Data Management Plan – Discussion

Led by Ram Seshadri, UC Santa Barbara and Xiaoqing Pan, UC Irvine

"*Datasets underpinning published research findings are expected to be shared with other researchers, at no more than incremental cost and within a reasonable time*." [from a 2022 NSF "Dear Colleague" letter]

Annual NSF Director's Meeting, Alexandria VA, 20240201

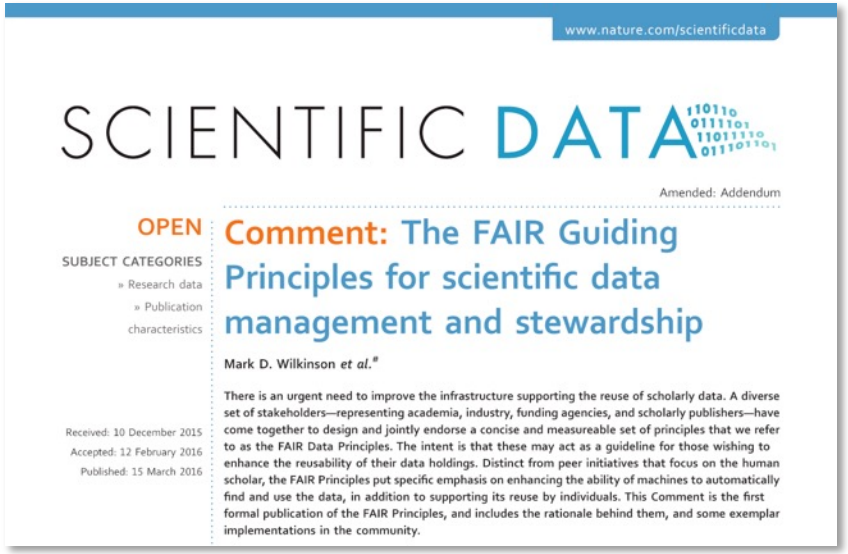
Not our first rodeo

NSF MRSEC Facilities Satellite + MRFN Workshop 2018

Thursday Oct 18th , Holiday Inn, 2460 Eisenhower Ave., Alexandria, VA 22314

Session I: Data		
2:30 pm to 2:40 pm	**Ram Seshadri,** University of California, Santa Barbara	Welcome and outline
2:40 pm to 3:00 pm	**Eva Campo**, NSF Program Director, NSF, Division of Materials Research. Crosscutting Activities in Materials Research (XC); Partnerships between Science and Engineering Fields and the NSF TRIPODS Institutes	Cross-Program Efforts on Materials Research Data at the NSF
3:00 pm to 3:45 pm	**Robert Hanisch**, NIST Director, Office of Data and Informatics https://www.nist.gov/people/robert-hanisch	Data Management, Curation, and Dissemination Strategies for Materials Science
3:45 pm to 4:00 pm	*break*	
4:00 pm to 4:45 pm	**Dana Vanderwall,** Bristol-Myers Squibb Director of Biology and Preclinical IT and Chair, Allotrope Foundation: https://www.allotrope.org	A Community and Framework to Harmonize the Scientific Data Landscape
Session II: Materials Research Facilities Network (MRFN) and Other Topics		
4:45 pm to 5:15 pm	**Divya Abhat** and **Ashish Tonse**, KZN Consulting	The MRFN Website
5:15 pm to 5:30 pm	**Jerry Hunter**, University of Wisconsin, Director of Research Facilities, College of Engineering	A New Reporting Tool for Facilities Online Manager (FOM)

The literature makes clear the need for better data handling/dissemination



www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

OPEN

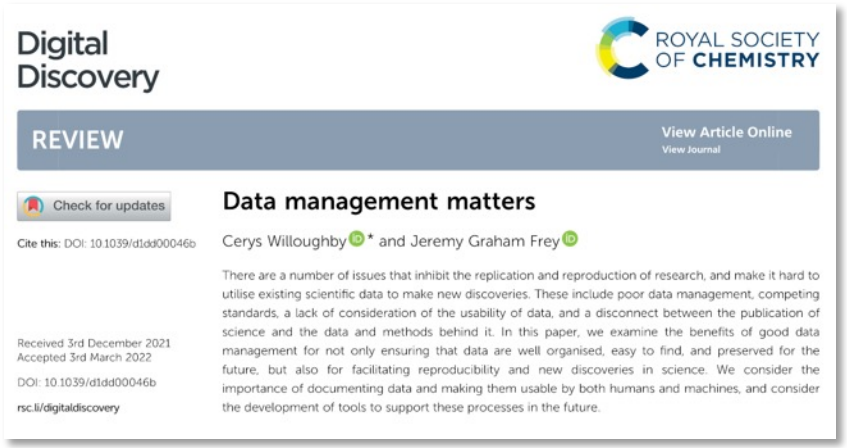
SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016



Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d1dd00046b

Data management matters

Cerys Willoughby * and Jeremy Graham Frey

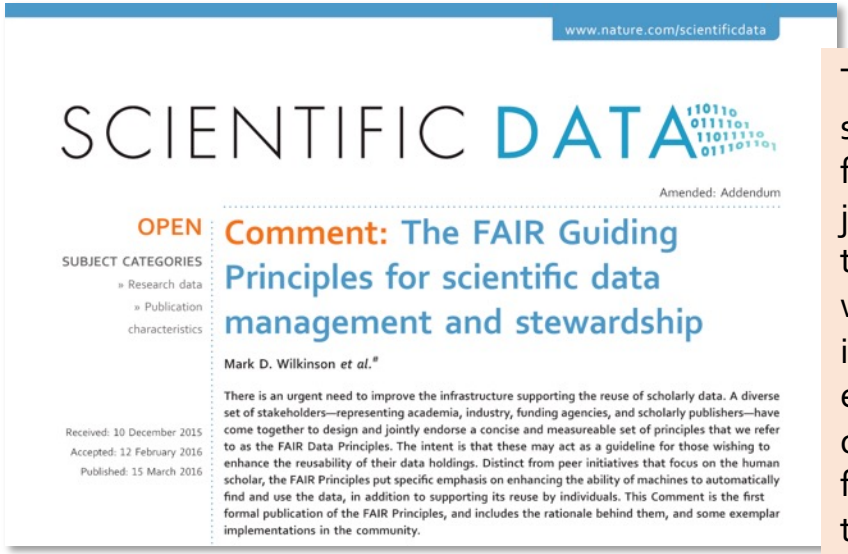
There are a number of issues that inhibit the replication and reproduction of research, and make it hard to utilise existing scientific data to make new discoveries. These include poor data management, competing standards, a lack of consideration of the usability of data, and a disconnect between the publication of science and the data and methods behind it. In this paper, we examine the benefits of good data management for not only ensuring that data are well organised, easy to find, and preserved for the future, but also for facilitating reproducibility and new discoveries in science. We consider the importance of documenting data and making them usable by both humans and machines, and consider the development of tools to support these processes in the future.

Received 3rd December 2021
Accepted 3rd March 2022

DOI: 10.1039/d1dd00046b

rsc.li/digitaldiscovery

The literature makes clear the need for better data handling/dissemination



www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal

Check for updates

Data management matters

Cerys Willoughby * and Jeremy Graham Frey

Cite this: DOI: 10.1039/d1dd00046b

There are a number of issues that inhibit the replication and reproduction of research, and make it hard to utilise existing scientific data to make new discoveries. These include poor data management, competing standards, a lack of consideration of the usability of data, and a disconnect between the publication of science and the data and methods behind it. In this paper, we examine the benefits of good data management for not only ensuring that data are well organised, easy to find, and preserved for the future, but also for facilitating reproducibility and new discoveries in science. We consider the importance of documenting data and making them usable by both humans and machines, and consider the development of tools to support these processes in the future.

Received 3rd December 2021
Accepted 3rd March 2022

DOI: 10.1039/d1dd00046b

rsc.li/digitaldiscovery

There are a number of issues that inhibit the replication and reproduction of research, and make it hard to utilise existing scientific data to make new discoveries. These include poor data management, competing standards, a lack of consideration of the usability of data, and a disconnect between the publication of science and the data and methods behind it. In this paper, we examine the benefits of good data management for not only ensuring that data are well organised, easy to find, and preserved for the future, but also for facilitating reproducibility and new discoveries in science. We consider the importance of documenting data and making them usable by both humans and machines, and consider the development of tools to support these processes in the future.

Focused IRG meetings on data management (October 9, 2023)

Adopting Open Science Principles at MRSEC

Renata Curty & Julien Brun
Research Data Services, UCSB Library
rds@library.ucsb.edu
10/09/2023

UC SANTA BARBARA

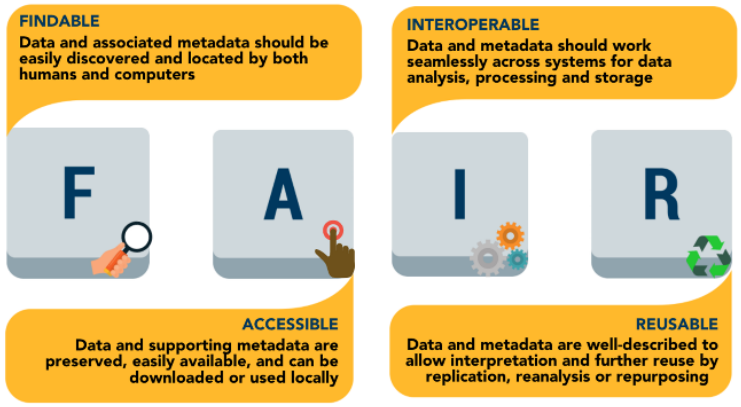
Open Science

A set of principles and practices to increase the ***transparency***, ***reproducibility*** and ***accessibility*** of scientific research by fostering **sharing** and **collaboration** as part of the **research process**

Reproducibility

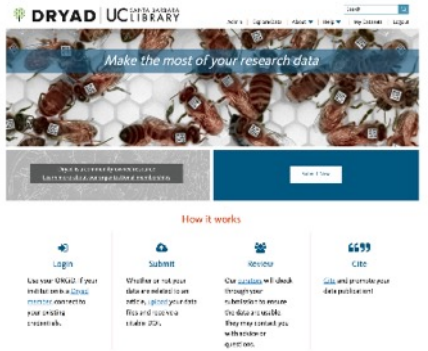
Convey clear, specific, and complete information about the *data*, study *methods*, and *computational tools* and *environment* used

FAIR Principles



Be FAIR to your data!

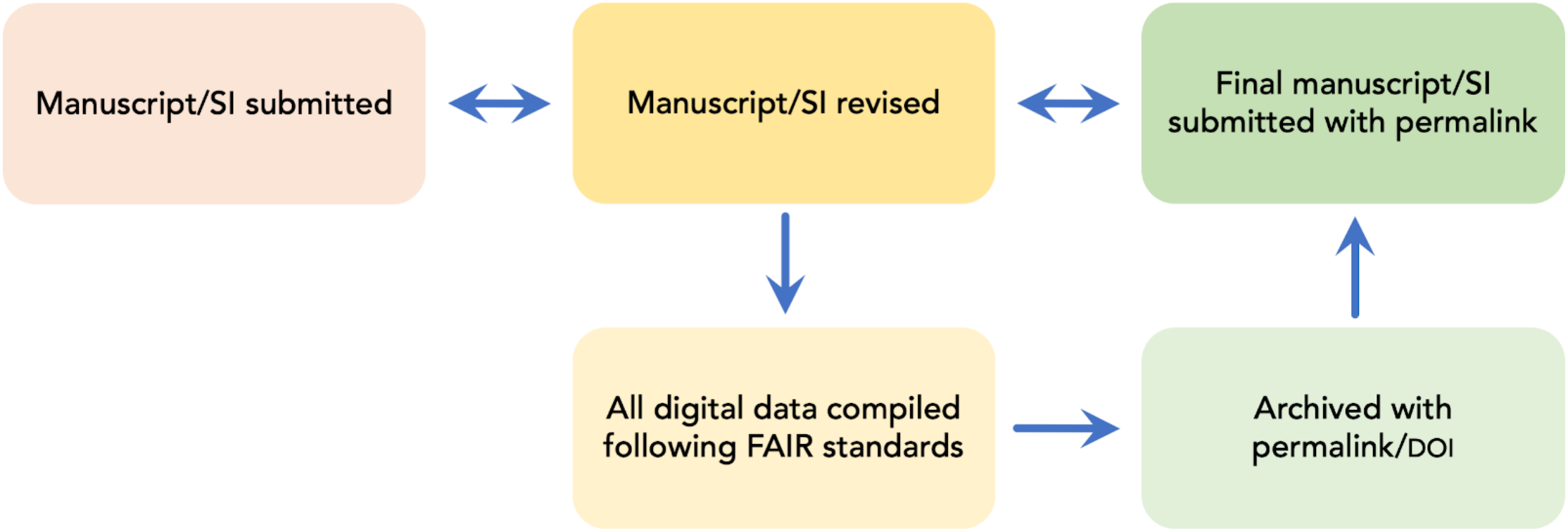
Dryad (datadryad.org)



~ 100 K Users
~ 75 K Datasets

UC SANTA BARBARA

A flow-chart for data associated with published work



In practice: A single manuscript (with just text files): 475 MB and 474 items

Macromolecules

pubs.acs.org/Macromolecules Article

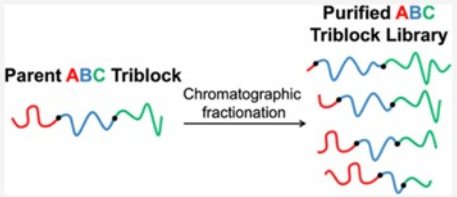
Efficient Creation and Morphological Analysis of ABC Triblock Terpolymer Libraries

Elizabeth A. Murphy, Yan-Qiao Chen, Kaitlin Albanese, Jacob R. Blankenship, Allison Abdilla, Morgan W. Bates,* Cheng Zhang,* Christopher M. Bates,* and Craig J. Hawker*

Cite This: *Macromolecules* 2022, 55, 8875–8882 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Multiblock copolymers with increasingly complex block sequences—for example, triblock terpolymers—offer unique opportunities to create nanostructured materials, but this potential has been hindered by a vast design space that complicates the explor… property relationships. Here, we report a versatile and scal… t ABC and isomeric ACB triblock terpolymers … les spanning a wide range of compositio… lymerization and automated chromat… n 10 ABC and ACB parent materia… . Separations follow systematic and predictable trends in volume fraction resulti… rich in non-polar blocks elute first, followed by more polar derivatives, yieldi… molar-mass dispersity. As evidenced by small-angle X-ray scattering, fract… red to as-synthesized parent materials and allows for the definitive iden… endly separation strategy significantly increases the availability of well-defin… munity while also improving sample quality and accelerating discovery.



- Figures
  - Figure2
    - Figure2_L_NMR.csv
    - Figure2_LD_NMR.csv
    - Figure2_LDF-22_NMR.csv
  - Figure3
  - Figure6b
  - Figure7
  - Figure17b
  - Figure35b
  - FigureS1-39
  - README.md
- Tables
  - TableS1
    - TableS1_LDF-17_SEC.txt
    - TableS1_LDF-18_SEC.txt
    - TableS1_LDF-19_SEC.txt
    - TableS1_LDF-43_SEC.txt
  - TableS4
  - TableS5
  - TableS6
  - TableS7
  - TableS9
  - TableS10
  - TableS11

Creating the necessary metadata is a nightmare !

What about unpublished data?

nature

nature › letters › article

Letter | Published: 04 May 2016

Machine-learning-assisted materials discovery using failed experiments

Paul Raccuglia, Katherine C. Elbert, Philip D. F. Adler, Casey Falk, Malia B. Wenny, Aurelio Mollo, Matthias Zeller, Sorelle A. Friedler ✉, Joshua Schrier ✉ & Alexander J. Norquist ✉

Nature **533**, 73–76 (2016)

Hope out there? Materials-relevant ELNs and LIMS (potentially all-in-one)

UNIVERSITY OF HOUSTON

Department of Chemistry

Faculty Profile

Joshua D. Bocarsly

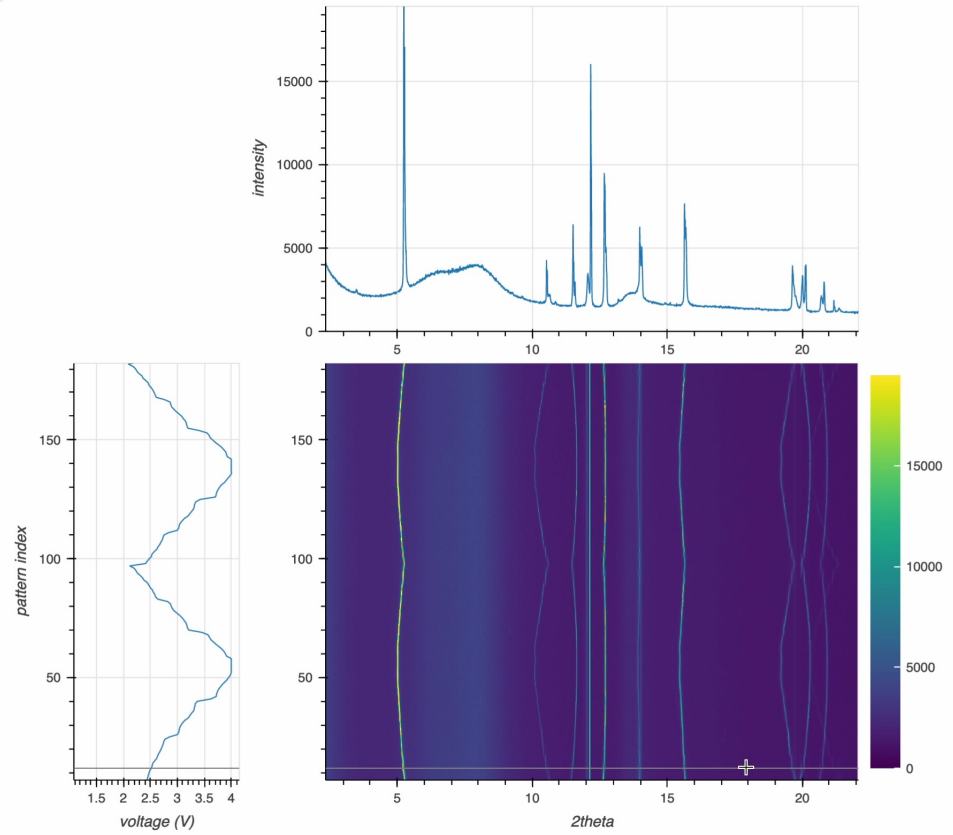
Assistant Professor

Robert A. Welch Foundation Professor in High Temperature Superconducting & Chemical Materials Science

Department of Chemistry

Office: STL, 447

Contact: jdbocars@central.uh.edu - 713-743-4053



Datalab: Automatically managing experimental data in context

"Datalab" is a new data management software that seeks to frictionlessly keep track of disparate data and store it with context in a self-describing manner.

Collaboration with Dr. Matthew Evans, funded by EU BIGMAP-SI initiative

M.L Evans and J.D. Bocarsly (2023). datalab (v.2.5). Zenodo. https://doi.org/10.5281/zenodo.8127783

NSF MRSEC

Annual NSF Director's Meeting, Alexandria VA, 20240201

To ponder …

- ❑ Data collection and documentation
- ❑ Data storage and preservation (raw versus processed)
- ❑ Data sharing and accessibility
- ❑ Compliance with legal and ethical standards
- ❑ Data security and privacy
- ❑ Data quality and integrity
- ❑ Interdisciplinary collaboration and integration
- ❑ Budgeting for data management
- ❑ Training and support
- ❑ Monitoring and evaluation
- ❑ Impact on research outcomes
- ❑ Feedback mechanisms

The highest barrier is culture !