Data Stewardship at the UZH: Investing in Data Quality

Dr. Susanna Weber

Open Science Services, University Library

Susanna.weber@ub.uzh.ch
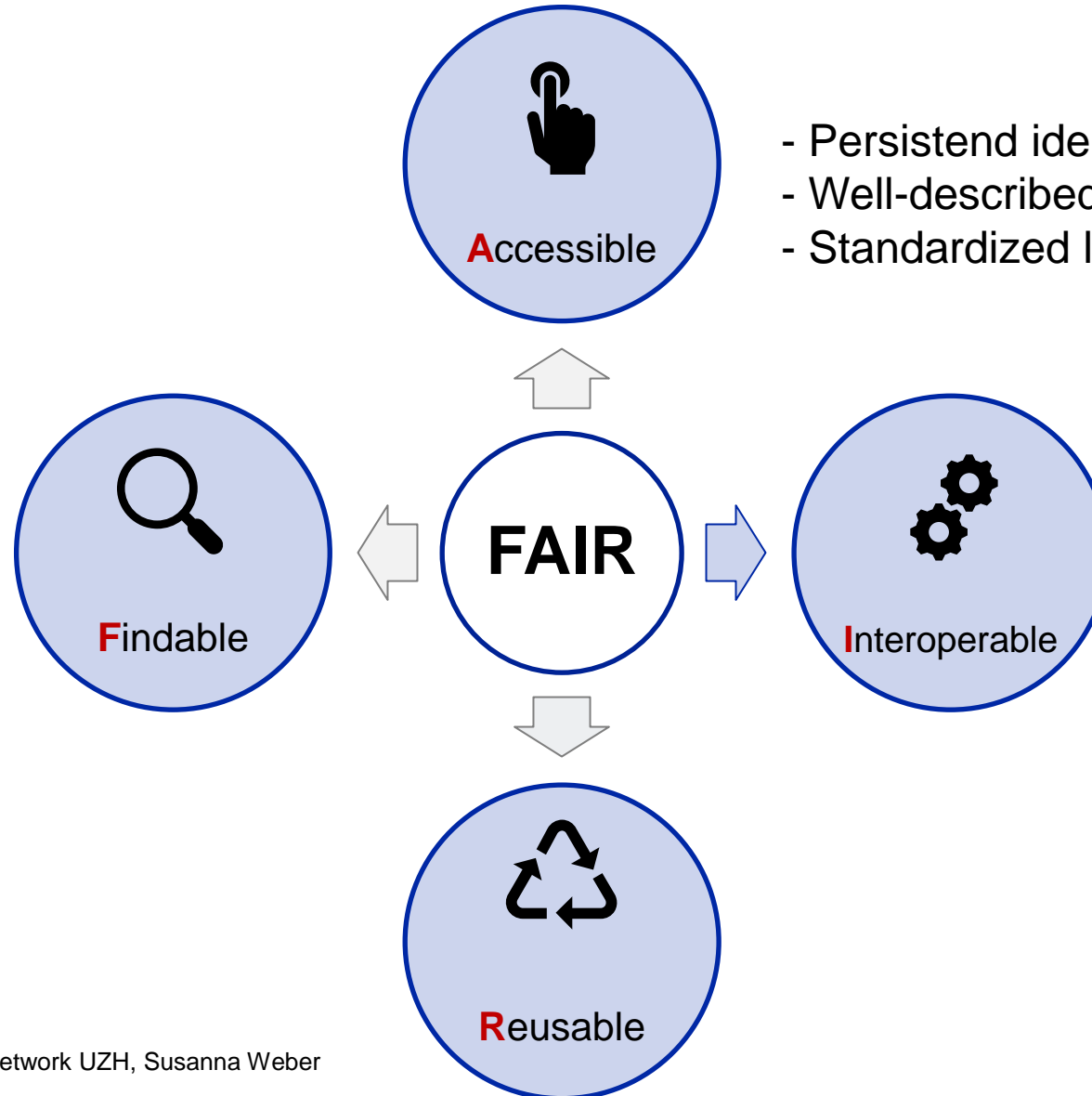
Outline

– FAIR and Open Research: Benefits and Barriers

– Data Stewardship as an Open Science Driver

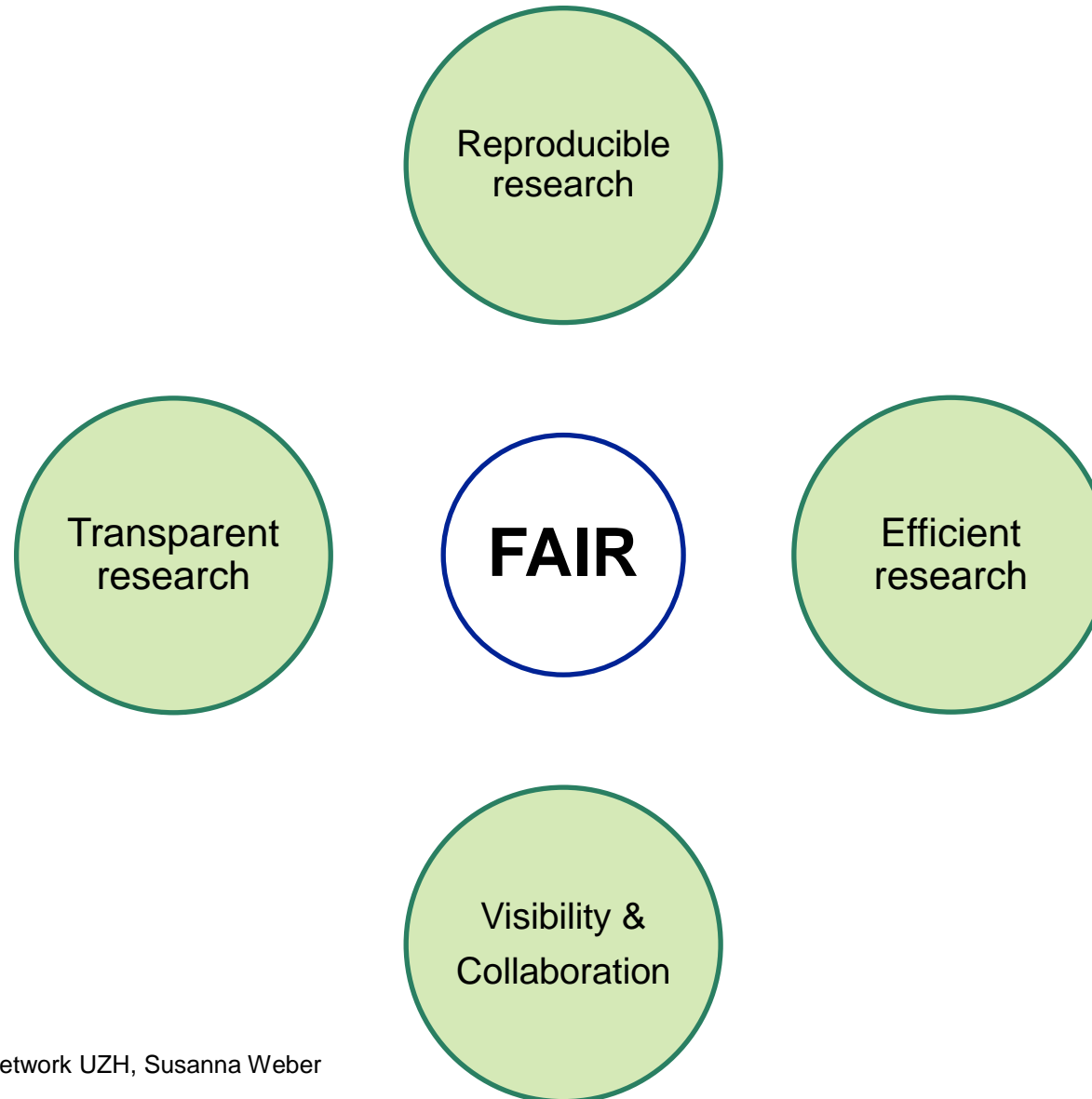– Data Stewardship at the UZH

– Outlook
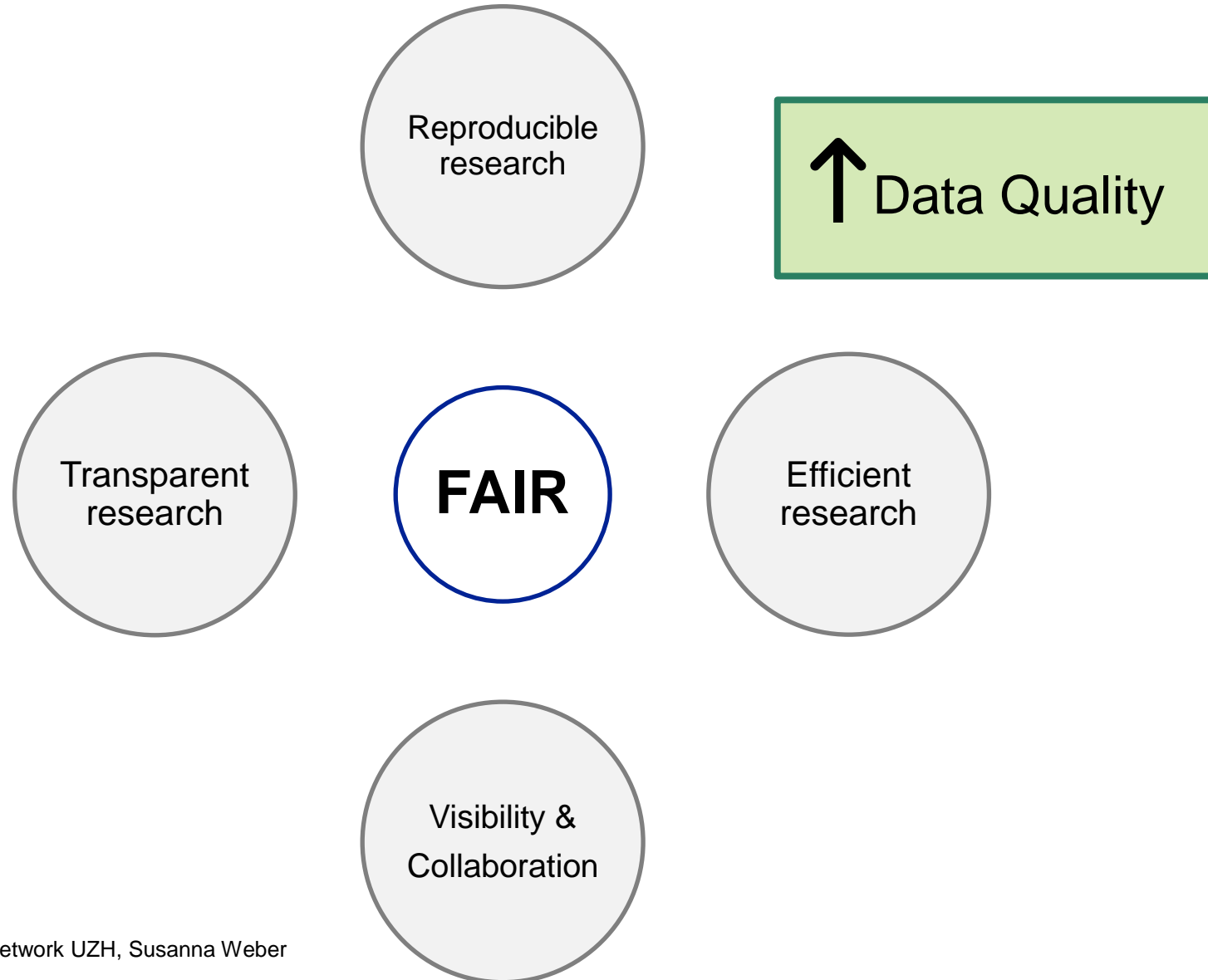
FAIR & Open Research Data

FAIR Data

- Persistend identifiers
- Well-described machine-readable metadata
- Standardized language and formats

FAIR Data

Reproducible research

Transparent research

FAIR

Efficient research

Visibility & Collaboration

FAIR Data

Reproducible research

↑ Data Quality

Transparent research

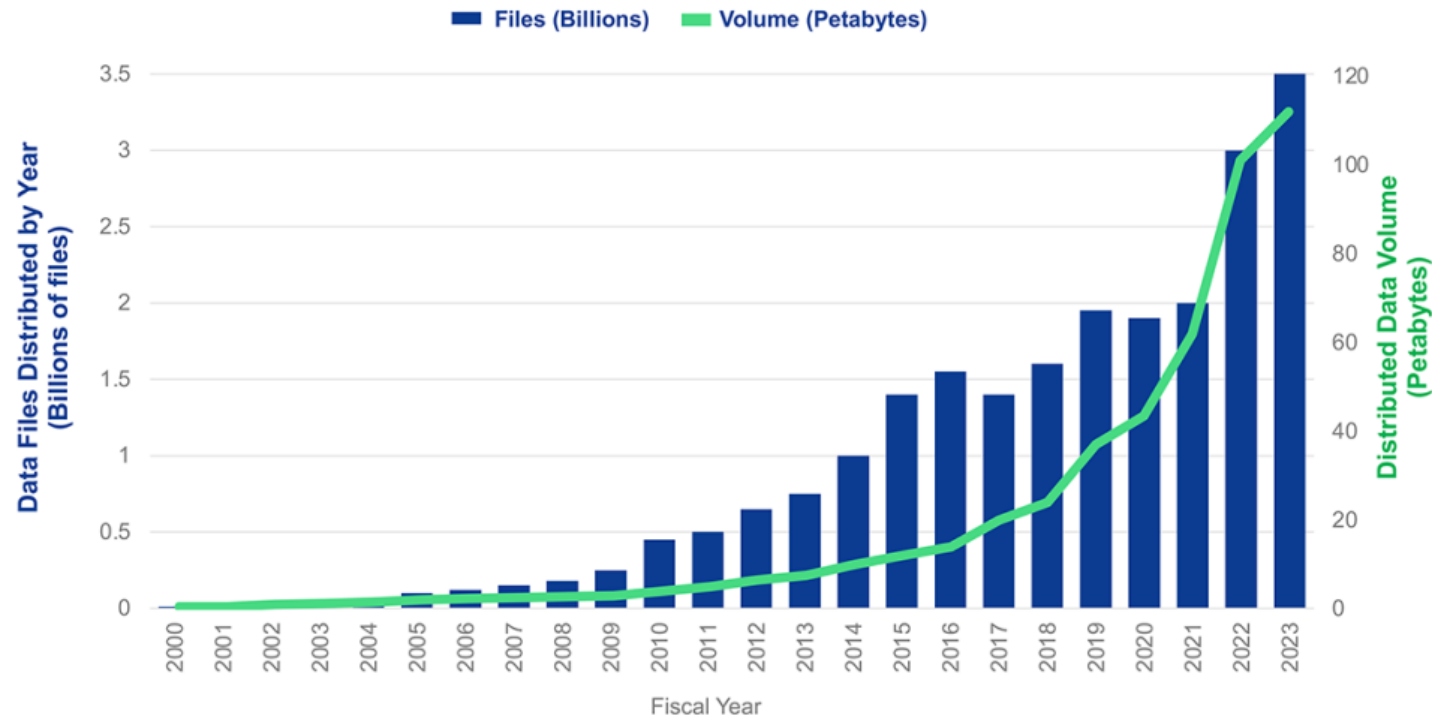**FAIR**

Efficient research

Visibility & Collaboration

FAIR & Open Research : Easy in theory but not in practice
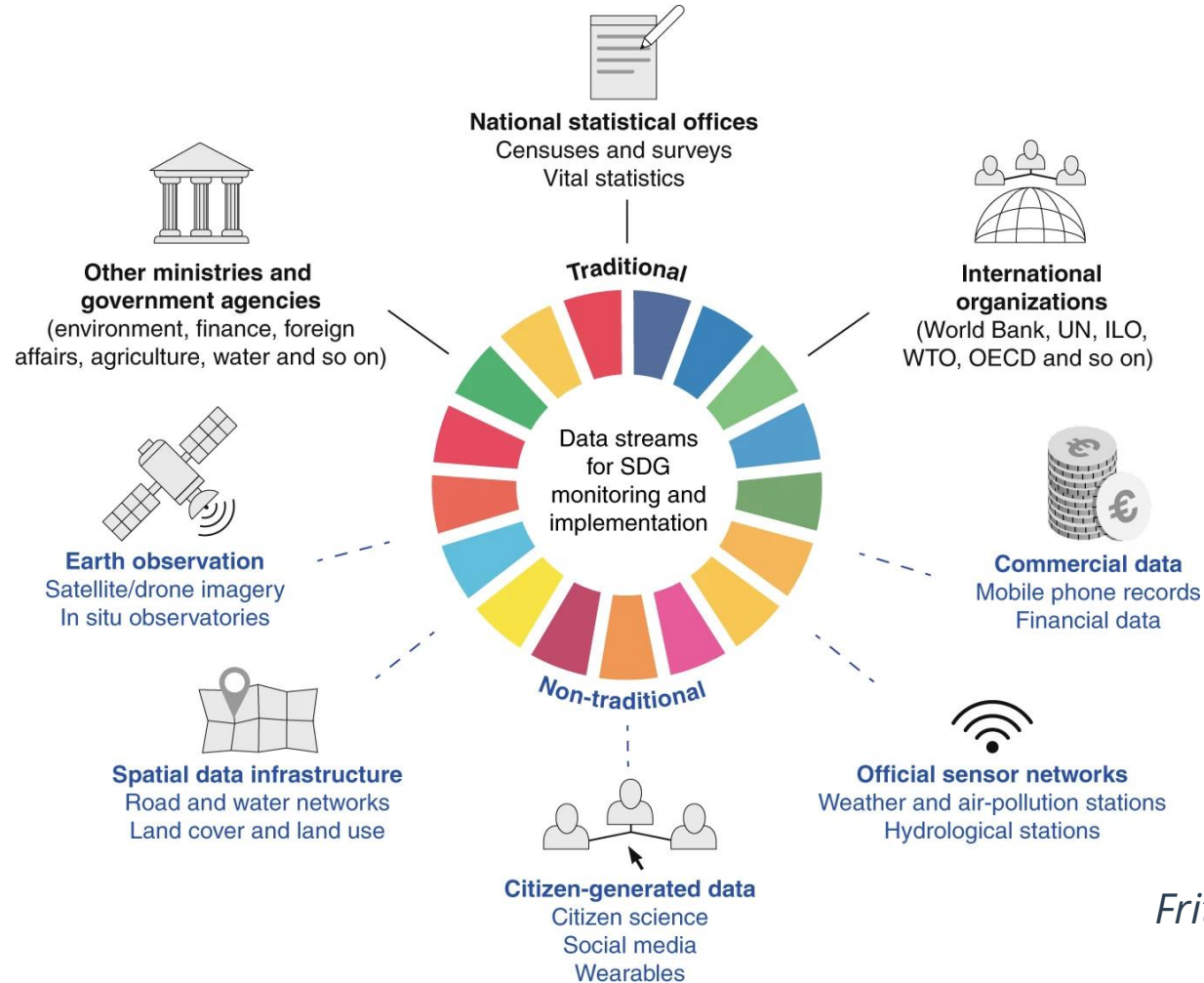
Research Data is becoming more complex

Research Data is becoming more complex

NASA Earth observation data distribution by volume and by number of data files 2000 to 2023.
Credit: NASA ESDIS Project Metrics System (EMS)

Research Data is becoming more complex

National statistical offices
Censuses and surveys
Vital statistics

Other ministries and government agencies
(environment, finance, foreign affairs, agriculture, water and so on)

International organizations
(World Bank, UN, ILO, WTO, OECD and so on)

Traditional

Data streams for SDG monitoring and implementation

Earth observation
Satellite/drone imagery
In situ observatories

Commercial data
Mobile phone records
Financial data

Non-traditional

Spatial data infrastructure
Road and water networks
Land cover and land use

Official sensor networks
Weather and air-pollution stations
Hydrological stations

Citizen-generated data
Citizen science
Social media
Wearables

Fritz et al (2019)

Research Data is becoming more complex

growth in data
volumes

(inter)national
collaborations

digitalization of
data and
processes

non-traditional data

sensitive data

Open Data Barriers

sharing rights

scooping

transient storage

lack of time

sensitive content

lack of incentives

reuse concerns

disincentives

data and code
sharing

perceived barriers

inappropriate use

insecurity

knowledge barriers

data too large

unclear value

unclear process

complex workflow
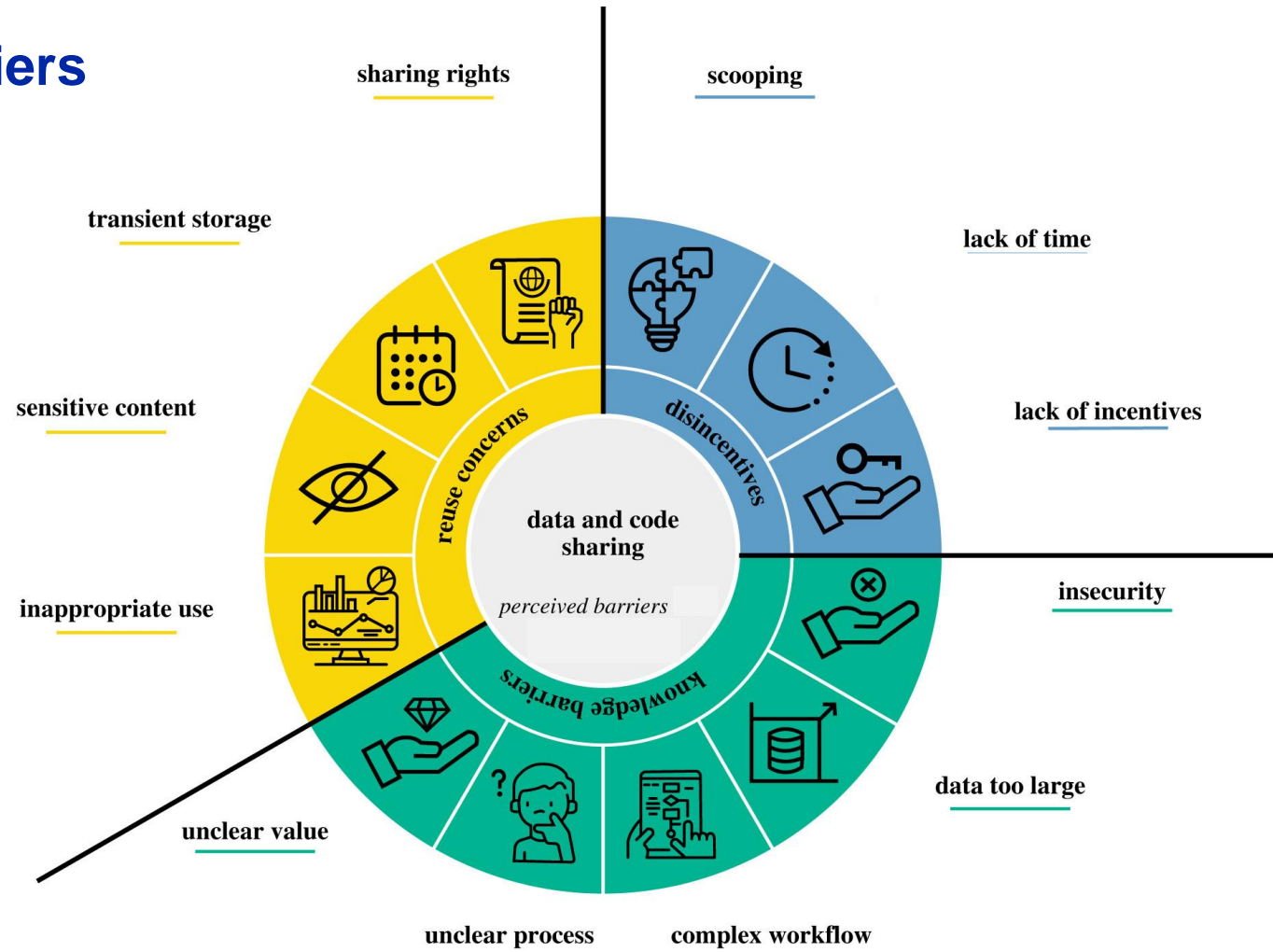
Gomes et al (2022)

Data Stewardship as a possible solution

Data Stewards

– Combine general research data management (RDM) knowledge with **specific scientific expertise**

– FAIR **Support**: The first point of contact for researchers with research data management (RDM) questions

– FAIR **Ambassadors**: Convey news, best practices and RDM knowledge to colleagues

– FAIR **Bridge**: Identify needs of researchers and translate them for service units and the university administration

– FAIR **Enhancers**: Create FAIR Resources & Tools, define Guidelines, Processes, Strategies

Data Stewardship at the UZH

Needs assessment

We interviewed 13 researchers and employees of the UZH:

- Current RDM practices

- Needs and concerns relating to RDM

Needs assessment

Comments:

"Information about data sharing and data management, for example about support and trainings, is difficult to find. There are so many websites even just within the UZH and no unified place to access all relevant information."

"When I am unsure of something I tend to ask my colleagues or just google it. I have never reached out to research support services at the UZH, and honestly wouldn't really know where to start."

Information and support is difficult to find and access

Needs assessment

Comments:

"Data management across the department varies a lot and is not discussed much"

RDM varies within faculties, departments and even research teams

"I use the same data management procedures for most projects. It's not automatized and there is for the most parts no common strategy for the department."

"A documented standard procedure [for data management] would be great. Right now, every researcher does it their own way, and most of the time, FAIR and open data are not top priorities."

Needs assessment

Comments:

"It would be beneficial to have community data standards and adhere to them for each discipline or similar area of research to improve data interoperability. In some cases, it is more of a decision-making problem than a problem of available standards at hand."

Need for guidelines and standards

"I think it should become regular practice to publish metadata at least, but it would also be great to develop a guideline for how to do this that has buy-in at the department level head."
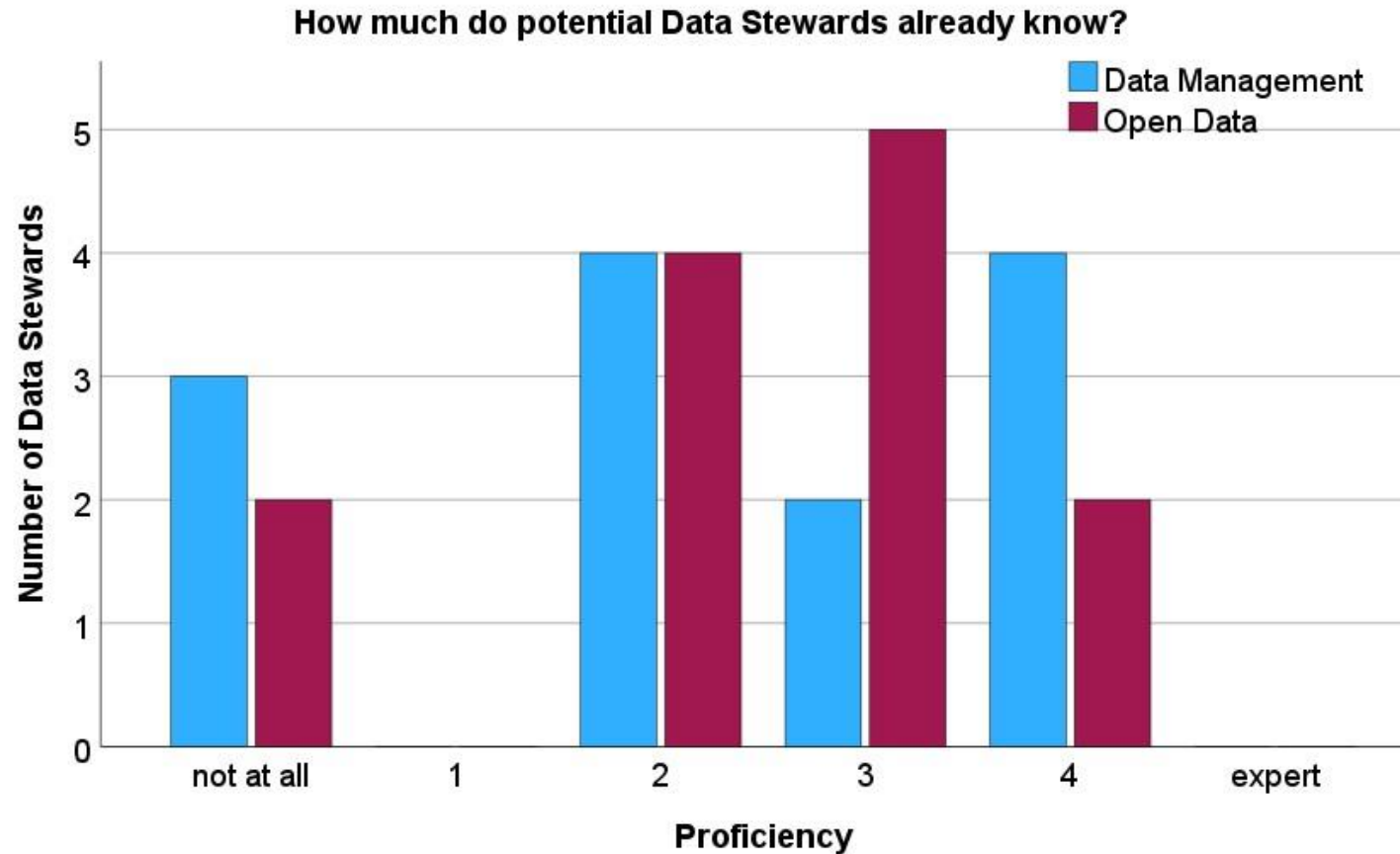
"I am very committed to the idea of Open & FAIR data principles but have struggled with implementing them, given the sensitive and identifiable nature of the medical data [I work with], especially when combined with geolocation data. I recently became aware that we could at least publish the metadata, but have been lacking the knowledge and time to do so, although I would like to very soon."

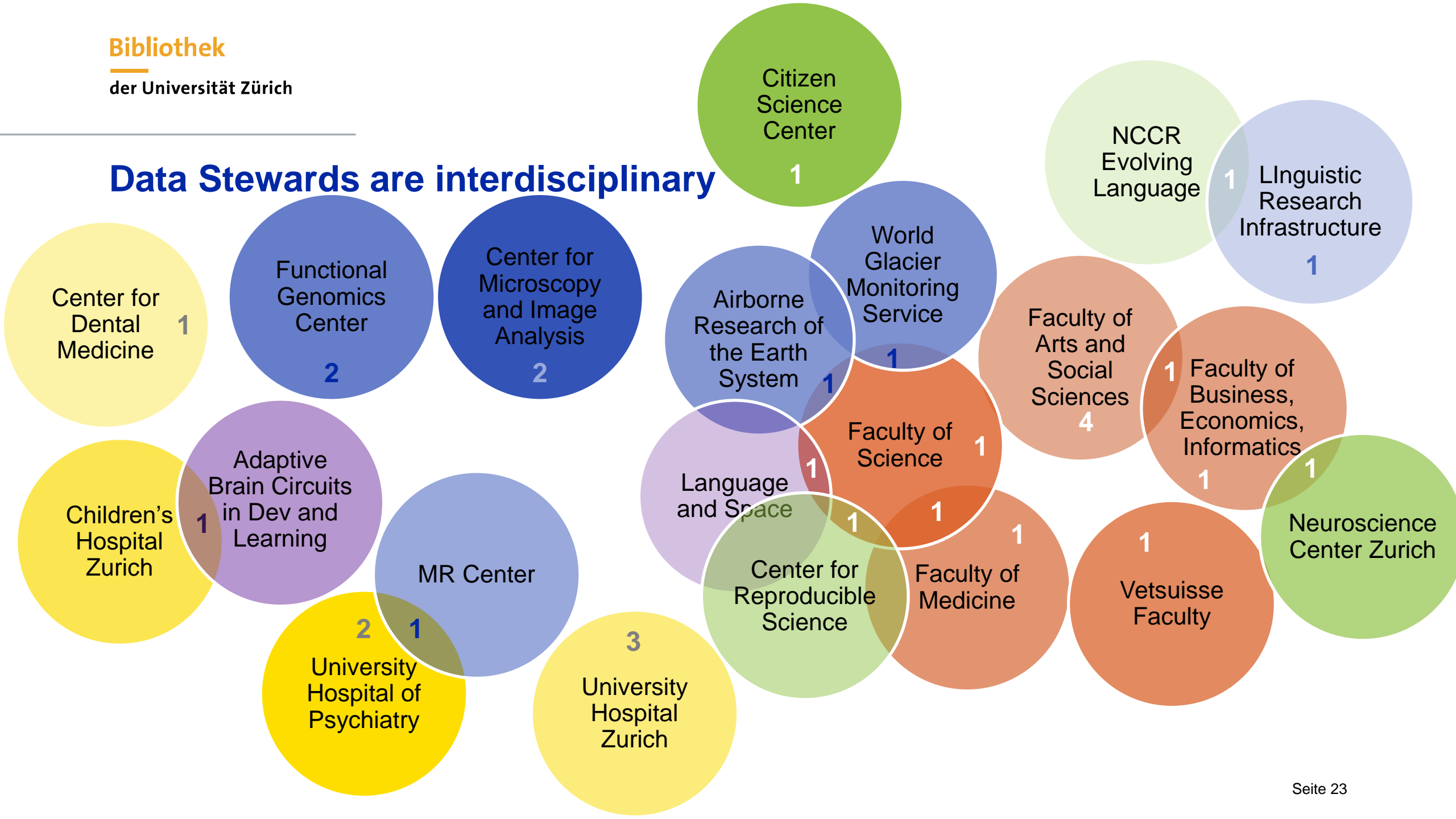# Data Stewardship at the UZH : Kick-Off

Data Stewards are heterogeneous

– 30 Data Stewards make up the UZH Data Stewards Network

– About half of the Data Stewards are researchers and the other half scientific or technical staff

– All Data Stewards work with quantitative (n=30) and some with qualitative data (n=6)

– Data Stewards work with a variety of data types: *Imaging, Genomics, Proteomics, Sequencing, Text, Audio, Physiological, Medical, Survey, Satellite, Geodata, Historical*

– Data Stewards have been at the university from 1 year to up to 22 years.

Data Stewards are experts in the making

How much do potential Data Stewards already know?

Data Stewards are interdisciplinary

Bibliothek
der Universität Zürich

Citizen Science Center
1

NCCR Evolving Language

LInguistic Research Infrastructure
1

Center for Dental Medicine
1

Functional Genomics Center
2

Center for Microscopy and Image Analysis
2

World Glacier Monitoring Service

Airborne Research of the Earth System
1

Faculty of Arts and Social Sciences
4

Faculty of Business, Economics, Informatics
1

Faculty of Science
1

Adaptive Brain Circuits in Dev and Learning
1

Children's Hospital Zurich

Language and Space
1

MR Center

Neuroscience Center Zurich
1

Center for Reproducible Science
1

Faculty of Medicine
1

University Hospital of Psychiatry
2 1

University Hospital Zurich
3

Vetsuisse Faculty
1

Data Stewardship at the UZH in practice

Data Stewards Network

- **Monthly** meetings
  - Online & in-person
  - Topics: repositories, metadata, IT services and solutions, research centers and infrastructure at the UZH, and data protection and ethics
  - Training & Networking

- **Sharing** of resources and information internally

Data Stewards

– The network increases knowledge of RDM, FAIR & Open Data

Data Stewards

- **Functional Genomics Center Zurich**

  - Integrating FAIR RDM in their trainings

  - Updating the procedures of how to share data generated through their platform

- **Center for Microscopy and Image Analysis**

  - Creating documents/guidelines to pass on to researchers

  - Integrating FAIR information into trainings

  - Working on a conversion tool to convert proprietary data to open data

FAIR
Enhancers

Data Stewards

– **LInguistic Research Infrastructure**

  – Creating documents/primers to pass on to researchers

  – Workinging on tools to check file naming, check for duplicates and author metadata

– **PUK MR Center**

  – Evaluating database management tools for the PUK IT Infrastructure

  – Organizing a FAIR training workshop for users of the PUK MR Center

– **World Glacier Monitoring Service & Airborne Research of the Earth System (ARES)**

  – Evaluating possible platforms for ARES data

  – Creating an action plan on how to make the data publicly accessible

FAIR
Enhancers

Data Stewardship at the UZH – Outlook

Outlook

– <u>Next Meetings:</u>

– **Wednesday June 19th from 9 – 10am**: *Sensitive Data, Data Protection & De-identification of research data* **(online)**

– **Monday July 8th from 3 – 4pm**: *Reproducibility at the UZH and abroad* **(online)**

– To find out more or come to our events please contact <u>susanna.weber@ub.uzh.ch</u>

Thank you for your time

"Data stewardship offers excellent returns of investment."
Mons, 2020

"Data stewardship is […] aiming at optimising data management in line with the FAIR principles."
Wendelborn et al, 2023

References

Fritz, S., See, L., Carlson, T. *et al.* Citizen science and the United Nations Sustainable Development Goals. *Nat Sustain* **2**, 922–930 (2019). https://doi.org/10.1038/s41893-019-0390-3

Gomes Dylan G. E., Pottier Patrice, Crystal-Ornelas Robert, Hudgins Emma J., Foroughirad Vivienne, Sánchez-Reyes Luna L., Turba Rachel, Martinez Paula Andrea, Moreau David, Bertram Michael G., Smout Cooper A. and Gaynor Kaitlyn M. Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proc. R. Soc. B.* 28920221113 (2022). http://doi.org/10.1098/rspb.2022.1113

Christian Wendelborn, Michael Anger, Christoph Schickhardt. What is data stewardship? Towards a comprehensive understanding. *Journal of Biomedical Informatics*, Volume 140, 104337, ISSN 1532-0464 (2023). https://doi.org/10.1016/j.jbi.2023.104337.

Barend Mons. Invest 5% of research funds in ensuring data are reusable. *Nature*, 578, p. 491 (2020) https://doi.org/10.1038/d41586-020-00505-7