Open Research Data – Bright New Future or Just a Flash in the Pan?

Workshop in the context of the Frankfurt Open Science Initiative

1st July 2024, Goethe University
https://doi.org/10.21248/gups.85527

Dr. Yves Vincent Grossmann
University Library Frankfurt
y.grossmann@ub.uni-frankfurt.de

Speaker

- Yves Vincent Grossmann
- Team Lead since March 2024 for
  1. Research Data Management and
  2. Library of Natural Sciences at the Goethe University
- 2020 to 2024 Research Data Management Officer at the Max Planck Digital Library
- October 2022 RDM Consultant with Goethe Institute fellowship at Ilia state University Tbilisi, Georgia
- (Art) Historian by training
- y.grossmann@ub.uni-frankfurt.de
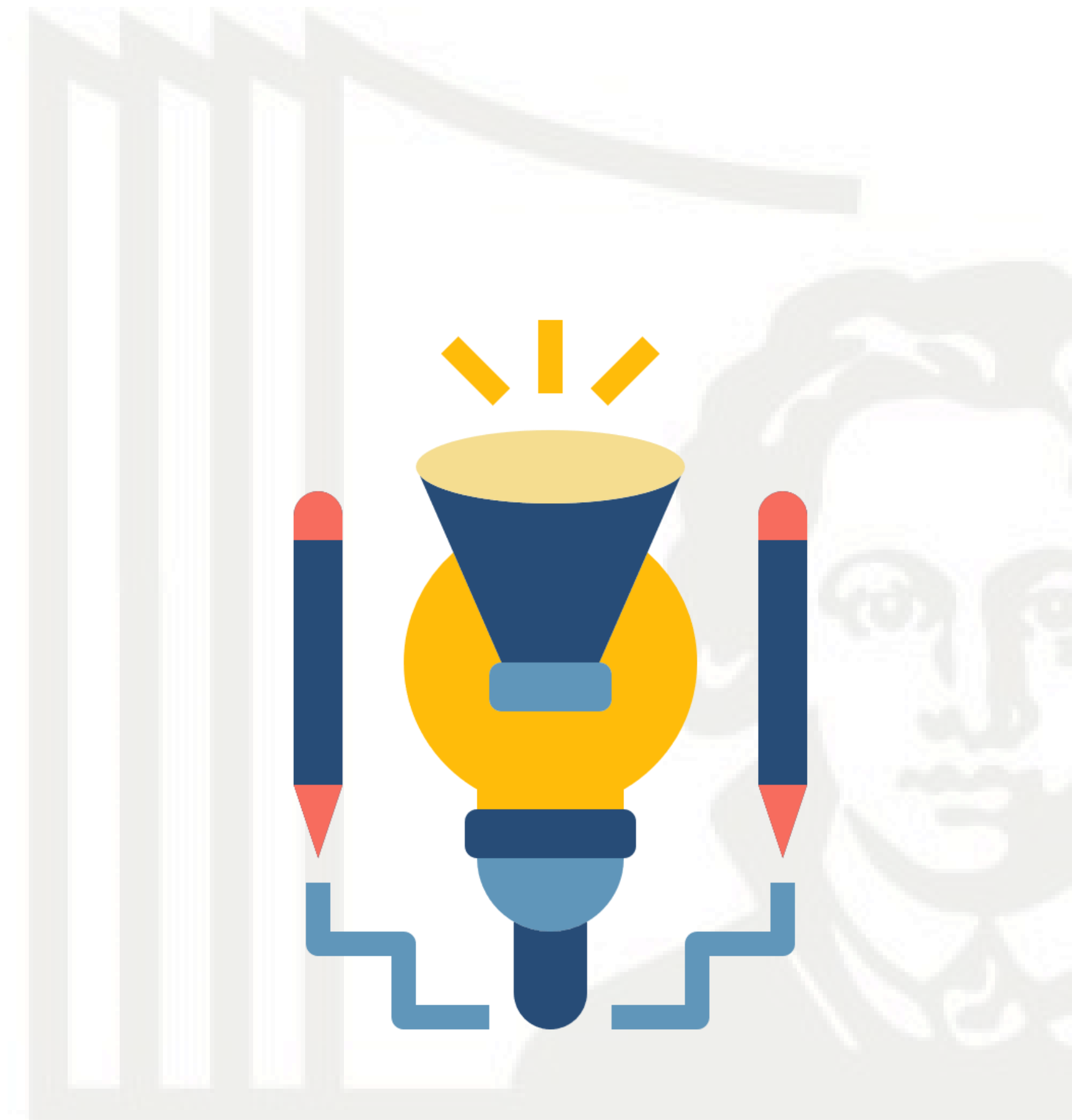
Objectives

- Understanding Open (Research) Data
- Recognizing that Open Research Data is not rocket science (except in rocket science)
- Discuss challenges connected to Open Research Data

Slido

1. Have you published research data openly?

2. Have you re-used research data (either your own or from others)?

Join at slido.com
Code: 3373 209

UB

UNIVERSITÄTS
BIBLIOTHEK
FRANKFURT AM MAIN

Open Science

UNESCO Recommendation on Open Science

- Adopted in November 2021

- https://www.unesco.org/en/open-science

- Increases scientific collaborations and sharing of information for the benefit of science and society

- Makes multilingual scientific knowledge openly available, accessible and reusable for everyone

- Opens the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community

UNESCO, https://unesdoc.unesco.org/ark:/48223/pf0 000379949.locale=en, CC BY-SA 3.0.

Hypothesis: Open Science is a Reaction

Reaction to high publication costs	Open Access
Duration of the publication process and lack of transparency	Open Peer Review
Digitisation and interaction with non-scientists	Citizen Science
Reproducibility and possibility of digitisation	Open Data
High costs for educational materials	Open Education Resources

|	
Lack of space for a detailed discussion of the methods	Open Methodology
No cross-linking of data	Linked Open Data
Crisis of reproducibility	Pre-Registration
Non-scientific experiences; reuse options	Open Source

Changing "World View"

- Similar to the Copernican revolution (shifting from the geocentric to the heliocentric world view)

- Transferred to Open Science:

  - Shift from an **article-centered view** (publication at the end of research)

  - to a **process-orientated view** (transparency and reproducibility of all steps, from the beginning with preregistration to the final publication)
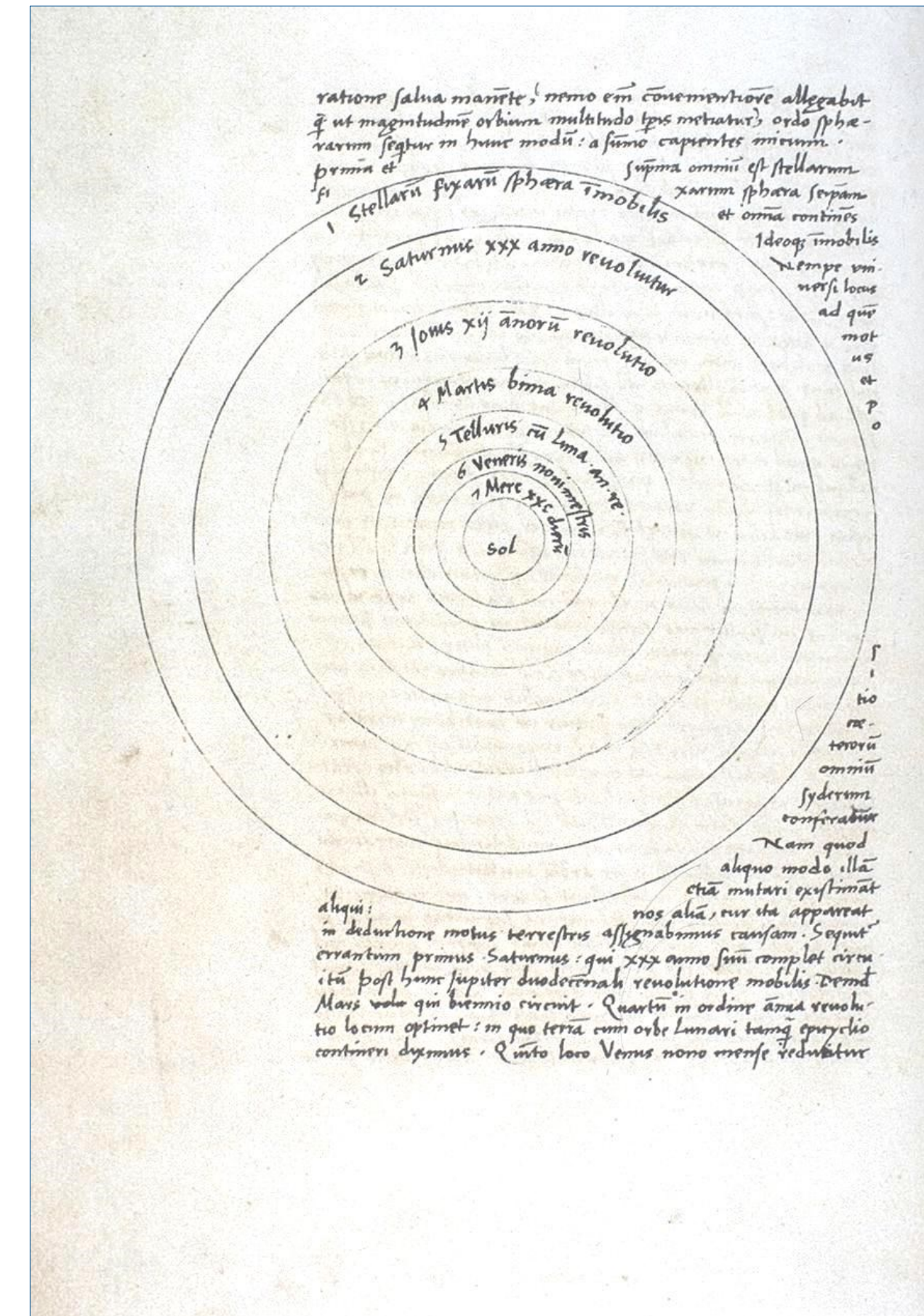
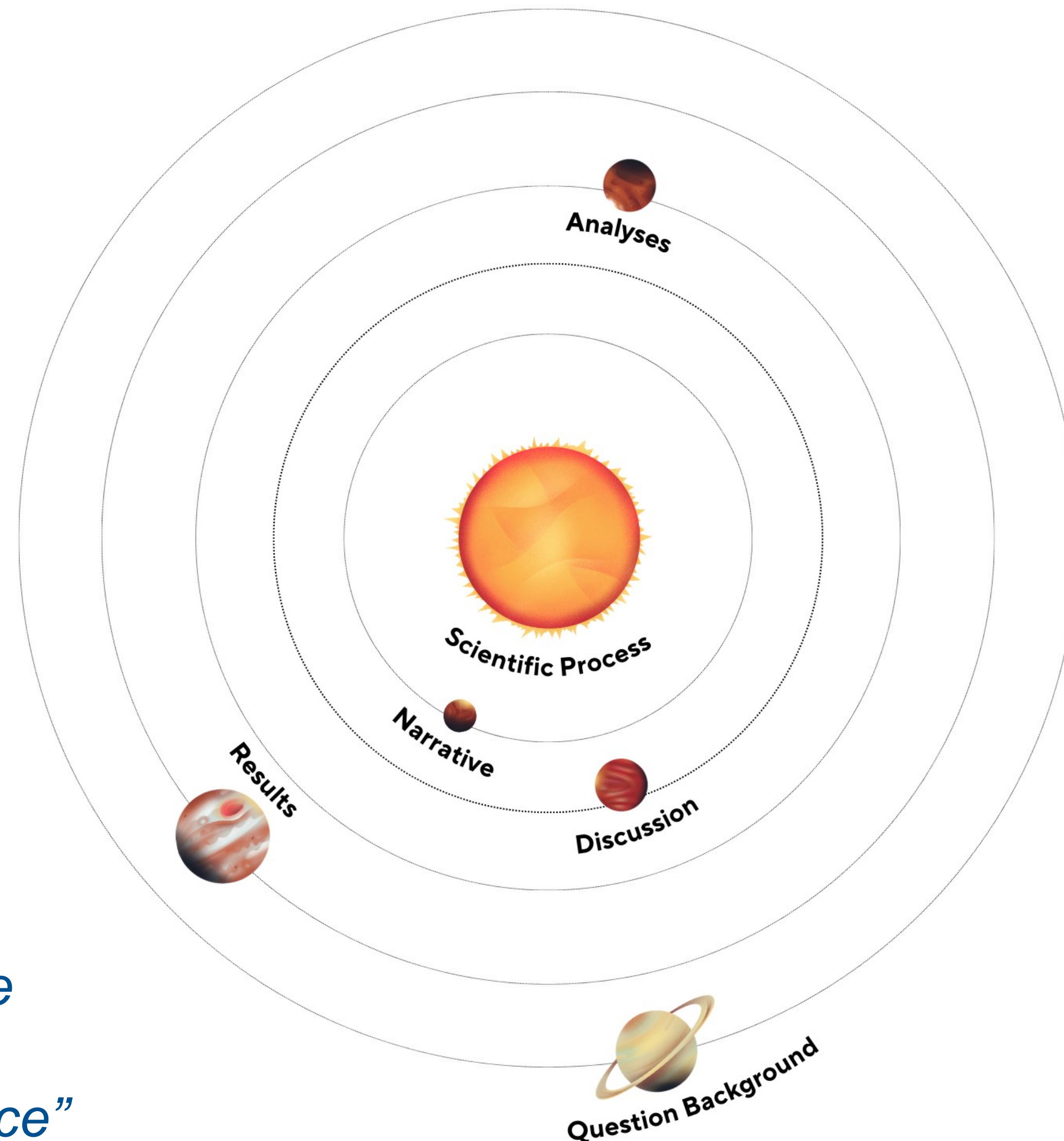Nicolaus Copernicus (1473–1543), "De revolutionibus", fol. 9 verso, Jagiellonische Bibliothek, Krakau, public domain.

Comparison 1: Geocentric Worldview
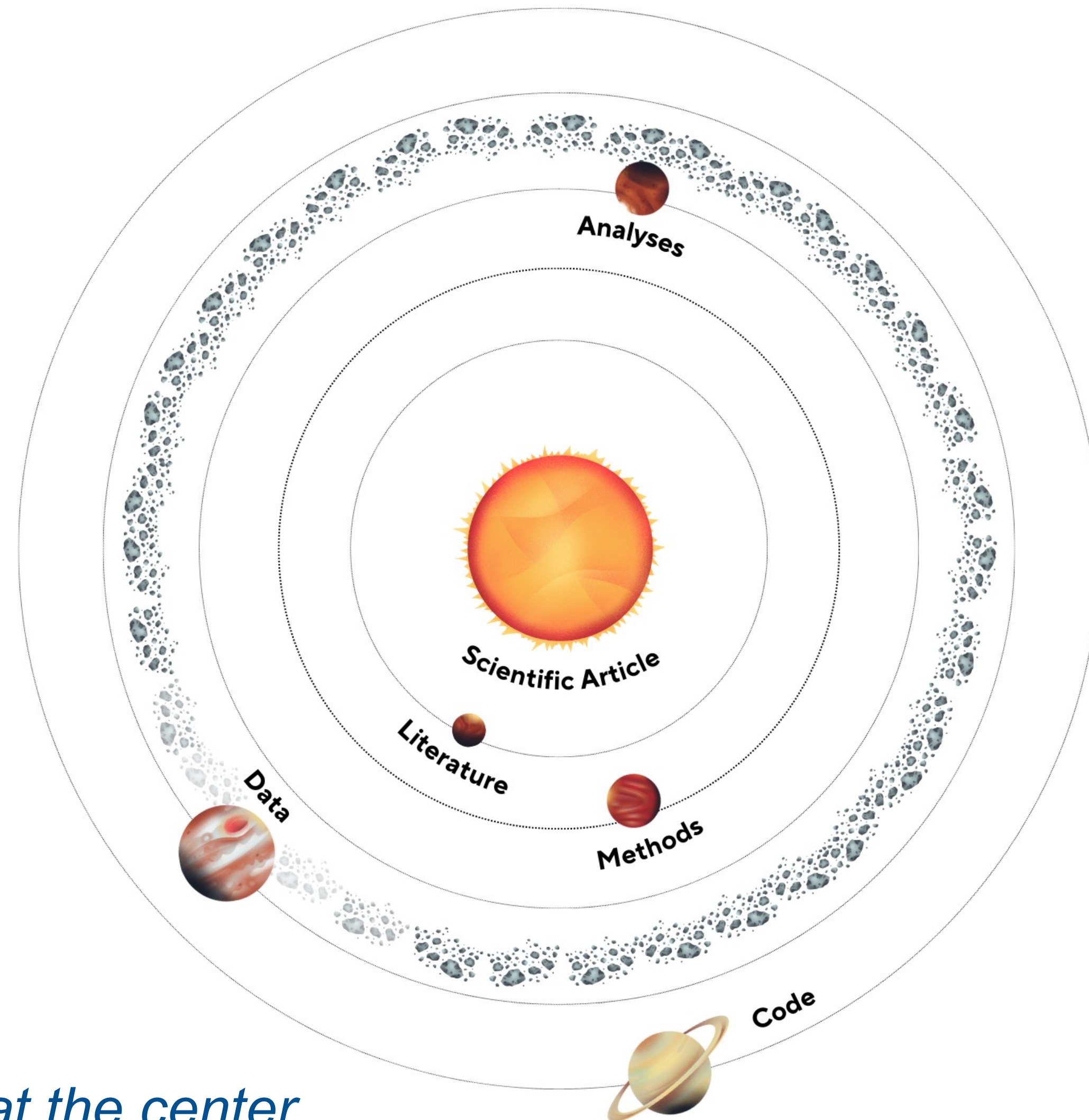
Scientific Article is at the center

Heliocentric Worldview

Scientific process moved to the centre → transition maybe called "Open Science"
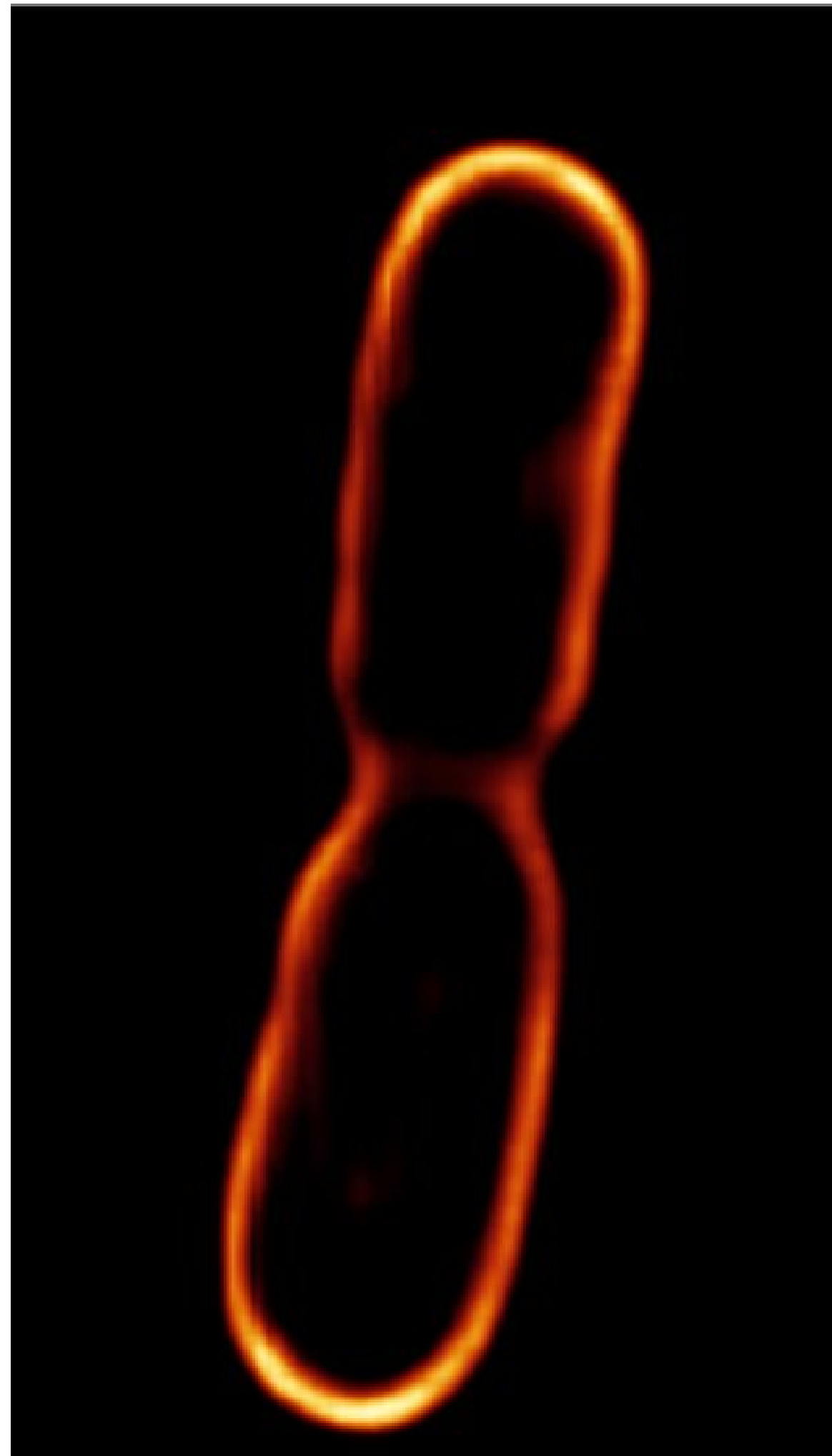
UB

UNIVERSITÄTS
BIBLIOTHEK
FRANKFURT AM MAIN

Short Data Quiz

- Escherichia coli membrane
- Spahn, C., & Heilemann, M. (2021). DeepBacs – Artificial labeling of E. coli membranes dataset and fnet/CARE models [Data set]. Zenodo. https://doi.org/10.5281/zenodo.5551123
- From the Institute of Physical and Theoretical Chemistry, GU Department 14

Short Data Quiz

- File "MDH-KF00016 _DSC_5329" from

- Döpper et al. (2023) in "MDH_SmallFindPhotos", GUDe, CC BY 4.0, https://doi.org/10.25716/gude.07ct-v20q.

- The object is an arrowhead from Oman discovered during archaeological excavations.

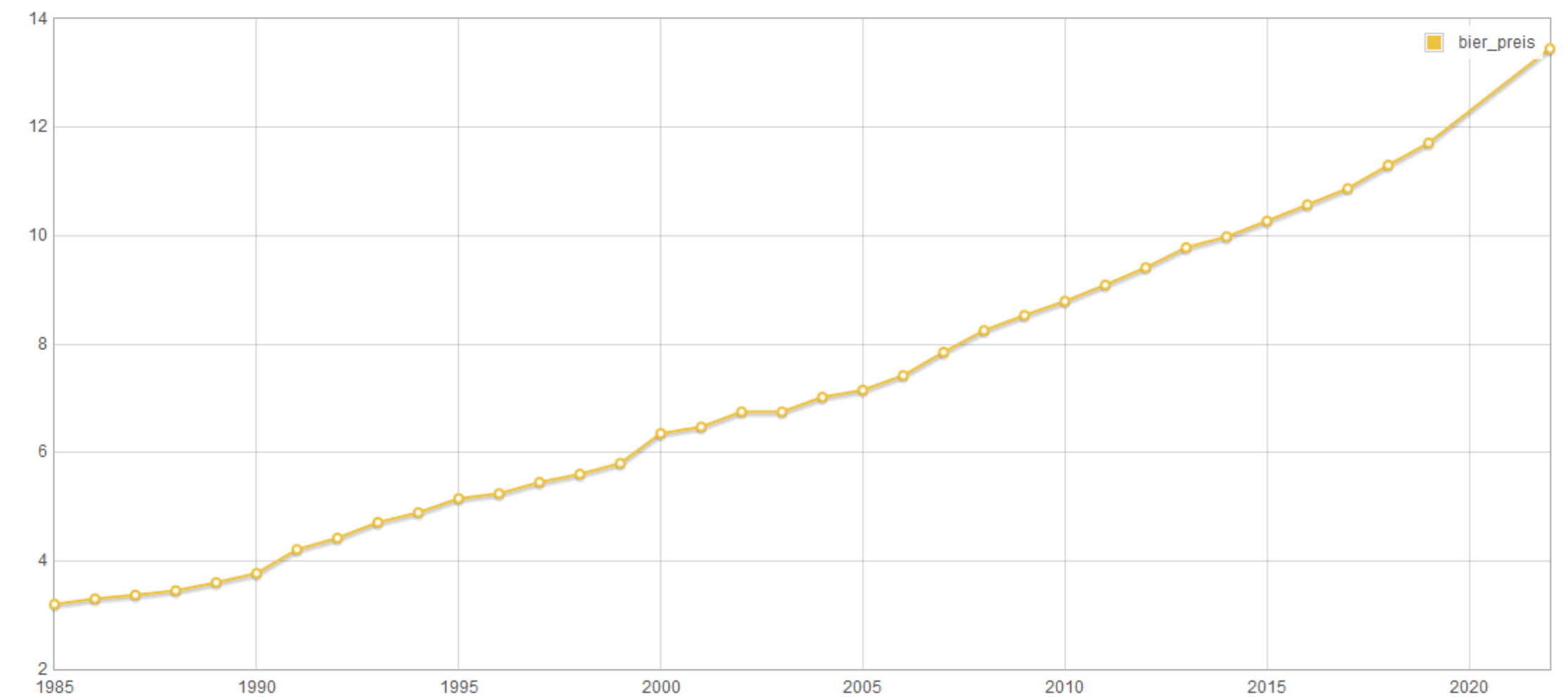- From the Institute for Archaeological Sciences, Department 09, GU

Short Data Quiz
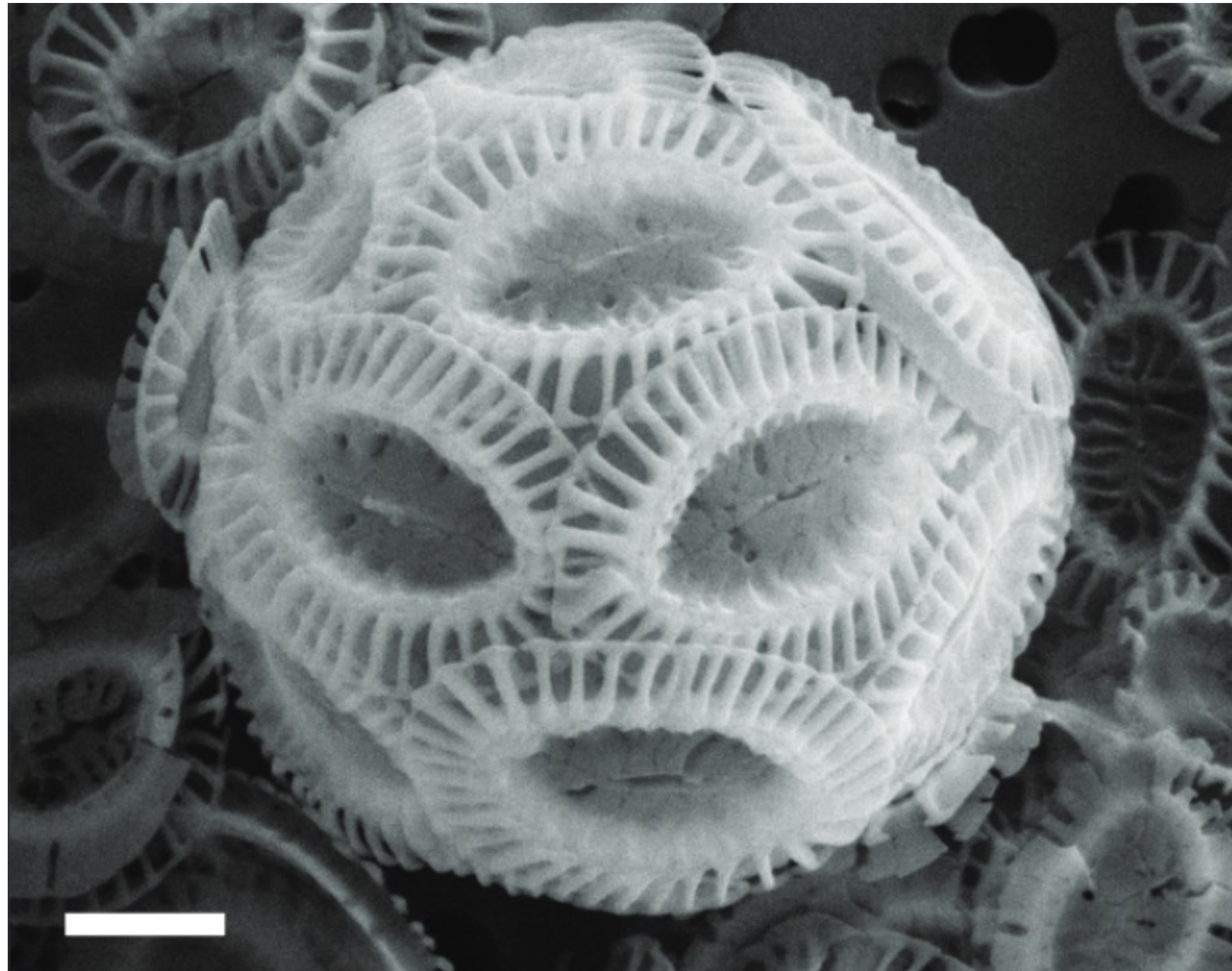
- Gebühr, C., Sheward, R., & Herrle, J. O. (2024). Growth and morphological measurements of the coccolithophore Emiliania huxleyi under three salinity conditions [Data set]. Zenodo. https://doi.org/10.5281/zenodo.11186906, CC BY 4.0.

- Image detail of an alga (coccolithophore Emiliania huxleyi)

- From the working group "Micropaleontology & Paleoceanography Group" in Department 11, GU

Short Data Quiz

_id	jahr	dauer	besuch...	besuch...	bier_pr...	bier_ko...	hendl_p...	hendl_k...
1	1985	16	7.1	444	3.2	54541	4.77	629520
2	1986	16	6.7	419	3.3	53807	3.92	698137
3	1987	16	6.5	406	3.37	51842	3.98	732859
4	1988	16	5.7	356	3.45	50951	4.19	720139
5	1989	16	6.2	388	3.6	51241	4.22	775674
6	1990	16	6.7	419	3.77	54300	4.47	750947
7	1991	16	6.4	400	4.21	54686	4.81	807710
8	1992	16	5.9	369	4.42	48888	5.11	725612
9	1993	16	6.5	406	4.71	51933	5.25	733517
10	1994	16	6.6	413	4.89	52108	5.39	663135
11	1995	16	6.7	419	5.15	50162	5.54	628890
12	1996	16	6.9	431	5.24	52622	5.45	595580
13	1997	16	6.4	400	5.45	55891	5.38	699562
14	1998	16	6.5	406	5.6	54782	5.34	603746
15	1999	16	6.5	406	5.8	60054	5.38	609661
16	2000	18	6.9	383	6.35	64591	7.85	681242
17	2001	16	5.5	344	6.47	48698	8.12	351705
18	2002	16	5.9	369	6.75	57614	8.1	459259
19	2003	16	6.3	394	6.75	62294	8.14	487487
20	2004	16	5.9	369	7.02	61163	8.28	481649
21	2005	17	6.1	359	7.15	61772	8.35	479610
22	2006	18	6.5	361	7.42	68617	8.5	494135
23	2007	16	6.2	388	7.85	69406	8.87	521872

Beer prices at the Oktoberfest 1985-2022, Datenlizenz Deutschland Namensnennung 2.0, https://opendata.muenchen.de/dataset/8d6c8251-7956-4f92-8c96-f79106aab828/resource/e0f664cf-6dd9-4743-bd2b-81a8b18bd1d2/download/oktoberfestgesamt19852022.csv.

Some Definition Attempts

- Open Data = Open Governmental Data
- Open Data ↔ Open Research Data
- Open Research Data ↔ Research Data with Restricted Access
- Open Research Data ↔ FAIR Research Data
- Open Research Data ↔ Linked Open Data

Open (Governmental) Data

EU Directive: "*Open data as a concept is generally understood to denote data in an open format that can be freely used, re-used and shared by anyone for any purpose.*"

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, http://data.europa.eu/eli/dir/2019/1024/oj.

Examples for Open Governmental Data:

- Open Data Portal Offene Daten Frankfurt
- Kompetenzzentrum für Open Data des Bundes
- Portal for European data
- …

1. Changing "Views" with Open Science

2. Looking Closer at Research Data

3. Why Sharing Data Open?

4. Start Sharing Data Open

5. Struggle and Repeat

Reasons for Open Research Data

- Intrinsic motivation
- Best practices within the research community
- Instructions from supervisors
- Requirements from publishers, funders, and stakeholders
- Open Research Data policies
- …

Horizon Europe

"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]."

Volkswagen Foundation

- Open Science Policy since 2021
  - "The Volkswagen Foundation encourages researchers to store the data generated within funded projects in public, non-commercial repositories [..]" (p. 4)
  - "In the application process, data generating and data using projects working in disciplines lacking a clear data workflow are asked to upload a digital concept with a data management plan". (pp. 4-5)
  - "In the application process, the applicants are asked to mark separately the already generated Open Data in the CV for the review and decision process." (p. 5)
- (New) funding opportunity: "Data Reuse – Additional Funding for the Preparation and Storage of Research Data (Open Science)"
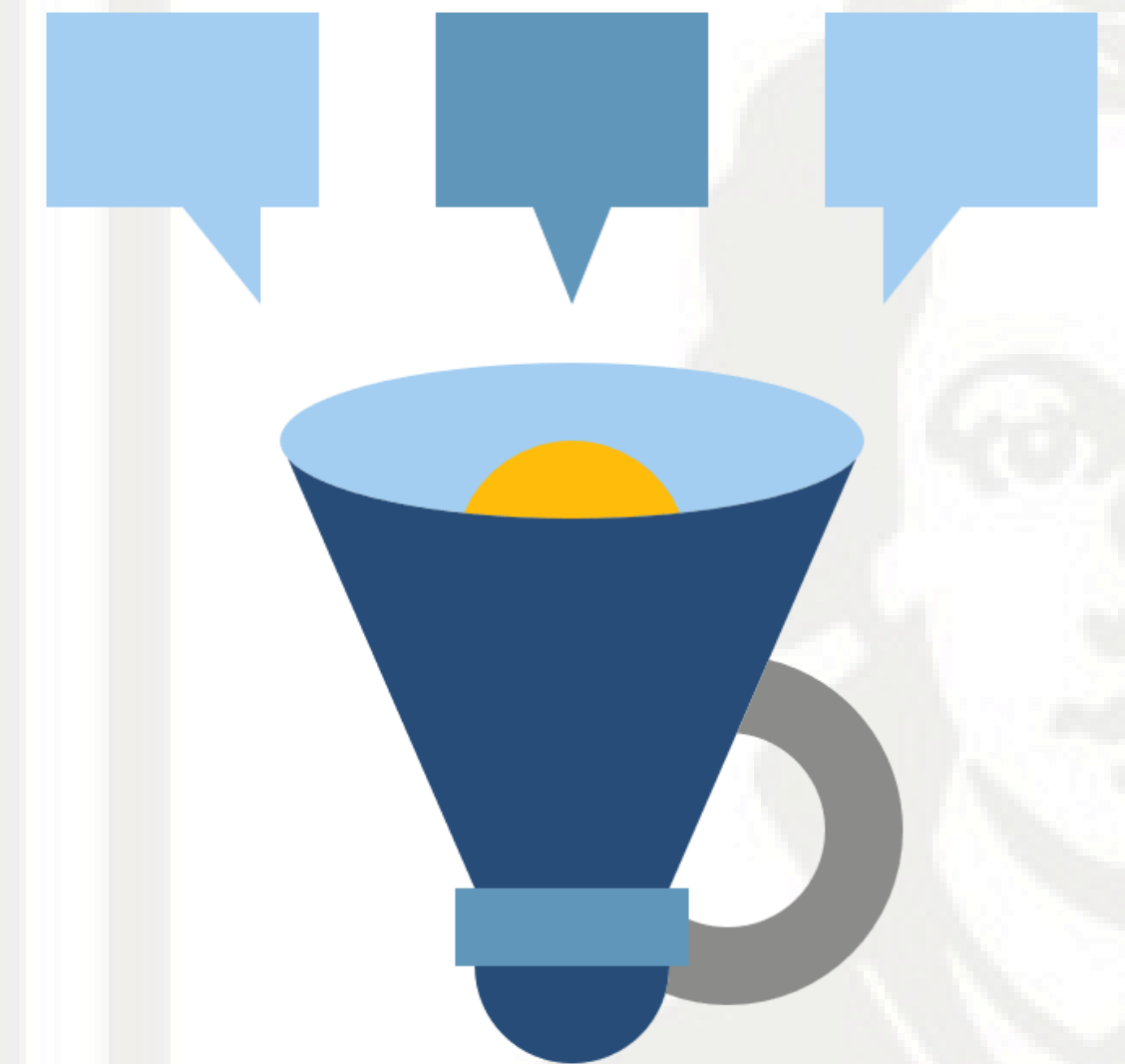
Deutsche Forschungsgemeinschaft (DFG)

"*Making research data available, developing methods and standards and building data infrastructures are important contributions to the re-use of research findings as well as integrated part of the good research practice and* **should be listed** *as part of a researcher's preliminary work or academic profile.*"

DFG Handling of Research Data, www.dfg.de/proposal_process/research_data.

"*Aspects relevant to ensuring that open science is advantageous to research and scholarship include* **open access to and legally protected re-use** *of publications and research data, and also if possible and reasonable research and infrastructure software (open code).*"

DFG (2022): Open Science as Part of Research Culture. Positioning of the German Research Foundation, https://doi.org/10.5281/zenodo.7194537, p. 5.

Helmholtz Association

„**Open Research Data**
All Centers will establish detailed procedures for managing research data in publicly available policies, and will regularly examine and if necessary adapt these procedures.

In 2023, a basic indicator for the presentation of citable research data publications will be established as an incentive within the framework of the PoF ["program-oriented funding"].

By 2024, a Helmholtz quality indicator for research data publications will be developed and established, which will be deployed within the framework of the PoF and will replace the aforementioned basic indicator."

HELMHOLTZ
Open Science

Helmholtz Open Science
Policy

Version 1.0

Helmholtz-Gemeinschaft (2022): Helmholtz Open Science Policy. Version 1.0, CC BY 4.0, https://doi.org/10.48440/os.helmholtz.056.

Leibniz Association

Guidelines on the Handling of Research Data within the Leibniz Association

"[T]*here has been only a slow increase in open access options for research data in recent years. The reasons often given are a lack of (research) policy requirements, the pressure of competition in academia and a lack of incentive systems*". (p. 4.)

"*The Leibniz institutions support the reusability of research data, provided there are no legal or ethical constraints, in line with the principle 'as open as possible, as closed as necessary'* ". (p. 4)

Guidelines on the Handling of Research Data within the Leibniz Association

Adopted by the General Assembly of the Leibniz Association on 29 November 2018.

I. Preamble

Research data forms the basis of scientific work. The increasing digitalisation of scientific processes calls for new approaches in the area of research data handling. Simply publishing the conclusions resulting from the analysis of collected research data is no longer sufficient. Instead, access to research data is becoming an increasingly important resource for researchers making use of the new opportunities offered by digital tools. Ensuring data access and interpretability creates a range of challenges on a number of different levels (for research funding bodies, research institutions and researchers and research support staff) and often requires a specific approach for each discipline. As a network of German research institutions in a range of different disciplines, the Leibniz Association emphasises the importance of a responsible and transparent approach to handling research data as part of a sustainable and quality-oriented research process. The proposals and measures set out below aim to promote the development of structures and processes within the Leibniz Association that meet this requirement.

II. Research data

Research data includes all data that is produced during research and that can be processed digitally. It takes a range of forms – depending in particular on the domain in question. It may include, for instance, measurement, (ongoing) survey and observational data, process-produced data, texts, data from polls, graphic visuals, software and simulations. Among other things, this research data includes raw data, aggregated data, metadata and descriptions of data structures.

Research data management is an active task and supports the activities linked to scientific discovery: the planning, generation, documentation, processing, archiving and, where relevant, publishing, of the data. Research data management is part of good scientific practice and ensures the quality, reproducibility, availability and reuse of research data and results. This kind of systematic management of research data is therefore a basis for scientific discourse and for maximising the data's reuse potential.

Leibniz Association, 2018, https://www.leibniz-gemeinschaft.de/fileadmin/user_upload/Bilder_und_Downloads/Forschung/Open_Science/Leitlinie_Forschungsdaten_2018_EN.pdf

Open Research Data Dashboards

Charité Dashboard on Responsible Research (https://quest-dashboard.charite.de)
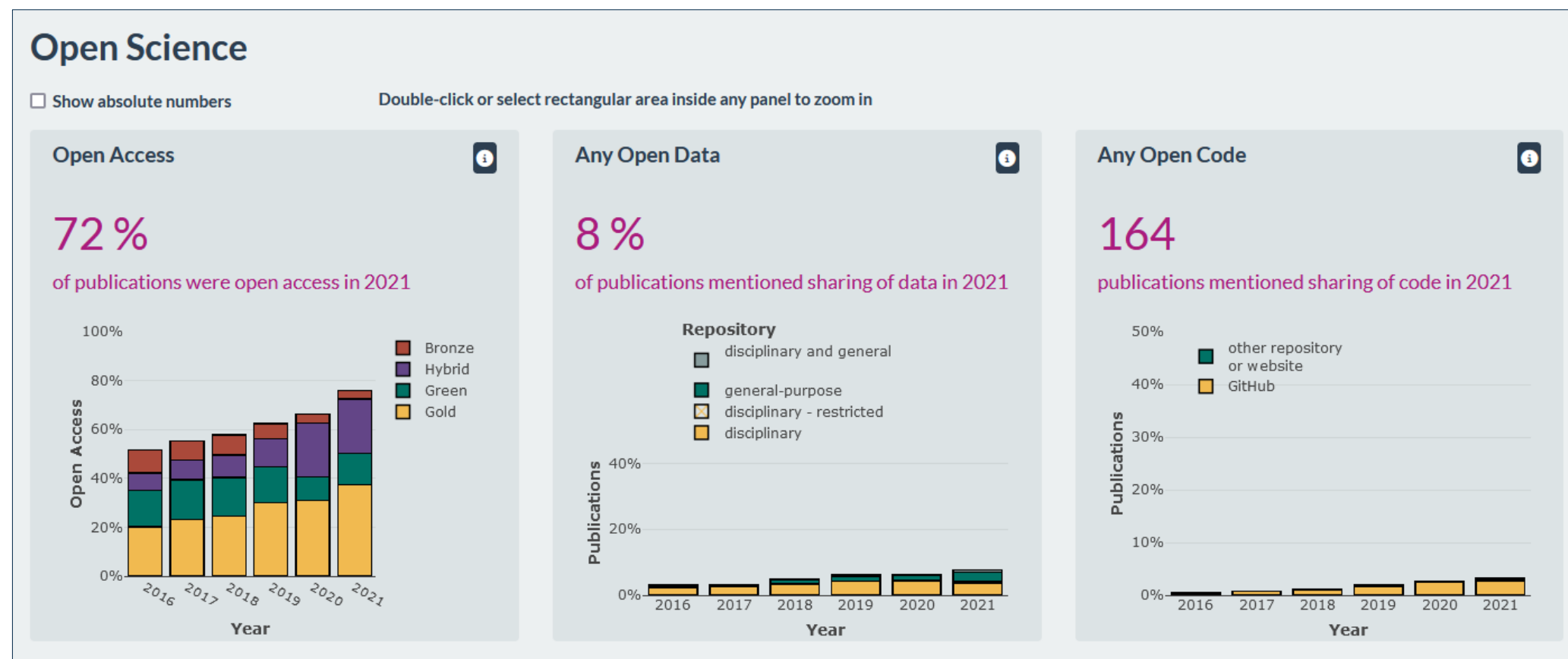
French Open Science Monitor (https://frenchopensciencemonitor.esr.gouv.fr)

OPEN LICENCE 2.0/LICENCE OUVERTE 2.0, https://github.com/etalab/licence-ouverte/blob/master/open-licence.md.
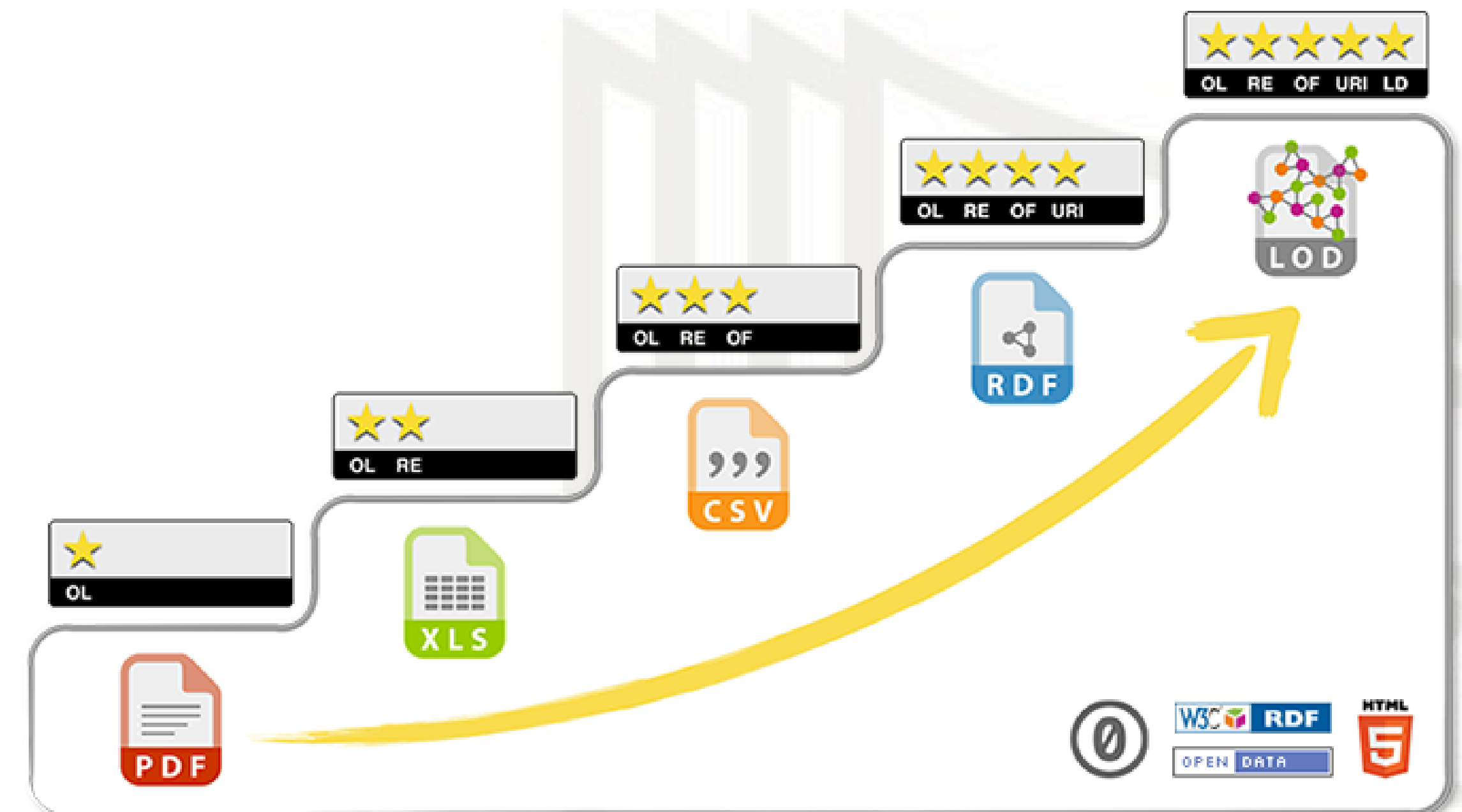
UB

UNIVERSITÄTS
BIBLIOTHEK
FRANKFURT AM MAIN

5 ★ Model for Open Data by Tim Berners-Lee

★ Data online

★★ Data online available as structured data

★★★ Data online available in a non-proprietary open format

★★★★ Use identifiers to denote things, so that people can point at your data

★★★★★ Link your data to other data to provide context

https://5stardata.info

https://5stardata.info/en/, CC0.

Find a (Open) Research Repository

- Registry of Research Data Repositories (= re3data)
- https://www.re3data.org
- best starting point for finding a suitable place for data publication
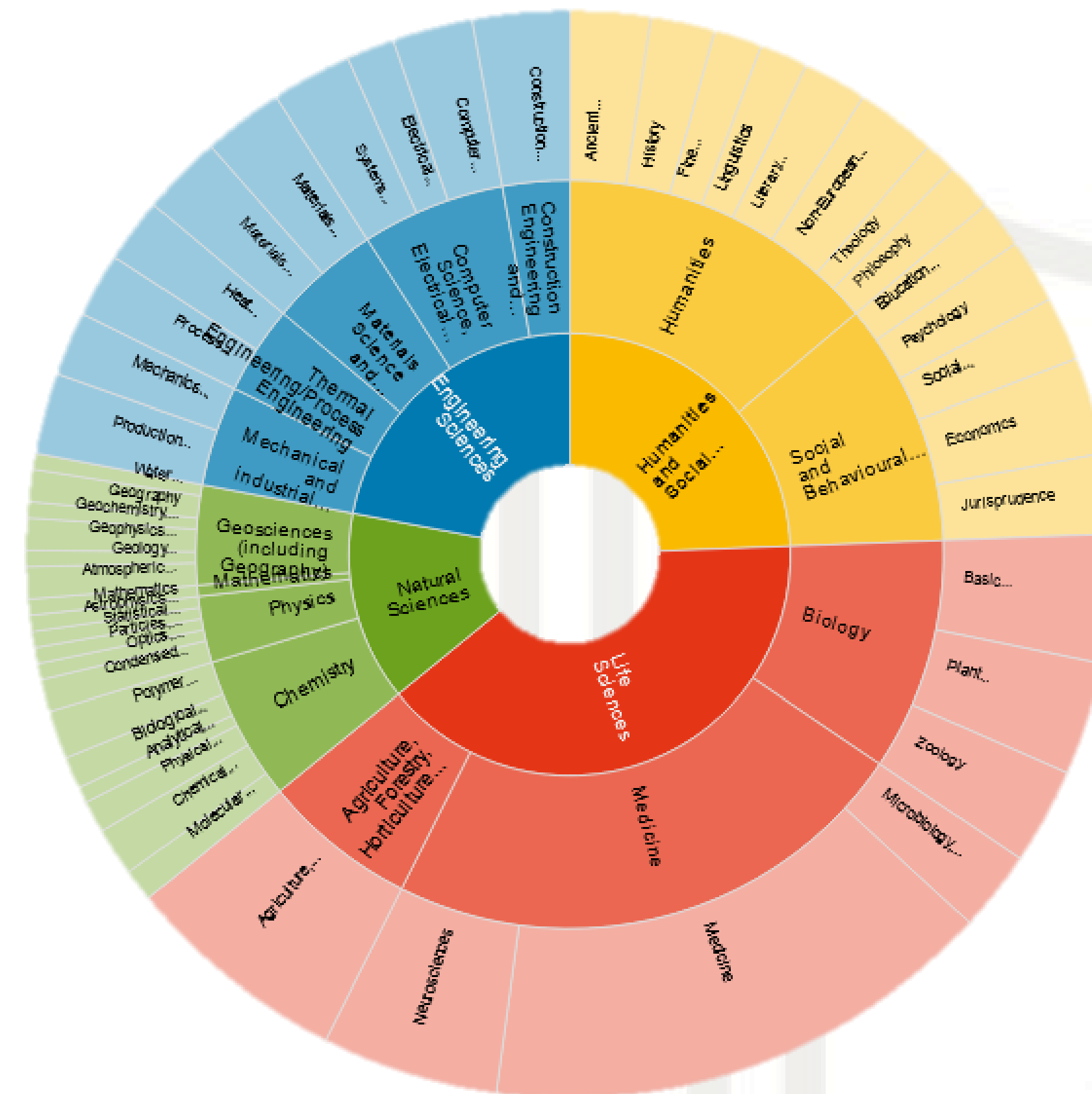
re3data: Search by Subject, CC BY 4.0, https://www.re3data.org.

Recommendation for Data Repositories

1. Community Specific
- i.e. Pangea for Earth & Environmental data
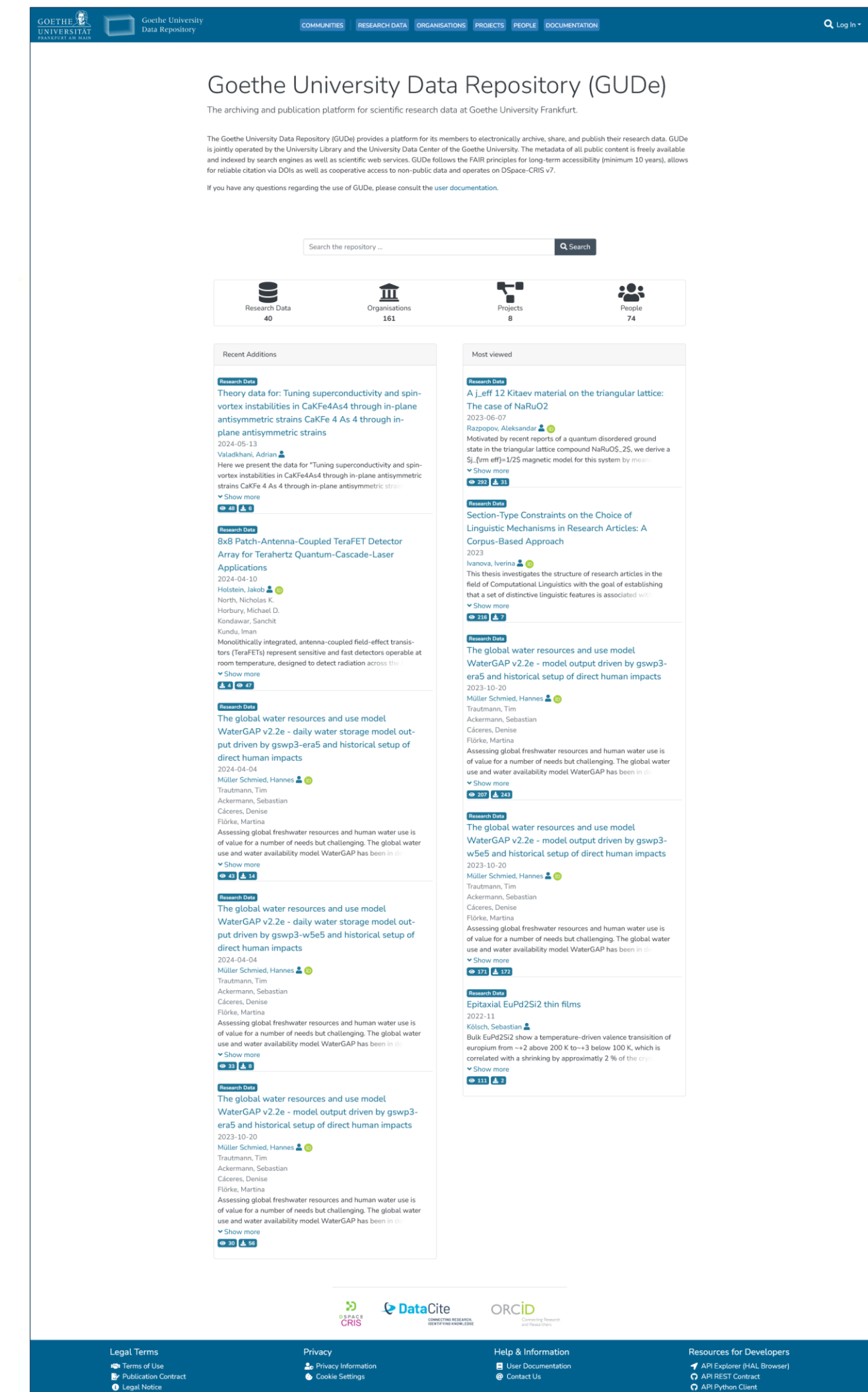
2. Institutional
- i.e. GUDe

3. Generic
- i.e. Zenodo

GUDe

- Goethe University Data Repository
- Hosted at the GU
- https://gude.uni-frankfurt.de
- A service to:
    - Publish research data (open)
    - Meet the FAIR Data Principles
    - Keep research data for at least ten years
    - Share your research data with enriched metadata and PIDs
    - …

Data Journals

Examples:

Psychology
- [Journal of Open Psychology Data](#)

Humanities
- [Journal of Open Humanities Data](#)
- Research Data Journal for the Humanities and Social Sciences

Life Sciences
- [Gigascience](#)
- [Biodiversity Data Journal](#)
- [Open Health Data](#)

Software
- [Journal of Open Research Software](#)

For more see https://www.forschungsdaten.org/index.php/Data_Journals

FAIR Data Principles

- [FAIR Data Principles](#) are a quasi standard for the science community
- It makes sense to apply these principles to your own data publication

- https://www.f-uji.net/?action=test
- F-uji: A web service to programatically assess FAIRness of published research data sets

Anusuriya Devaraju, & Robert Huber. (2020). F-UJI - An Automated FAIR Data Assessment Tool, MIT license, https://doi.org/10.5281/zenodo.4063720.

FAIR Data Principles

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

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

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

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://force11.org/info/the-fair-data-principles/

What to do for Open Research Data?

1. Check whether you like, can and/or must publish your research data openly

2. Structure your data in a human- and machine-readable manner

3. Make your data available via e.g. a data repository

4. Use non-proprietary open formats, if possible

5. Use identifiers

6. Provide sufficient metadata

7. Write a documentation or readme

8. Add a (open) license

Steps to Open Research Data

Four steps 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
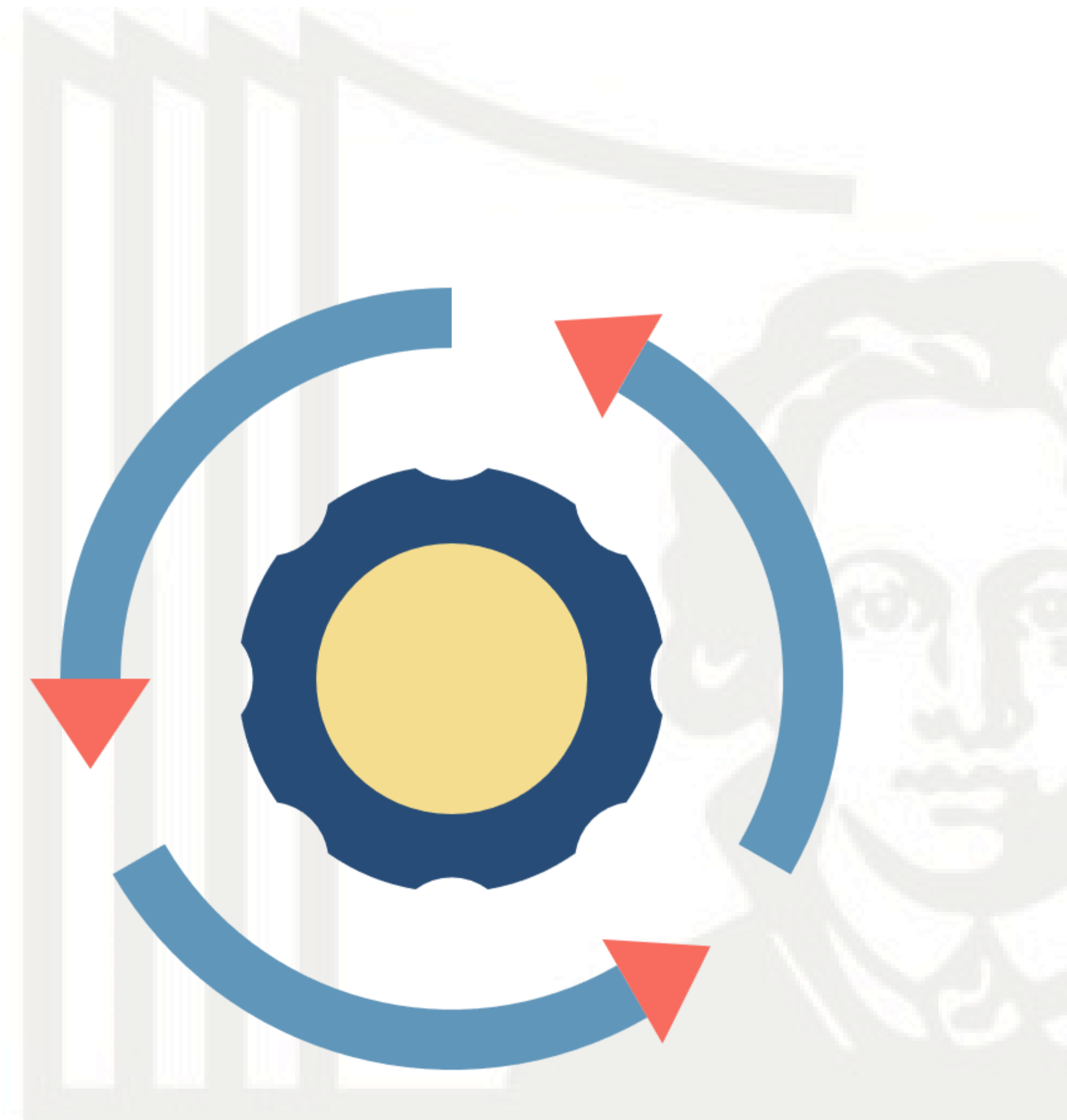
UB

UNIVERSITÄTS
BIBLIOTHEK
FRANKFURT AM MAIN

Reproducibility

- Open research data are desirable because:

  - science is often dependent on the reproducibility of results

  - Own research results are more visible (advantage in competition for attention)

  - the genre of data publication is likely to receive more recognition in the future (a thesis)

  - …

Higher Visibility of Research

- Visibility of own research
- Visibility of own research data
- Better chances with applications e.g. DFG
- Visibility that creating and editing research data is time-consuming → recognition!
- Visibility of the research process through Open Science
- …

Re-Use of Research Data

- Re-use of research data
- Re-use for the future self
- Re-use by community → recognition!
- Better on- and offboarding in working groups etc.
- Legal certainty in the case of onboarding or subsequent use in the event of a change of institution
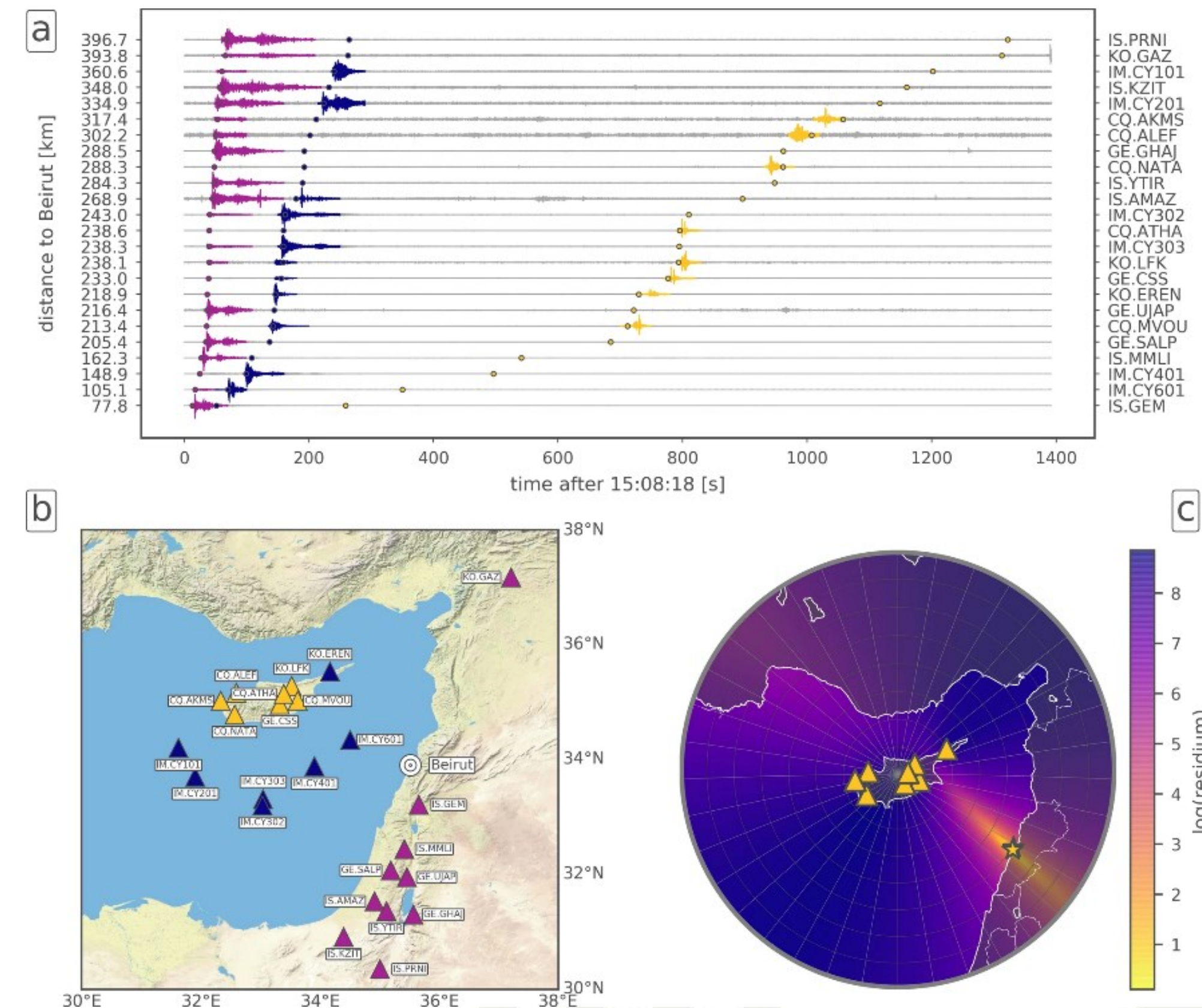
Example for Re-using Open (Research) Data

Explosion in the port of Beirut

- Combination of different open data sources to reconstruct the explosions

- 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.

Research Data with Restricted Access

However, there are also good arguments against open publishing:

- Working with sensitive data, e.g. patient data

- Dual-use data (keyword export control)

- Industry cooperation with contractual regulation of an embargo period for data publication

- Purchased data as a basis (prohibition of disclosure through licensing provisions)

- ...
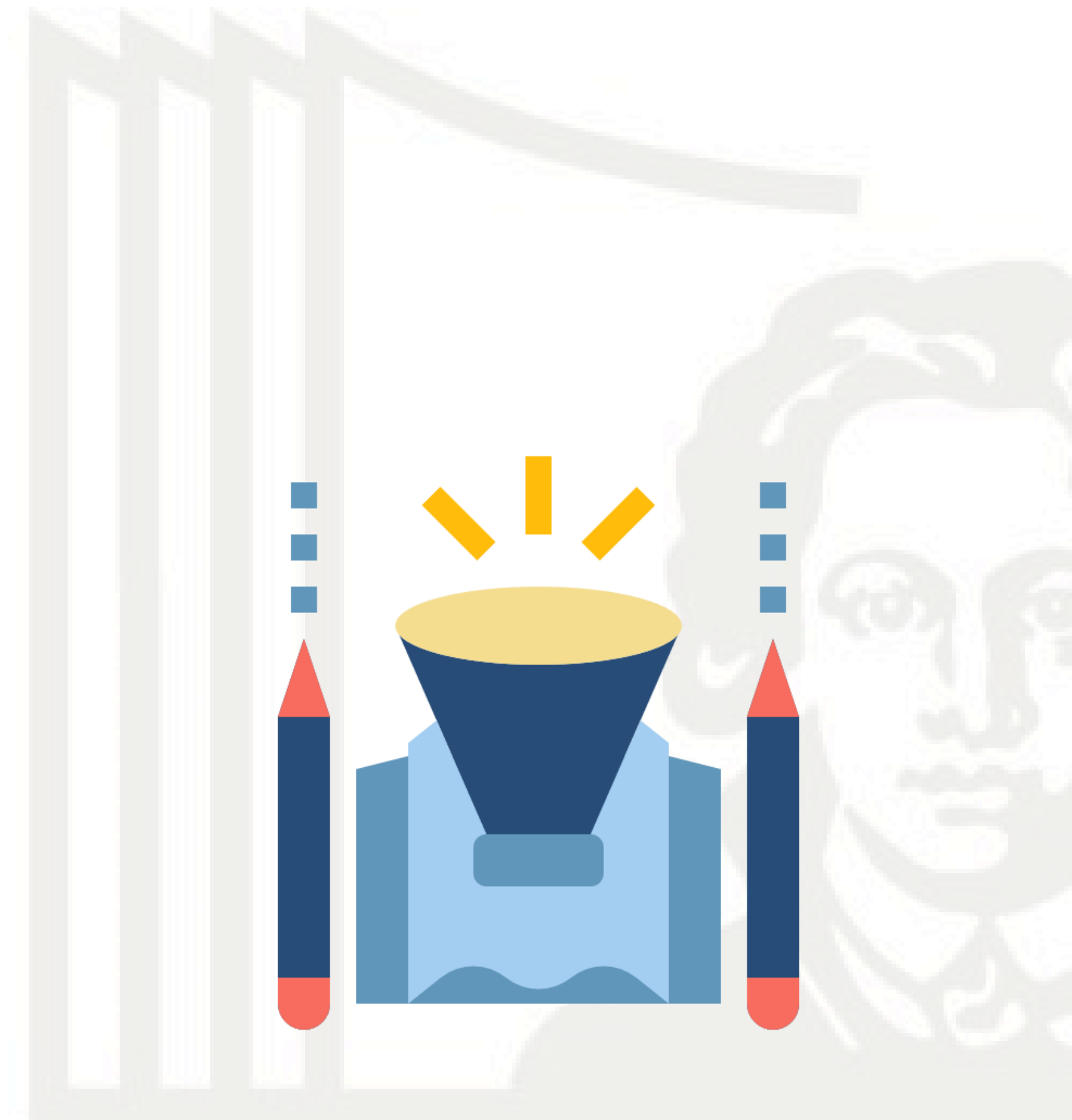
Disabvantages

- "Fits somehow" no longer works for research data

  - If the process of data generation and processing is transparent, you have to work much better than if it is only for you

- Danger of pseudo-transparency

  - You create real data and, in parallel, "open science" data. Everything that fits into the overall picture is taken, the rest is discreetly put in the background

- The standards of science through Open Science are increasing

  - the active participation of scientists from the global South is becoming even more difficult (more passive consumption, see Open Access)
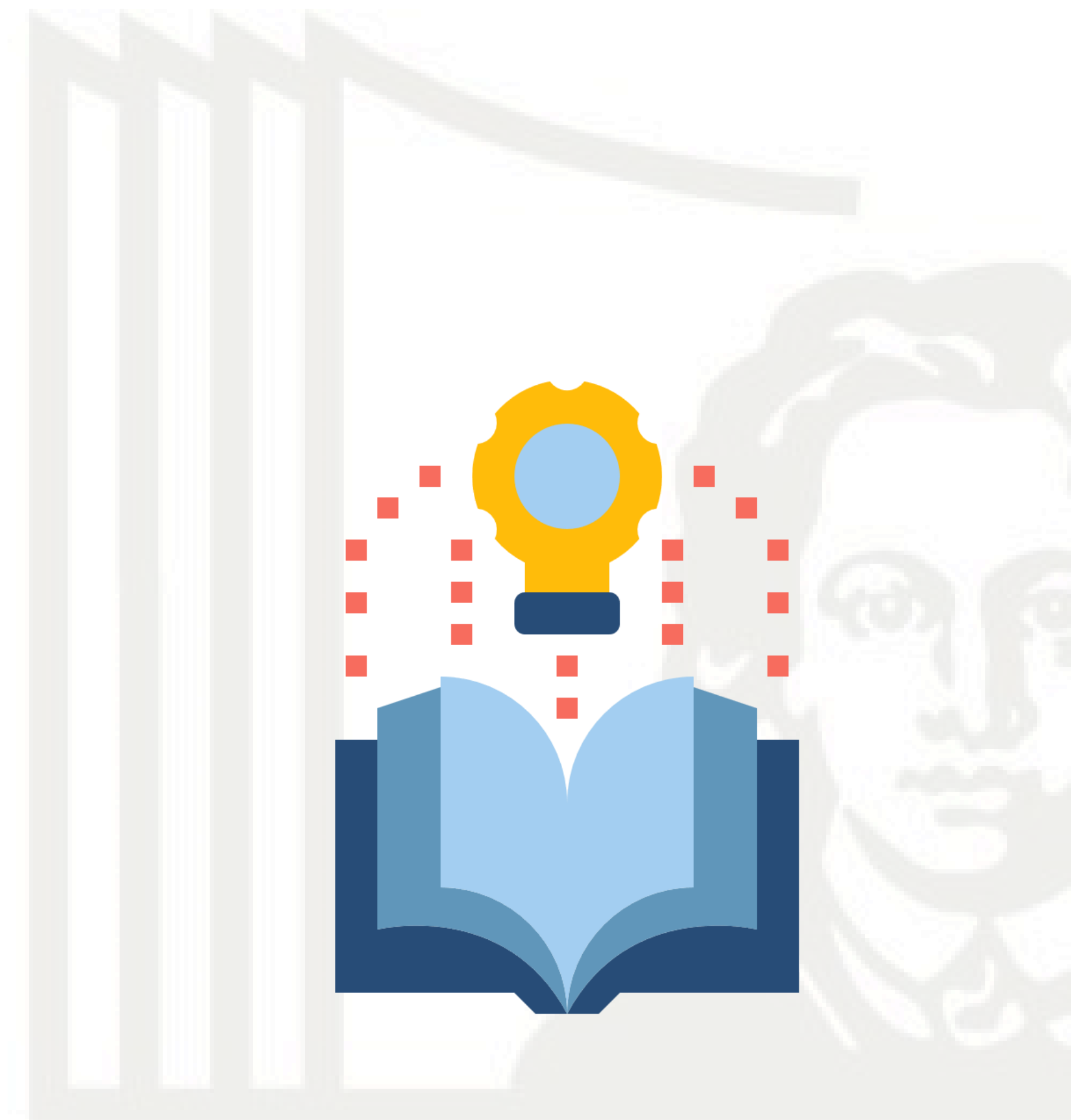
- ...
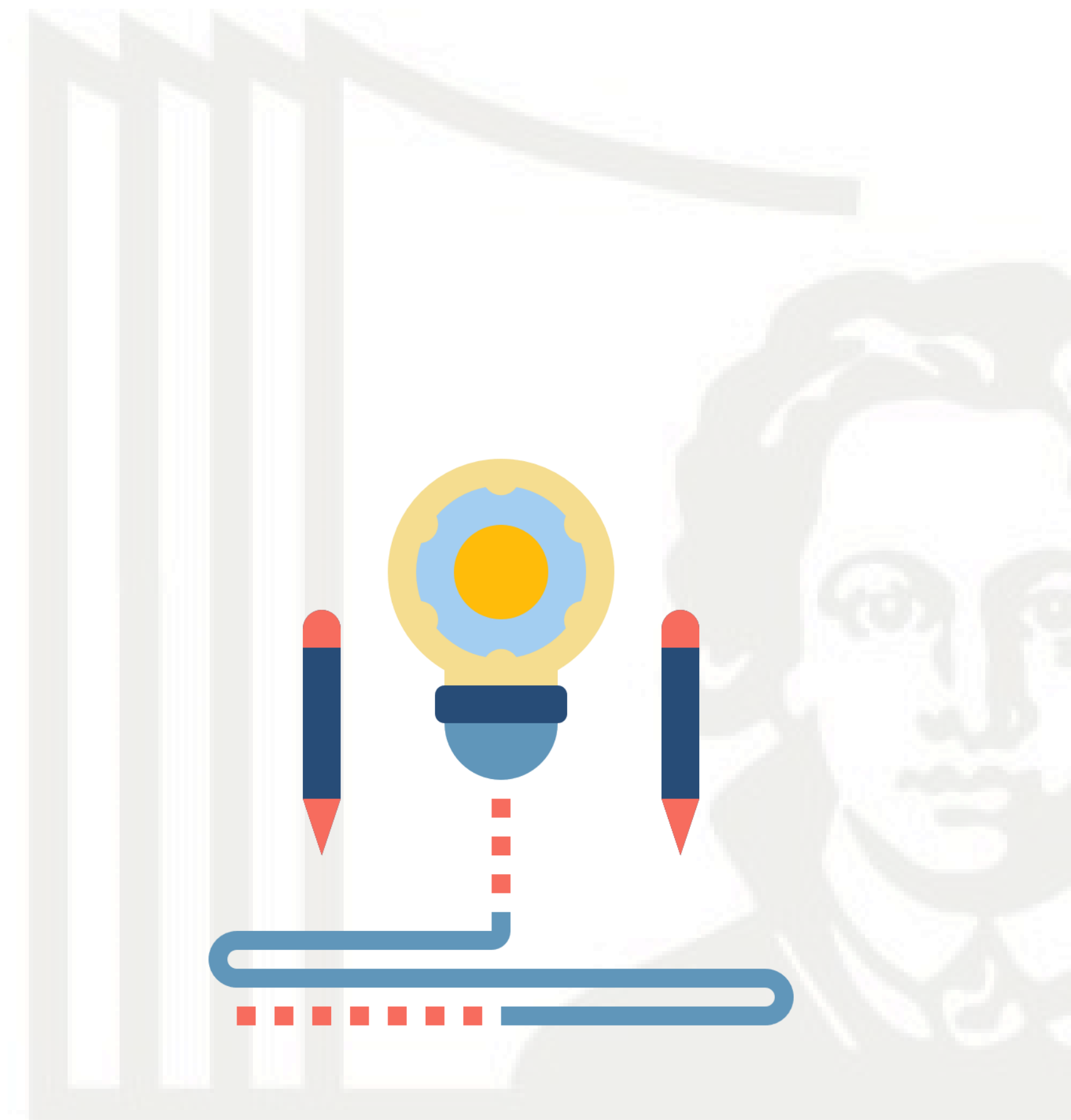
Open Societies?

- Since 2022 it's clear that other concepts against Open Societies exist

- Fundamental discussion around Open Science

- Sharing all data openly can be questionable (more enforcement on the dual-use aspect)

  - Do we want that as a society?

- …

Slido

3. What is your opinion on open and closed research data?

**Join at slido.com
Code: 3373 209**

Thanks for the attention!