ESSD, https://essd.copernicus.org/articles/ (ERC documents)
- eLife https://elifesciences.org/about/aims-scope (ERA documents)

Platforms

- CurveNote: https://curvenote.com/ (Jupyter derivative)
- Stenci.la: https://stenci.la/ (ERA documents)

WHERE TO PUBLISH JUPYTER ELSE?

GitHub

- Example: https://github.com/martius-lab/EQL

HPC CLUSTERS AND JUPYTER

Some examples

- Pangeo: https://gallery.pangeo.io/contributing.html (Jupyter documents)
- NeuroLibre: https://www.neurolibre.com/about/ (Jupyter documents)

FURTHER READING ON JUPYTER

- Nüst, D. (2021). A Web service for executable research compendia ... (Version 2), https://doi.org/10.5281/zenodo.5108218.
- Guizzardi, G. et al. (2021). Announcing the next phase of Executable Research Articles, https://elifesciences.org/labs/a04d2b80/announcing-the-next-phase-of-executable-research-articles.
- Feature Discussion (2021) on Jupyter forums: https://discourse.jupyter.org/t/feature-idea-jupyterhub-binderhub-jupyter-book-as-a-publishing-platform/8359

PART 4

Applications

STRUCTURE OF PART 4

1. NFDI
2. EOSC
3. GOSCI
4. GOSC & GAIA-X
5. Research Software
6. Electronic Laboratory Notebooks

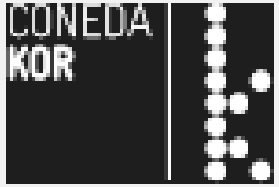
QUESTION FROM SESSION 2: SOFTWARE SOLUTION FOR COLLABORATIVE WORKING ON IMAGES AND TEXT ITEMS

DARIAH (DIGITAL RESEARCH INFRASTRUCTURE FOR THE ARTS AND HUMANITIES)

- European Research Infrastructure Consortium (ERIC)
- Aim: enhance and support digitally-enabled research and teaching across the arts and humanities
- Tools and services: www.dariah.eu/tools-services/tools-and-services/

CONEDAKOR

- a web-based database system with a graph-based architecture
- DARIAH-DE service
- aim: administrate and presentation academic object collections from image-based cultural and human sciences
- https://de.dariah.eu/en/conedakor
- Demo: https://conedakor-demo.de.dariah.eu
- GitHub: https://github.com/coneda/kor



Conedakor, https://github.com/coneda/kor/, GNU AGPLv3

RECOGITO

- Semantic annotation tool for texts and images
- https://recogito.pelagios.org
- Tutorial: https://recogito.pelagios.org/help/tutorial
- GitHub: https://github.com/pelagios/recogito2

GERMAN NATIONAL RESEARCH DATA INFRASTRUCTURE (NFDI)

GERMAN NATIONAL RESEARCH DATA INFRASTRUCTURE (NFDI)

"*The national research data infrastructure (NFDI) is intended to systematically develop, sustainably secure and make accessible the data holdings of science and research and to network them (inter)nationally. It will be established in a process driven by the scientific community as a networked structure of consortia acting on their own initiative.*" (DFG website)

- NFDI: https://www.nfdi.de
- DFG: https://www.dfg.de/foerderung/programme/nfdi/index.html
- Mailinglists: https://lists.nfdi.de/postorius/lists/

SOME FACTS ABOUT THE NFDI

- NFDI = Nationale Forschungsdateninfrastruktur
- Germany wide
- permanent
- application for five years
- funded by BMBF/DFG
- at the end 85mio € per year for 30 consortia

SOME FACTS ABOUT THE NFDI

- discipline-specific research data management
- Community-driven (i.e. Research + Infrastructure)
- models of engagement: applicant, co-applicant, participant
- General Aims:
  - Infrastructure
  - Standards
  - Curricula
  - …
- also interferences with EOSC and GAIA-X

VIDEO ABOUT THE NFDI

M-Create, Düsseldorf für DFG, CC BY 4.0, https://youtu.be/x3Cvn1vNQ98

1ST ROUND OF NFDI CONSORTIA 2019

Acronym	Name	Further information
DataPlant	Data in PLANT research	http://nfdi4plants.de/
GHGA	German Human Genome-Phenome Archive	https://ghga.dkfz.de/
KonsortSWD	Konsortium für die Sozial-, Bildungs-, Verhaltens-und Wirtschaftswissenschaften	https://www.konsortswd.de/
NFDI4BioDiversity	NFDI4BioDiversity: Biodiversity, Ecology & Environmental Data	https://www.nfdi4biodiversity.org/
NFDI4Cat	NFDI for Catalysis-Related Sciences	http://gecats.org/NFDI4Cat.html
NFDI4Chem	Fachkonsortium Chemie für die Nationale Forschungsdateninfrastruktur	https://www.nfdi4chem.de/
NFDI4Culture	Consortium for research data on material and immaterial cultural heritage	https://nfdi4culture.de/
NFDI4Health	National Research Data Infrastructure for Personal Health Data	https://www.nfdi4health.de/
NFDI4Ing	Nationale Forschungsdateninfrastruktur für die Ingenieurwissenschaften	https://nfdi4ing.de/

2ND ROUND OF NFDI CONSORTIA 2020

Acronym	Name	Further information
BERD@NFDI	Business, Economic and Related Data @ NFDI	https://www.berd-nfdi.de/
DAPHNE4NFDI	Data from Photon and Neutron Instruments for NFDI	https://www.sni-portal.de/de/daphne-nfdi
FAIRmat	FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids	https://www.fair-di.eu/fairmat/fairmat_/consortium
MaRDI	Mathematical Research Data Initiative	https://www.mardi4nfdi.de/
NFDI4DataScience	National Research Data Infrastructure for Data Science	https://www.nfdi4datascience.de
NFDI4Earth	NFDI Consortium Earth System Science	http://www.nfdi4earth.de
NFDI4Microbiota	National Research Data Infrastructure for Microbiota Research	http://nfdi4microbiota.de
NFDI-MatWerk	National Research Data Infrastructure for Materials Science & Engineering	https://nfdi-matwerk.de/
PUNCH4NFDI	Particles, Universe, NuClei and Hadrons for the NFDI	https://www.punch4nfdi.de/
Text+	Text+: Language- and Text-Based Research Data Infrastructure	https://www.text-plus.org

NFDI CONSORTIA APPLICATIONS 2021

1. DeBioData
2. FAIRagro
3. InnoMatSafety
4. METHODS
5. NFDI4BIOIMAGE
6. NFDI4Energy
7. NFDI4Immuno
8. NFDI4Memory
9. NFDI4Mobility
10. NFDI4Objects
11. NFDI4Patho
12. NFDI4Phys
13. NFDI-Neuro
14. NFDIxCS
15. TheoRes

Last update 14.09.2021, https://www.dfg.de/foerderung/programme/nfdi/absichtserklaerungen_2021/index.html

NFDI APPLICATION STATISTICS BY THE DFG

DFG: Statistische Übersicht zu den Förderentscheidungen in beiden Ausschreibungsrunden, June 2021, https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_auswertung_2021.pdf, only as pdf on DFG website.

DFG: NFDI – Statistische Übersichten zum Antragseingang, November 2019, https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_auswertung_2019.pdf, only as pdf on DFG website.

DFG: Statistische Übersichten zum Antragseingang – Zweite Ausschreibungsrunde, September 2020, https://www.dfg.de/download/pdf/foerderung/programme/nfdi/nfdi_auswertung_2020.pdf, only as pdf on DFG website.

all only in German…

WHY MIGHT THE NFDI ALSO IMPORTANT FOR GEORGIAN SCIENTISTS?

– idea of setting community standard, which might apply

– German cooperation partners from this context, i.e. using this infrastructure

EXAMPLE: KONSORTSWD

- Data access via data centres

https://www.konsortswd.de/en/konsortswd/tasks/data-access/

EXAMPLE: NFDI4ING VOCABULARIES

- NFDI4Ing Ontology Service: https://terminology.nfdi4ing.de/ts4ing/ontologies

EXAMPLE: NFDI4CULTURE

Task Area 3

- Software Consulting Agency
- Registry for Research Tools
- Development of Research Tools

(https://nfdi4culture.de/what-we-do/task-areas/task-area-3.html)

EUROPEAN OPEN SCIENCE CLOUD

EUROPEAN OPEN SCIENCE CLOUD

For the European Commission the EOSC "is a trusted digital platform for the scientific community, providing seamless access to data and interoperable services that address the whole research data cycle, from discovery and mining to storage, management, analysis and re-use across borders and scientific disciplines".

European Commission, 2019, European open science cloud: A new paradigm for innovation and technology. Publications Office, S. 2, https://data.europa.eu/doi/10.2759/016783.

EUROPEAN OPEN SCIENCE CLOUD

The central goal of the EOSC is nothing less than global EU leadership in research data management. This goes hand in hand to ensure that European researchers have access to all the benefits of data-driven research.

Achim Streit und Jos van Wezel (2021): Deutschland in der European Open Science Cloud, in: M. Putnings, H. Neuroth, & J. Neumann (Hrsg.), Praxishandbuch Forschungsdatenmanagement, S. 32, https://doi.org/10.1515/9783110657807-003.

EUROPEAN OPEN SCIENCE CLOUD

EOSC is not a cloud "made in Brussels", it should help Europe to build a "Schengen Area for digital Data".

Budroni, P., Claude-Burgelman, J., & Schouppe, M. (2019): Architectures of Knowledge: The European Open Science Cloud, in: ABI Technik 39(2), S.140, https://doi.org/10.1515/abitech-2019-2006.

EUROPEAN OPEN SCIENCE CLOUD

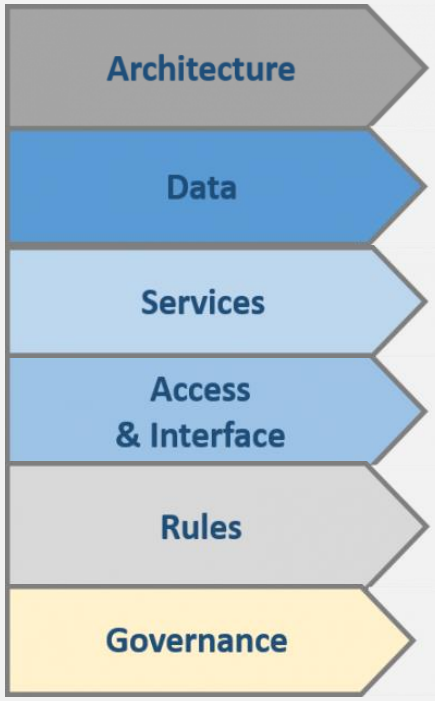
From a more technical perspective EOSC is an integration layer. It aims at establishing interoperability among the existing and forthcoming European Research Infrastructures. So, by federating existing research data infrastructures, the EOSC leverages national investments and adds value in terms of scale, interdisciplinary and faster innovation.

European Commission, https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud.

EUROPEAN OPEN SCIENCE CLOUD

- European Commission 2016
- EOSC Declaration (2017)
- EOSC implementation roadmap (2018)
- Strategic Research and Innovation Agenda of EOSC (2021)
- EOSC main background documents (2021)

EUROPEAN OPEN SCIENCE CLOUD



Architecture of the federated infrastructures as the solution to the current fragmentation in research data infrastructures which are insufficiently interoperable.

FAIR data management and tools. A common data language to ensure data stewardship across borders/disciplines based on FAIR principles.

Available services from a user perspective. A rich environment offering a wide range of services covering the needs of the users.

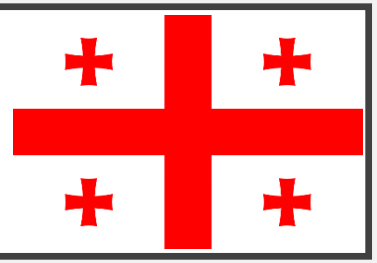
Mechanisms/interfaces for accessing EOSC. A simple way to deal with open data obligations, or to access research data across different disciplines.

Rules of participation for different EOSC actors. An opportunity to comply with existing legal and technical frameworks and increase legal certainty & trust.

Governance of the EOSC, aiming at ensuring EU leadership in data-driven science but requiring new governance frameworks.

GOSCI

GEORGIAN OPEN SCIENCE CLOUD INITIATIVE

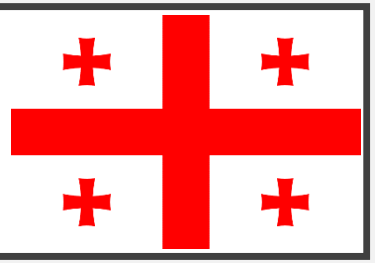


"The main outcome of the GOSCI will be the support towards research and education community of Georgia to have access to scientific data, services and facilities already available and the ones which will be implemented at EOSC in future. **The main purpose will be to support Georgian research teams with integration in European Research Area.**"

https://ni4os.eu/15-national-osc-initiatives/georgia/

https://rustaveli.org.ge/eng/siakhleebi/metsnierebastan-dakavshirebuli-sakitkhebis-gaziarebis-ghonisdzieba

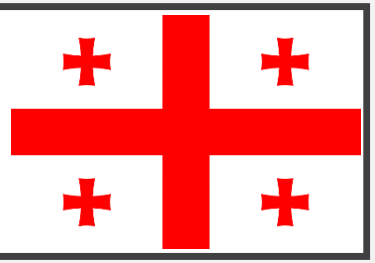
GEORGIAN OPEN SCIENCE CLOUD INITIATIVE



Main purpose:
The main outcome of the GOSCI will be the support towards research and education community of Georgia to have access to scientific data, services and facilities already available and the ones which will be implemented at EOSC in future. The main purpose will be to support Georgian research teams with integration in European Research Area.

Garavelli, Sara, Märkälä, Anu, & Liinamaa, Iiris. (2021). EOSC National Structures: an overview of the national EOSC coordination and engagement mechanisms in Europe. Zenodo. https://doi.org/10.5281/zenodo.5668275, p 60.

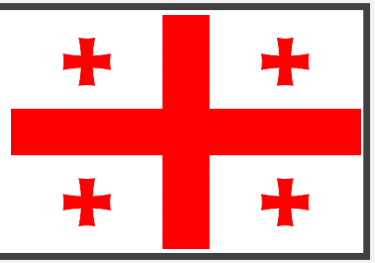
GEORGIAN OPEN SCIENCE CLOUD INITIATIVE



Governance structure:
A consortium. Organizations that expressed interest in participation include research infrastructure, electronic infrastructure, academic libraries, universities, research centers. Currently several institutions have clearly expressed their interest towards establishment of GOSCI: High Energy Physics Institute of Tbilisi State University, Ivane Beritashvili Center of Experimental Biomedicine, Georgian Research and Educational Networking Association GRENA, National Science Library.

Garavelli, Sara, Märkälä, Anu, & Liinamaa, Iiris. (2021). EOSC National Structures: an overview of the national EOSC coordination and engagement mechanisms in Europe. Zenodo. https://doi.org/10.5281/zenodo.5668275, p 60.

GEORGIAN OPEN SCIENCE CLOUD INITIATIVE



Are all the EOSC Association members/observers in the country participating in the EOSC national structure and how?

Currently none of the organizations from Georgia holds status of EOSC Association members/observers, however it is expected that GOSCI member organization will apply for observer status

Garavelli, Sara, Märkälä, Anu, & Liinamaa, Iiris. (2021). EOSC National Structures: an overview of the national EOSC coordination and engagement mechanisms in Europe. Zenodo. https://doi.org/10.5281/zenodo.5668275, p 61.

DISCUSSION: WHAT IS THE ROLE OF SCIENCE CLOUDS IN GEORGIA?

GOSC & GAIA-X AND MORE

GLOBAL OPEN SCIENCE CLOUD

"The Global Open Science Cloud (GOSC) initiative will encourage cooperation, alignment, and ultimately interoperability, between existing and emerging Open Science Clouds"

CODATA, S. 3, https://codata.org/wp-content/uploads/2021/06/GOSC-Introduction-Event-Objectives-and-Agenda.pdf.

GAIA-X

- Franco-German Initiative 2020 for a Data Cloud, with a special focus on the economy
- The declared goal of GAIA-X is to strengthen the digital sovereignty of business, science, government and society by promoting the development of innovation ecosystems.
- In addition, GAIA-X is also about big politics and marketing: it is about European values, about data protection and, above all, about independence.

AND MORE

- China Science and Technology Cloud (CSTCloud)
- African Open Science Platform
- LA Referencia for Latin America,

RESEARCH SOFTWARE

UNDERSTANDING RESEARCH SOFTWARE

- In-house developed research software
- software applications for research
- Infrastructure software/services

In discussion: Is it really helpful to distinguish between research software and "non-research software" (i.e. MS Word etc.)?

RESEARCH SOFTWARE AND DATA

Similarities of software and research data in the research **process**:

- both play an increasingly important role in research
- both are necessary to make research results reproducible
- both represent a potential value for downstream users
- both are not yet (or only to a limited extent) recognised as independent research achievements.

RESEARCH SOFTWARE AND DATA

Similarities of software and research data in **management**:

- both can be metadata tagged and archived
- both need to be curated to ensure usability over time
- both often lack the necessary knowledge for adequate management

RESEARCH SOFTWARE AND DATA

Differences between software and research data:

- curation of software is much more complex
- metadata for software https://codemeta.github.io/ are much more homogeneous than that for research data
- for software there is a lot of prior experience with open/commercial licences https://opensource.org/
- an accurate versioning plays a much greater role for software

RESEARCH SOFTWARE POLICIES

- can help to establish standard procedures
- can help to establish publication processes
- can help by selecting licenses
- can help to clarify the ownership
- can help by commercializing the software
- can help to maintain the software
- can help to give the software into a community
- can improve the acceptance of software publication as scientific value

EXAMPLE FOR RESEARCH SOFTWARE POLICIES

TU Delft

https://doi.org/10.5281/zenodo.4629635

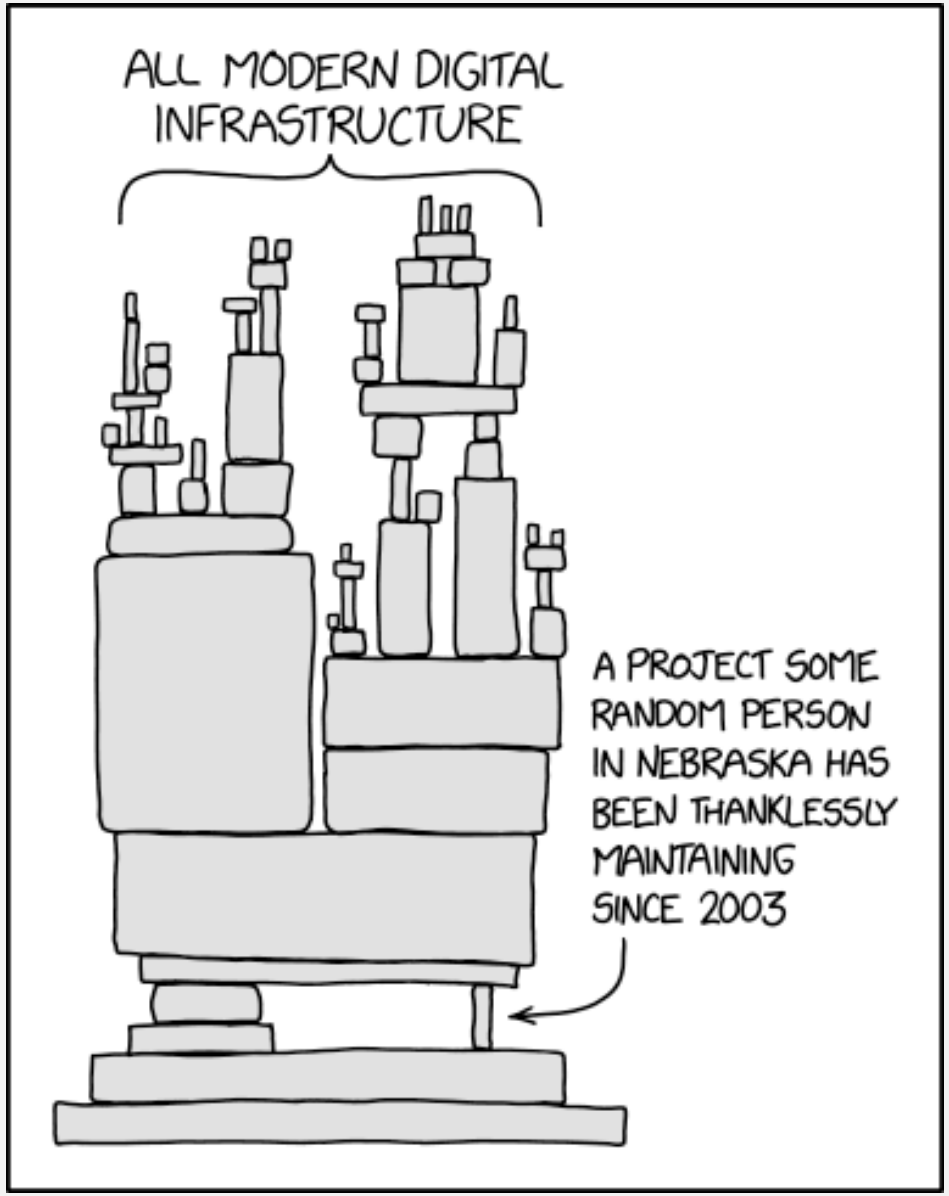
- Licensing
- Registration
- Commercialisation

SECOND FRENCH PLAN FOR OPEN SCIENCE

Path Three "Opening up and promoting source code produced by research"

French Ministry of Higher Education, Second French Plan for Open Science. Generalizing open science in France 2021-2024, July 2021, https://www.ouvrirlascience.fr/wp-content/uploads/2021/10/Second_French_Plan-for-Open-Science_web.pdf pp. 1 and 16-19, CC BY 4.0.

OPEN SOURCE SOFTWARE



xkcd, Dependency, https://xkcd.com/2347/, CC BY-NC 2.5

OPEN SOURCE SOFTWARE

Virtues of Open Source

1. you can re-check the software
2. development flexibility (host for your own, local or as cloud)
3. reduced costs
4. open
5. software publication

ALTERNATIVE OS SOFTWARE

- OSS Directory (https://www.ossdirectory.com)
- https://alternativeto.net
- …

SOFTWARE CARPENTRY

- Teaching basic lab skills for research computing
- take part: https://software-carpentry.org/workshops/
- request a workshop: https://software-carpentry.org/workshops/request/

- Current Example Centre for Digital Life Norway: https://korbinib.github.io/2021-11-15-DLN-swc-online/

LIBRARY AND DATA CARPENTRY

- Library Carpentry: Around software within libraries (https://librarycarpentry.org/)
- Example ZB Med. https://zbmed.github.io/2021-06-07-fdm_nrw-online/

- Data Carpentry: Data skills for research outcome (https://datacarpentry.org)
- Example University of Oslo: https://uio-carpentry.github.io/2021-11-22-uio/

QUESTION FROM SESSION 2: USE LOCAL STORED FILES IN JUPYTER NOTEBOOK

Perhaps, best way by running Jupter Notebook locally:
https://geohackweek.github.io/wiki/running_jupyter_notebooks.html

ELECTRONIC LABORATORY NOTEBOOKS

ASPECTS OF LABORATORY NOTEBOOKS

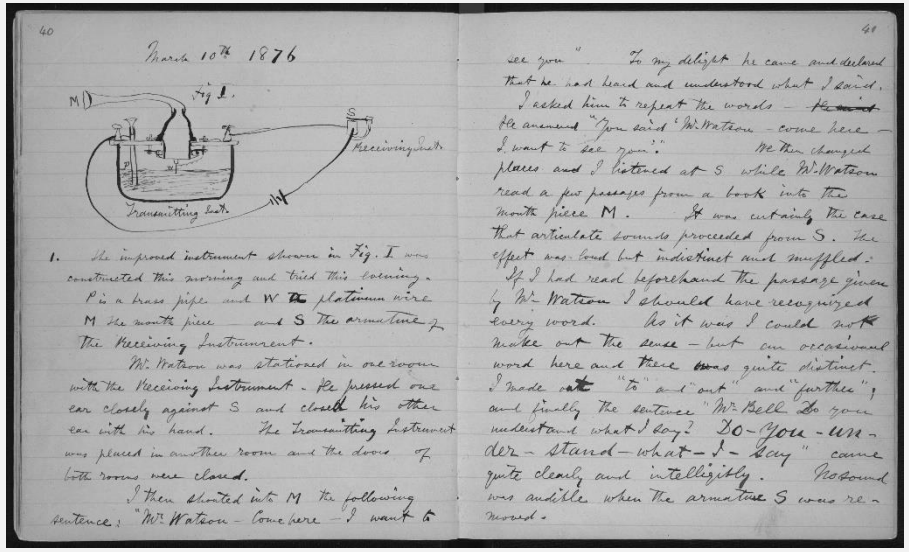
- permanent records of research ideas, concepts, data and observations
- a laboratory notebook is a **legal** record (i.e. evidence at applying for a patent)
- protection of your intellectual property rights

- but only, if the laboratory notebook is used in a correct way

Mende, Michael (2021): Notebook Origins – why document? Talk presented at Digital Workshop "Living with Electronic Laboratory Notebooks", Munich, 21st September 2021, http://hdl.handle.net/21.11116/0000-0009-3F16-9.

USE OF LABORATORY NOTEBOOKS

- never remove pages
- do not write outside the document
- never us non-permanent pen/ink
- sequentially numbered pages
- cross out free space
- do not start a new page until the previous is full
- never remove entries (errors must still be visible)
- MPG rule: Only English



Mende, Michael (2021): Notebook Origins – why document? Talk presented at Digital Workshop "Living with Electronic Laboratory Notebooks", Munich, 21st September 2021, http://hdl.handle.net/21.11116/0000-0009-3F16-9.

Alexander Graham Bell, pp. 40-41 of Alexander Graham Bell Family Papers in the Library of Congress' Manuscript Division, public domain.

TRANSFORMATION FROM ANALOG TO DIGITAL

- normally no delete-function in an ELN
- tracking changes etc. to one user
- only one account per user

ARE ELNS IN USE AT ILIA?

ELECTRONIC LABORATORY NOTEBOOKS

Lists of ELN Software Systems Overviews

- Harvard Medical School (https://datamanagement.hms.harvard.edu/analyze/electronic-lab-notebooks)
- Gurdon Institute at the University of Cambridge (https://www.gurdon.cam.ac.uk/institute-life/computing/elnguidance)
- Wikipedia (https://en.wikipedia.org/wiki/List_of_electronic_laboratory_notebook_software_packages)

ELECTRONIC LABORATORY NOTEBOOKS

JoplinApp

- Website: https://joplinapp.org
- Demo Video: https://www.youtube.com/watch?v=VAAA6uNPxec
- Software Type: Open Source

ELECTRONIC LABORATORY NOTEBOOKS

Labfolder

- Website: https://labfolder.com
- Demo: https://www.labfolder.com/free-version/
- Software Type: Proprietary

ELECTRONIC LABORATORY NOTEBOOKS

eLabFTW

- Website: https://www.elabftw.net
- Demo: https://demo.elabftw.net
- Software Type: Open Source

ELN SURVEY 2021 WITHIN THE MPG

Seven key findings:

- A smooth integration of basic data analysis and laboratory work must be possible with an ELN system.
- A Standard Operating Procedure for ELN handling is helpful for the concrete application at the institutes.
- The compatibility to and with other software solutions is a central component of scientific work with an ELN.
- Aspects of open data and the FAIR principles should be considered from the beginning of the ELN usage.
- The need for staff and IT support for local ELN usage should not be underestimated.
- There is a need within the Max Planck Society for an ELN community and an exchange on ELN topics.
- There are still some legal uncertainties in the operation of ELN systems.

https://rdm.mpdl.mpg.de/2021/08/02/survey-results-on-the-usage-of-elns-in-the-max-planck-society-2021/

BREAK (7 MINUTES)

PART 5

Open Science

STRUCTURE OF PART 5

1. Open Science
2. Open Research Data
3. Linked Open Data
4. Citizen Science
5. Pre-Registration
6. Open Peer Review
7. Open Educational Resources
8. Open Access

OPEN SCIENCE

ASPECTS OF OPEN SCIENCE

Aspects of Open Science

- **Open Access** Public access to research results in the form of publications
- **Open Research Data**: Free availability of research data/raw data
- **Linked Open Data*** Accessibility and networking of public data repositories
- **-Open Review**: Search for alternatives to the traditional review processes of journal publishers to ensure greater transparency in this area

OPEN SCIENCE

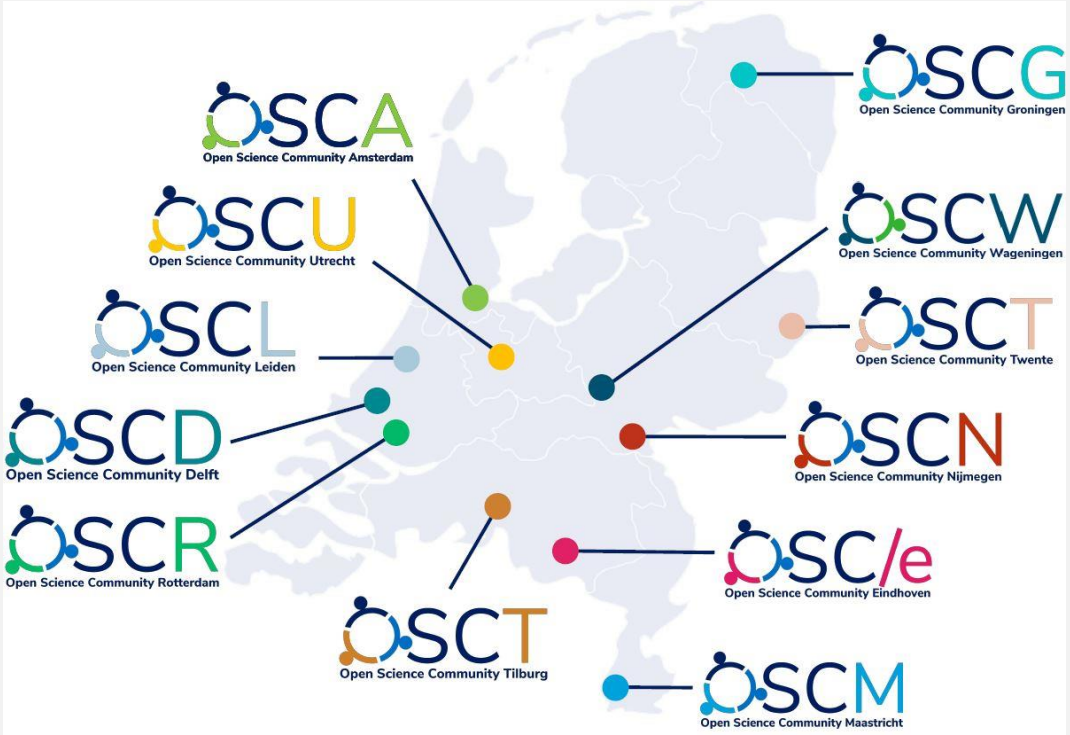
Aspects of Open Science

- **Open Source**: Development and use of open source software for science
- **Citizen Science**: Involving non-professional scientists in the scientific process
- **Open Education**: Free access to digital teaching material and (recordings of) courses
- **Pre-registration**: Deposit and peer review of a project and method plan including hypotheses to be tested on a suitable platform

OPEN SCIENCE COMMUNITIES

Example The Netherlands

- https://www.osc-nl.com
- https://www.openscience.nl



https://www.osc-nl.com, CC BY 4.0

DISCUSSION: OPEN SCIENCE ADVANTAGE OR DIS-ADVANTAGE FOR GEORGIA?

Sesame street effect (i.e. https://doi.org/10.1207/s1532785xmep0102_5)

OPEN RESEARCH DATA

OPEN RESEARCH DATA

4 Steps to Open Data by the European Commission:

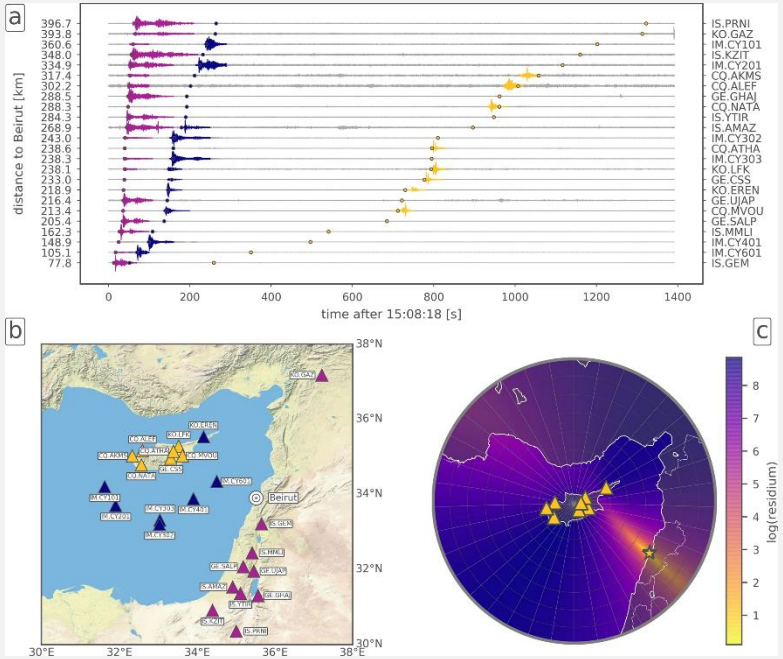
1. Prepare your data for sharing
2. Select a repository
3. Add a Data Availability Statement to your article
4. Link your datasets to your article

https://think.f1000research.com/open-research-europe-submission/opendataguide/?utm_source=CPB&utm_medium=cms&utm_campaign=JQC19229

OPEN RESEARCH DATA

Example: Explosion in the harbor of Beirut

Pilger, C., Gaebler, P., Hupe, P. et al. (2021): Yield estimation of the 2020 Beirut explosion using open access waveform and remote sensing data, Scientific Reports 11, 14144, https://doi.org/10.1038/s41598-021-93690-y.



Pilger, C., Gaebler, P., Hupe, P. et al., 2021, CC BY 4.0, https://www.nature.com/articles/s41598-021-93690-y/figures/1

OPEN DATA

Open Data Further Reading

- Open Data Handbook: https://opendatahandbook.org

OPEN DATA: EXAMPLES

OECD Data

- https://data.oecd.org
- https://www.oecd.org/countries/georgia/

OPEN DATA: EXAMPLES

Bibliothèque nationale de France

- https://data.bnf.fr/en/

OPEN DATA: EXAMPLES

OCHA Centre for Humanitarian Data

- https://centre.humdata.org

LINKED OPEN DATA

A SPECIAL TYPE OP OPEN DATA: LINKED OPEN DATA (LOD)

- major advantage of processing open research data is that it is freely accessible
- this makes data readily available and easier to re-use for other researchers
- overarching goal of LOD is to weave a so-called network of knowledge, whereby structures, connections and contexts become visible and machine-readable

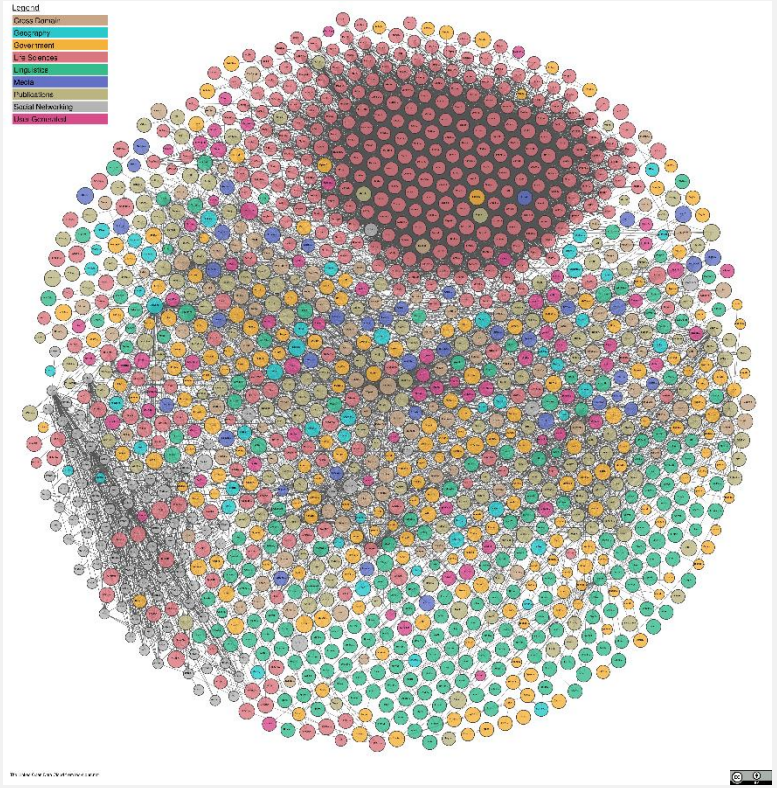
LINKED OPEN DATA

- LOD puts an emphasis on the structure of the data using triples
- each statement is divided into three elements: **subject**, **predicate** and **object**
- the description is especially based on RDF (Resource Description Framework, https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/) to make the data not only accessible on the internet but also linkable to other scientific representations
- Example: https://www.nobelprize.org/about/linked-data-examples/

Jannis Charalabidis and others, 'The Multiple Life Cycles of Open Data Creation and Use', in The World of Open Data: Concepts, Methods, Tools and Experiences, ed. by Yannis Charalabidis and others, Public Administration and Information Technology (Cham: Springer International Publishing, 2018), p. 14. https://doi.org/10.1007/978-3-319-90850-2_2.

LINKED OPEN DATA

- LOD cloud website (https://www.lod-cloud.net) shows datasets, which have been published as linked open data
- gives a good impression of the diversity of linked open data



www.lod-cloud.net, C BY 4.0

LOD EXAMPLES

- Wikidata: https://www.wikidata.org
- VIAF: https://viaf.org
- GND: https://gnd.network/Webs/gnd/DE/Home/home_node.html

CITIZEN SCIENCE

CITIZEN SCIENCE

- involvement of non-scientists in research
- i.e. counting thinks, measuring something, making photos, recording voices

Aims:

- support of research
- better communication of science, especially the way of scientific findings

EXAMPLE

- Penguin Watch: https://www.zooniverse.org/projects/courtaulddigital/world-architecture-unlocked/classify
- Europeana: https://europeana.transcribathon.eu
- NASA Citizen Science, https://science.nasa.gov/citizenscience

- general EU platform: https://eu-citizen.science

PRE-REGISTRATION

REPRODUCIBILITY CRISES

- "reproducibility crisis" or "replication crisis"
- around 2015/16 in medical, life and behavioural sciences

- five aspects of the crises:
  1. absence of replication studies
  2. widespread failure to reproduce results of published studies
  3. evidence of publication bias
  4. questionable research practices
  5. lack of transparency and completeness in the reporting of methods, data and analysis in scientific publication

Fidler, Fiona and John Wilcox, "Reproducibility of Scientific Results", The Stanford Encyclopedia of Philosophy, 2021, https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/.

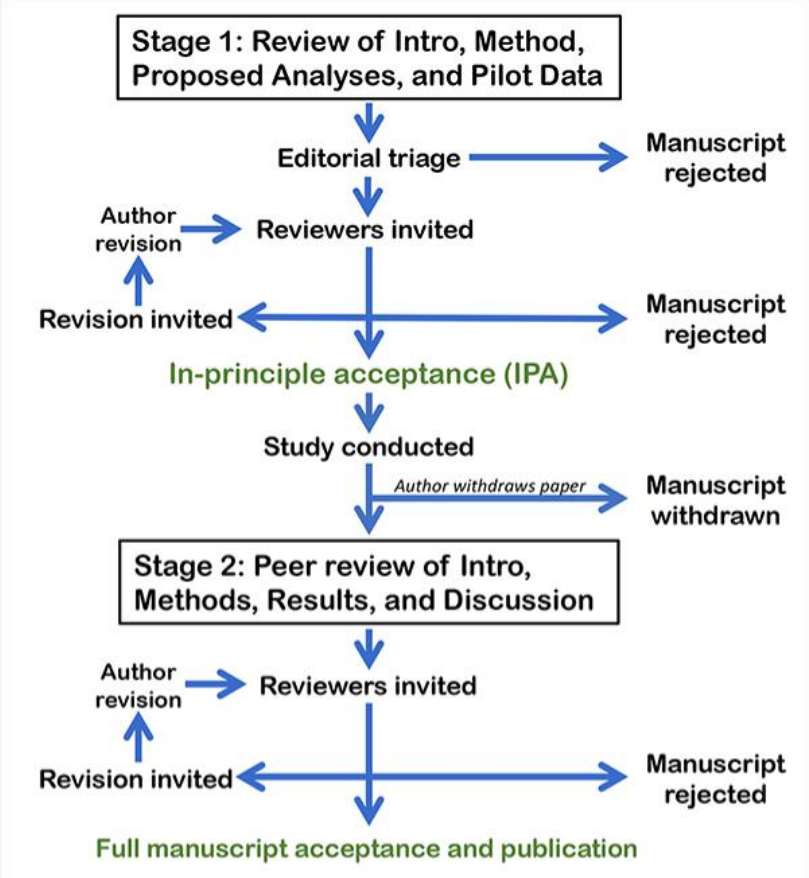
PRE-REGISTRATION

- registering the hypotheses, methods and analyses of a research project, **before** it is started
- aim is to reduce problematic research practices (i.e. p-harcking/data fishing)
- already supported by some journals (https://www.cos.io/initiatives/registered-reports#journals)

VERY SCHEMATIC "CLASSICAL" PUBLICATION PROCESS

PRE-REGISTRATION WORKFLOW

PRE-REGISTRATION WORKFLOW



Center for Open Science, https://www.cos.io/initiatives/registered-reports#journals, CC BY 4.0

EXAMPLE: PCI REGISTERED REPORTS

– https://rr.peercommunityin.org

Example:

Corina Logan and Chris Chambers (2021): Does incorporating open research practices into the undergraduate curriculum decrease questionable research practices?. Peer Community in Registered Reports, 100002, https://doi.org/10.24072/pci.rr.100002.

PCI Registered Reports, https://rr.peercommunityin.org, CC BY 4.0

EVENT HINT ON PRE-REGISTRATION

Workshop on Pre-Registration: Why, How, and Where? by Open Science Community Maastricht on 1st December 2021

QUESTION FROM SESSION 1:

WHAT ARE THE ADVANTAGES FOR QUALITATIVE DATA SHARING IN SOCIAL SCIENCES?

ADVANTAGES FOR QUALITATIVE DATA SHARING IN SOCIAL SCIENCES

"Standard" answer:

- ensure reproducibility
- increase the visibility of own research
- necessary due to funding

ADVANTAGES FOR QUALITATIVE DATA SHARING IN SOCIAL SCIENCES

Specific answers by **Tarrant** and **Hughes**:

- Testing new methodological techniques with existing data
- Familiarisation existing datasets, bringing them together and theoretically sampling from them
- Developing new empirically-driven research questions
- Sustaining and extending existing study samples

+ teaching: Students can learn and develop new questions by re-using data

ADVANTAGES FOR QUALITATIVE DATA SHARING IN SOCIAL SCIENCES

- Anna Tarrant and Kahryn Hughes: The re-use of qualitative data is an under-appreciated field for innovation and the creation of new knowledge in the social sciences, 8th June 2020, https://blogs.lse.ac.uk/impactofsocialsciences/2020/06/08/the-re-use-of-qualitative-data-is-an-under-appreciated-field-for-innovation-and-the-creation-of-new-knowledge-in-the-social-sciences.
- Bishop, Libby, and Arja Kuula-Luumi. 'Revisiting Qualitative Data Reuse: A Decade On'. SAGE Open, vol. 7, no. 1, Jan. 2017, p. 215824401668513, https://doi.org/10.1177/2158244016685136.

OPEN PEER REVIEW

OPEN PEER REVIEW

no common standard, i.e.:

- the colleagues know the articles in advance
- reviews will be published with the article
- there are preprints
- there is a comment function on the article
- …

REPROHACK

- Call for hacking the own paper
- Aim: Improve reproducibility
- https://www.reprohack.org

OPEN EDUCATIONAL RESOURCES

OPEN EDUCATIONAL RESOURCES (OER)

- training and teaching material, which is under public domain or an open license
- UNESCO Recommendations: https://en.unesco.org/themes/building-knowledge-societies/oer
- https://open-science-training-handbook.gitbook.io/book/open-science-basics/open-educational-resources

Jonathasmello, OER logo, CC BY 3.0

ONLINE TUTORIAL ON RDM

- https://datacarpentry.org/lessons/
- https://zenodo.org/communities/dcc-rdm-training-materials?page=1&size=20
- https://mantra.ed.ac.uk/
- https://datasupport.researchdata.nl/en/start-the-course
- https://www.coursera.org/learn/data-management

- Biernacka, Katarzyna, et al.: 'Adaptable Methods for Training in Research Data Management'. Data Science Journal, vol. 20, no. 1, 1, April 2021, p. 14. datascience.codata.org, https://doi.org/10.5334/dsj-2021-014.

ONLINE TUTORIALS ON OPEN SCIENCE

- https://www.oclc.org/research/events/oclc-liber-open-science-series.html
- https://www.openaire.eu/frontpage/webinars
- https://opensciencemooc.eu
- http://catalogue.openaire.eu
- https://www.fosteropenscience.eu/resources
- European Space Agency: https://www.youtube.com/channel/UCPnL3aynCQxTOjPttxMiS3Q

OPEN ACCESS

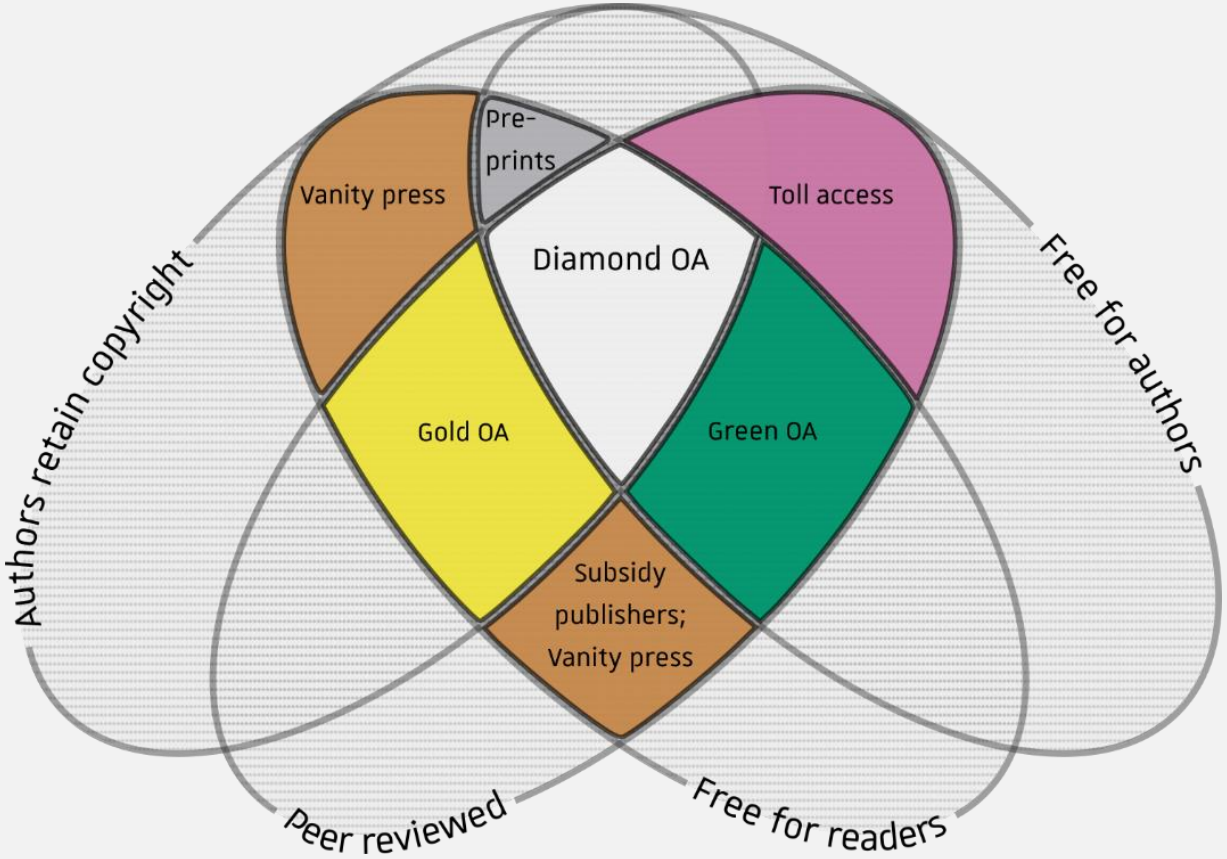
GENERAL ABOUT OPEN ACCESS

- Open access to scientific literature
- main focus on digital publications
- includes also metadata of publications

OPEN ACCESS VARIANTS

- Gold Model: Content is immediately available
- Green Model: Self-archiving by authors is possible (i.e. institutional repository)
- Diamond Model: Upgrade of gold model, no cost for authors either
- Black Model: Unauthorized copying

OPEN ACCESS VARIANTS



Farquharson, Jamie (2018): Diamond open access venn, https://doi.org/10.6084/m9.figshare.6900566.v1

OPEN ACCESS TRANSFORMATION

- OA2020 (https://oa2020.org)
- Aim to transform funding streams for scientific quality journals
- payment will no longer be made for access to articles, but for their publication in open access
- can be achieved through contracts between libraries and publishers
- allows all researchers at an institution participating in the contract to publish open access

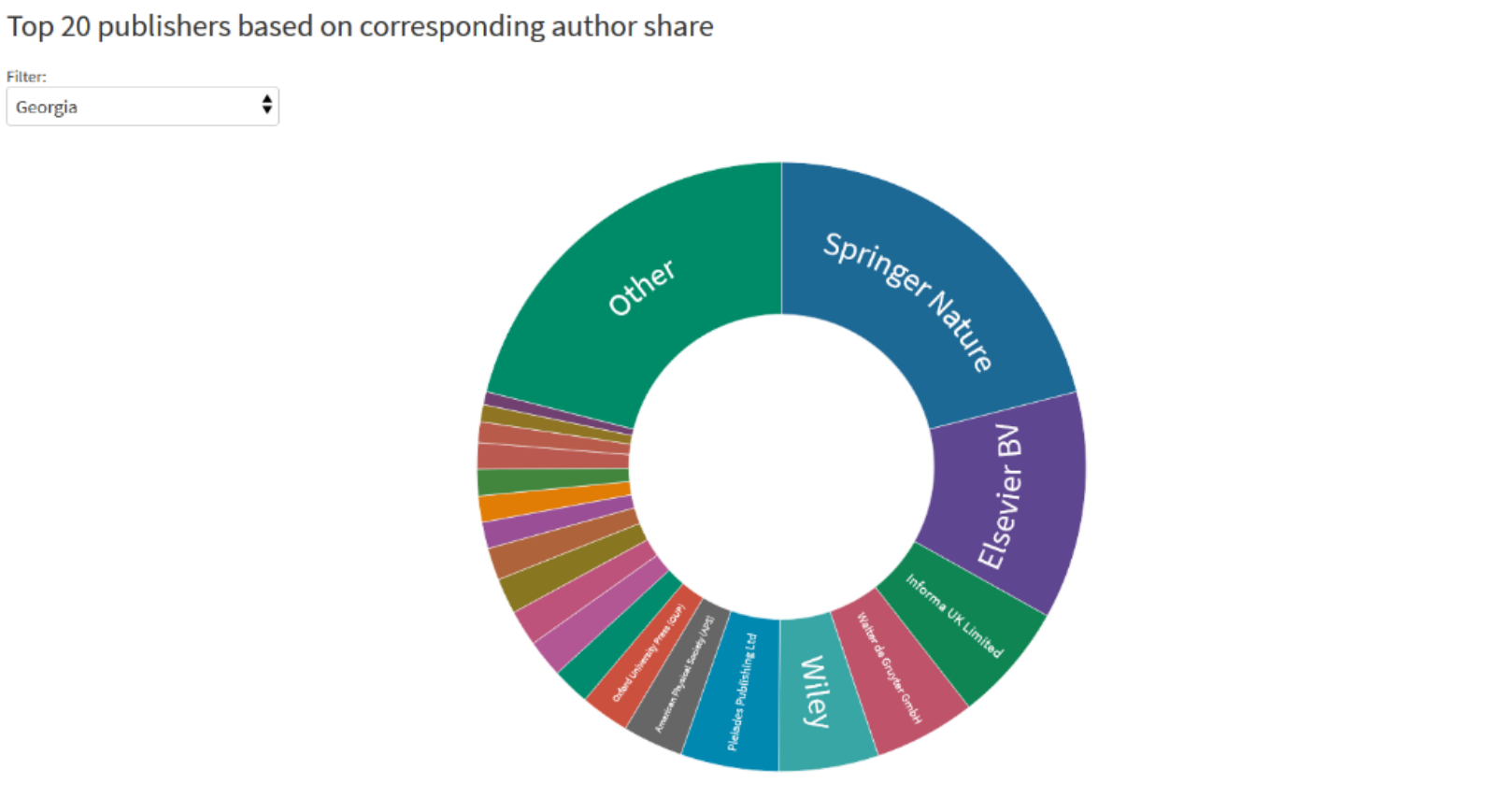
- *Author pays for publishing* or *Not pay to read, but pay to say*

ESAC INITIATIVE

- Efficiency and Standards for Article Charges Initative
- Data about the Open Access Market
- https://esac-initiative.org/market-watch/

- Open Access Transformative Agreement in Georgia? Please uncover a publishing profile: https://esac-initiative.org/about/data-analytics/publishing-profile/
- Questions or problems: contact@esac-initiative.org

ESAC INITIATIVE



Top 20 publishers based on corresponding author share: Georgia

DISCUSSION: OPEN ACCESS IN GEORGIAN SCIENTIFIC INSTITUTIONS

დიდი მადლობა!