Insights into the practicalities of collaboration, data and code sharing across the globe

The PARSEC team, chairs Alison Specht and Shelley Stall

PARSEC is a project sponsored by the Belmont Forum as part of its Collaborative Research Action (CRA) on Science-Driven e-Infrastructures Innovation (SEI), with funding from FAPESP, the ANR, JST and the NSF, with collaborators from Australia, and support from the synthesis centre CESAB of the French Foundation for Research on Biodiversity.

We acknowledge the Traditional Owners and Custodians of the land and sea in all nations. We honour their profound connections to land, water, biodiversity and culture and pay our respects to their Elders past, present and emerging.

time	speaker/topic
8h30	Welcome and Introduction (Belmont Forum and PARSEC data goals)
PARSEC partner talks : presenting ≤ 5 past data challenges that have been met (or not) and 3 main data/code priorities for the future	
8h40	ORCID (Brian Minihan, Engagement Lead: 0000-0001-8412-717X)
8h50	Scholix (Rachael Lammey, product director, community outreach: 0000-0001-5800-1434)
9h	CrossREF (Martyn Rittman, Product Manager: 0000-0001-9327-3734)
9h10	DataCite (Matt Buys, Executive Director of DataCite: 0000-0001-7234-3684)
9h20	EDI (Margaret O'Brien, Data Manager: 0000-0002-1693-8322)
9h30	WDS (David Castle, Chair of the Scientific Committee of the International Science Council's World Data System): 0000-0002-6884-0001)

BELMONT FORUM

PARSEC

ORCID

Brian Minihan, Engagement Lead

0000-0001-8412-717X

Insights into the practicalities of collaboration, data and code sharing across the globe—ORCID

20 March, 2023

Brian Minihan, *Engagement Lead Consortia*

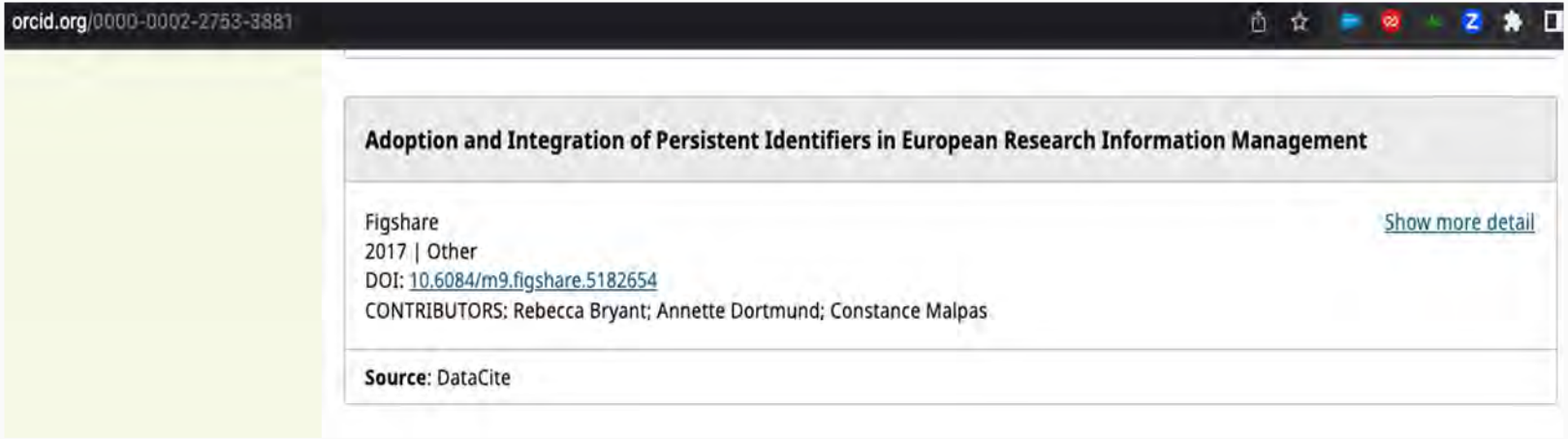
https://orcid.org/0000-0001-8412-717X

Past 5 Data Challenges that have been met (or not)

Other systems actually connected to the ORCID Registry!

- Initially publishers, but this was an important first step to get ORCID going, as a viable persistent identifier for individuals.

Provenance. Opt-In. Source. Trust Markers.



- The decision to a) have all records be user-initiated was very important because that meant that b) as externally added data grew c) that external data was valuable and permission from the user was understood

https://info.orcid.org/the-importance-of-opt-in/

Interoperability. Funding, Peer Review, Membership, Service

- API v 3.0 was groundbreaking because it added so many of the invisible aspects of scholarly research to a central place for a researcher.

- Because of the earlier steps these data aspects are an individual's activities added from external sources, via a trusted connection

Schema triumphs and frustrations

Contributor Roles Taxonomy adopted

- The adoption of this great work recently finalized by NISO, has given an additional impetus for the use of ORCID ID

The other work type.

- One weakness of of current metadata schema is that a number of increasingly important output work types fall under the umbrella category of “other”
- Data management plans, datasets, artistic performances, physical objects and software all appear as “other”

Bringing the data from the Registry to light

- The new Member Portal offers report dashboards of details from the registry--your ORCID integration, affiliations with your institution. One great side effect has been a way to illustrate the importance of organisation ID PIDs (ROR)

Past Data Challenges that have been met, or not

The Top 5

- Other systems connected to registry (publishers were first)!
- Provenance. Source. Opt-In. Trust Markers
- Interoperability. Funding, Peer Review, Membership, Service
- Schema triumphs and frustrations
- Bringing data from the registry to light

3 Main data challenges for the future

Keep working with other PIDs

- Partnership and connection with other systems increases value to our own systems

https://doi.org/10.23640/07243.16687207.v1

Increase visibility and quantity of trust marker data

- Data from external sources should increasingly be the norm.

Data stewardship

- Ensuring that the data we handle is secure and has room to grow

Future Data Priorities for the Future

The Top 3

- Keep working with other PIDs
- Increase visibility and quantity of trust markers
- Data stewardship

Thanks!

b.minihan@orcid.org

@MinihanBrian

BELMONT FORUM

PARSEC

SCHOLIX

Rachael Lammey, product director, community outreach

0000-0001-5800-1434

RESEARCH DATA ALLIANCE

Data challenges & opportunities: Scholix perspective

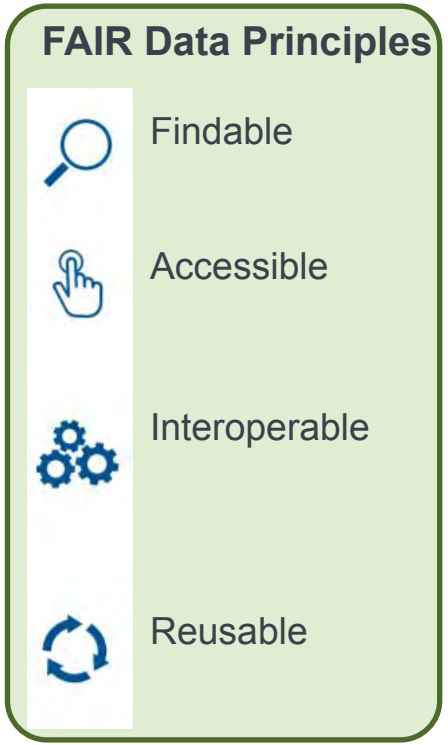
Rachael Lammey, Director of Product, Crossref & Scholix co-chair

PARSEC colocated event

research data sharing without barriers

rd-alliance.org

Scholix is a framework



SCHOLIX

- *Scholarly Link Exchange*

- A framework for standardizing the exchange of *scholarly link* information between scholarly infrastructure providers
  - Focus on articles and datasets
  - Information Model for scholarly links representation
  - Recommendation and provision of exchange formats and protocols

See also http://www.scholix.org/guidelines
https://doi.org/10.1038/sdata.2016.18

5 data challenges: achievements/failures

- **Letting perfect be the enemy of good:** a good, better, best approach would probably have served us all better
- **Clarity:** ‘how do I join Scholix’? rather than the what and why via supporting organizations
- **Participation:** getting data citation on or at the top of publisher agendas and increasing adoption at publishers
- **Standardization:** being able to rely on relationships and identifiers gets us a lot further than hoping everyone will provide the same set of metadata, but we want to be careful not to ‘flatten’ data and software
- **Teamwork:** walking people through step by step (policy, workflows, documentation, working examples) and banging the adoption drum

3 visions for the future

- Achieving greater visibility of research data and software and making more links between research outputs
- Embedding the provision and collection of data and software into publishing workflows
- Better tooling to realise the first point (including machine learning)

BELMONT FORUM

PARSEC

Martyn Rittman, Product Manager

0000-0001-9327-3734

Crossref and software citation

Martyn Rittman, Ph.D.
Rachael Lammey

@CrossrefOrg

Our mission

Crossref makes research outputs easy to find, cite, link, assess, and reuse.

We’re a not-for-profit membership organization that exists to make scholarly communications better.

Background

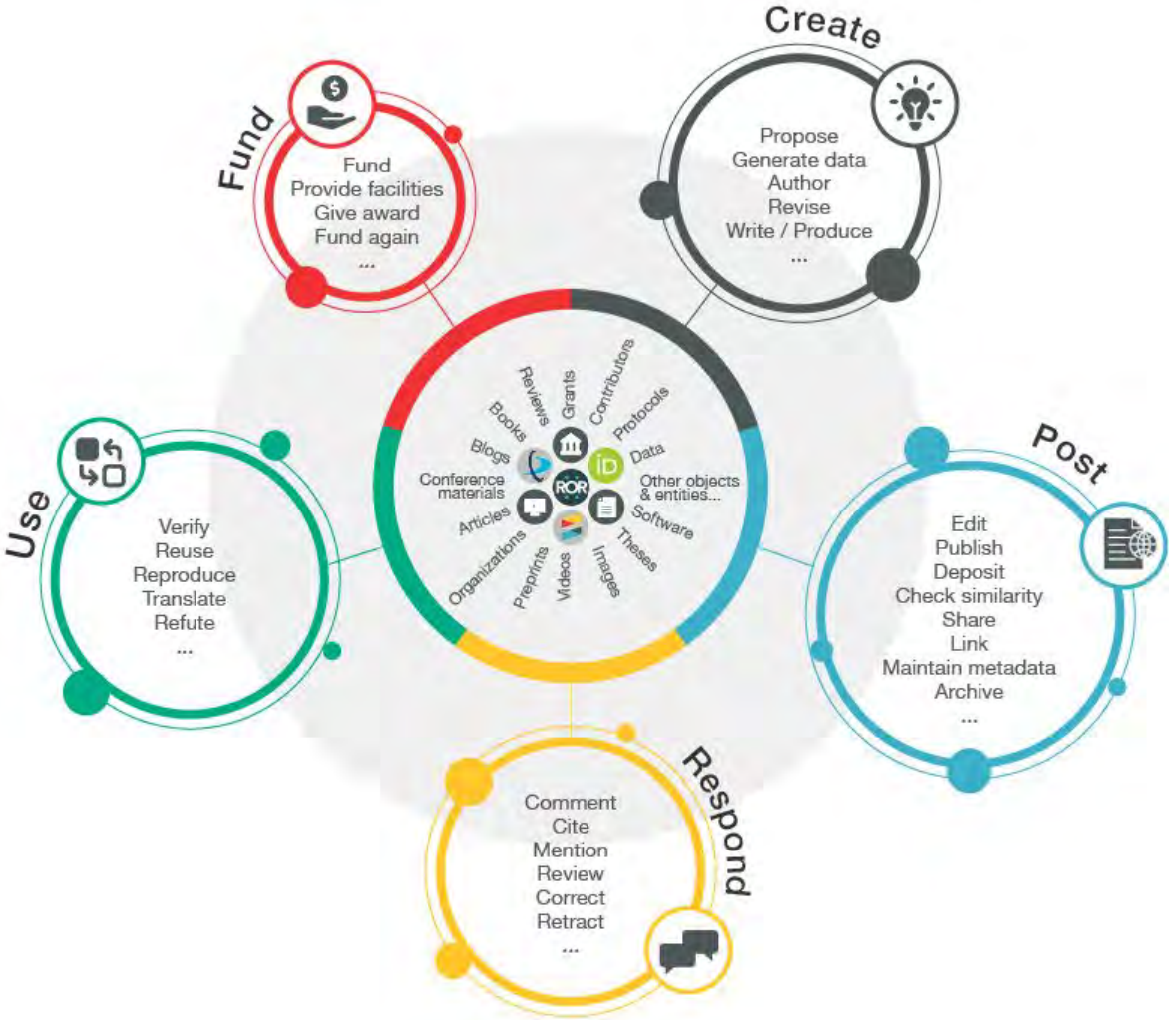
We collect metadata from:

- Members
- Partner organisations (e.g. DataCite)
- The Internet (Wikipedia, Hypothesis, Reddit, etc.)

We link metadata records where a relationship exists.

All our metadata is publicly and freely available.

The Research Nexus



Challenges

- **Garbage in, garbage out:** We need high quality metadata from the community.
- **Give me a DOI:** A focus on identifiers, not metadata.
- **Data models:** A diversity of data types meets a rigid schema.

Challenges

- **Identifying identifiers:** Not all data/software has an identifier. If they have, they’re not always used.
- **What is a publisher?** The makeup of our membership has changed. A minority identify primarily as publishers. There is a shift towards small, globally-based members.

Our vision

- Diverse data in, connected metadata records out.
- One flexible data model, multiple output formats.
- A rich and diverse scholarly communication community that recognises the importance of identifiers with rich metadata.

Thank-you

mrittman@crossref.org
https://chat.rittmanchat.org/@martyn
https://mastodon.online/@crossref

Crossref

@CrossrefOrg

BELMONT FORUM

PARSEC

DataCite

FIND, ACCESS, AND REUSE DATA

Matt Buys, Executive Director of DataCite

0000-0001-7234-3684

DataCite

CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE

DataCite

Insights into the practicalities of collaboration, data and code sharing across the globe.

Matt Buys

March 20, 2023

RDA Co-Located Event, Sweden

@datacite
@mjbuys

DataCite
Working together!

We are a global community that share a common interest: to ensure that research outputs and resources are openly available and connected so that their reuse can advance knowledge across and between disciplines, now and in the future.

As a community, we make research more effective with metadata that connects research outputs and resources–from samples and images to data and preprints. We enable the creation and management of persistent identifiers (PIDs), integrate services to improve research workflows, and facilitate the discovery and reuse of research outputs and resources.

Cross stakeholder collaboration

DataCite

Our challenge

- Data and software citation is not difficult and it works, but the difficulty is that the citation metadata is not always exposed for reuse in an open community corpus because it **does not always end up in the persistent identifier (PID) metadata.**
- Data and software citation infrastructure and services should **never be manual and should always be simple**.
- For the community's benefit, many stakeholders have created and applied potent technologies or strategies to find data and software citations that are not included in PID metadata. **Many of these participants are searching for a place to store this metadata.**
- Use of identifiers varies. **The absence of identifiers and varying abilities to extract useful metadata limits the usefulness of citation metadata.**
- Metadata is key to addressing downstream use cases. In particular, **we lack key metadata fields such as subject classification.**

Strategies for change

1. Leverage existing technologies to accelerate the availability and reuse of data and software citation metadata
2. Provide aggregate open corpuses of key citation metadata for reuse and application globally.
3. Implement technologies that support enhanced PID metadata.

Adoption of Open Research
by Hendricks & Buys (derived from Nosek: cos.io/blog/strategy-for-culture-change)

Path forward
Data Citation Corpus

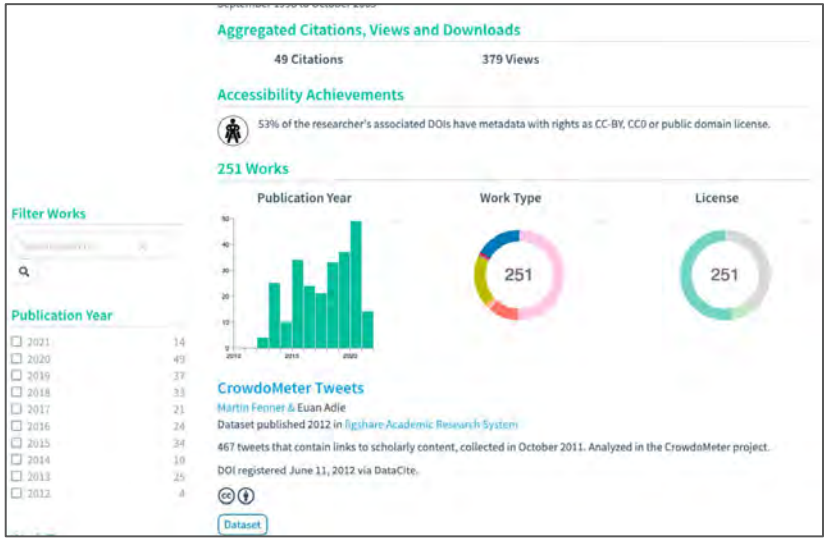
DataCite, with support from the Wellcome Trust, has announced the creation of a Data Citation Corpus, a comprehensive dataset of citations to research data. The Corpus will be developed through a collaborative effort among community stakeholders and will leverage tools that extract data citations from scholarly articles with a high precision rate and relational links in existing PID metadata. The resulting dataset will be made freely available to the global research community, providing a new way to measure the impact of research data and enabling the community to track and analyze data citation patterns. The Data Citation Corpus has the potential to enhance the discoverability and reuse of research data and contribute to the development of new metrics for research assessment. (see https://doi.org/10.5438/vjz9-kx84)

Fundamentally, the corpus addresses the major issue that known data citations exist in third-party systems but are not compiled into a comprehensive, publicly accessible corpus that the community can use.

Applying open citation metadata
Researcher profiles

To access Research Profiles navigate to:
https://commons.datacite.org/orcid.org/
+RESEARCHER_ORCID_ID

Example Researcher Profile:
https://commons.datacite.org/orcid.org/0000-0003-1419-2405



This work was supported by the PARSEC project funded by the Belmont Forum (Collaborative Research Action on Science-Driven e-Infrastructures Innovation) managed through the National Science Foundation (Grant ID 1929464)

DataCite

CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE

info@datacite.org

pidforum.org

datacite.org
blog.datacite.org

support.datacite.org
support@datacite.org

@datacite

DataCite

@datacite

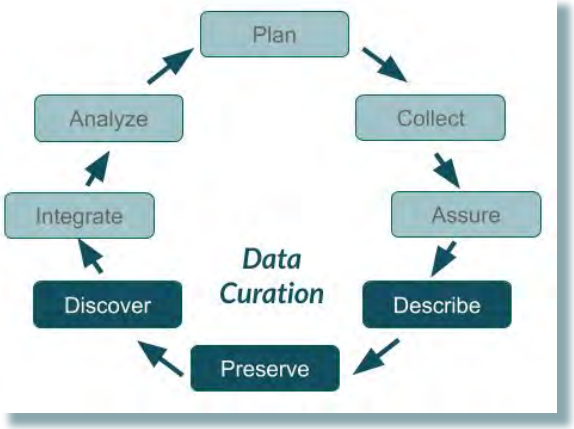
Margaret O’Brien, Data Manager
0000-0002-1693-8322

Challenges and Visions - Environmental Data Initiative

Margaret O'Brien[1], Paul Hanson[2], Bob Waide[3], Corinna Gries[2], Mark Servilla[3]

[1]*University of California, Santa Barbara*
[2]*University of of Wisconsin, Madison*
[3]*University of New Mexico*

Environmental Data Initiative



Semantics, etc

Schema.org

EDI

Digital Object Identifiers Minted

DataONE Member Node

Repository Infrastructure

40 years of LTER data management experience

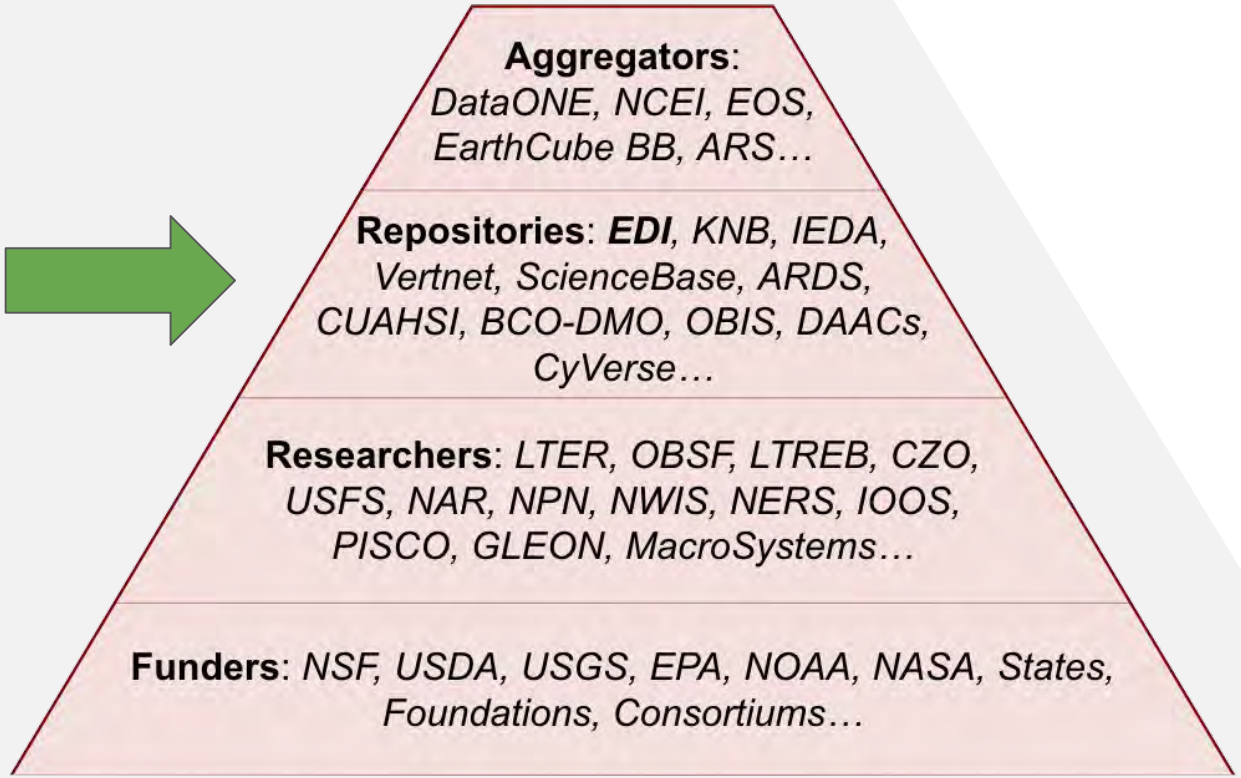
1980s

2010

2015

2020

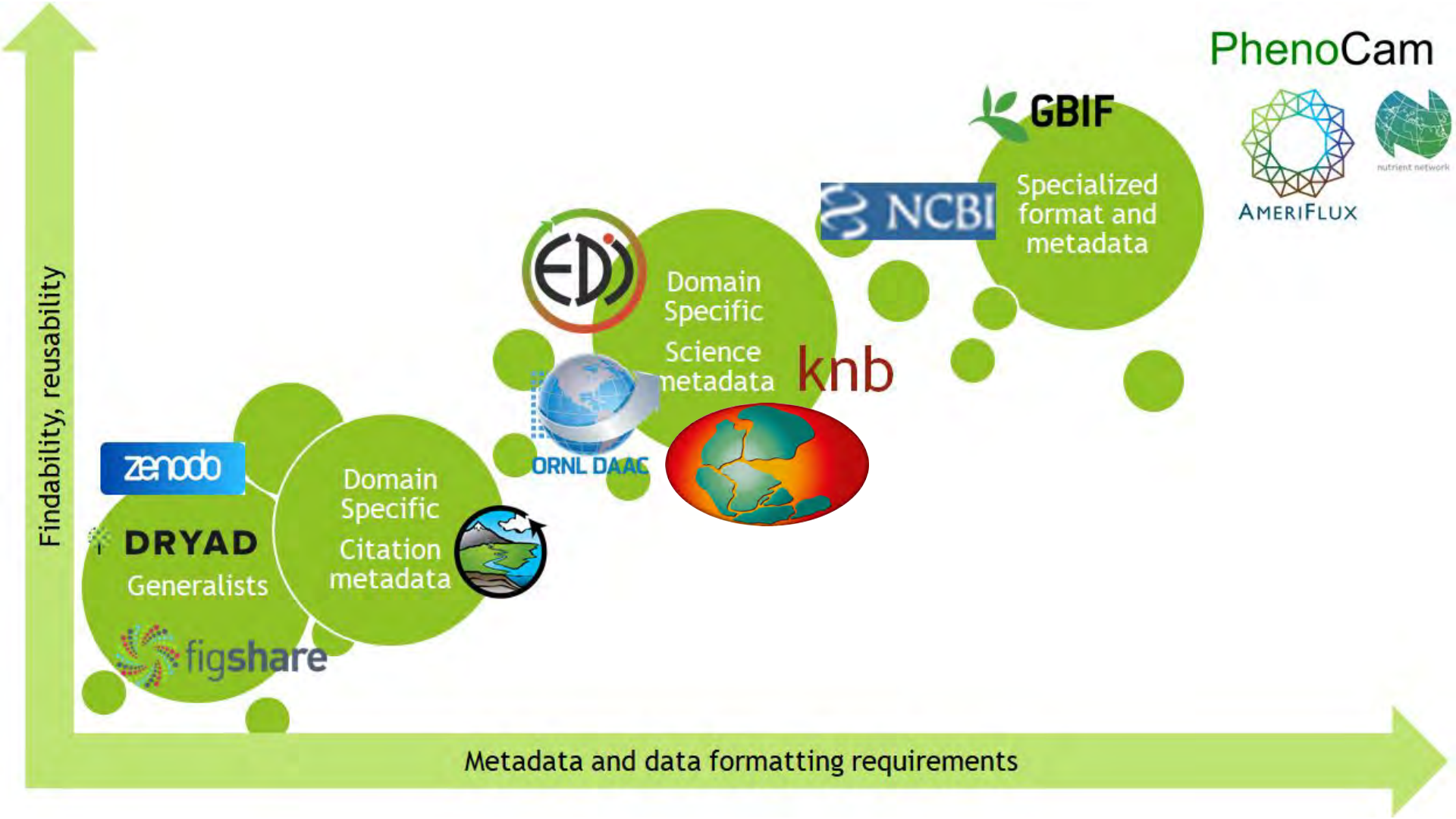
Environmental Data Initiative



Repository software

Assist researchers with data curation

Collaborate with and leverage other work in environmental data management


Findability, reusability
Metadata and data formatting requirements
zenodo
DRYAD
Generalists
figshare
Domain
Specific
Citation
metadata
ORNL DAAC
Domain
Specific
Science
metadata
knb
NCBI
GBIF
Specialized
format and
metadata
PhenoCam
AMERIFLUX
nutrient network

Challenges

1. Culture of data sharing is still developing

Challenges

1. Culture of data sharing is still developing
2. Licensing is inconsistently adopted

Challenges

1. Culture of data sharing is still developing
2. Licensing is inconsistently adopted
3. Existing vocabularies are incomplete and difficult to use

Challenges

1. Culture of data sharing is still developing
2. Licensing is inconsistently adopted
3. Existing vocabularies are incomplete and difficult to use
4. Technical collaboration among repositories

Challenges

1. Culture of data sharing is still developing
2. Licensing is inconsistently adopted
3. Existing vocabularies are incomplete and difficult to use
4. Technical collaboration among repositories
5. Ad hoc research data are not “Analysis Ready”

Visions

Improved Discoverability

Visions

Improved Discoverability

AI Readiness - Data Standardization

Visions

Improved Discoverability

AI Readiness - Data Standardization

Sustainability, and Sustainable Data

Thank you

Contact - info@environmentaldatainitiative.org

Website - https://environmentaldatainitiative.org/

Data portal - https://portal.edirepository.org

Twitter - @EDIgotdata

Slack - edi-got-data

GitHub - https://github.com/EDIorg https://github.com/PASTAplus/PASTA

David Castle, Chair of the Scientific Committee of the International Science Council's World Data System

0000-0002-6884-0001

PARSEC
and the
World Data System

PARSEC in the Eyes of the WDS

- Synthesis Strand (the scientific researchers)
- Data Strand (the data stewards)
- Work on Data & Digital Outputs MP, knowledge/practices, relevance to other Belmont funded projects
- Central, term-limited primary funder (Belmont Forum) 2019
- Integration of natural and social science research themes
- Potential positive impact on global south
- Community, indigenous, and citizen-science led initiatives

Meredith Goins wds-ipo@utk.edu worlddatasystem.org

WORLD DATA SYSTEM

ACTION PLAN

This Action Plan focuses on making progress on **four objectives in the next two years:**

1. Provide services and support to existing and new members
2. Develop value narratives for WDS members
3. Provide global leadership and agenda setting
4. Enhance access, quality, and accessibility of data worldwide

Meredith Goins wds-ipo@utk.edu worlddatasystem.org

WORLD DATA SYSTEM

WDS Action Plan (2022-2024)

1. Provide services and support to existing and new members

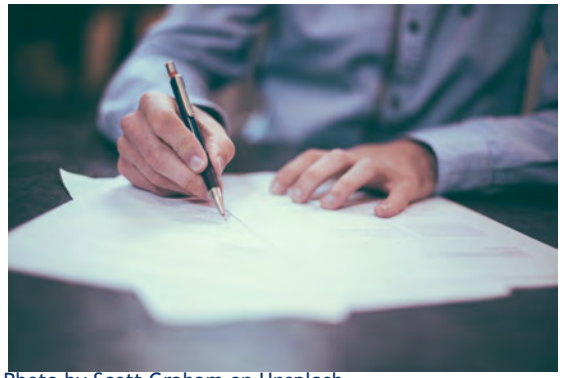
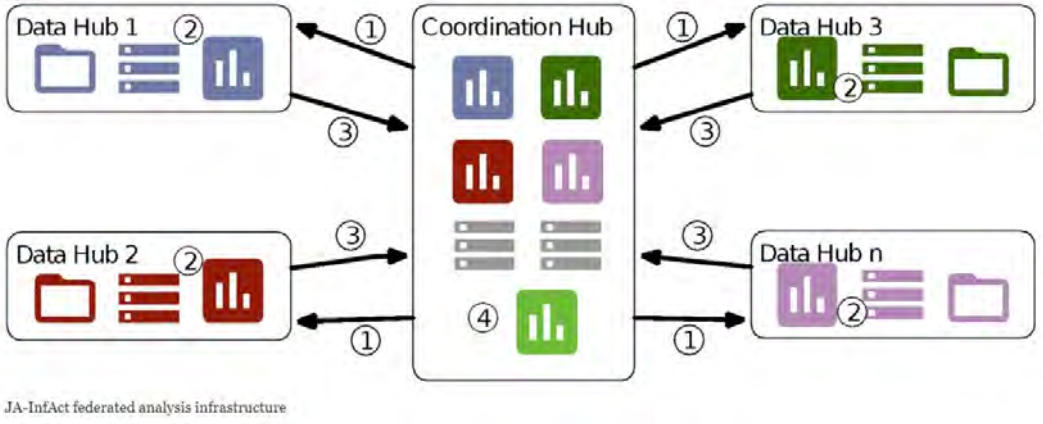


Photo by Scott Graham on Unsplash

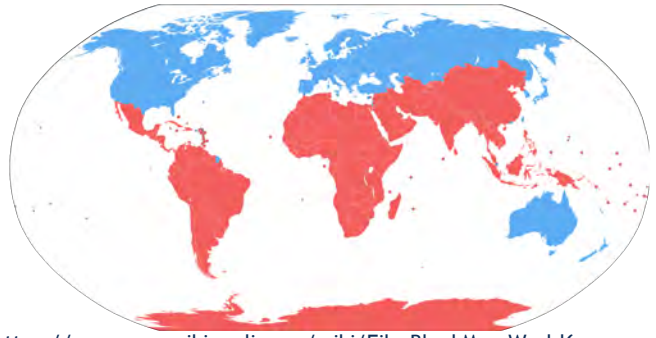
Fig. 3

From: Coping with interoperability in the development of a federated research infrastructure: achievements, challenges and recommendations from the JA-InfAct



JA-InfAct federated analysis infrastructure

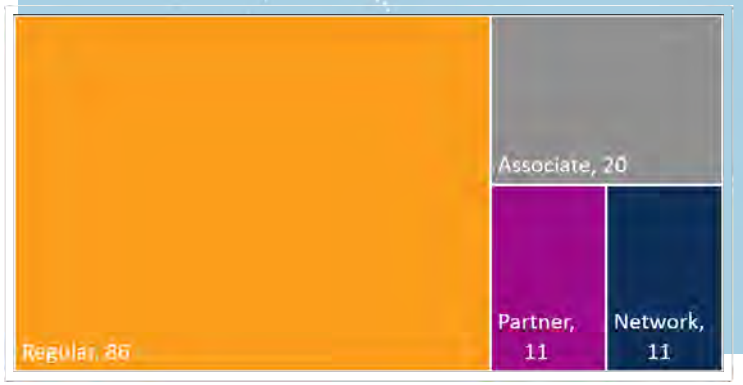
González-García, J., Estupiñán-Romero, F., Tellería-Orriols, C. *et al.* Coping with interoperability in the development of a federated research infrastructure: achievements, challenges and recommendations from the JA-InfAct. *Arch Public Health* **79**, 221 (2021). https://doi.org/10.1186/s13690-021-00731-z



https://commons.wikimedia.org/wiki/File:BlankMap-World6.svg

Meredith Goins wds-ipo@utk.edu worlddatasystem.org

WDS Members & The Global South



Meredith Goins wds-ipo@utk.edu worlddatasystem.org

WORLD DATA SYSTEM

WDS Action Plan (2022-2024)

2. Develop value narratives for WDS members

Photo by Panos Sakalakis on Unsplash

Photo by Jon Tyson on Unsplash

Meredith Goins wds-ipo@utk.edu worlddatasystem.org

WORLD DATA SYSTEM

WDS Action Plan (2022-2024)

3. Provide global leadership and agenda setting

Meredith Goins wds-ipo@utk.edu worlddatasystem.org

WDS Action Plan (2022-2024)

3. Provide global leadership and agenda setting - CONTINUED

WORLD DATA SYSTEM

Early Career Researchers and Scientists Network

Photo by Chris Montgomery on Unsplash

WDS Data Sharing Principles

- Data, metadata, products, and information should be fully and openly shared, subject to national or international jurisdictional laws and policies, including respecting appropriate extant restrictions, and in accordance with international standards of ethical research conduct.
- Data, metadata, products, and information produced for research, education, and public-domain use will be made available with minimum time delay and free of charge, or for no more than the cost of dissemination, which may be waived for lower-income user communities to support equity in access.
- All who produce, share, and use data and metadata are stewards of those data, and have responsibility for ensuring that the authenticity, quality, and integrity of the data are preserved, and respect for the data source is maintained by ensuring privacy where appropriate, and encouraging appropriate citation of the dataset and original work and acknowledgement of the data repository.
- Data should be labelled 'sensitive' or 'restricted' only with appropriate justification and following clearly defined protocols, and should in any event be made available for use on the least restrictive basis possible.

Meredith Goins wds-ipo@utk.edu worlddatasystem.org

WORLD DATA SYSTEM

WDS Action Plan (2022-2024)

4. Enhance access, quality, and accessibility of data worldwide

Photo by Ronda Dorsey on Unsplash


Photo by FLY:D on Unsplash

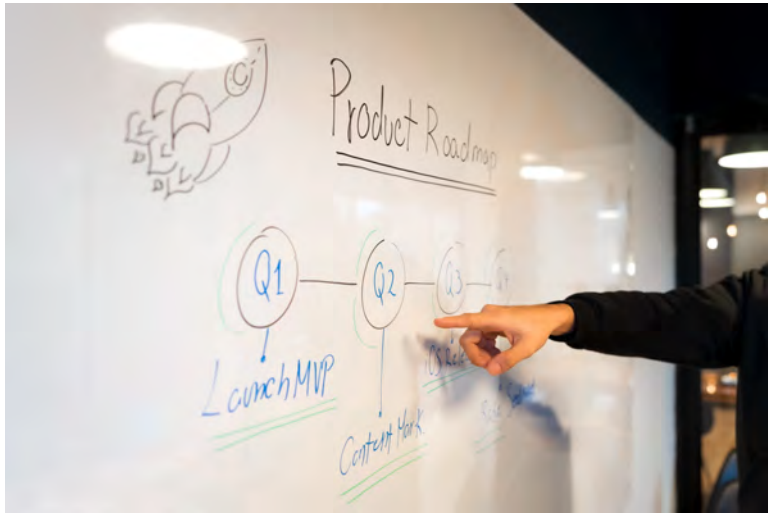


Photo by Slidebean on Unsplash

Meredith Goins wds-ipo@utk.edu worlddatasystem.org

23-26 OCT

2023

SALZBURG

International Data Week

A FESTIVAL OF DATA

https://www.rd-alliance.org/plenaries/international-data-week-2023-salzburg

Organised by

COMMITTEE ON DATA CODATA INTERNATIONAL SCIENCE COUNCIL

RDA RESEARCH DATA ALLIANCE

WORLD DATA SYSTEM

Hosted by:

SALZBURG CONVENTION BUREAU

PARIS LODRON UNIVERSITY SALZBURG

SALZBURG CONGRESS

Meredith Goins wds-ipo@utk.edu worlddatasystem.org

Insights into the practicalities of collaboration, data and code sharing across the globe

Morning tea

PARSEC practitioner talks : presenting ≤ 5 past data challenges that have been met (or not) and 3 main data/code priorities for the future	
10h (remote)	Rodolphe Devillers (IRD: 0000-0003-0784-847X)
10h10	Romain David (ERINHA, EOSC Life: 0000-0003-4073-7456)
10h20	Laurence Mabile (University of Toulouse and SHARC : 0000-0002-7724-1721)
10h40 (remote)	Jeaneth Machicao (University of Sao Paulo : 0000-0002-1202-0194)
10h50	Pedro Correa (University of Sao Paulo: 0000-0002-8743-4244)
11h (remote)	Lesley Wyborn (ANU, NCI : 0000-0001-5976-4943)
11h10	Hayashi Kazuhiro (NISTEP: National Institute for Science and Technology : 0000-0003-1996-4259)

Rodolphe Devillers

0000-0003-0784-847X

Marine conservation and spatial analysis – some data challenges and priorities

Rodolphe Devillers &
Gaétan Morand

IRD, France

March 20, 2023

Challenge #1 – Metadata are often of little use to actual users

- Data analysis typically requires re-using data produced by someone else for another context

Challenge #1 – Metadata are often of little use to actual users

- Data analysis typically requires re-using data produced by someone else for another context
- Creates various challenges (accessibility, reusability, reliability, etc.)

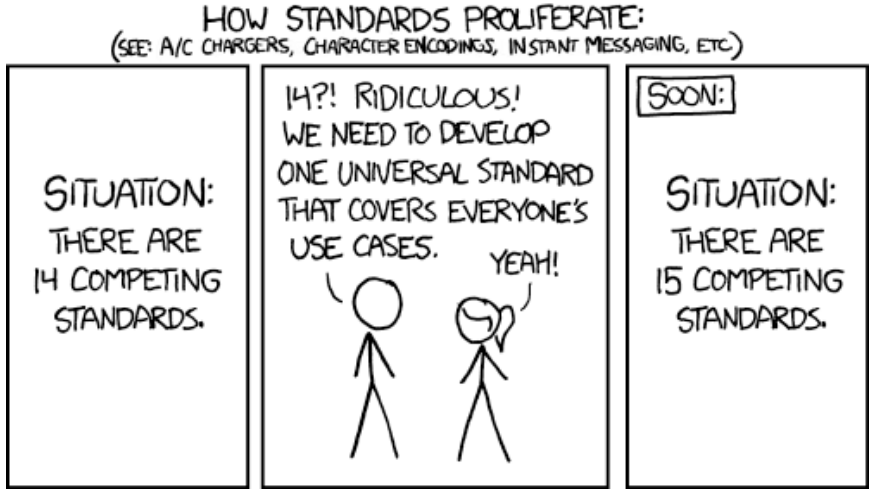
Challenge #1 – Metadata are often of little use to actual users

- Data analysis typically requires re-using data produced by someone else for another context
- Creates various challenges (accessibility, reusability, reliability, etc.)
- End-users face a deluge of imperfect data = what data to use? Is dataset X good enough for my work? etc.

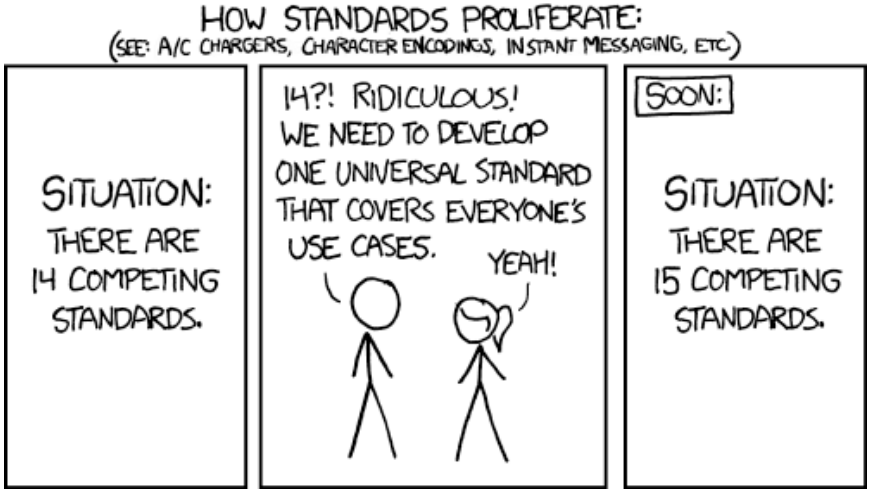
Challenge #1 – Metadata are often of little use to actual users

- Data analysis typically requires re-using data produced by someone else for another context
- Creates various challenges (accessibility, reusability, reliability, etc.)
- End-users face a deluge of imperfect data = what data to use? Is dataset X good enough for my work? etc.
- While core « discovery metadata » are useful for accessing/cataloguing data, they fall short to efficiently inform users and reduce potential risks of misuse

Challenge #2 – Too many standards kills the standards

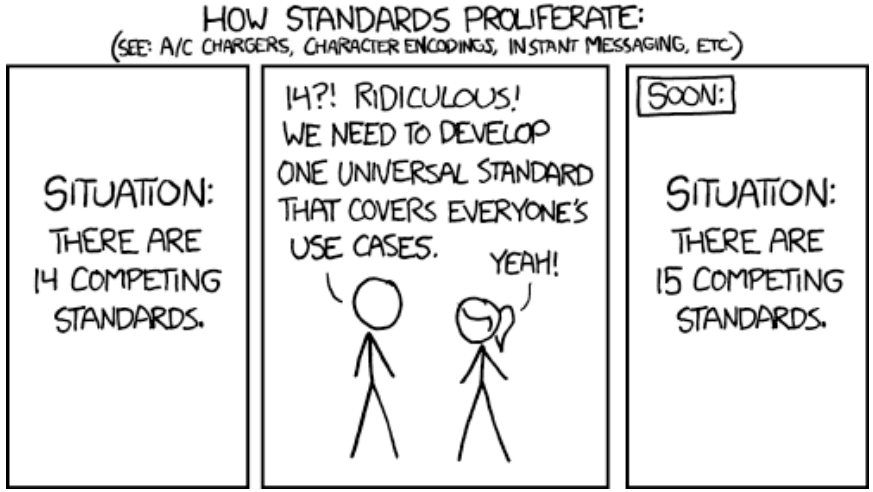


Challenge #2 – Too many standards kills the standards



- "*The nice thing about standards is that you have so many to choose from; furthermore, if you do not like any of them, you can just wait for next year's model* » (A. Tanenbaum)

Challenge #2 – Too many standards kills the standards



- "*The nice thing about standards is that you have so many to choose from; furthermore, if you do not like any of them, you can just wait for next year's model* » (A. Tanenbaum)
- Too many standards, too many data repositories, too many tools, etc. – confusing for too many scientists (where should I catalog/archive my data, etc.)

Challenge #3 – Data science solutions seem suitable… for data scientists

- The way most of you see the world… is not shared by the rest of the world

Challenge #3 – Data science solutions seem suitable… for data scientists

- The way most of you see the world… is not shared by the rest of the world
- What seems simple and obvious to you can be cumbersome and a pain for people to do outside of the data science circles

Challenge #3 – Data science solutions seem suitable… for data scientists

- The way most of you see the world… is not shared by the rest of the world
- What seems simple and obvious to you can be cumbersome and a pain for people to do outside of the data science circles
- Real end-users should be more often involved in the conceptualization of tools/methods developed in data/open-science

Challenge #4 – Data integrators need… to really integrate data

- Data integrators (e.g. GBIF, OBIS with biological data) see their role as providing a service to store and share scientific data, but not to ensure data are usable/of quality

Challenge #4 – Data integrators need... to really integrate data

- Data integrators (e.g. GBIF, OBIS with biological data) see their role as providing a service to store and share scientific data, but not to ensure data are usable/of quality
- Integrating is more than piling up data! Simple data verifications remain too often missing (ex. how can whales be spotted by scientists in the middle of Africa, or thousands of years ago? – the tip of the data issues iceberg?)

Challenge #4 – Data integrators need… to really integrate data

- Data integrators (e.g. GBIF, OBIS with biological data) see their role as providing a service to store and share scientific data, but not to ensure data are usable/of quality
- Integrating is more than piling up data! Simple data verifications remain too often missing (ex. how can whales be spotted by scientists in the middle of Africa, or thousands of years ago? – the tip of the data issues iceberg?)
- Possible legal ramifications in case of data misuse (a duty to inform/protect users)

Challenge #5 – Why bother sharing if nobody looks at it?

- Curating data vs effective and efficient promoting data reuse (dataset download vs citation statistics)

Challenge #5 – Why bother sharing if nobody looks at it?

- Curating data vs effective and efficient promoting data reuse (dataset download vs citation statistics)
- In science, reviewers rarely bother to review appendices. Should we require journals to systematically call upon data science reviewers who could dig into the appendices/code/data?

Main data/information priorities

1. **Providing tools that truly help/guide users select and use existing data** (ex. could something like ChatGPT act as an interface between data and users?) – that can qualify various aspects of the data and raise users' awareness of trade-offs between different data sources

Main data/information priorities

1. **Providing tools that truly help/guide users select and use existing data** (ex. could something like ChatGPT act as an interface between data and users?) – that can qualify various aspects of the data and raise users' awareness of trade-offs between different data sources
2. **Providing user-centric** (not data-centric) **open-data solutions**
3.

Main data/information priorities

1. **Providing tools that truly help/guide users select and use existing data** (ex. could something like ChatGPT act as an interface between data and users?) – that can qualify various aspects of the data and raise users' awareness of trade-offs between different data sources

2. **Providing user-centric** (not data-centric) **open-data solutions**

3. **Promoting data science positions** on large projects and in organizations (and not simply expect scientists will adopt those new practices). Much like science communication is a job on its own, FAIR-ising science also requires specific skills.

Romain David

0000-0003-4073-7456

Data Challenges in Life Sciences

(PARSEC: Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas)

Romain David

ERINHA-AISBL - FR

romain.david@erinha.eu

ORCID : 0000-0003-4073-7456

romain_david_13

https://hal.science/tel-01839376

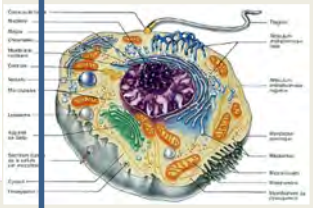
Multi-scale – Multi-formats – Multi-sources…

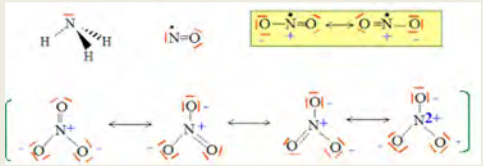
Landscape, regional and human interactions levels

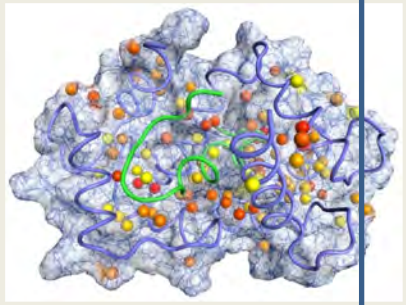
Habitats – population levels

Living organism levels

Physical and Chimical levels







Genomic, cells, transcriptomic

Proteomic, metabolimic, epigenetic

Phylogeography

,

Ecosystem

Sociology,

Economy,

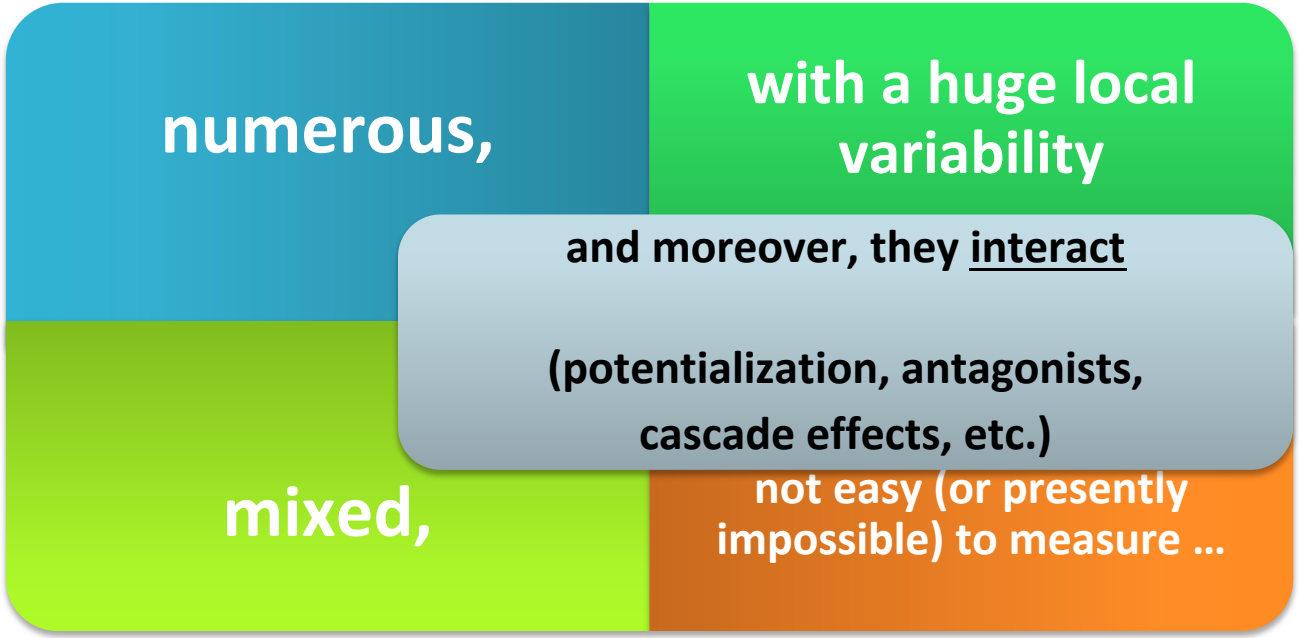
Law and well being…

Which holistic approach for a patchwork of contexts at each level of organization?

Challenge **1**: Highly heterogeneous, Highly linked

- *Biological and ecological functioning is linked to a huge number of factors, not easy to measure,*
- *Indication « value » of factors is very dependant on the* ***context***

Factors are:



History LS Data challenge (**2**):

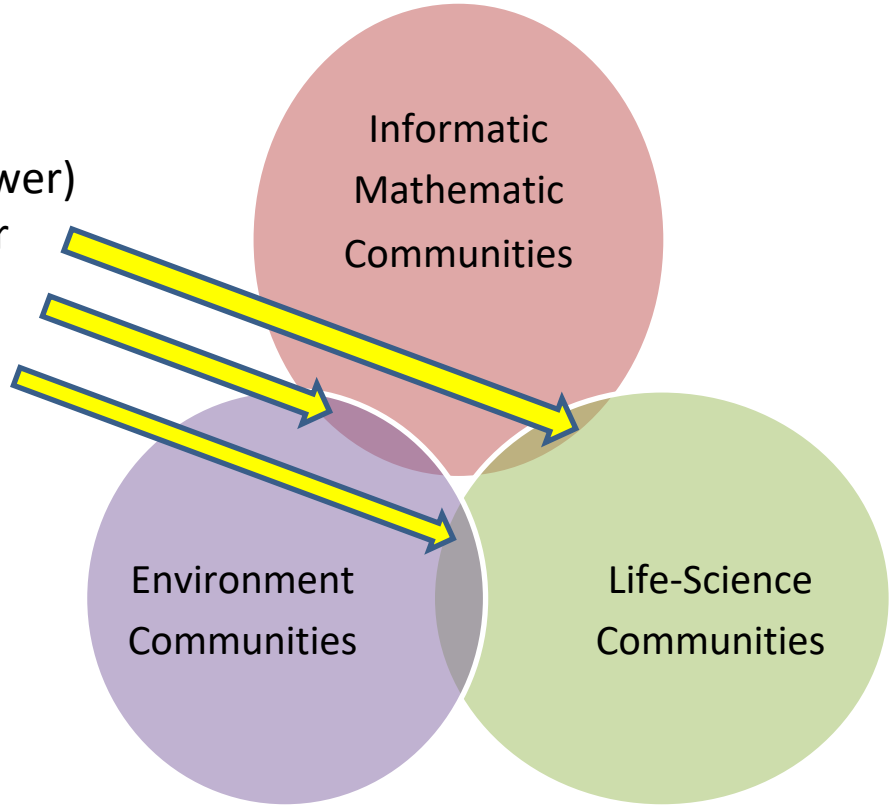
LS Data challenge **(2)** : The weight of history

More often that!?

romain_david_13

Human LS Data challenge **(3)**: Communities skills interoperation

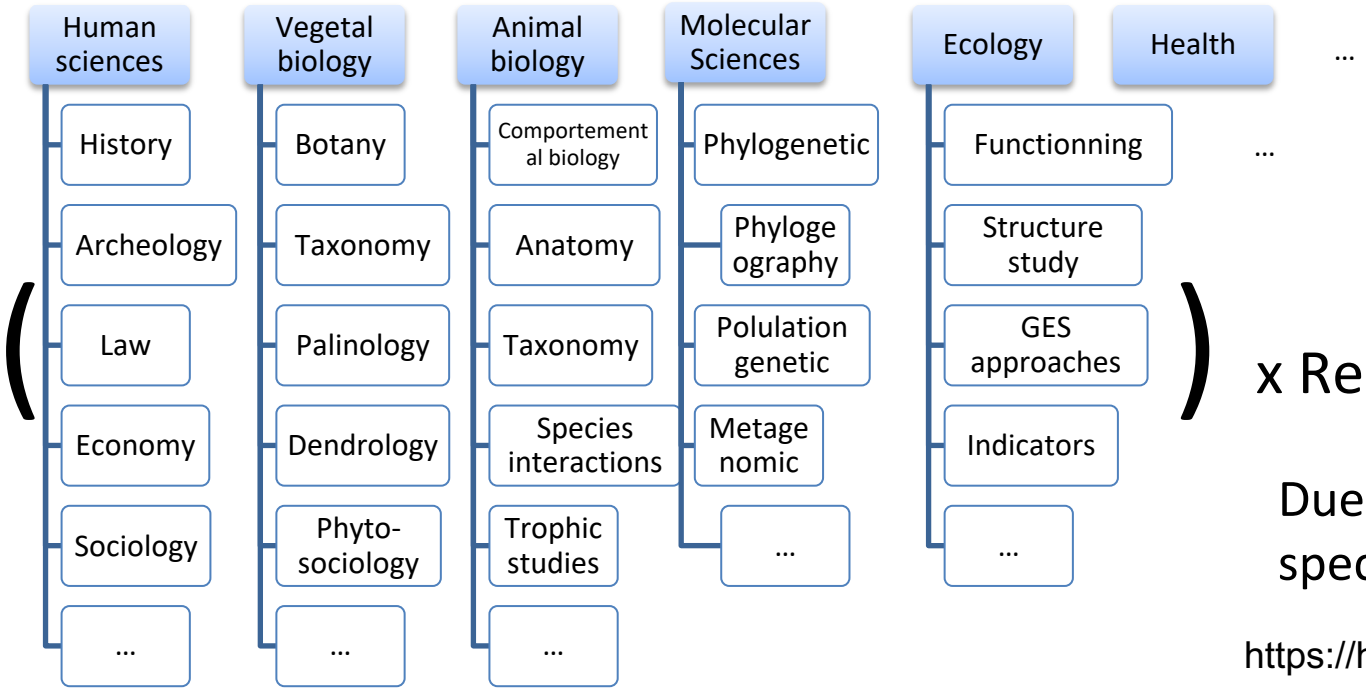
Only a little
of time (man power)
is really spent for
Interdisciplinary
approach



Silo challenge **(4)**: community-specific e-infrastructure

Scope

Mono-scale – Mono-disciplinar,
Not well connected



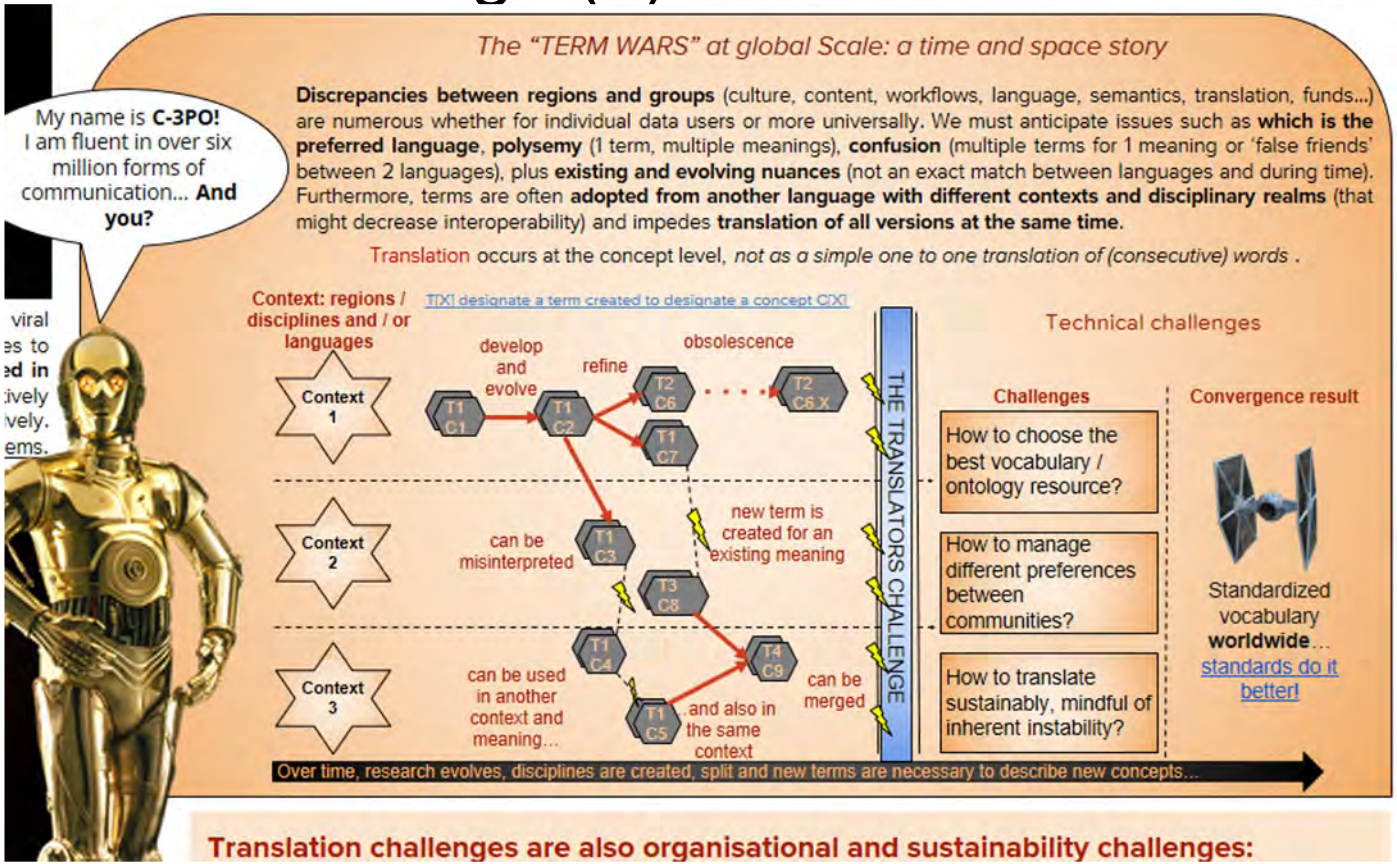
x Region x Models…

Due to an oversectored specialization?

https://hal.science/tel-01839376

https://doi.org/10.5281/zenodo.6588167

LS Data challenge (**5**): The TERM wars

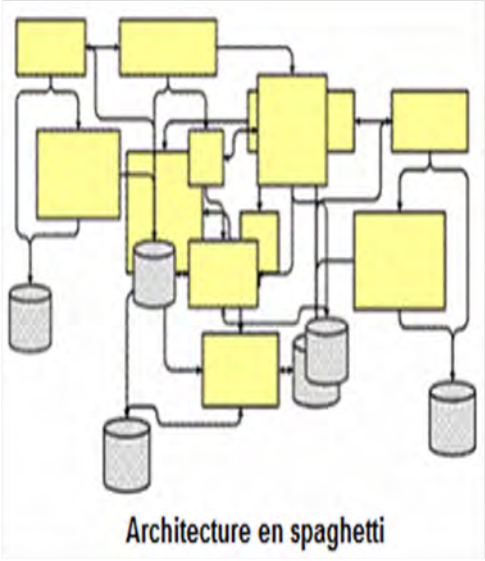


romain_david_13

Data sharing challenges in LS

My PREEEECIOUSS DATABASE

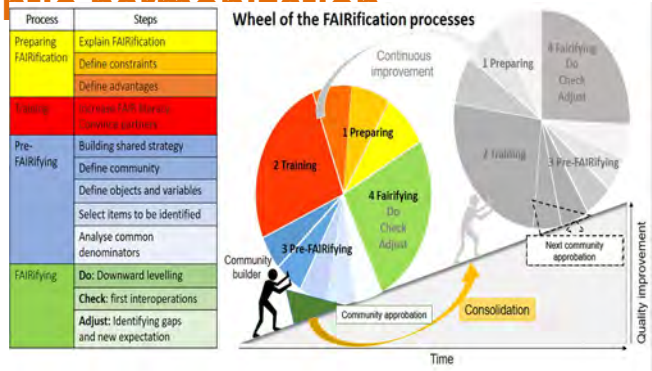
Without organisation of SI and data productions (and humans):



3 main LS data sharing priorities for the future

- **PROPOSE attractive INCENTIVES and REWARDS for FAIR Data Sharing**
  - FAIR Literacy as a start, inclusiveness, **collaborations more than competition**
- **INCREASE and SHARE SKILLS - organise them in strong sustainable networks**
  - Inter disciplinary **LS Data Competence Centers** - e.g. Cluster EOSC-Life
  - **Young researchers as targets for Data Stewardship**
- **INTEROPERATE! Iterative and progressive semantic harmonization**
  - Promote community approval processes* -
  - From data dictionaries to **linked data**

EOSC-*Life*



* http://doi.org/10.5334/dsj-2020-032

romain_david_13

References

- Romain David. De la conception d'un système d'observation à large échelle au déploiement et à l'exploitation de son système d'information : application à l'observation des habitats coralligènes et à la colonisation de récifs artificiels (ARMS). Biodiversité et Ecologie. Aix Marseille Université, 2018. Français. ⟨tel-01839376⟩
- Romain David, Laurence Mabile, Alison Specht, Sarah Stryeck, Mogens Thomsen, et al.. FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles. *CODATA Data Science Journal*, Committee on Data for Science and Technology (CODATA), 2020, 19 (32), pp.1-11. ⟨10.5334/dsj-2020-032⟩. ⟨hal-02483307v2⟩
- Daniel Jacob, Romain David, Sophie Aubin, Yves Gibon. Making experimental data tables in the life sciences more FAIR: a pragmatic approach. *GigaScience*, Oxford Univ Press, 2020, 9 (12), pp.giaa144. ⟨10.1093/gigascience/giaa144⟩. ⟨hal-02883355v4⟩
- Romain David, Laurent Bouveret, Lorraine Coché, Pedro Pizzigatti Corrêa, Rorie Edmunds, et al.. Data dictionary cookbook for research data and software interoperability at global scale. *Research Data Alliance Plenary 17 (RDA P17)*, Apr 2021, Edinburg (virtual), United Kingdom. , Research Data Alliance Plenary 17 (RDA P17), Edinburg, remotely, 20-22 april 2021 (Session poster session), 2021, ⟨10.5281/zenodo.4683066⟩. ⟨hal-03214743⟩
- David, Romain, Specht, Alison, O'Brien, Margaret, Wyborn, Lesley, Drummond, Christina, Edmunds, Rorie, Filippone, Claudia, Machicao, Jeaneth, Miyairi, Nobuko, Parton, Graham, Pignatari Drucker, Debora, Shelley Stall, & Niklas Zimmer. (2022). Multilingual Data Challenges in Professionalizing Data Stewardship worldwide (V1.1). RDA 19th Plenary Meeting, Part Of International Data Week, 20–23 June 2022 (RDA Plenary 19th), Seoul, South Korea. Zenodo. https://doi.org/10.5281/zenodo.6588167

Inter-University Research Institute Corporation
National Institutes for the Humanities
Research Institute for Humanity and Nature

TOKYO METROPOLITAN UNIVERSITY

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Laurence Mabile

0000-0002-7724-1721

Data challenges: insight into social epidemiology

Laurence Mabile

CERPOP - Centre for Epidemiology and Research in POPulation Health, Toulouse, France

Université de Toulouse

Université Paul Sabatier TOULOUSE III

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inserm
La science pour la santé
From science to health

CERPOP

PARSEC

Overview of social epidemiology

Social epidemiology is :

- the study of how the social world (social structures, institutions, living conditions and relationships) influences — and in many cases defines — health at the population level.
- Interdisciplinary public health field that overlaps with economics, medical anthropology, medical sociology, health psychology and medical geography

Data usage in practice:

- lots of health data potentially, not enough social data

- traditional methods for data collection (e.g. cohort studies and surveys) become more problematic due to high cost and low response rates or attrition, reuse of existing data has become the common practice…

- Reuse from 1 or several harmonized db: cohort-based studies
- Reuse from multi-sources databases and linkage of individual-level data

Data & databases overview

- Big health & social data is quite fragmented and the information collected is increasingly heterogeneous due to their:
  - **nature:** health & social

    Health data: genomics, physiological, biological, clinical, pharmacy, imaging, medico-economic, epidemiological,

    Social data: psychological, social, cultural, geographical
  - **format:** text, numerical values, signals, 2D and 3D images, genomic sequences, etc.
  - **distribution across** several information systems: healthcare establishments, research laboratories, public databases, etc.
  - **sensitivity:** must comply with GDPR

- Ongoing data standardization:

  For health data, standards are being developed, such as i2b2 (Informatics for Integrating Biology and the Bedside), used to compile all the data collected in biomedical data repositories, which can be queried by researchers via web interfaces.

- Data organization: integrated platforms

  Integrated platforms have been set to match databases and aggregate health data with those of cohorts

  They provide complex computer and statistical programs and algorithms to analyze large volumes of information.

  Ex: Health Data Hub: evolutive registry of health databases with IT services & tools to use them for researcher;

  EDP Santé

Data challenges

Lack of social data

Social data characterizing people's living conditions & ethically collected data on race/ ethnicity are often not collected in many European countries for historical and cultural reasons,

>>> health inequalities and monitoring cannot happen regarding for ex. racial discrimination

Discoverability

Social data are scattered in many different governmental establishments and in disciplinary siloed places

>>> difficult to know what's available and to find it

Databases linkage

Quality quantitative social data available at the individual level (ie. an individual's occupation, education level, income) is often not linked or linkable to quality health data across countries

Multiple reasons:

- non-interoperable datasets,
- regulatory problems, data hogging etc…)
- For cross-countries comparisons: heterogeneity of data from different sources:

  Large structural and lexical data heterogeneity

  heterogenous formats not always easy to use for research purposes

  Heterogenous data quality

Data challenges

Heavy data preparation for access and reuse

Discouraging thick legal processes:

Due to the sensitivity of health & social data and to GDPR, linking together information across multiple data sources require approvals that are subject to long delays that are difficult to align with project schedules and funding timelines.

Heterogeneous unclear data management

No open sharing practice as an established practice: only if mandated by the funder or publisher;

No collective thought at the team level that could help ease some processes

« Tools to facilitate data archiving and script/ code sharing for research teams and simple systems to do this at a team level rather than individually would be useful. For example, a research team space on github »

Priorities

To centralise multi-sources & multi-disciplinary data relevant for social epidemiology

Better visibility of data:

more catalog-like and integrated platforms

Facilitating/simplifying data access processes

Access rules & practicalities

Need to balance both privacy (for the individual) and quality (for research purposes) of linked data is a priority for research in data linkage methods

Dedicated support

- Human resources:

  data steward / managers to help in accessing and pre-processing data

- Material resources:

  Tools to help access & usage

  Training for Data standards/ Storage/ Legal issues/ Data science

Thank you for your attention!

Any questions or information, please:

Laurence.mabile@univt-tlse3.fr

Université de Toulouse

Université Paul Sabatier TOULOUSE III

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inserm
La science pour la santé
From science to health

CERPOP

PARSEC

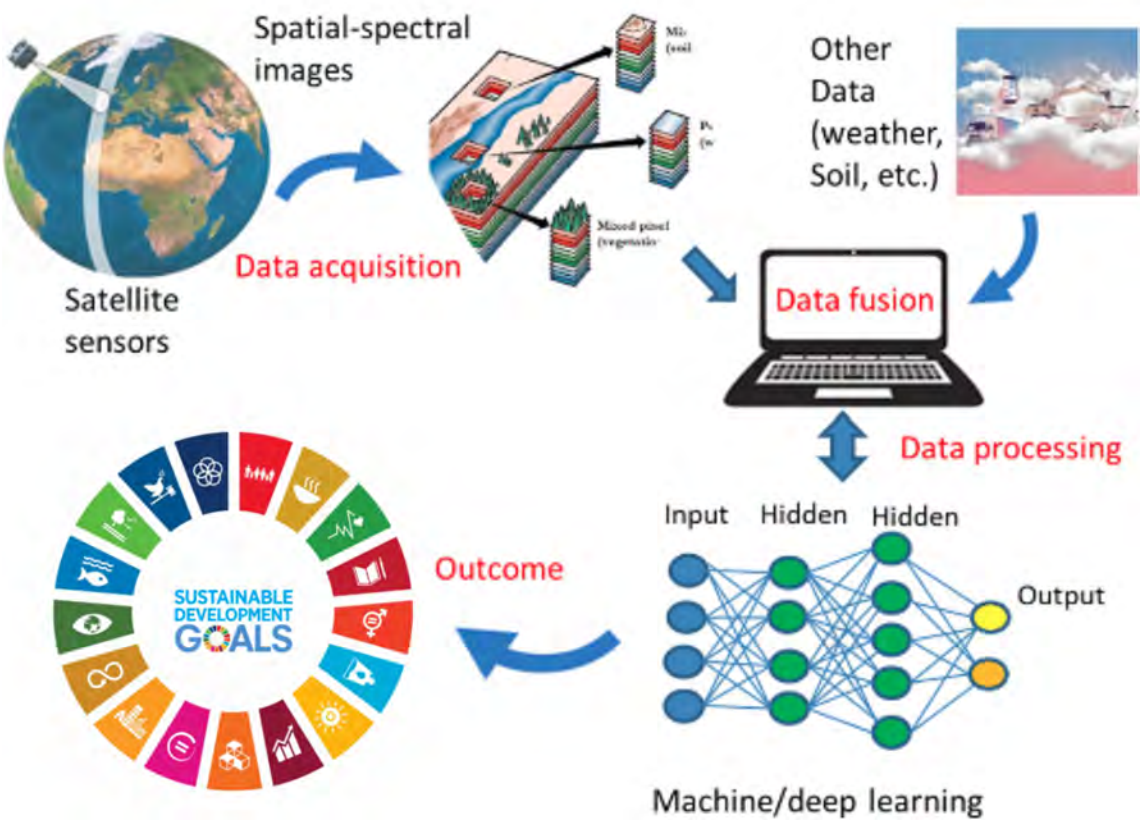
Jeaneth Machicao
0000-0002-1202-0194

Challenges in Machine Learning

(PARSEC: Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas)

Jeaneth Machicao, Ali Ben Abbes

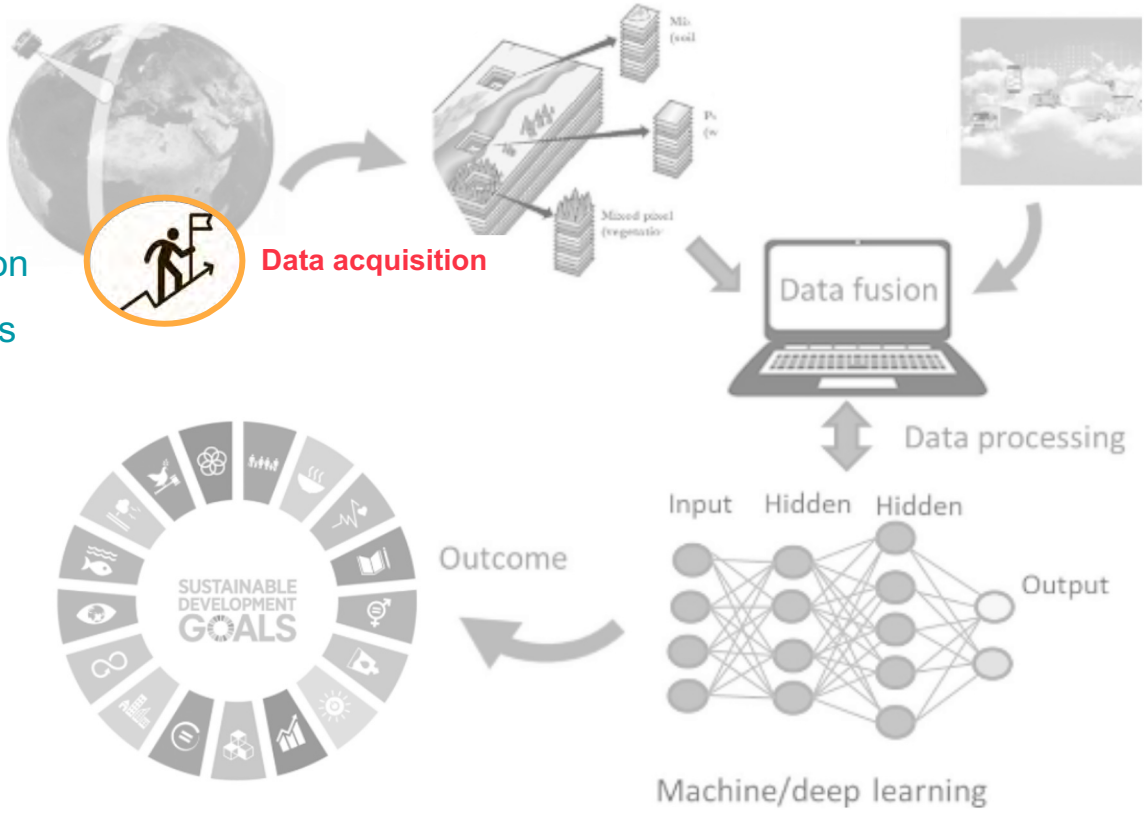
Context:


Spatial-spectral images
Other Data (weather, Soil, etc.)
Data acquisition
Satellite sensors
Data fusion
Data processing
Input
Hidden
Hidden
Output
Machine/deep learning
Outcome
SUSTAINABLE DEVELOPMENT GOALS

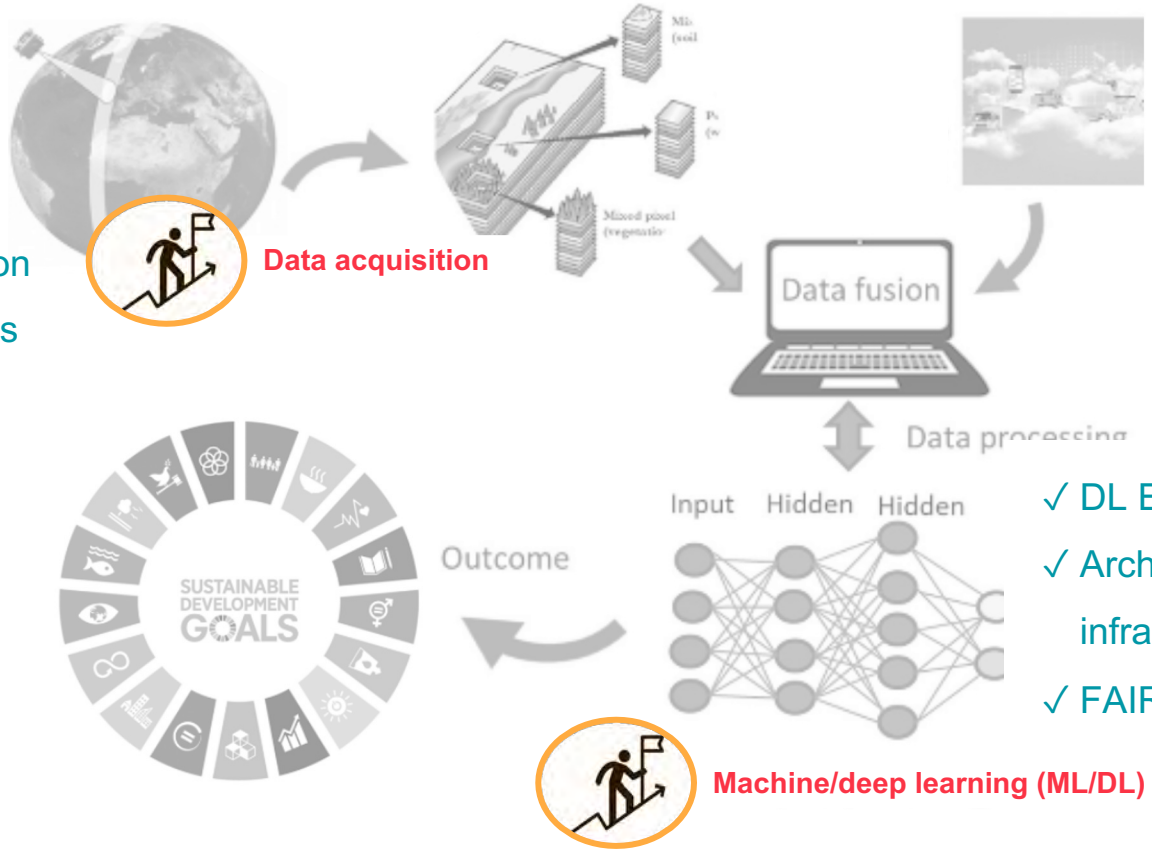
Challenges:

✓ Data acquisition

✓ FAIR principles



Challenges:



✓ Data acquisition

✓ FAIR principles

Data acquisition

✓ DL Experiment

✓ Architecture and infrastructure

✓ FAIR principles

Machine/deep learning (ML/DL)

Past data/code sharing challenges

Data

Data [& metadata] acquisition

- To find publicly available data and to reuse it.
  - e.g. DHS: needs authorization, incomplete on some countries, and metadata is not clear. Remains a limitation.
- Data generated, metadata, checkpoints trained model are also [**output**] **dataset** and **need to be shared as well.**

Past data/code sharing challenges

Data

Data [& metadata] acquisition

- To find publicly available data and to reuse it.
  - e.g. DHS: needs authorization, incomplete on some countries, and metadata is not clear. Remains a limitation.
- Data generated, metadata, checkpoints trained model are also [**output**] **dataset** and **need to be shared as well.**

EO imagery acquisition lacks of custom APIs

- Non standardization yet, dependency on the package, hardware configuration, etc.

Past data/code sharing challenges

Data

Data [& metadata] acquisition

- To find publicly available data and to reuse it.
  - e.g. DHS: needs authorization, incomplete on some countries, and metadata is not clear. Remains a limitation.
- Data generated, metadata, checkpoints trained model are also [**output**] **dataset** and **need to be shared as well.**

EO imagery acquisition lacks of custom APIs

- Non standardization yet, dependency on the package, hardware configuration, etc.

Code

To share ML experiment

- Challenges in describing ML experiments according to FAIR principles.
- **Not all FAIR** principles are **applicable** to **ML**.
  - Findability (F1, F2), Accessibility (A1), Reusability (R1, R1.2), and Interoperability (I3).

Making data/code shareable

- Sharing data and code, not just reusing.

https://github.com/PARSECworld

Reproducibility [& Replicability] of ML experiments

- Open science requires reproducible papers

Past data/code sharing challenges

Data

Data [& metadata] acquisition

- To find publicly available data and to reuse it.
  - e.g. DHS: needs authorization, incomplete on some countries, and metadata is not clear. Remains a limitation.
- Data generated, metadata, checkpoints trained model are also [**output**] **dataset** and **need to be shared as well.**

EO imagery acquisition lacks of custom APIs

- Non standardization yet, dependency on the package, hardware configuration, etc.

Code

To share ML experiment

- Challenges in describing ML experiments according to FAIR principles.
- **Not all FAIR** principles are **applicable** to **ML**.
  - Findability (F1, F2), Accessibility (A1), Reusability (R1, R1.2), and Interoperability (I3).

Making data/code shareable

- Sharing data and code, not just reusing.

https://github.com/PARSECworld

Reproducibility [& Replicability] of ML experiments

- Open science requires reproducible papers

A checklist and guidelines to achieve R & R

(Machicao et al., 2022)

This  and recommended actions can make this possible!

Earth and Space Science

Research Article | Open Access

Mitigation Strategies to Improve Reproducibility of Poverty Estimations From Remote Sensing Images Using Deep Learning

J. Machicao, A. Ben Abbes, L. Meneguzzi, P. L. P. Corrêa, A. Specht, R. David, G. Subsol, D. Vellenich, R. Devillers, S. Stall, N. Mouquet, M. Chaumont, L. Berti-Equille, D. Mouillot

First published: 06 August 2022 | https://doi.org/10.1029/2022EA002379

3 main data/code sharing priorities for the future

- **Standardization** of DL experiments [using EO imagery]
  - Promote reproducibility with community guidelines (Machicao et al., 2022).
- **Simplifying** data management for teams and researchers
  - DDOMP is a good start! (Stall et al., 2022)*
- **Framework** to create end-to-end **reproducible** ML experiments
  - Supporting researchers in data acquisition, processing, training experiments until reporting aiming reproducibility.

* https://doi.org/10.5281/zenodo.4910115

Inter-University Research Institute Corporation
National Institutes for the Humanities
Research Institute for Humanity and Nature

TOKYO METROPOLITAN UNIVERSITY

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Pedro Correa
0000-0002-8743-4244

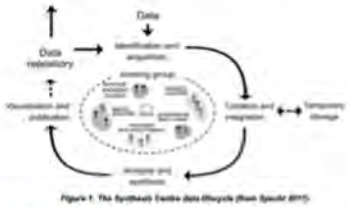
Challenges in data suitability for use in Artificial Intelligence/Data Science Experiments

(PARSEC: Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas)

Pedro Luiz Pizzigatti Corrêa
Escola Politécnica da Universidade de São Paulo

Why is a challenge to manage the Data Science/Artificial Inteligence Experiments ?

- PARSEC Data and Digital Output Management Plan and Workbook
  - https://zenodo.org/record/3891427#.YUurmyZv9Gp
- Current way of organizing and managing the datasets: Excel workbook
  - https://docs.google.com/spreadsheets/d/1fVC_IU35tMZ1ZNpRz0p



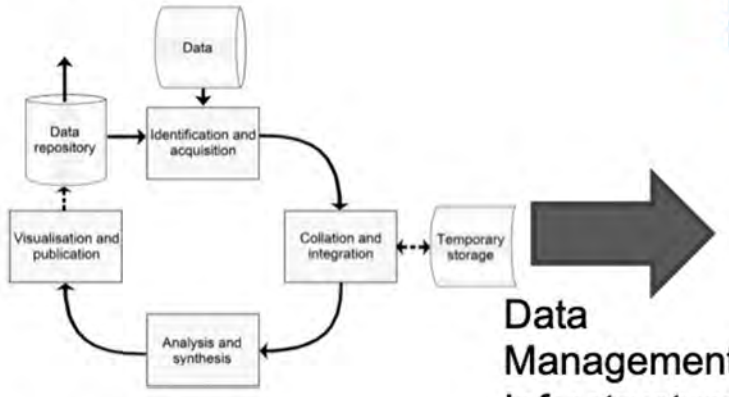
Is this really necessary?

In order to ensure compliance with policies and procedures described in the PARSEC DDOMP document so as the community standards and best practices, a tool that could **automate**, organize and **manage** the **datasets**, attributes, **metadata** and rules (especially with a huge amount of data) it would be helpful and welcome. Not to mention that it would significantly increase efficiency and productivity as well as reduce the chances of human error when dealing with spreadsheets manually.

In addition to these main reasons, we have other features in using a tool to manage datasets such as API Integrations, privacy & security, multilingual support that could help to avoid communication noise in terms of definitions (eg: synonyms with different semantics), user friendly interface and better user experience.

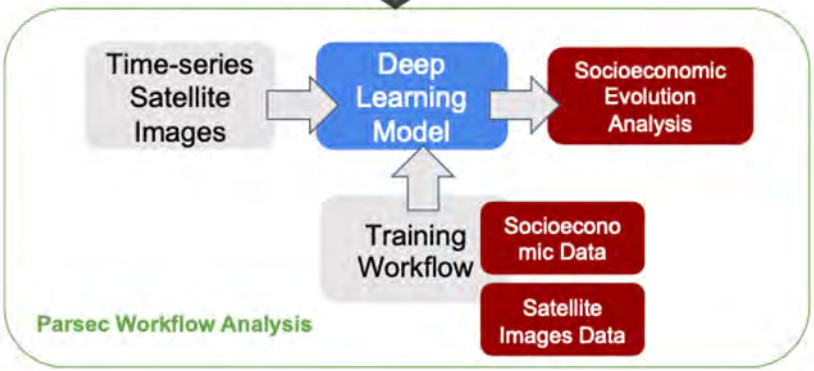
Data Science/AI Experiments in PARSEC

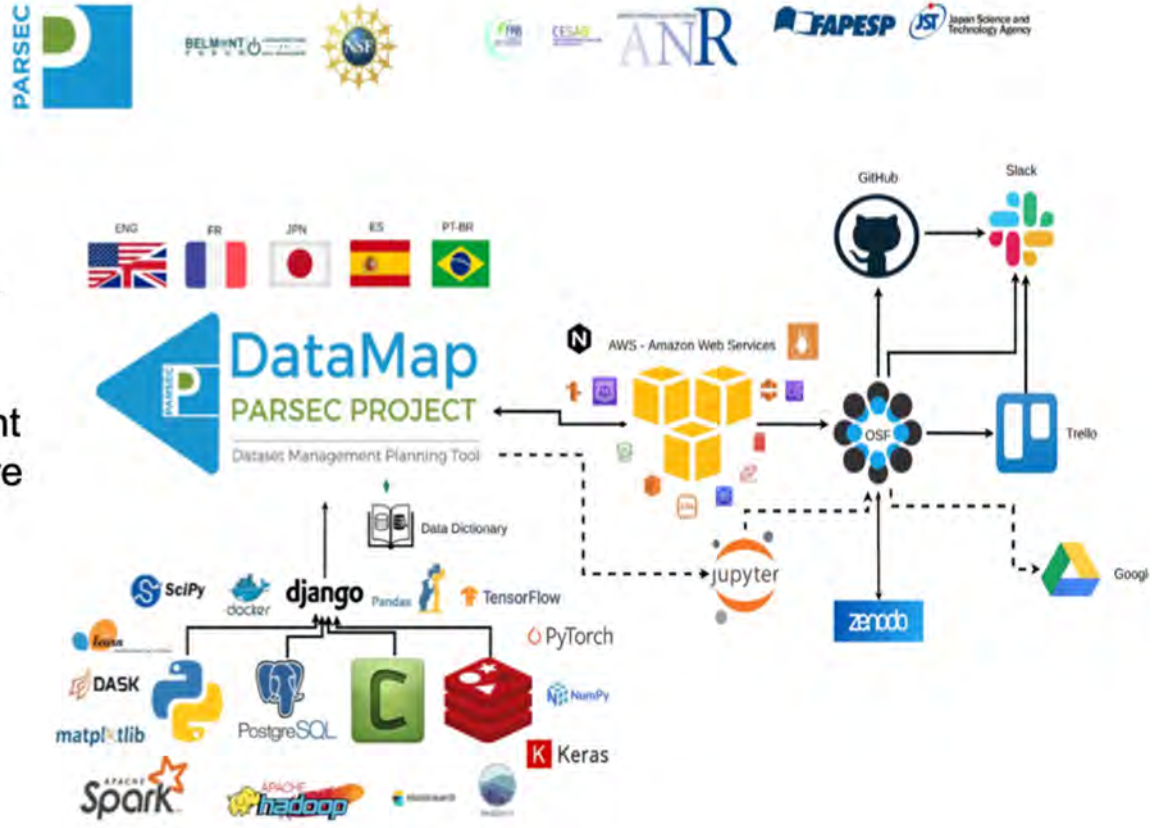
Synthesis Model (Spetcht, 2015)



Data Management Infrastructure

Workflow of AI





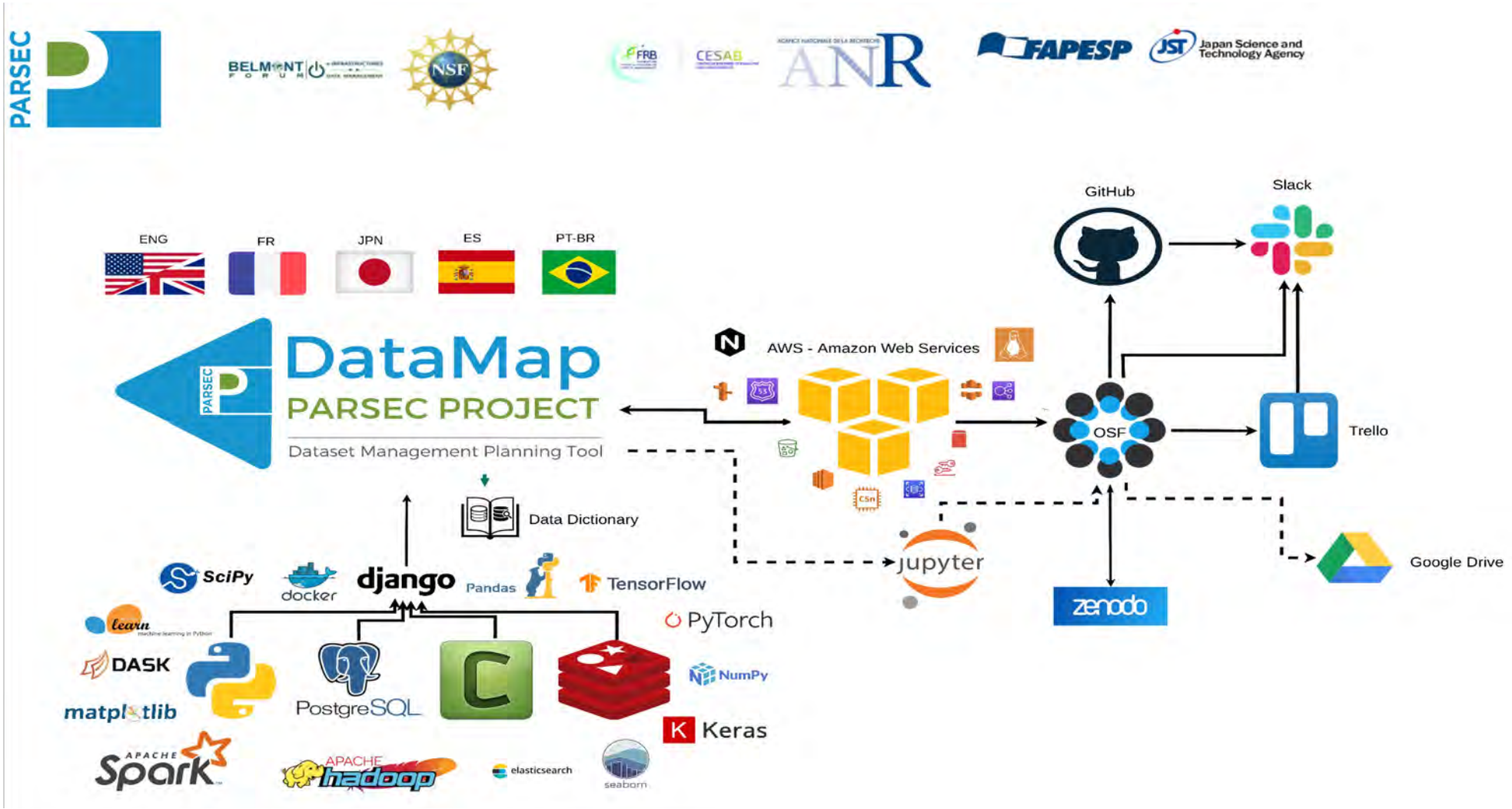
Challenges:

- Multidisciplinary teams involving areas of knowledge in Computing and Information Science: be transparent in methods and data plataforms – all Project Team must at the same page – (trainning/workshops)

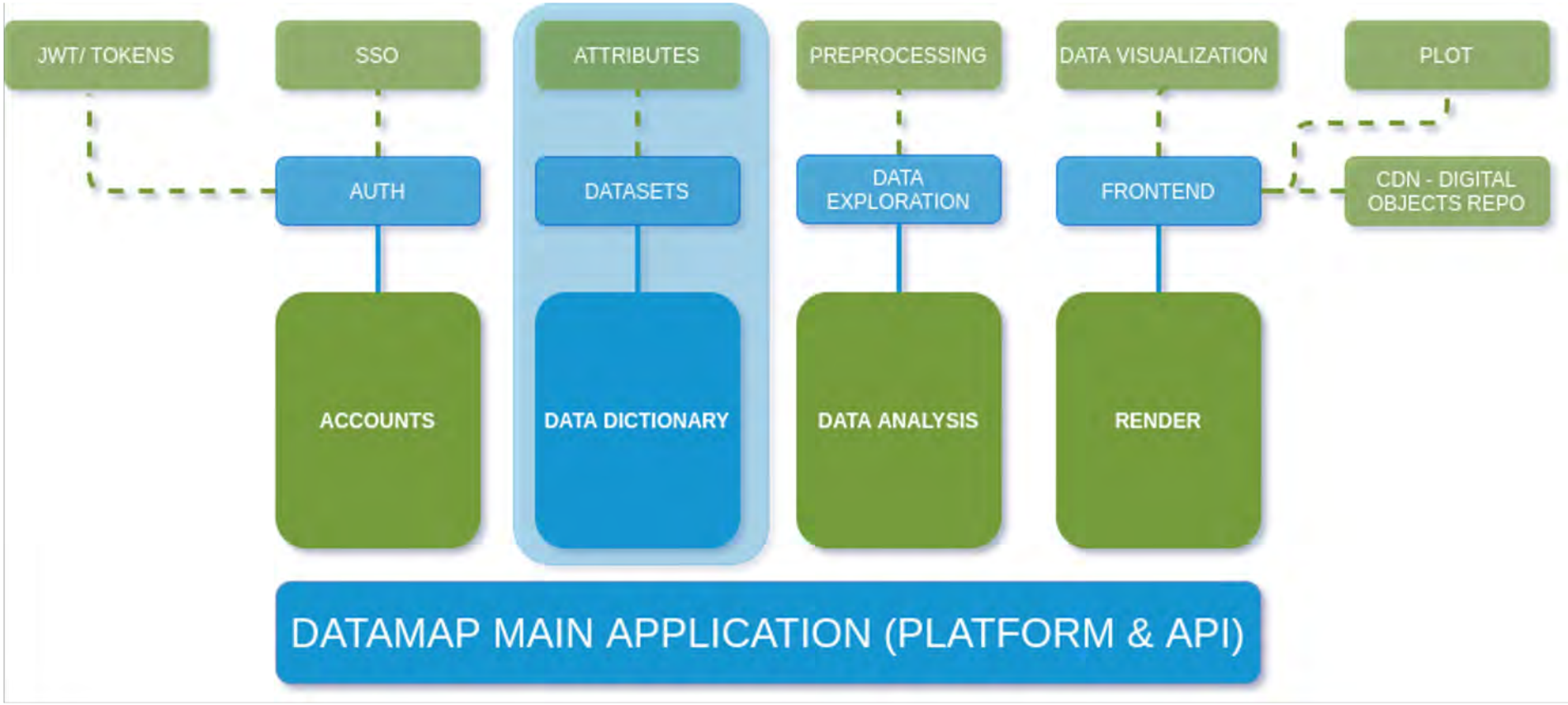
- Data plataforms that support the integration between Data Science Experiments and the Project Management (no silos);

- Data plataform must be flexible/prepared with the Open Science inatives (service computing approach).

Our approach

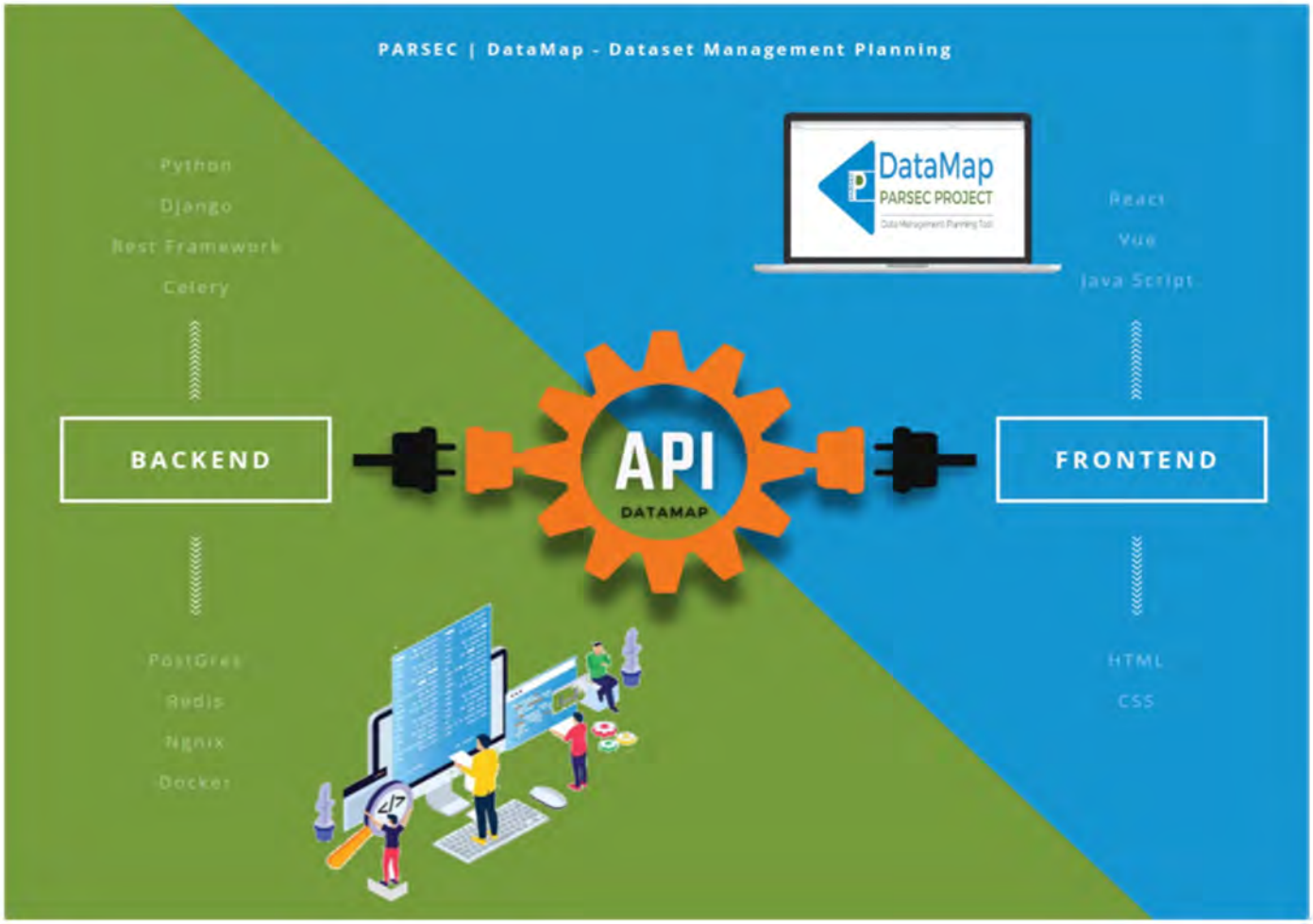


SIMPLE AND GENERIC VIEW OF DATAMAP STACK



DATAMAP Isn't just a single tool but a platform which hosts tools. So, the Data Dictionary is a tool hosted by datamap.

BASIC TOP LEVEL VIEW



3 main priorities for the future

- **Data Infrastructure that supports Data Science/IA Experiments**
  - Research initiatives of Data Infrastructure Services that support all steps of a Data Science Experiment in the Open Science Framework
- **Improve the transparency and collaboration of the multidisciplinarity data science teams**
  - Promote training and workshops between different domains of knowledge are nice approaches
- **Framework** to create end-to-end Data Science/IA experiments
  - Supporting researchers from data acquisition, processing, training experiments until reporting, aiming for reproducibility.


NICT
国立研究開発法人
情報通信研究機構
National Institute of Information and Communications Technology
RIHN
Inter-University Research Institute Corporation
National Institutes for the Humanities
Research Institute for Humanity and Nature
TOKYO METROPOLITAN UNIVERSITY
UNIVERSITÉ TOULOUSE III PAUL SABATIER
UNIVERSITE MONTPELLIER
USP
Southern OREGON UNIVERSITY
IRD
Institut de Recherche pour le Développement
FRANCE
cnrs
EDI
DataCite
FIND, ACCESS, AND REUSE DATA
ORCID
SciELO
AGU 100
ADVANCING EARTH AND SPACE SCIENCE
WORLD DATA SYSTEM
marbec
marine biodiversity exploitation & conservation
RDA
RESEARCH DATA ALLIANCE
Powell Center
erinha
European Research Infrastructure on Highly Pathogenic Agents
BELMONT FORUM
The Nature Conservancy
NSF
EOSC-Life
ESIP
PARSEC
FRB
FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ
CESAB
CENTRE FOR THE SYNTHESIS AND ANALYSIS OF BIODIVERSITY
AGENCE NATIONALE DE LA RECHERCHE
ANR
FAPESP
FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO
JST
THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

Lesley Wyborn
0000-0001-5976-4943

Data Challenges in the Geosciences

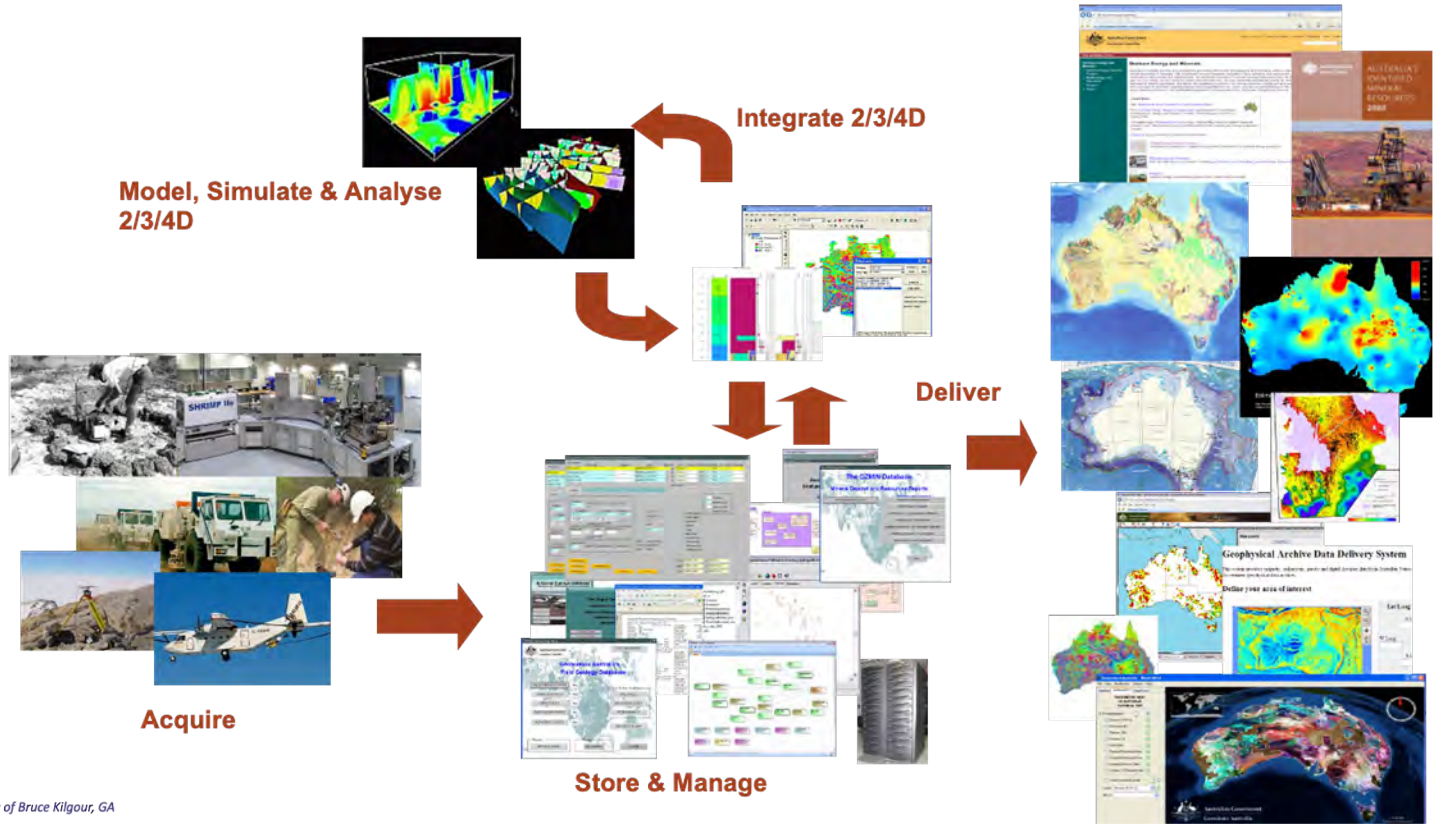
(PARSEC: Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas)

Lesley Wyborn
National Computational Infrastructure, AuScope, ARDC

PARSEC is a project sponsored by the Belmont Forum as part of its Collaborative Research Action (CRA) on Science-Driven e-Infrastructures Innovation (SEI), with funding from FAPESP, the ANR, JST and the NSF, with collaborators from Australia, and support from the synthesis centre CESAB of the French Foundation for Research on Biodiversity.

We acknowledge the Traditional Owners and Custodians of the land and sea in all nations. We honour their profound connections to land, water, biodiversity and culture and pay our respects to their Elders past, present and emerging.

The Diversity of Geoscience Data

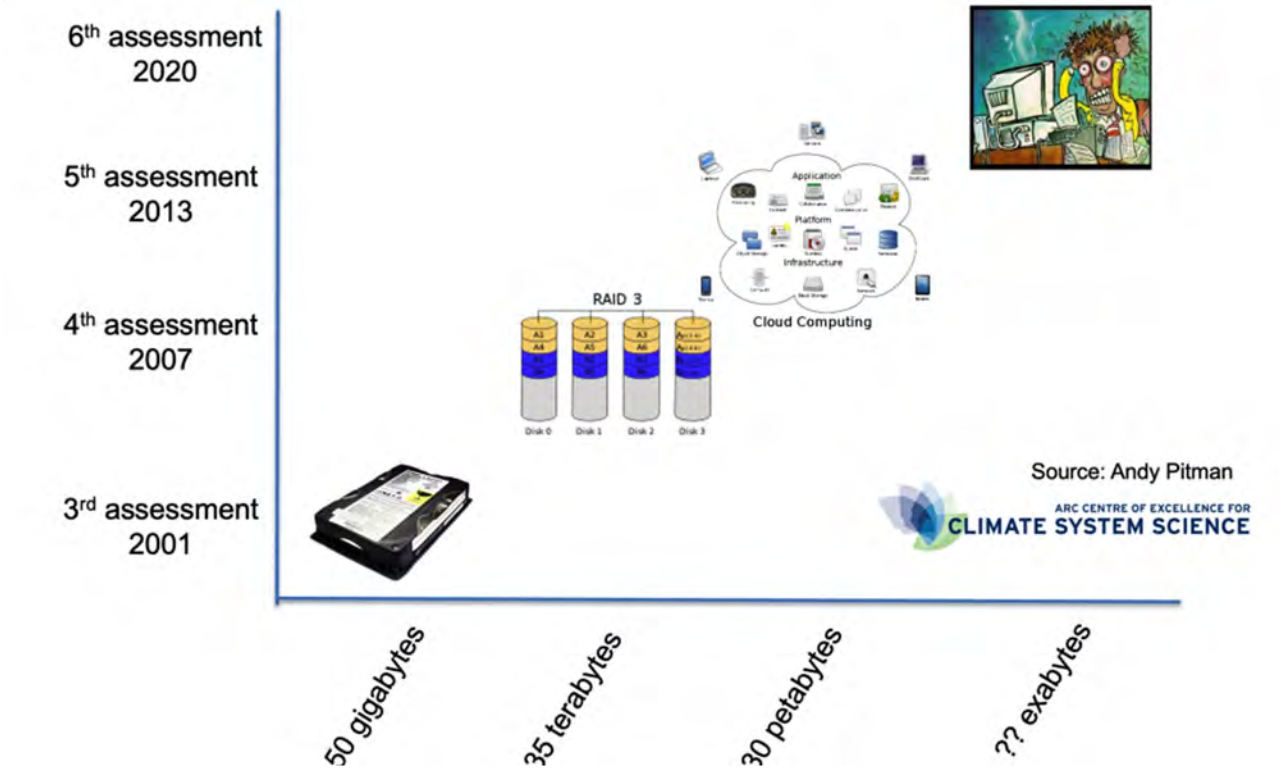


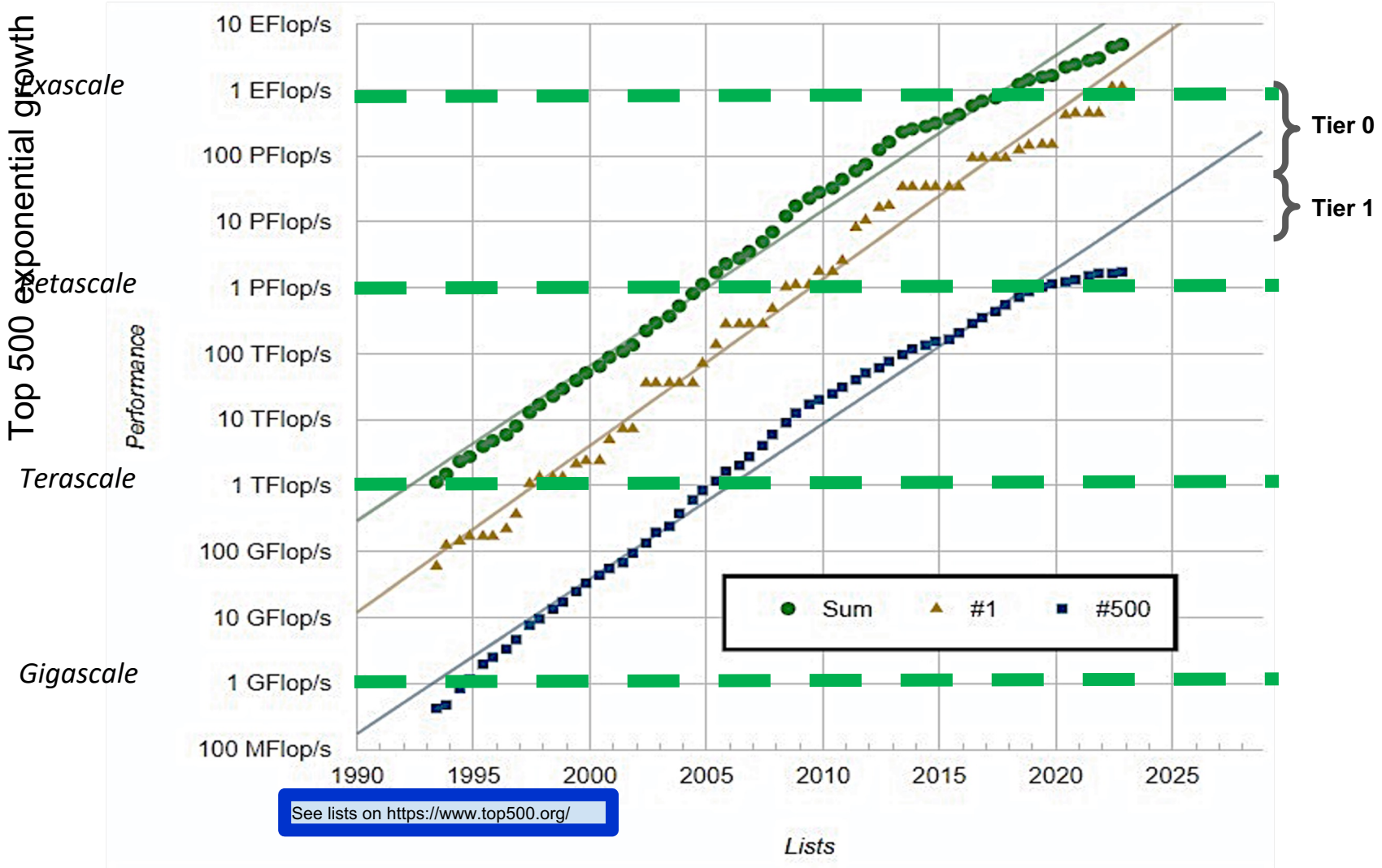
Slide Courtesy of Bruce Kilgour, GA

Geoscience Data Can be of 3 types:

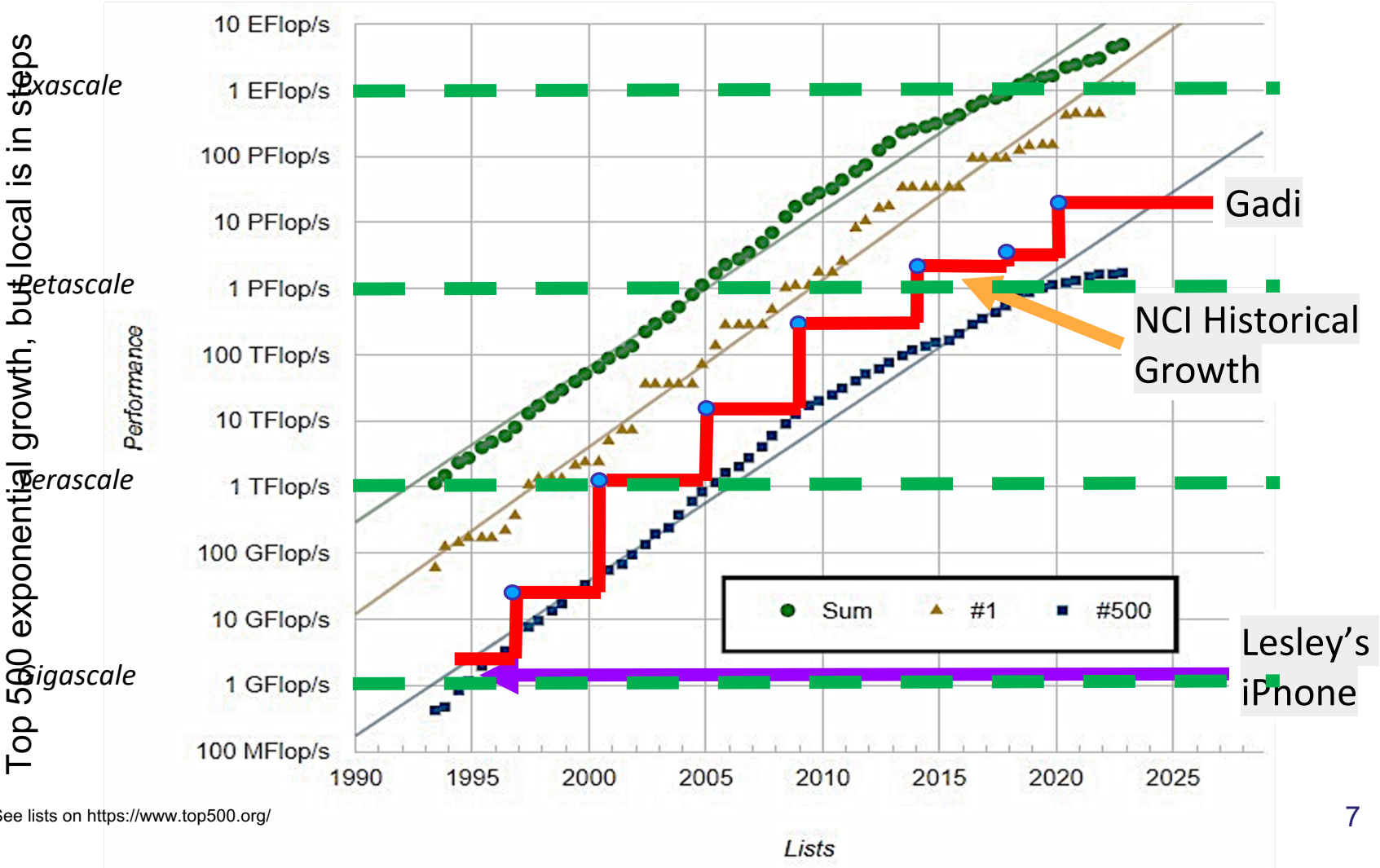
1) **Geological** - can be dominated by qualitative, descriptive science: lack of agreed semantics and vocabularies make interoperability difficult.
2) **Geochemical** - quantitative data, but very long tail, collected by thousands of laboratories and research groups, multiple data types and multiple instrument types for each: no global agreement on standards or minimum variables (hence the new OneGeochemistry Initiative)
3) **Geophysics** - quantitative data, usually well structured, but as instrumentation improves and resolution of surveys collected increases, data volumes are approaching petascale and are hard to manage.

As we scale, can we learn from climate data?



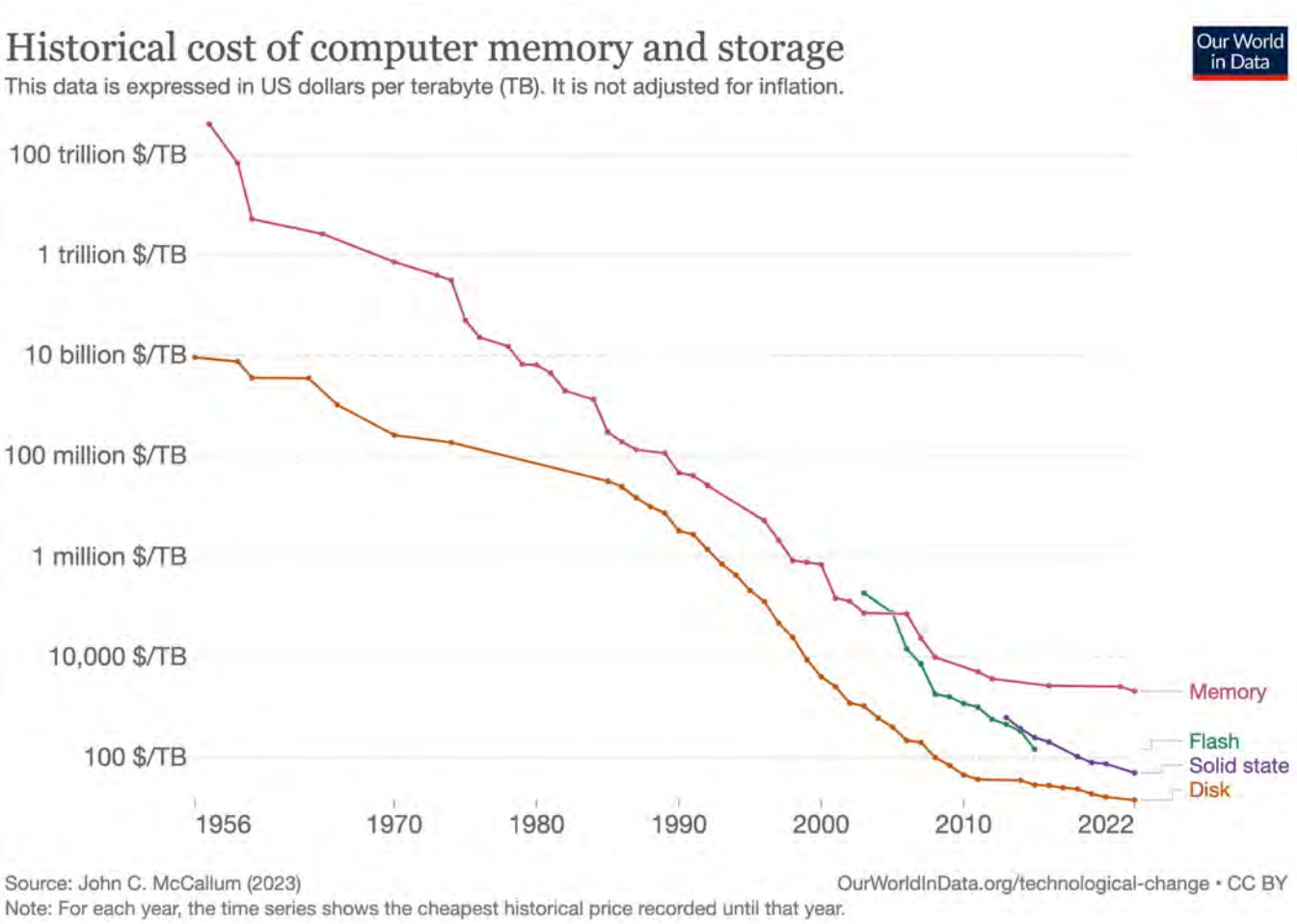

Top 500 exponential growth
Exascale
Petascale
Terascale
Gigascale
Performance
10 EFlop/s
1 EFlop/s
100 PFlop/s
10 PFlop/s
1 PFlop/s
100 TFlop/s
10 TFlop/s
1 TFlop/s
100 GFlop/s
10 GFlop/s
1 GFlop/s
100 MFlop/s
1990
1995
2000
2005
2010
2015
2020
2025
Sum
#1
#500
Tier 0
Tier 1
See lists on https://www.top500.org/
Lists

Top 500 exponential growth, but local is in steps



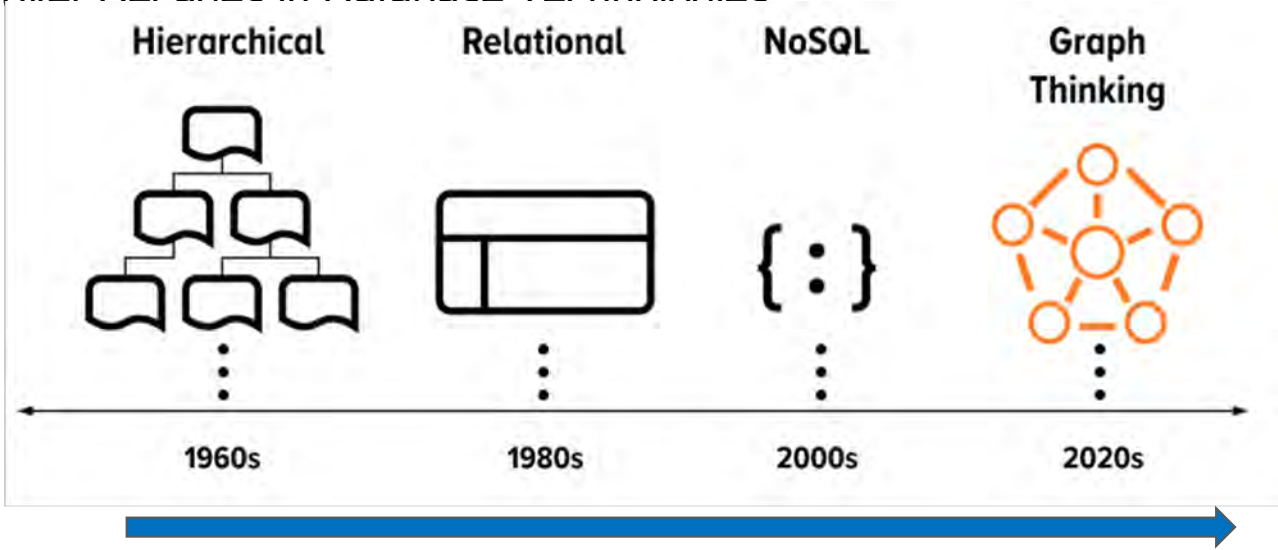
See lists on https://www.top500.org/

Costs in the past Limited What we Stored



https://ourworldindata.org/technological-change

Key Change over Decades in Database Technologies

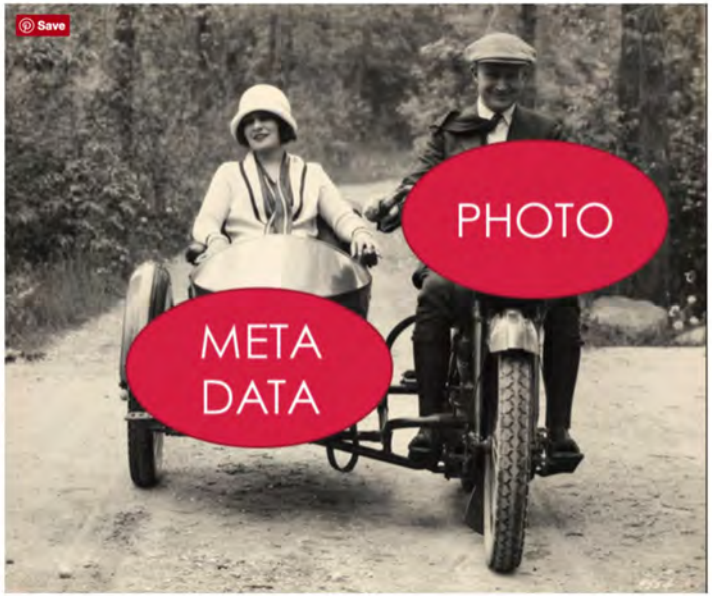


Shift from efficiently managing data to needing to extract value from it.

Gosnell, D, 2020. **Graph Databases Part 1: What does craft beer have in common with database evolution?**

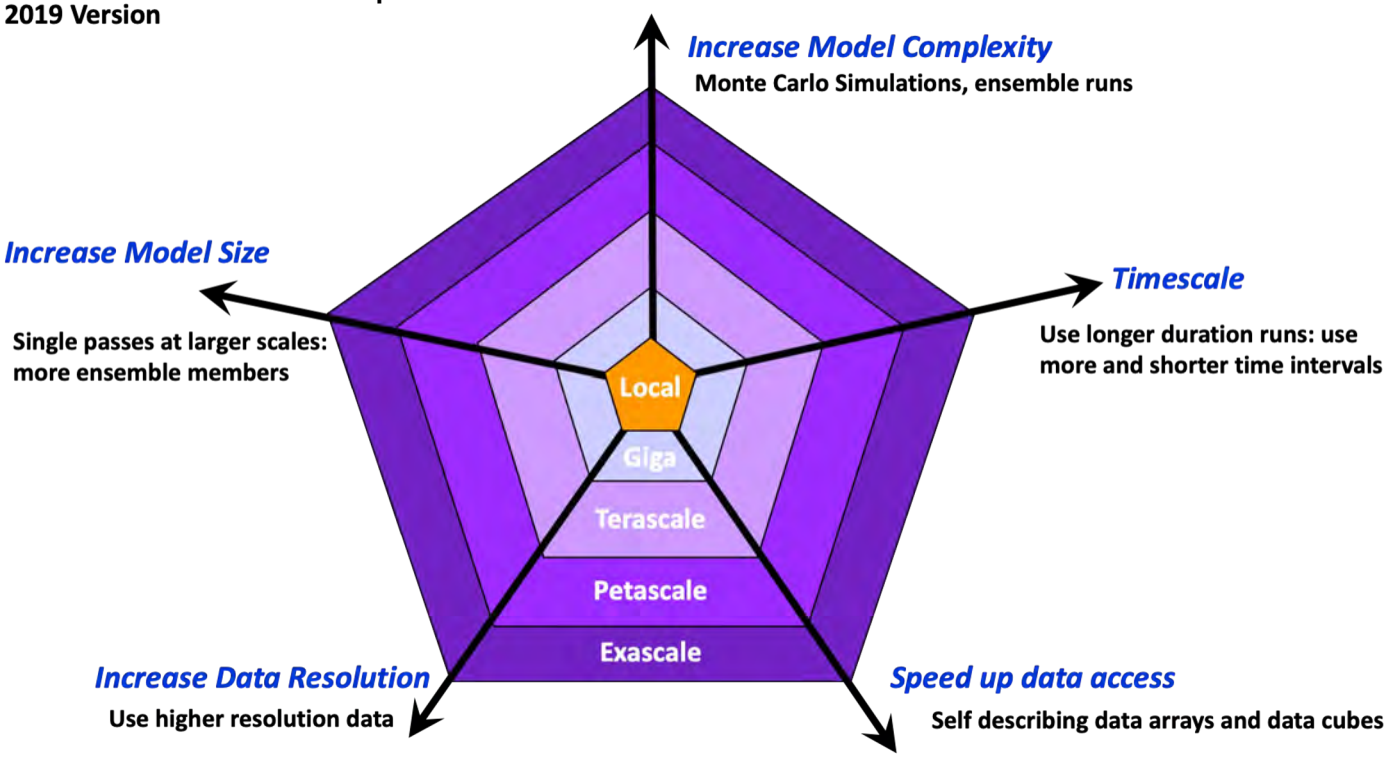
‘Sidecar ‘metadata no longer suitable in FAIR Machine Actionable Data (including AI and ML)

- Rich, self-describing data models that accommodate multi-variate/multidimensional datasets
- Consistent packaging of data with embedded metadata to help facilitate other good data management practices (e.g. versioning, provenance tracking, archiving, citation, etc).
- Internal indexing for rapid subsetting (e.g. line extraction, spatial subsetting, etc.) supporting large-scale parallel processing



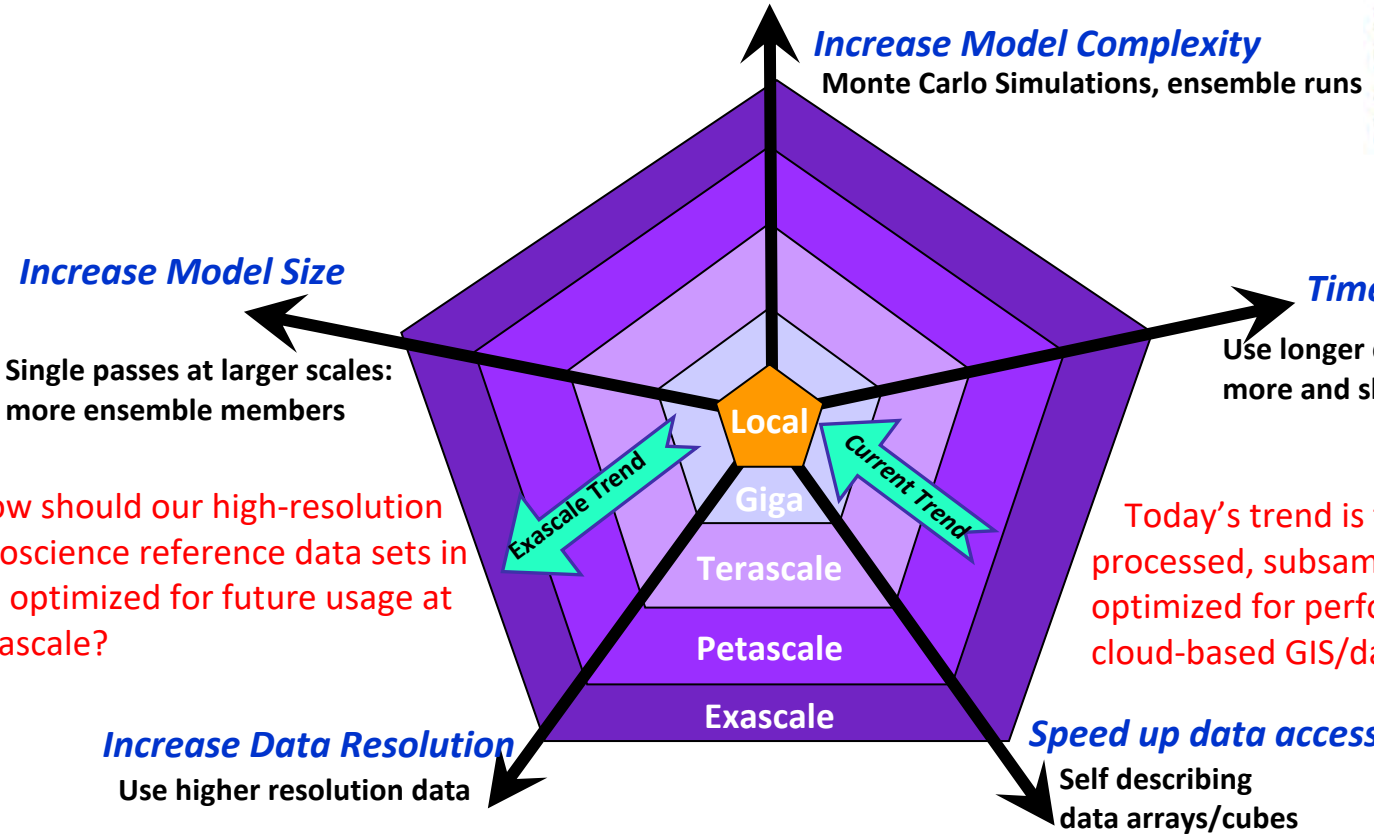
https://www.picturesandstories.com/news/2017/2/13/metadata-writing-on-the-back-of-a-digital-photo

What Does More Computational Power Mean to our Science?



Based on European Climate Computing Environments, Bryan Lawrence (http://home.badc.rl.ac.uk/lawrence/blog/2010/08/02)

Which direction should the geosciences go in?

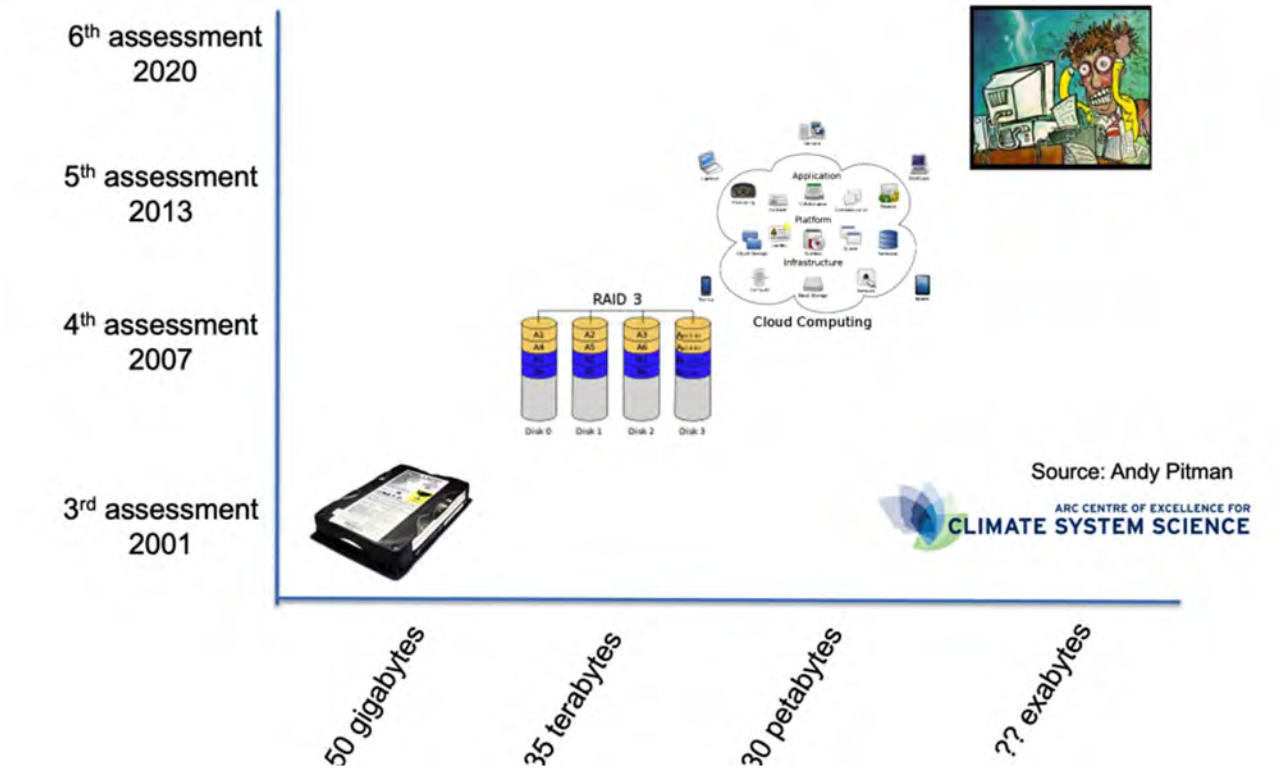


How should our high-resolution geoscience reference data sets in be optimized for future usage at exascale?

Today's trend is to highly processed, subsampled data that is optimized for performance in online, cloud-based GIS/dashboard systems

Based on European Climate Computing Environments, Bryan Lawrence (http://home.badc.rl.ac.uk/lawrence/blog/2010/08/02)

As we scale, can we learn from climate data?



Past Data Problems That Have Been Solved

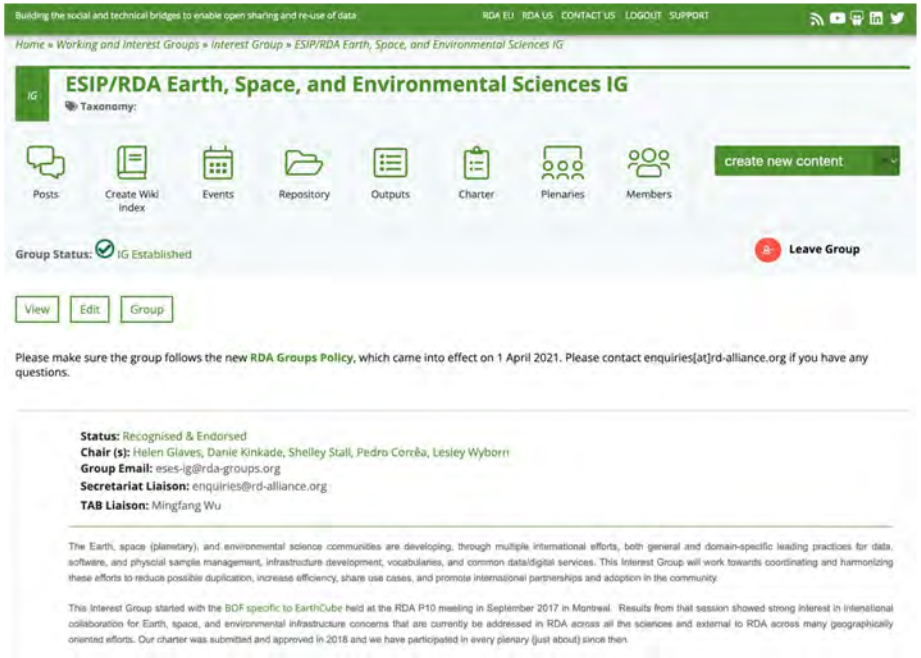
1) Data Management Plans
2) Starting to get researchers to realise that data management needs to be done at the start of the project
3) Better appreciation of what proper data management and data sharing means.

3 challenges that still need to be solved

1) FAIR means for both humans AND machines
2) Standardised vocabularies and semantics (and multilingual)
3) We need to build code and data systems that can be reusable as new technologies and research infrastructures scale and come on line

Introducing the Earth Space and Environmental Sciences Interest Group (ESES-IG)

- Co-Chairs are Shelley Stall (AGU, USA), Pedro Corrêa (Universidade de Sao Paulo, Brazil), Danie Kincaide (Woods Hole, USA), Helen Glaves (BGS, UK), Lesley Wyborn (ANU, Australia).
- Meets to spread the word on ESES projects anywhere.
- We maintain two catalogues:
  - A Data Infrastructure Catalogue of groups building systems anywhere in the world; and
  - A Semantic Resources Catalogue which lists vocabularies, ontologies, semantic models relevant to ESES.
- Please add your projects and resources to these catalogues.



Join us on https://www.rd-alliance.org/groups/esiprda-earth-space-and-environmental-sciences-ig


NICT
国立研究開発法人
情報通信研究機構
National Institute of Information and Communications Technology


RIHN
Inter-University Research Institute Corporation
National Institutes for the Humanities
Research Institute for Humanity and Nature


TOKYO
METROPOLITAN
UNIVERSITY


UNIVERSITÉ
TOULOUSE III
PAUL SABATIER


UNIVERSITÉ
MONTPELLIER


USP


SO
Southern OREGON
UNIVERSITY


IRD
Institut de Recherche
pour le Développement
FRANCE


cnrs


EDI


DataCite
FIND, ACCESS, AND REUSE DATA


ORCID


SciELO


AGU
100
ADVANCING EARTH
AND SPACE SCIENCE


WORLD
DATA SYSTEM


marbec
marine biodiversity
exploitation & conservation


RDA
RESEARCH DATA ALLIANCE


Powell Center


erinha
European Research Infrastructure
on Highly Pathogenic Agents


BELMONT
FORUM


The Nature
Conservancy


NSF


EOSC-Life


SIP


PARSEC


FRB
FONDATION
POUR LA RECHERCHE
SUR LA BIODIVERSITÉ


CESAB
CENTRE FOR THE SYNTHESIS AND ANALYSIS
OF BIODIVERSITY


AGENCE NATIONALE DE LA RECHERCHE
ANR


FAPESP
FUNDAÇÃO DE AMPARO À PESQUISA
DO ESTADO DE SÃO PAULO


JST


THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA


Thankyou

Hayashi Kazuhiro
0000-0003-1996-4259

National Institute of
Information and Communications Technology

Domain variations in open science and open data

(from State of Open Science in Japan)

Kazuhiro Hayashi, Ui Ikeuchi
National institute of Science and Technology Policy, Japan
March 20, 2023

From Bottom-up to Top-down with bird's-eye view

Chemical Society of Japan (1995-2012)
- Journal Manager
- EJ development (with my IT Skill)
- OA implementation
- ALPSP Board Member (2011)

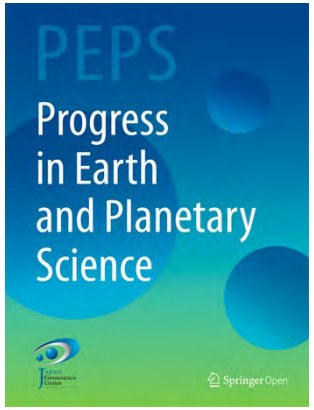
https://onlinelibrary.wiley.com/journal/24750328

https://www.cell.com/patterns/

National Institute of Science and Technology Policy (2012-)
- Open Science policy development

G7 Science and Technology Ministers' Meeting in Tsukuba, Ibaraki

G7 FRANCE



http://progearthplanetsci.org/

Advisory Board Member
Consultation

https://iupac.org/

GAP analysis
Translation
Consultation

OECD
BETTER POLICIES FOR BETTER LIVES

Expert Member, Advisory Committee

Governmental Researcher who directly experienced Publisher, EJ-development and OA-implementation

State of Open Science in Japan

Researchers

Institutions

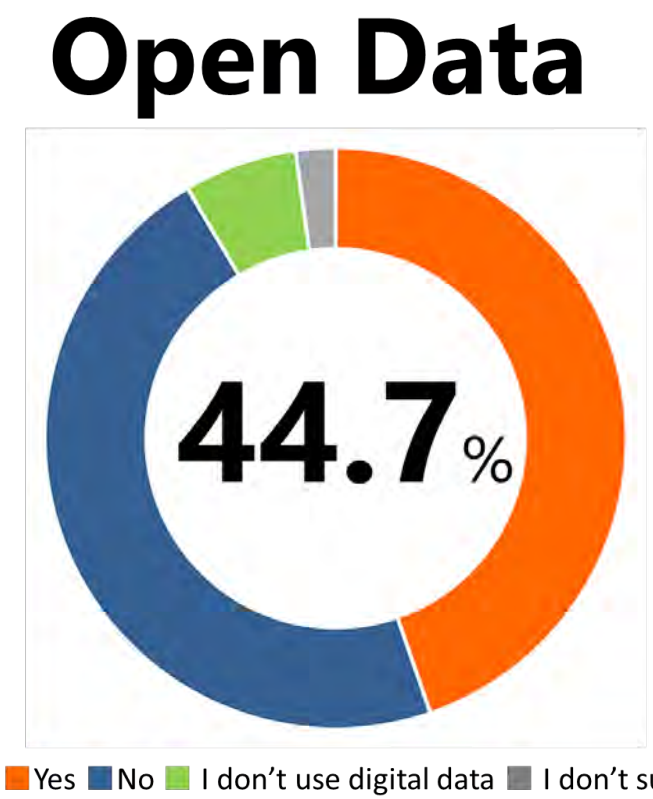
2016	2018	2020	2022
Open Data Open Access	Open Data Open Access	Open Data Open Access	Open Data
		Preprints	Open Access Preprints
RDM Data Policy	RDM Data Policy		
JPCOAR/AXIES	JPCOAR/AXIES		

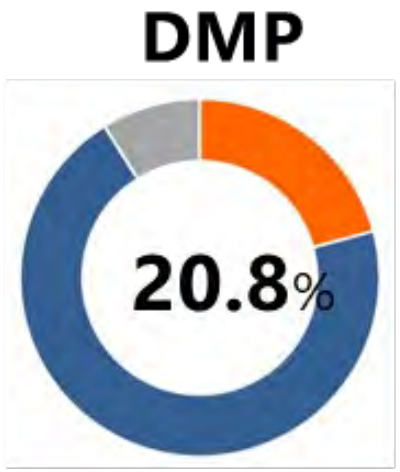
- 5th & 6th Science and Technology Basic Plan
- Integrated Innovation Strategy

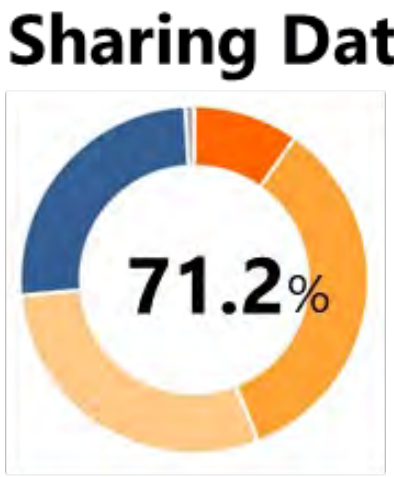
(Monitoring Framework)

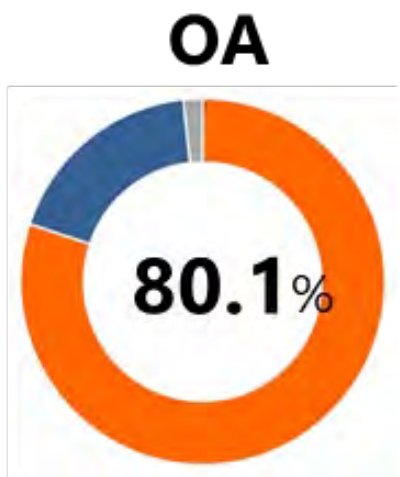
Introduction

The National Institute of Science and Technology Policy (NISTEP) conducted a web-based questionnaire survey in 2020, following on from 2016 and 2018, to identify the status and perceptions of open research data by researchers in Japan. This result shows the survey results from 1,268 respondents (66.2% response rate).



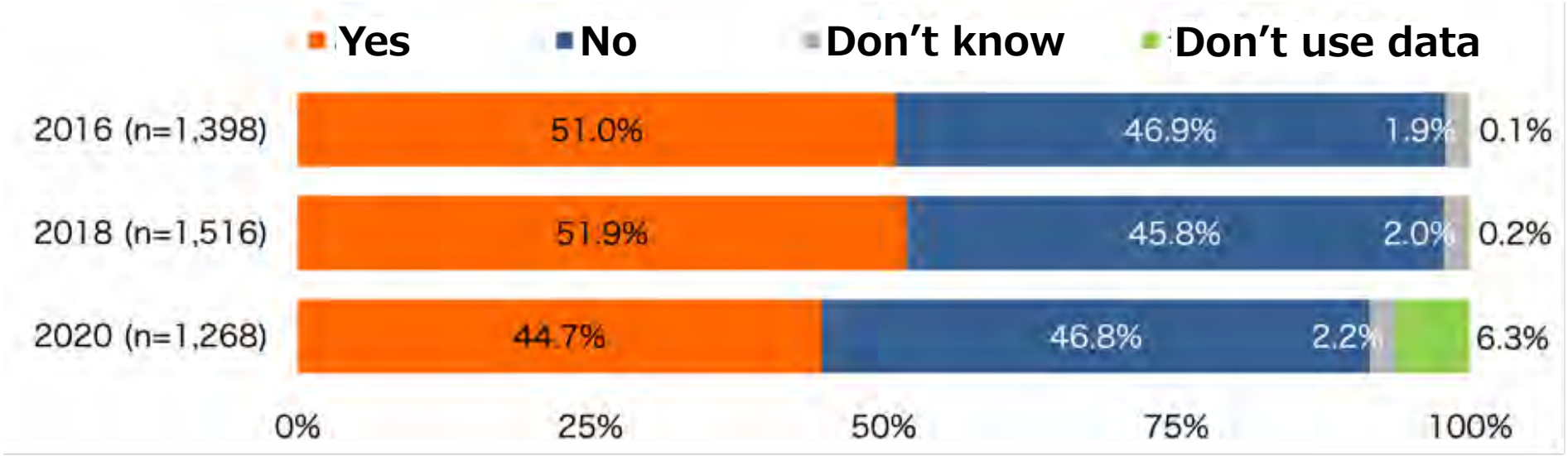




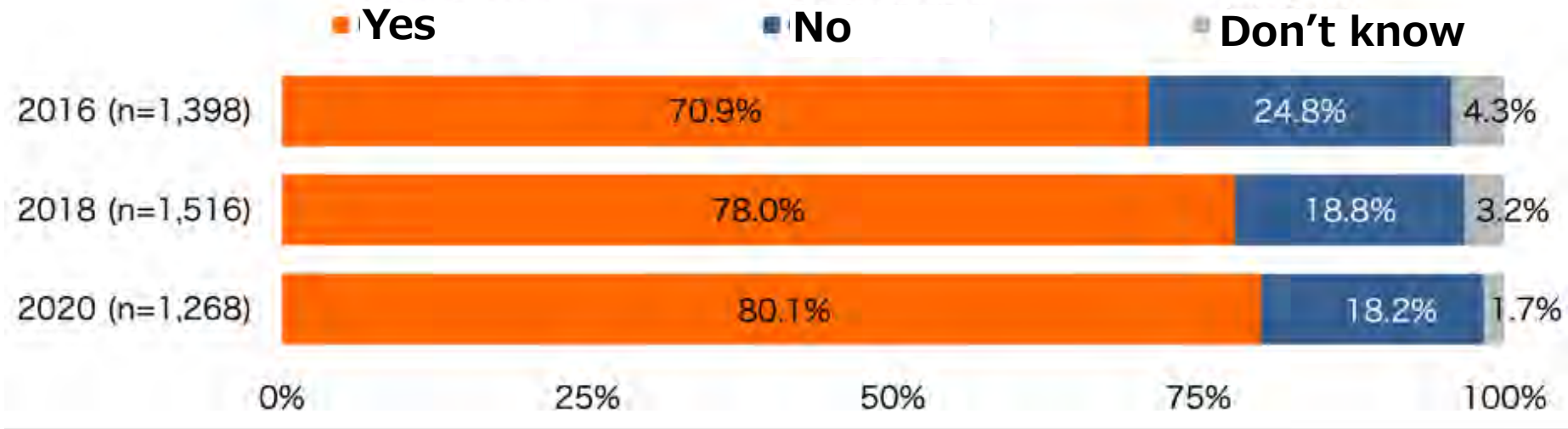


OD is not increasing, though OA is increasing

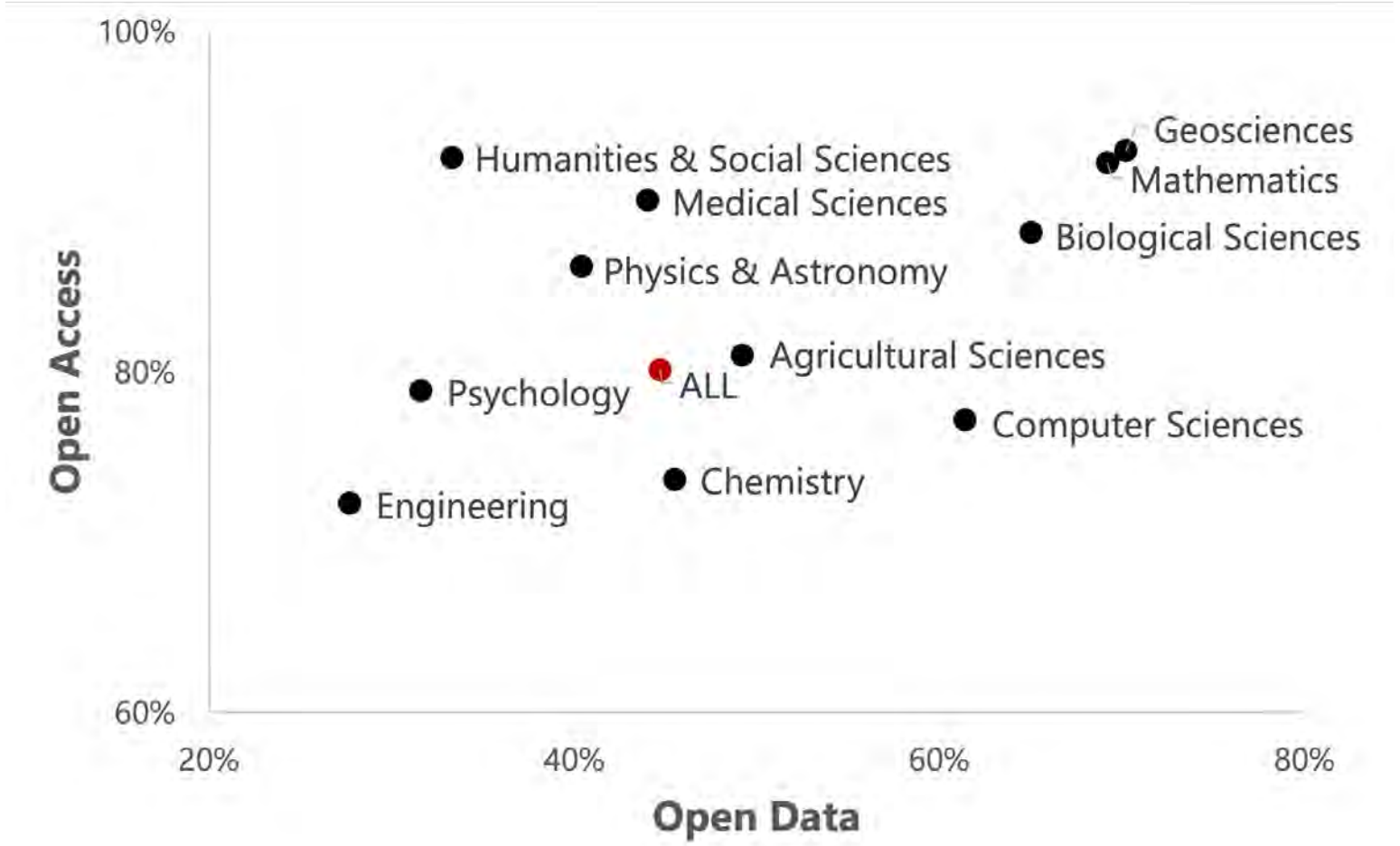
OD



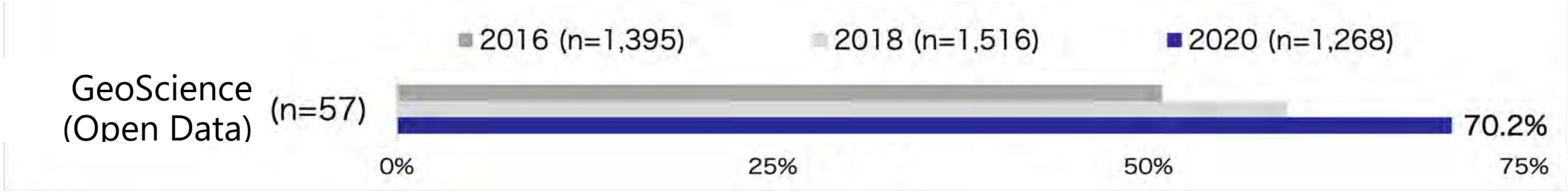
OA



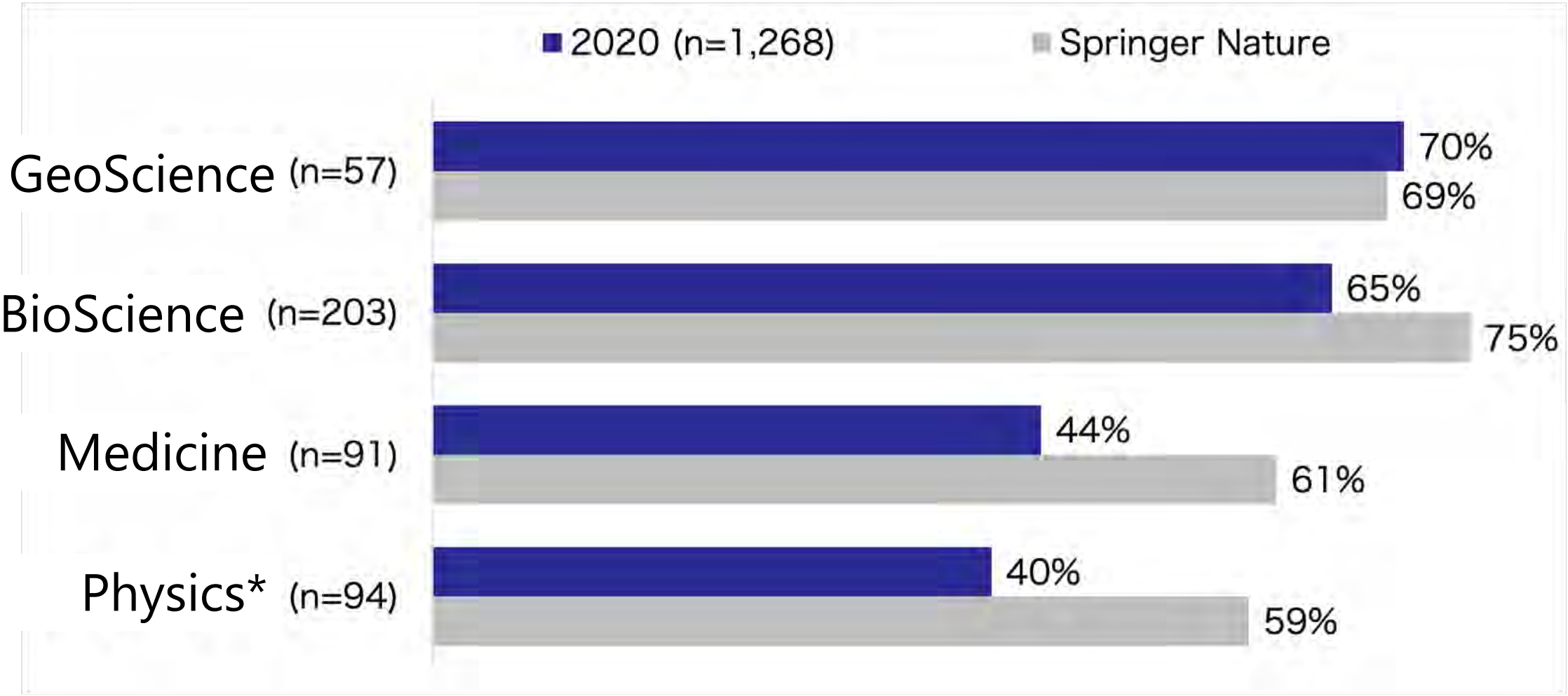
Open Data and OA practices by disciplines



N=57



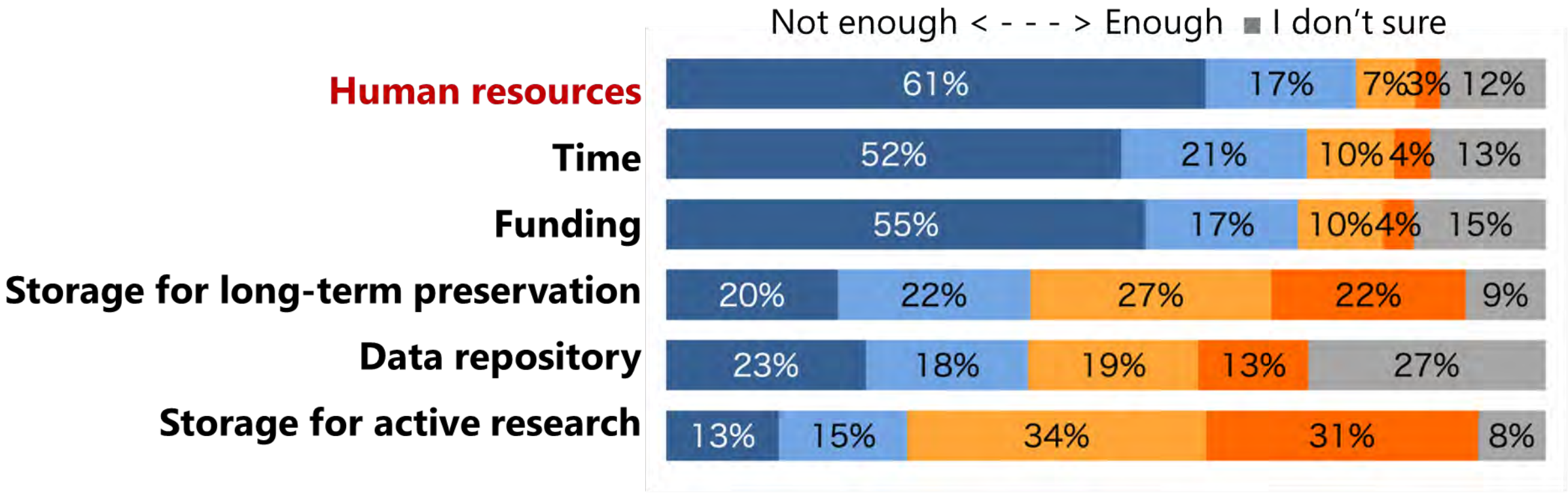
Comparison with other results



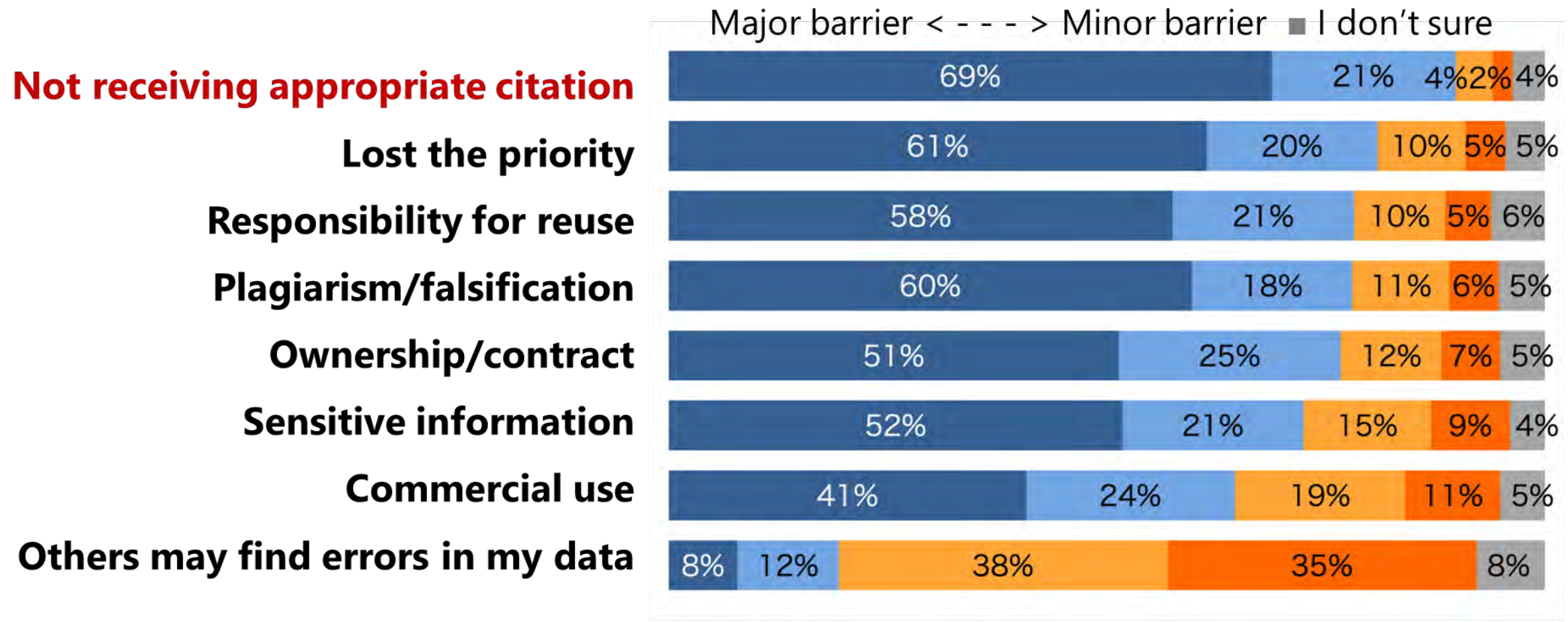
Whitepaper: Practical challenges for researchers in data sharing. Springer Nature. https://doi.org/10.6084/m9.figshare.5996786

*Physics and Astronomy (NISTEP), Physics and Chemical Science (SN)

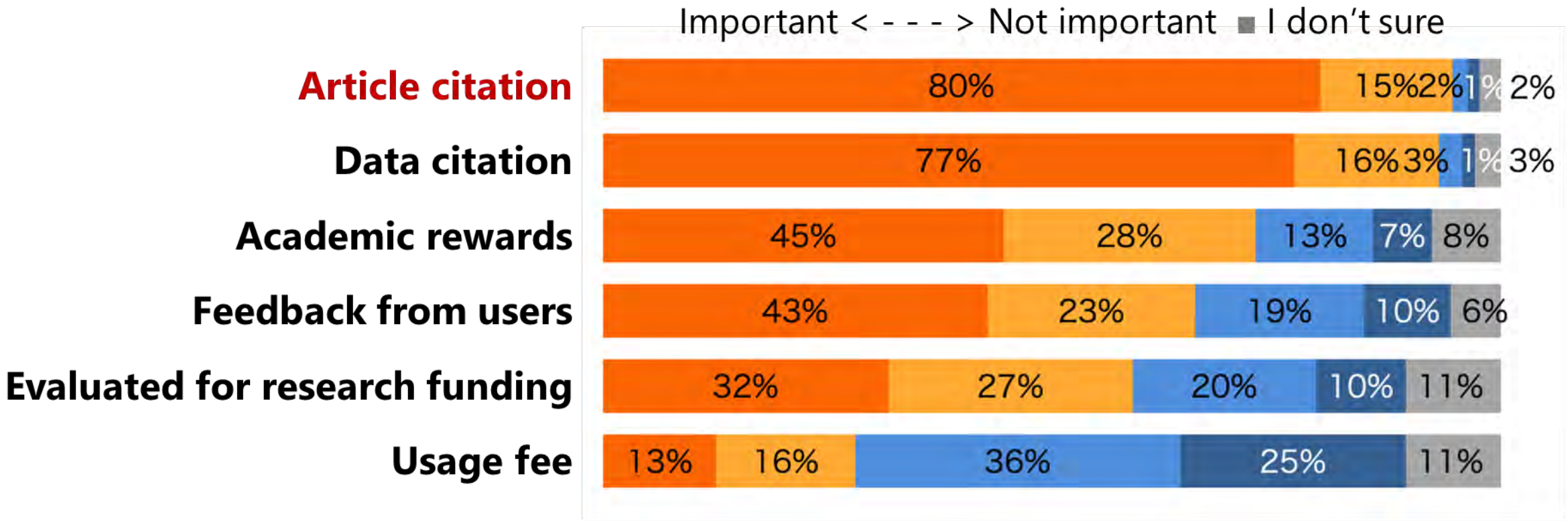
Sufficiency of resources for open data



Barriers to publish data as open data



Importance of open data incentives



11h20	Comments and questions
11h50	Summary and goal setting

Inter-University Research Institute Corporation
National Institutes for the Humanities
Research Institute for Humanity and Nature

TOKYO METROPOLITAN UNIVERSITY

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA


Thankyou