Training of Trainers

Title: Strengthening RDM through Data Stewardships and Training

Magdalena Szuflita-Żurawska & Gdańsk University of Technology, Poland

MINISTRY OF EDUCATION AND SCIENCE

UNITe

NGO Links

Research Data Management

About me

mgr Magdalena Szuflita-Żurawska

Employment

Lider Centrum Komeptencji Otwartej Nauki at Gdańsk University of Technology
Senior Librarian at Scientific and Technological Information Section
Head of GUT Library section at Scientific and Technological Information Section

Research fields

Scholarly Communication   Open Access   Open Data

Motivation   Academic Productivity

Profile   Biographical note   Organizations (7)   Projects (6)   Teaching (81)   **Publications (11)**   Achievements (4)   Scientific posters   Research data (42)

Publications

results on page: 20 **50** 100   year: ☑ newest first ☐ oldest first   title: **A-Z** Z-A   citation: descending ascending

Filters   total: 11

Category
☐ Conference activity
☐ Monographic publication
☐ Articles

Year ⌄

2025

Motivation and scientific productivity under pressure

publication M. Szuflita-Żurawska, B. Basińska - 2025

Scientific productivity is crucial in driving innovation and fostering social and economic growth. The research addresses three core questions: (1) what are the key motivators and demotivators influencing scientific productivity, (2) how do these factors vary across disciplines and career stages, and (3) what performance-oriented practices can support academic motivation and productivity in higher education institutions. The1

📄 Full text available

2022

MINISTRY OF EDUCATION AND SCIENCE

NGO Links

Slides content

Why Research Data Management Matters
– RDM in the research lifecycle, FAIR & CARE principles, Open Science context
Data Stewards as Key Enablers of RDM
– Roles, types of data stewards, and institutional models
From Policy to Practice
– Institutional drivers, funder requirements, and cultural change
Competence-Based RDM Training Framework
– Target groups, competence matrix, and learning pathways
Training Methodology
– Competence-based, problem- and case-based, blended learning, microlearning, learning by doing
Theoretical Foundations of RDM Training
– Change management, socio-technical perspectives, communities of practice
Designing FAIR & EOSC-Aligned Training Materials
– FAIR by design, Skills for EOSC, reusability and sustainability
Impact and Sustainability
– Train-the-trainer approach, continuous learning, long-term institutional capacity

Research Data Management

Practicies

- **Research Data Management (RDM)** refers to the **organisation, documentation, storage, and sharing** of research data.
- It covers the **entire research data lifecycle**, from planning to long-term preservation.
- RDM ensures that research data are **secure, usable, and reusable**.

Many funders and journals now require good data management. But beyond compliance, RDM simply makes research more efficient and trustworthy.

- Improves **research quality and reliability**
- Supports **reproducibility and transparency**
- Helps meet **funders' and institutional requirements**
- Reduces risk of **data loss and misuse**

What are research data?

- **Research data** are all data **collected, observed, generated, or reused** in the course of research.
- They are used to **produce, validate, and support research findings**.
- Research data can be **digital or non-digital**, **raw or processed**.
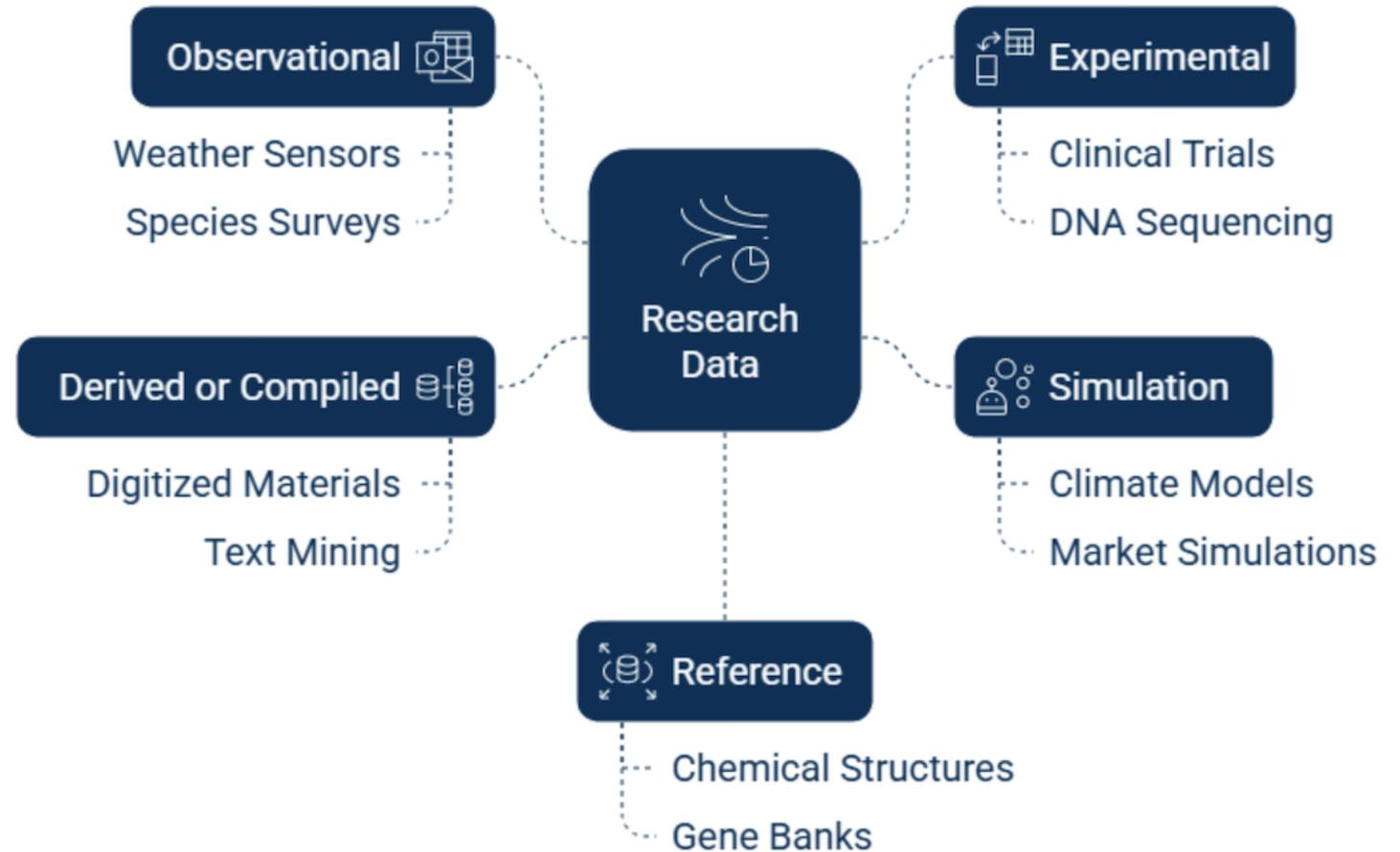
Research Life Cycle

· **Planning – DMP & FAIR by design**
→ FAIR principles are applied already at the data planning stage.

· **Collection – Quality & consent**
→ Data quality and ethical considerations are addressed during data collection.

· **Processing – Clean & structured**
→ Data are prepared for analysis and future reuse.

· **Analysis – Meaningful results**
→ Data are transformed into knowledge and scientific insights.

· **Preservation – Long-term storage**
→ Data are preserved beyond the lifetime of the research project.

· **Sharing – As open as possible**
→ Data are shared openly while respecting legal and ethical restrictions.

· **Reuse – FAIR impact**
→ Data achieve their highest scientific value through reuse.

Research Data Lifecycle (FAIR Data Perspective)
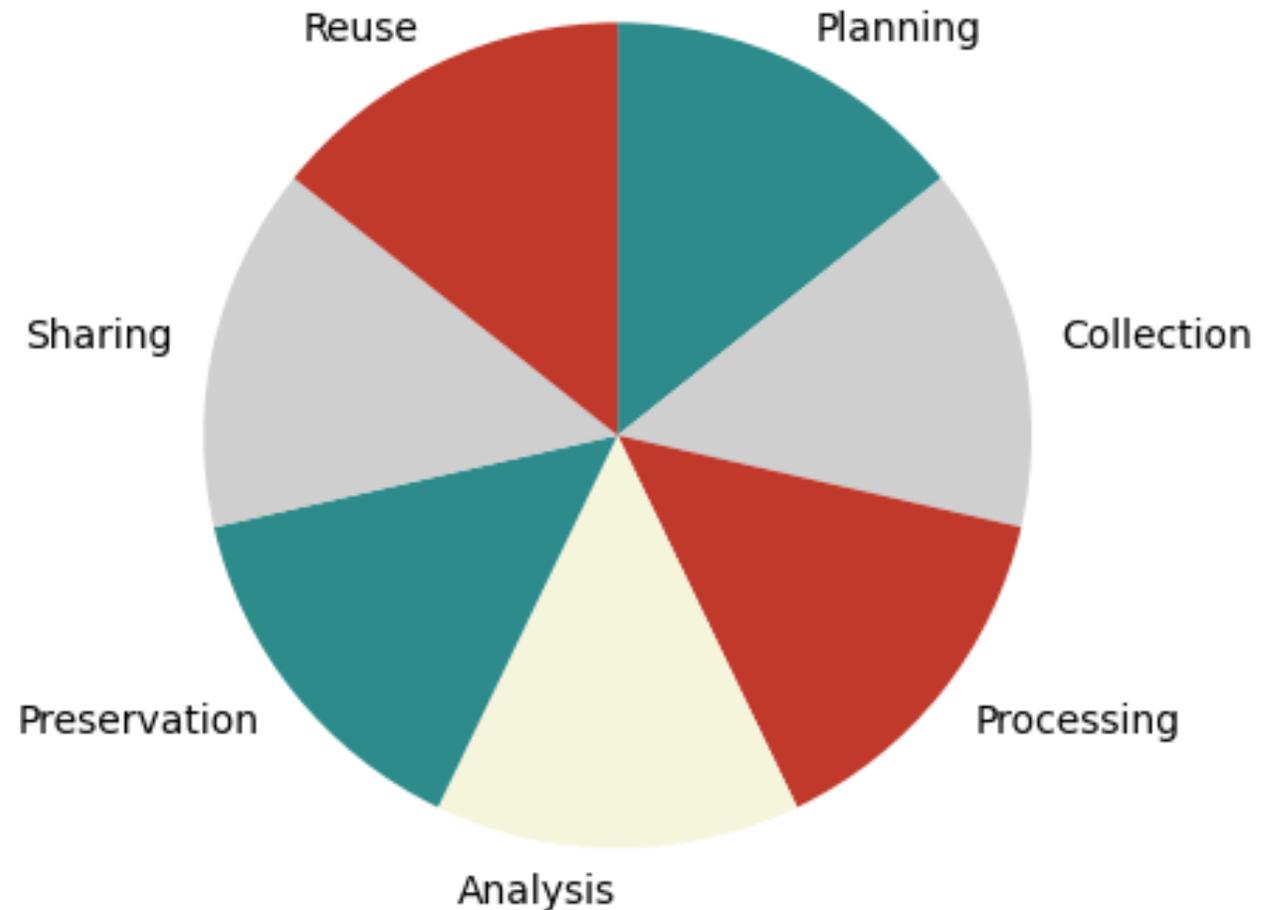
What is FAIR?

It is a set of 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

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

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

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

CARE Principles

The **CARE Principles** focus on the **ethical use of data**, especially data related to **Indigenous peoples and communities**.

C – Collective Benefit
Data use should benefit the communities from which the data originate.
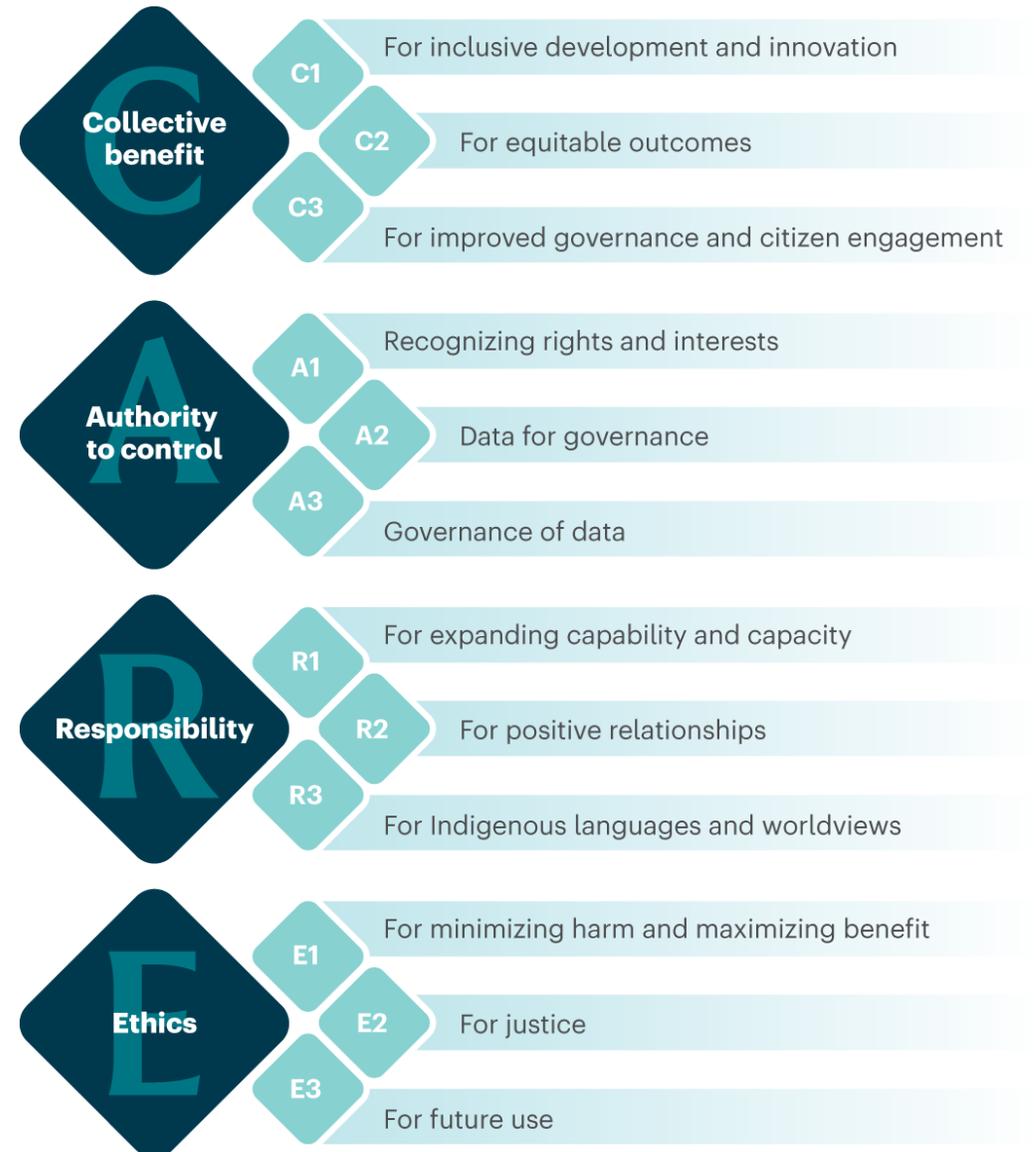
A – Authority to Control
Communities have the right to control how their data are collected, used, and shared.

R – Responsibility
Researchers must act responsibly and respectfully when using data.

E – Ethics
Data use should minimize harm and support justice and equity.

Collective benefit
- C1 For inclusive development and innovation
- C2 For equitable outcomes
- C3 For improved governance and citizen engagement

Authority to control
- A1 Recognizing rights and interests
- A2 Data for governance
- A3 Governance of data

Responsibility
- R1 For expanding capability and capacity
- R2 For positive relationships
- R3 For Indigenous languages and worldviews

Ethics
- E1 For minimizing harm and maximizing benefit
- E2 For justice
- E3 For future use

Aspect	FAIR Principles	CARE Principles
Main focus	Data and metadata	People and communities
Primary goal	Make data reusable and interoperable	Ensure ethical and responsible data use
Key question	How can data be shared and reused efficiently?	Who benefits from data and who controls them?
Perspective	Technical and practical	Ethical, social, and political
Target users	Researchers, machines, infrastructures	Communities, especially Indigenous peoples
Core concepts	Findable, Accessible, Interoperable, Reusable	Collective Benefit, Authority to Control, Responsibility, Ethics
Access	Encourages openness with clear access rules	Emphasizes appropriate control and consent
Licensing	Clear, machine-readable licences	Respect for community rights and governance
Relation to Open Science	Supports openness and reproducibility	Ensures openness does not cause harm
Typical application	Research repositories, DMPs, data standards	Sensitive, social, cultural, and Indigenous data

Data Management Plan – living document

- A **Data Management Plan (DMP)** describes how data will be **created, managed, stored, shared, and preserved**.
- It is usually required by **funders and institutions**.
- A DMP supports **FAIR and CARE principles** from the start of a project.

DMP Stage	What happens in the DMP	How AI can help
1. Planning	Define data types, formats, standards, ethics, FAIR strategy	AI can help draft DMP text, suggest data formats, metadata standards, and FAIR-compliant wording
2. Data Collection	Describe how data will be collected and documented	AI can help design surveys, improve questionnaires, and suggest consistent file naming and documentation
3. Processing & Documentation	Explain how data will be cleaned, organised, and described	AI can assist with data cleaning, detecting errors, generating README files and metadata descriptions
4. Storage & Security	Plan secure storage, backups, and access control	AI can support risk assessment, data classification, and suggest security measures for sensitive data
5. Analysis	Describe how data will be analysed and interpreted	AI tools can support data exploration, pattern detection, coding support, and reproducible workflows
6. Preservation	Define long-term preservation strategy	AI can help select repositories, preservation formats, and assess long-term usability
7. Sharing	Specify how and when data will be shared	AI can suggest repositories, licences, and help prepare datasets for publication
8. Reuse	Enable future reuse of data	AI can generate rich metadata, summaries, and usage documentation to support reuse

Metadata

- **Metadata standards** provide structured, consistent descriptions of research data.
They ensure that data are understandable, interoperable, and reusable across systems and disciplines.

| Discipline	Metadata Standard	Typical Use	Example Data Types
All disciplines (general)	Dublin Core	Basic description and discovery	Datasets, publications, repositories
Social Sciences	DDI (Data Documentation Initiative)	Survey and observational data	Questionnaires, interviews, statistics
Humanities / Cultural Heritage	TEI	Text encoding and annotation	Manuscripts, digital editions
Humanities / Cultural Heritage	CIDOC CRM	Cultural heritage data integration	Museum objects, archives
Life Sciences	MIAME	Microarray experiment metadata	Gene expression data
Life Sciences	Darwin Core	Biodiversity and species data	Occurrence records, specimens
Medicine & Health	CDISC	Clinical research data	Clinical trials
Earth & Environmental Sciences	ISO 19115	Geospatial and environmental data	Maps, satellite data
Chemistry	Chemical Markup Language (CML)	Chemical structure and experiments	Spectra, reactions
Physics / Astronomy	FITS	Astronomical data description	Telescope images
Computer Science	CodeMeta	Software and code metadata	Research software
Linguistics	OLAC	Language resources	Corpora, lexicons

Datacite vs schema.org
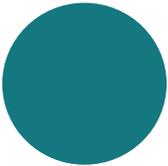
DataCite provides a widely used metadata schema for research data and supports the assignment of **persistent identifiers (DOIs)**.
It is designed to make data **Findable, Accessible, and Reusable (FAIR)** by enabling persistent identification, citation, and discovery.
Core metadata elements include information about the creator, title, publisher, publication year, resource type, identifier (DOI), and licence or rights.
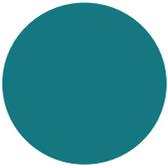
schema.org is a structured metadata vocabulary used on the web to improve **machine readability and discoverability**.
It enables research datasets to be indexed by search engines such as **Google Dataset Search**, connecting data with the wider web through standard properties like name, description, creator, publisher, publication date, licence, and identifiers.
Together, DataCite and schema.org strengthen FAIR data by combining persistent identification with web-based discovery.
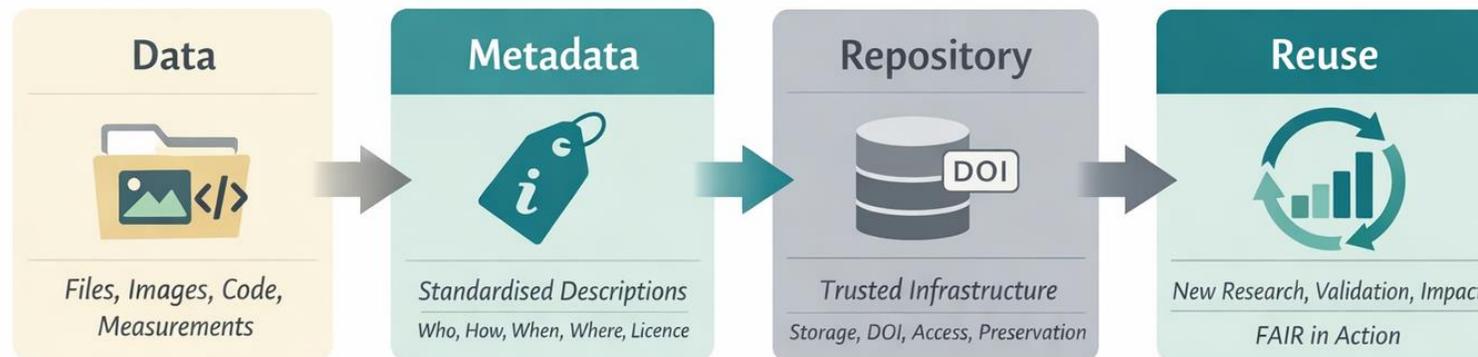
Aspect	DataCite	schema.org
Main purpose	Persistent identification and citation	Web discovery and indexing
Typical use	Repositories, DOIs, citations	Websites, search engines
FAIR focus	Findable, Reusable	Findable, Accessible
Machine-readable	Yes	Yes
Complementary?	✓ Yes	✓ Yes

DataCite and *schema.org* work together to make research data more discoverable and accessible.
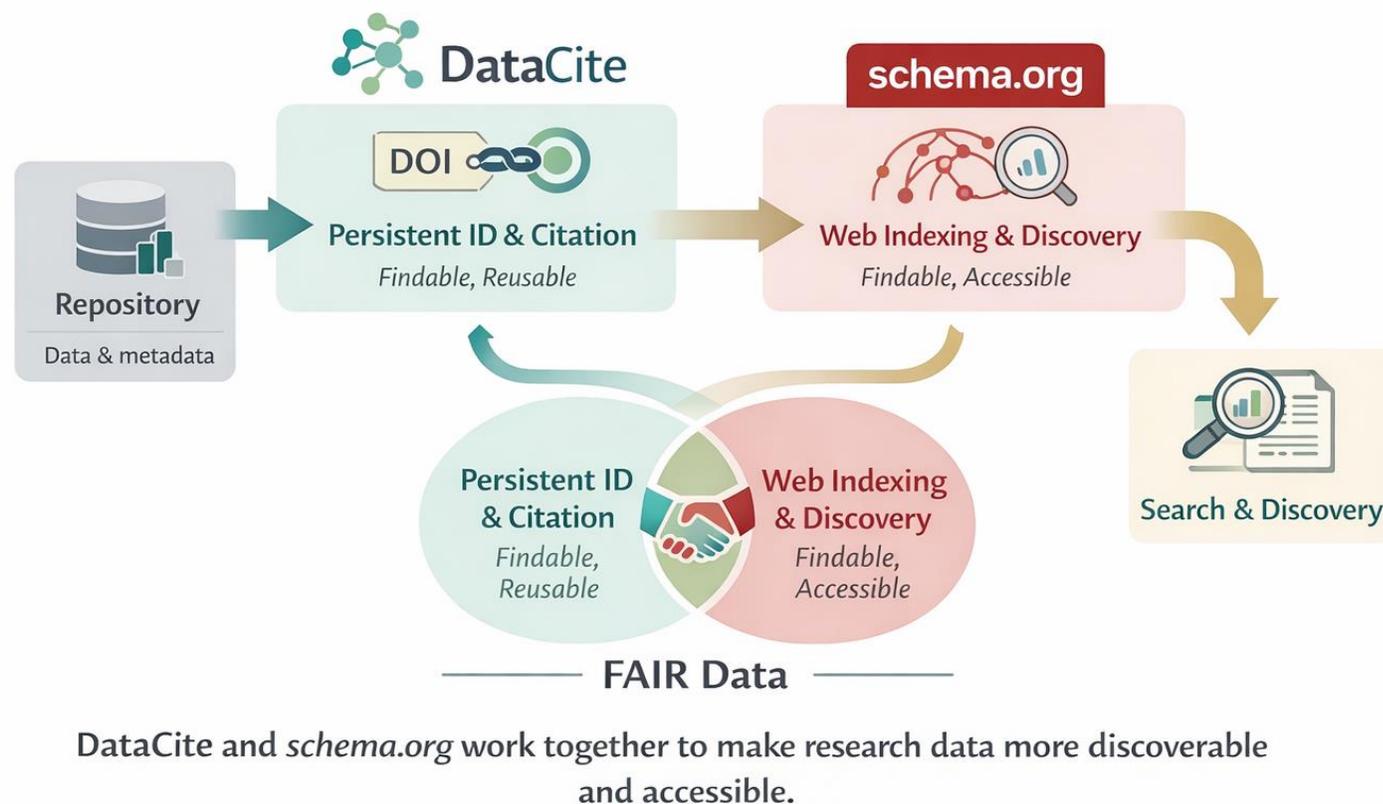
DataCite supports citation and persistence, while schema.org improves web visibility — together they strengthen FAIR data.

Metadata connections

Data alone are not enough. Metadata provide context and meaning, repositories provide infrastructure and trust, and together they make data reusable. This is the core idea behind FAIR data.

Data	Metadata	Repository	Reuse
Files, Images, Code, Measurements	Standardised Descriptions Who, How, When, Where, Licence	Trusted Infrastructure Storage, DOI, Access, Preservation	New Research, Validation, Impact FAIR in Action

Metadata connect data with repositories and enable reuse.

Repositories

Definitions and roles

- A **repository** is a structured system for storing, managing, and providing access to digital resources.
- Repositories rely on **metadata** to organize content and make it searchable and interoperable.
- They support secure storage, controlled or open access, and long-term preservation of digital objects.
- Repositories play a key role in research, digital libraries, and data management.

Key Functions of Repositories

- **Storage** – secure and long-term preservation of digital objects
- **Access** – controlled or open access for users
- **Management** – versioning, permissions, and content lifecycle
- **Preservation** – ensuring data integrity and long-term usability
- **Discovery** – enabling search and retrieval through metadata

TRUST and CoreTrustSeal

The **TRUST Principles** define key characteristics of **trusted digit**
repositories:
- **Transparency** – clear policies, governance, and workflows
- **Responsibility** – proper data stewardship and accountability
- **User focus** – serving the needs of designated user communitie
- **Sustainability** – long-term financial and organizational support
- **Technology** – reliable, secure, and well-maintained infrastructure

These principles help ensure that repositories are **reliable, reusable, and**
sustainable.

CoreTrustSeal is an international certification for trustworthy data repositories.
- Based on the **TRUST Principles**
- Focuses on organizational, technical, and legal requirements
- Demonstrates commitment to **long-term preservation and access**
- Widely recognized by research funders and institutions

CoreTrustSeal provides confidence that a repository meets **international standards of**
trustworthiness.

Types of repositories

Disciplinary repositories – focus on specific research domains
- **GenBank** (life sciences)
- **ICPSR** (social sciences)

Institutional repositories – managed by universities or research organizations
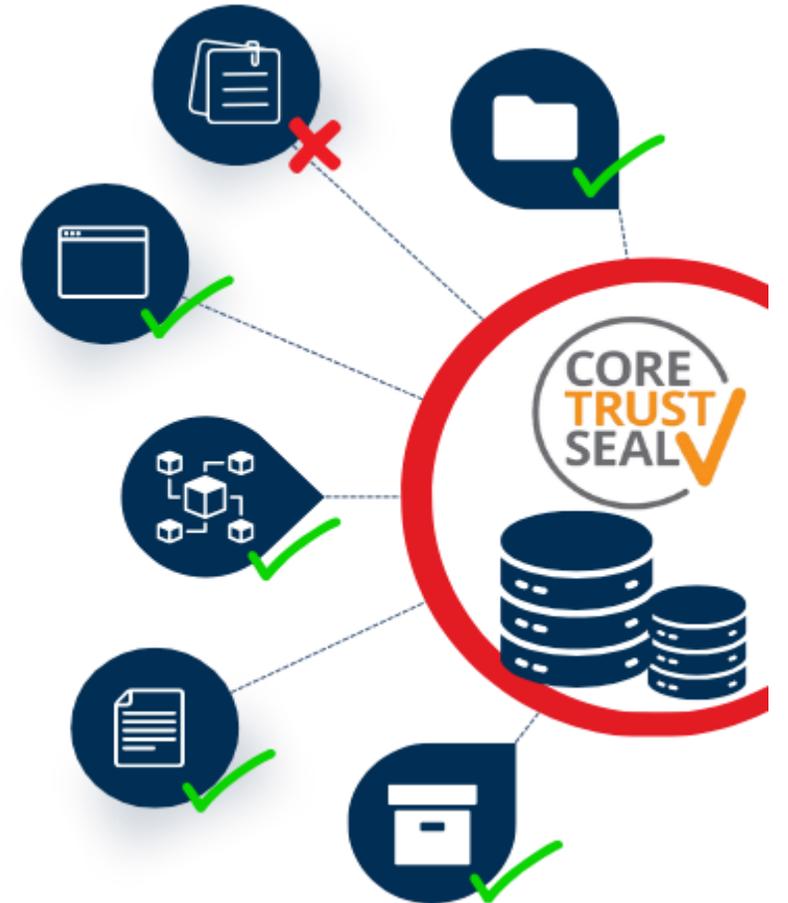- **DSpace** repositories at universities
- Bridge of Data from Gdańsk Tech

General-purpose repositories – accept data from many disciplines
- **Zenodo**
- **Figshare**

National or domain-specific repositories – support strategic or large-scale data
- **UK Data Service**
- **PANGAEA**

Repository indexing, metrics and TRUST

Repository indexing increases visibility, discovery, and reuse of research data. Standardized metadata and persistent identifiers (e.g. DOIs) enable repositories to be indexed by global services.

Make Data Count

- Promotes **standardized, open metrics** for research data (views, downloads, citations).
- Helps demonstrate the **impact and reuse** of datasets across repositories.
- Supports FAIR principles, especially **Reusable** and **Findable**.

DataCite

- Provides DOIs for datasets and infrastructure for citation and usage tracking.
- Connects datasets with publications, researchers, and metrics.

re3data

- Indexes and describes repositories worldwide.
- Helps researchers identify **trusted and FAIR-aligned repositories**.

Together, indexing services and data metrics strengthen **TRUST principles** (Transparency, Responsibility, Sustainability) and support **FAIR data practices** in repositories.

F-uji – FAIR Data Assessment Tool

F-UJI is an automated tool for assessing how well research data comply with the **FAIR principles**.

➤ Developed within the FAIRsFAIR project to support **objective and transparent FAIR evaluation**.
- Evaluates datasets based on:
- persistent identifiers (e.g. DOIs)
- rich and standardized metadata
- accessibility and reuse conditions

➤ Produces **machine-readable FAIR scores** and detailed feedback.
➤ F-UJI helps repositories and researchers **monitor, improve, and demonstrate FAIR compliance** of datasets.

Repository / Dataset → Metadata & PID → F-UJI Assessment → FAIR Indicators → FAIR Score & Recommendations

Support researchers

Role of the Open Science Competence Center
- Provides **guidance and training** on Open Science practices
- Supports researchers in **adopting FAIR principles**
- Promotes **openness, transparency, and reusability** of research outputs
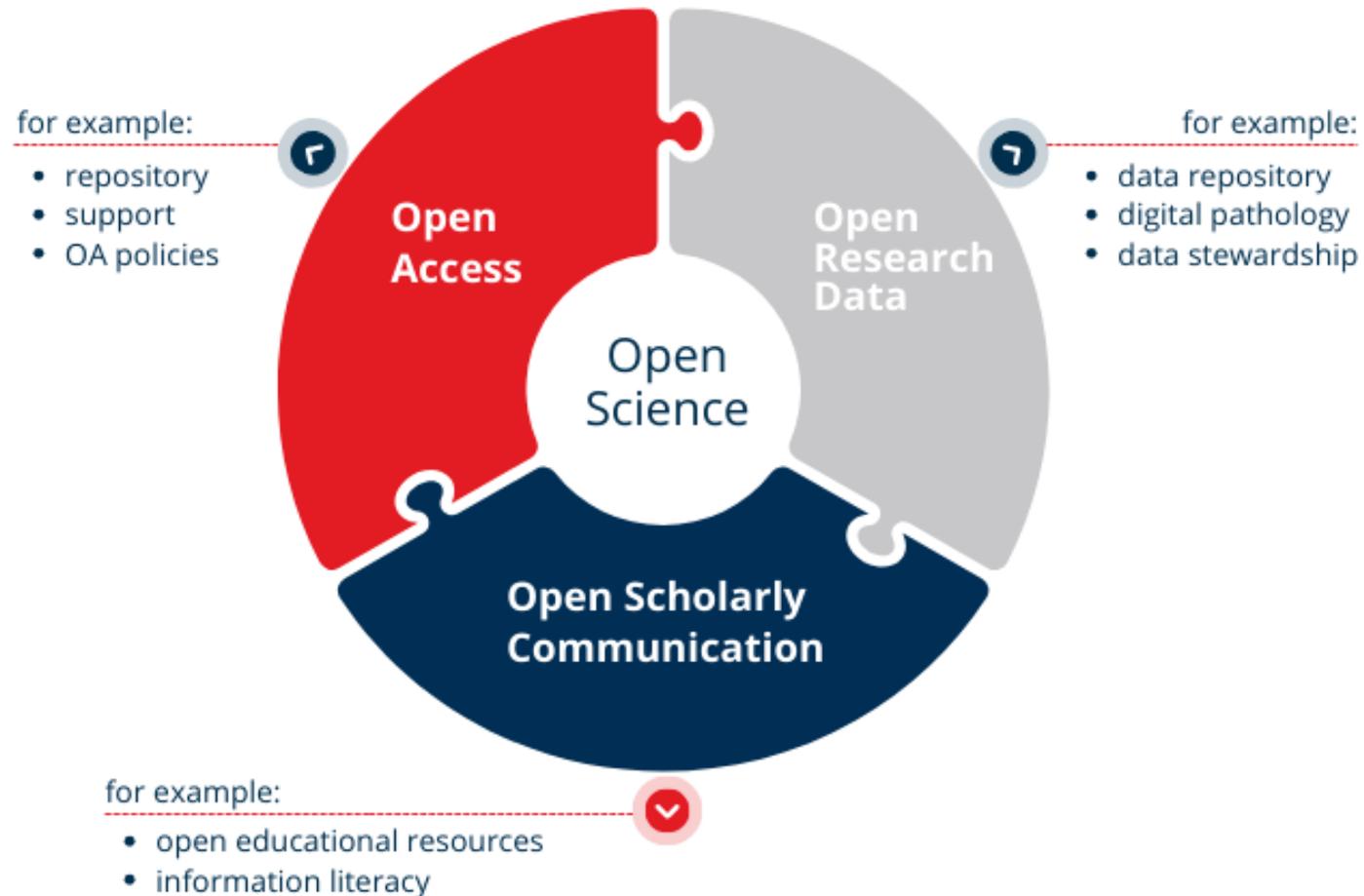
Relation to Research Data Management (RDM)
- RDM is a **core pillar** of Open Science
- The Center supports researchers across the **entire data lifecycle**:
- Data Management Plans (DMPs)
- Data organization, documentation, and metadata
- Data storage, backup, and security
- Data sharing, publishing, and reuse
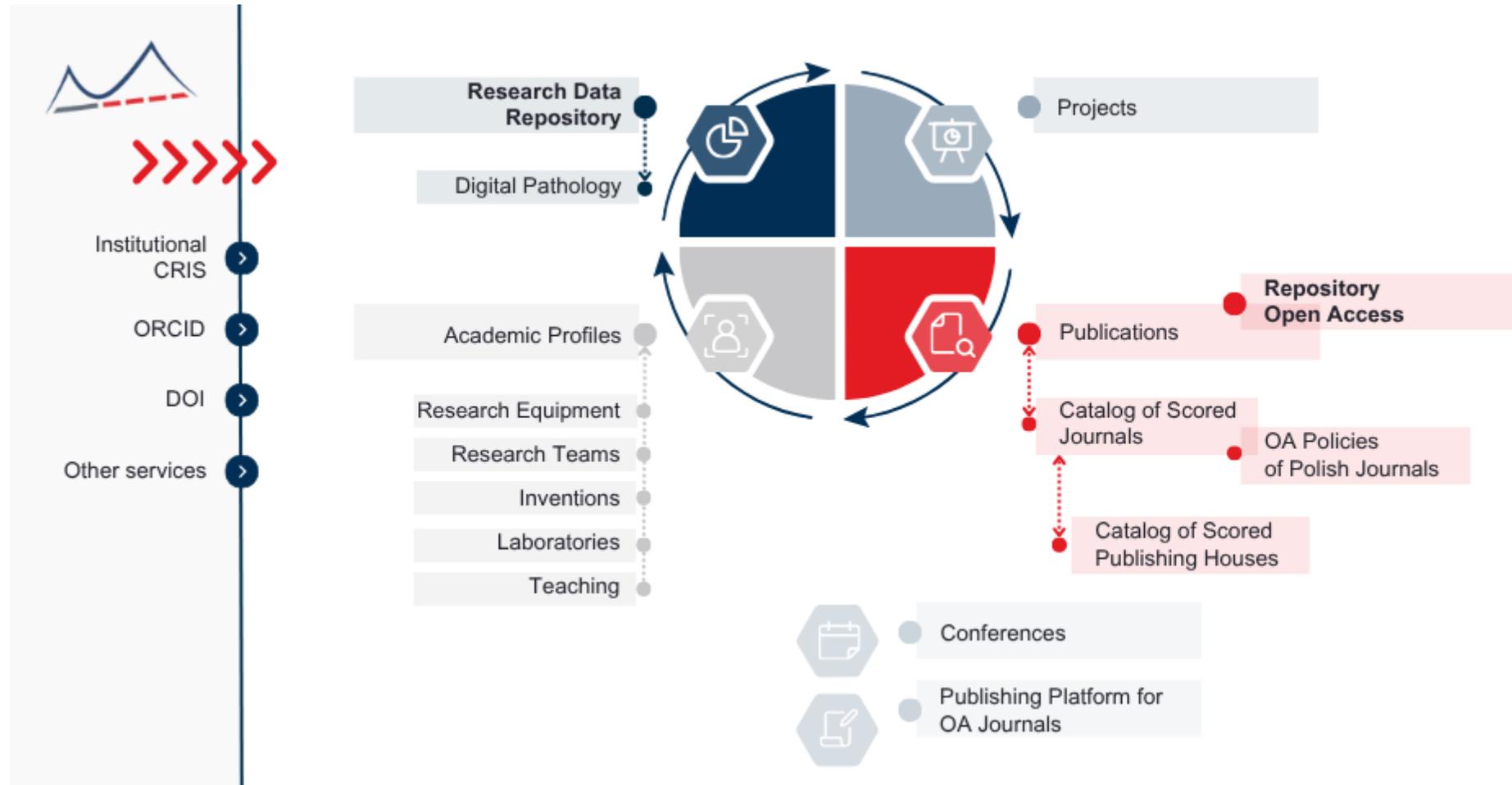- Advises on **policies, funder requirements, and legal/ethical aspects** (e.g. GDPR)

Value for Researchers
- Improves **research quality and reproducibility**
- Saves time through **best practices and standards**
- Increases **visibility and impact** of research data

Open Science Competence Center at Gdańsk Tech

Open Science Competence Center at Gdańsk Tech

Who is datasteward?

- Experts supporting researchers in **Research Data Management (RDM)**
- Act as a **bridge between researchers, IT services, and Open Science units**
- Provide **discipline-specific and practical suport**

Key Responsibilities
- Support creation and review of **Data Management Plans (DMPs)**
- Advise on **data organization, documentation, and metadata standards**
- Help implement **FAIR data principles**
- Guide researchers on **data storage, sharing, and long-term preservation**
- Support compliance with **funder, institutional, and legal requirements**
- **Value for Researchers & Institutions**
- Higher **data quality and reusability**
- Reduced risk related to **data loss and non-compliance**
- Stronger alignment with **Open Science practices**

Types of datasteward

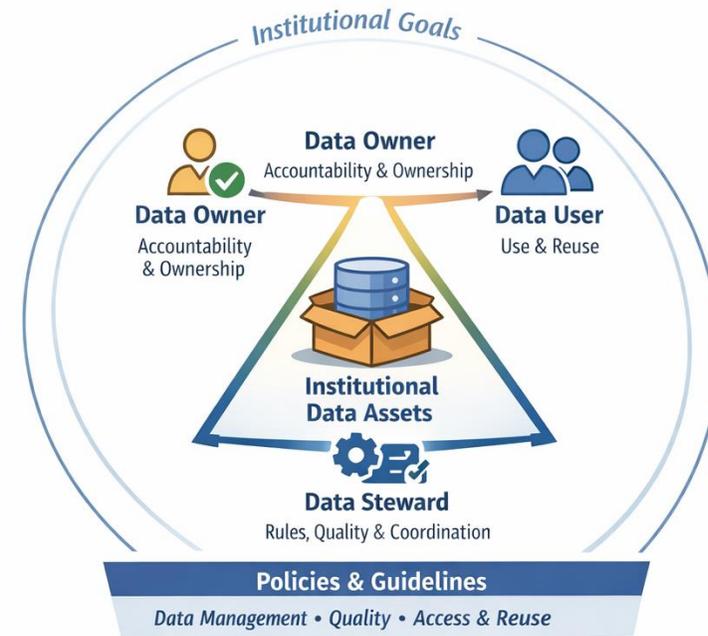| Type of Data Steward	Main Focus	Key Responsibilities	Typical Level
Policy / Strategic Data Steward	Strategy & governance	• Develop RDM and Open Science policies• Align with funder & legal requirements• Coordinate data stewardship networks	Institution / Organization
Research / Embedded Data Steward	Day-to-day research support	• Support DMPs• Advise on metadata & standards• Implement FAIR in research workflows	Project / Research group
Disciplinary Data Steward	Domain-specific expertise	• Advise on **discipline-specific standards & ontologies**• Support domain repositories & community practices• Bridge researchers with international infrastructures	Discipline / Research community
Infrastructure / Technical Data Steward	Systems & technology	• Support storage and computing services• Manage repositories & data publishing• Enable interoperability & automation	Institutional / National infrastructure

Minimal institutional model

A minimal institutional data stewardship model focuses on **clarity of roles, responsibilities, and processes**, without requiring large structural changes.

Key elements include:

- **Clear ownership of data** (who is responsible for which data assets)
- **Defined roles**, such as:
  - ➤ Data Owner
  - ➤ Data Steward
  - ➤ Data User
- **Basic policies and guidelines**, e.g.:
  - ➤ Data management principles
  - ➤ Data quality expectations
  - ➤ Access and reuse rules
- **Alignment with institutional goals**, such as research excellence, compliance, or digital transformation
- The minimal model should be **scalable** and adaptable over time.

Minimal Institutional Data Stewardship Model

Institutional Goals

Data Owner — Accountability & Ownership

Data Owner — Accountability & Ownership

Data User — Use & Reuse

Institutional Data Assets

Data Steward — Rules, Quality & Coordination

Policies & Guidelines
Data Management • Quality • Access & Reuse

✓ Clear Ownership   ✓ Defined Roles   ✓ Minimal Structure   ✓ Scalable Over Time

First steps

When introducing data stewardship, institutions should start small and focus on impact.
Recommended first steps:

- **Assess the current state**
  - ➤ Existing policies, practices, and pain points
- **Identify priority data domains**
  - ➤ Research data, administrative data, student data, etc.
- **Select pilot areas**
  - ➤ Choose motivated units or projects with clear needs
- **Define success criteria**
  - ➤ What does "better data stewardship" mean in practice?

Prioritisation should be based on:

- Institutional risk (legal, ethical, reputational)
- Strategic value of the data
- Available resources and readiness for change
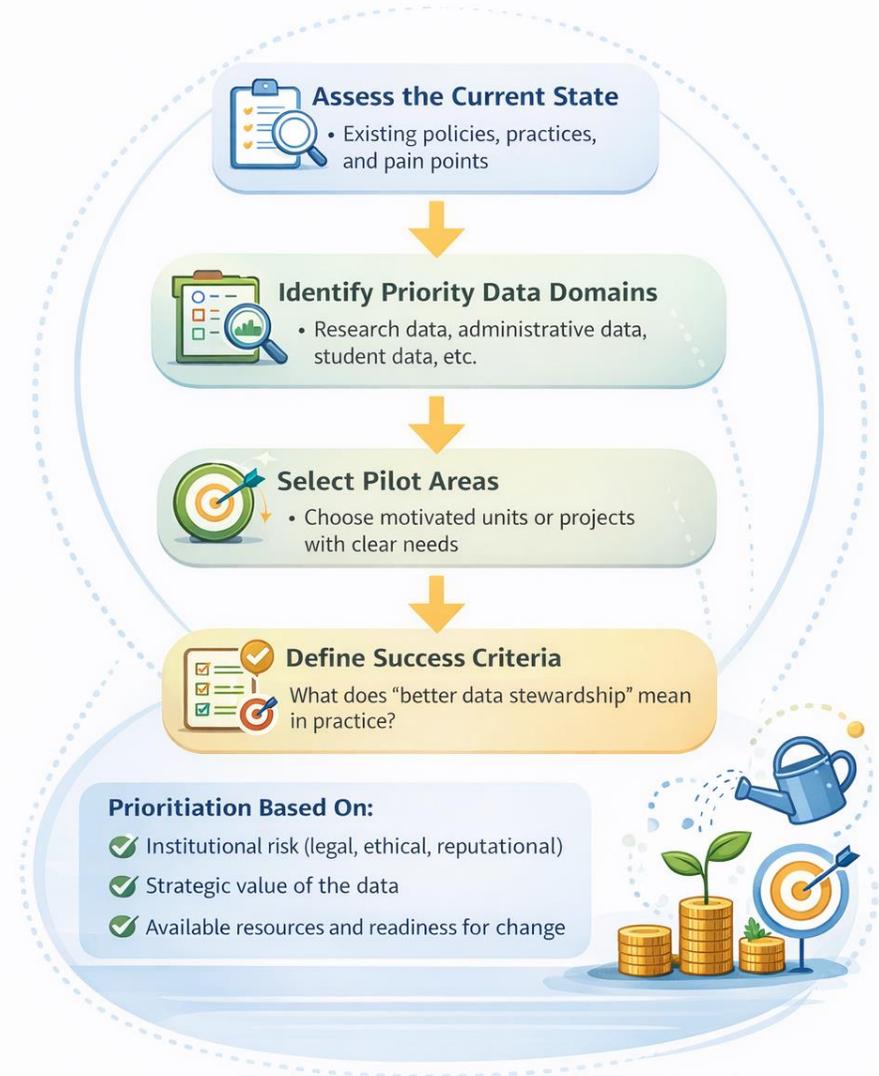

Starting Small with Data Stewardship Model

Assess the Current State
- Existing policies, practices, and pain points

Identify Priority Data Domains
- Research data, administrative data, student data, etc.

Select Pilot Areas
- Choose motivated units or projects with clear needs

Define Success Criteria
- What does "better data stewardship" mean in practice?

Prioritiation Based On:
- ✅ Institutional risk (legal, ethical, reputational)
- ✅ Strategic value of the data
- ✅ Available resources and readiness for change

Developing a short institutional plan (example)

Institutional Data Stewardship Plan

6~18 Months • Concise • Realistic • Action-Oriented

1. Vision & Objectives
- Why & what to achieve
- Institutional goals

2. Scope
- Data types
- Units / Projects

3. Roles & Responsibilities
- Data Owners, Stewards
- Accountability

6. Governance & Review
- Monitoring progress
- Annual updates

LIVING DOCUMENT
Review & Update

4. Actions & Milestones
- Priority actions
- 6–18 months timeline

5. Training & Awareness
- Build skills & engagement
- Workshops, Guides, Comms

Developing a short institutional plan – summary

A short institutional data stewardship plan should be:

- **Concise** – focused on priorities, not exhaustive detail
- **Realistic** – aligned with current capacity and resources
- **Action-oriented** – clearly defining what will be done next
- **Purpose of the plan**
- Provide clear direction for the **first implementation phase**
- Support decision-makers and operational staff
- Enable learning before scaling up

Key Components of the Plan (6–18 months)

- **Vision & objectives**
  Why data stewardship matters and what the institution wants to achieve
- **Scope**
  Which data types, units, or projects are covered initially
- **Roles & responsibilities**
  Clear ownership and accountability (even if provisional)
- **Key actions & milestones**
  A small number of concrete, time-bound actions
- **Training & awareness**
  Building skills, understanding, and engagement
- **Governance & review**
  Light mechanisms for monitoring progress and updating the plan

Training planning

Training and awareness activities are essential for building sustainable data stewardship capacity.

They support not only skills development, but also **shared understanding, responsibility, and cultural change** within the institution.

Key principles:

- Training should be **role-based** and aligned with institutional responsibilities
- Awareness-raising should reach **both decision-makers and operational staff**
- Learning should be **progressive**, combining basic understanding with practical application
- Activities should be embedded in **existing institutional structures** where possible
- Effective training helps translate data stewardship from a policy concept into **everyday practice**.

Target groups
- Institutional leadership and management
- Data stewards and data owners
- Researchers and project teams
- Support staff (IT, legal, research support, administration)

Training formats
- Introductory workshops (common understanding and terminology)
- Short internal courses (role-specific skills)
- Online self-learning materials (flexibility and scalability)

Awareness-raising activities
- Internal communication and guidance materials
- Events, seminars, and discussion forums
- Practical examples, pilots, and success stories

Training should be **aligned with the roles defined in the data stewardship plan** and designed to support **gradual, long-term cultural change**, rather than one-off compliance.

Training methodology

No.	Learning Approach	Description	Key Focus Areas / Activities	Learning Outcomes
1	**Competence-Based Learning**	Training structured around clearly defined RDM competencies aligned with researchers' roles and career stages (PhD students, senior researchers, data stewards, support staff).	• Data management planning• FAIR principles implementation• Data documentation & metadata• Data storage, security & ethics• Data sharing, reuse & preservation	Learners demonstrate role-relevant, practical RDM skills rather than only theoretical knowledge.
2	**Problem-Based Learning (PBL)**	Participants work on real-world RDM problems derived from research practice.	• Managing sensitive or personal data• Ensuring reproducibility in complex datasets• Compliance with funder and institutional RDM requirements	Ability to analyze problems, identify gaps, propose solutions, and reflect on outcomes.
3	**Case-Based Learning**	Use of discipline-specific RDM cases to illustrate best practices and common pitfalls.	• Data management failures and lessons learned• Successful data sharing and reuse cases• Ethical and legal challenges in research data	Improved contextual understanding and transfer of good practices across disciplines.
4	**Blended Learning**	Combination of multiple learning formats to maximize flexibility and engagement.	• Short e-learning modules• Live workshops (online/onsite)• Self-paced materials and toolkits• Guided practical assignments	Foundational knowledge combined with interactive and applied learning suited to different learning styles.

Training methodology

No.	Learning Approach	Description	Key Focus Areas / Activities	Learning Outcomes
5	**Train-the-Trainer Model**	Selected participants are trained as RDM trainers or champions to ensure scalability and long-term impact.	• Pedagogical skills for adult and academic learning• Adaptation of RDM content to disciplinary contexts• Facilitation of workshops and consultations• Access to reusable training materials	Sustainable institutional capacity building and multiplication of RDM expertise.
6	**Learning by Doing**	Hands-on activities embedded directly in participants' research workflows.	• Creating or reviewing real DMPs• Applying metadata standards to own datasets• Assessing datasets for FAIRness• Using institutional or disciplinary RDM tools	Immediate and effective application of RDM practices in real research contexts.
7	**Microlearning**	Delivery of content in short, focused learning units (5–15 minutes), each addressing a single RDM topic or skill.	• Writing a data management plan• Choosing the right repository• File naming and version control basics	Just-in-time learning through reusable, searchable, and easy-to-update modules.
8	**Continuous & Adaptive Learning**	RDM training treated as an ongoing, evolving process.	• Refresher modules and updates• Advanced or specialized training paths• Integration into onboarding and doctoral training• Feedback-driven improvement of materials	Continuous competence development aligned with evolving research, policy, and technical requirements.

Competence Areas vs Roles and Training Methods

RDM Competence Area	Core Competences (What learners can do)	R1	R2	R3	R4	Training Methods
RDM Fundamentals	Explain RDM concepts, roles, and responsibilities	✓	✓	✓	✓	Microlearning, e-learning
Data Management Planning (DMP)	Create, review, and update DMPs aligned with funder requirements	✓	✓	✓	✓	Learning by doing, case-based
FAIR Data Principles	Apply FAIR principles to datasets and workflows	✓	✓	✓	✓	Problem-based, practical exercises
Data Documentation & Metadata	Select and apply appropriate metadata standards	✓	△	✓	✓	Hands-on workshops, microlearning
Data Storage & Backup	Choose secure storage and backup solutions	✓	△	✓	✓	Case-based, practical demos
Data Security & Ethics	Handle sensitive data in compliance with ethical and legal requirements	✓	✓	✓	✓	Problem-based learning, cases
Legal & IP Aspects	Understand licensing, consent, and data ownership	△	✓	✓	✓	Case-based learning
Data Sharing & Reuse	Select repositories and prepare data for sharing	✓	✓	✓	✓	Learning by doing, blended
Data Preservation	Apply long-term preservation strategies	△	△	✓	✓	Workshops, expert-led sessions
Tools & Infrastructure	Use institutional and disciplinary RDM tools	✓	△	✓	✓	Hands-on labs
Data Quality & Reproducibility	Assess and improve data quality and reproducibility	✓	✓	✓	✓	Problem-based, real cases
Consultation & Support Skills	Advise researchers on RDM practices	–	–	✓	✓	Role-play, case discussions
Training & Facilitation Skills	Design and deliver RDM training	–	–	△	✓	Train-the-trainer, learning by doing

Theoretical Foundations for Research Data Management Training

Why use theory in RDM training design?
Research Data Management is not only a technical challenge, but a **social, organisational, and cultural change.**

Theoretical concepts help to:
- explain **why RDM adoption succeeds or fails**
- design training that fits **real research practices**
- ensure **sustainability, legitimacy, and scalability**

Combining organisational, socio-technical, and learning theories allows us to:

- align RDM training with **institutional and funder expectations**
- support **behavioural change**, not only knowledge transfer
- embed RDM skills into **everyday research workflows**

Linking Theoretical Concepts to RDM Training Modules

No.	Theory / Concept	Core Idea	Relevance to RDM	Practical Implications for RDM Training & Implementation
1	**Change Management Theory / Agents of Change**	Organizational change requires structured processes and individuals who actively promote and support change.	RDM implementation represents a cultural and behavioral change in research practices.	Identification and empowerment of RDM champions (data stewards, trainers) to act as agents of change; phased implementation and communication strategies.
2	**Science and Technology Studies (STS)**	Scientific knowledge and practices are co-produced with technologies and social structures.	RDM practices are shaped by interactions between researchers, tools, policies, and infrastructures.	Emphasis on socio-technical systems in RDM training; focus on workflows, tools, and social practices rather than tools alone.
3	**Institutional Theory**	Organizational behavior is influenced by norms, rules, and institutional pressures.	RDM adoption is driven by funder mandates, institutional policies, and professional norms.	Alignment of RDM training with institutional strategies, policies, and incentives to increase adoption and compliance.
4	**Diffusion of Innovation Theory**	Innovations spread through social systems over time via adopters and communication channels.	RDM practices and tools diffuse unevenly across disciplines and research communities.	Targeting early adopters, showcasing success stories, and using peer influence to accelerate RDM uptake.

Linking Theoretical Concepts to RDM Training Modules

No.	Theory / Concept	Core Idea	Relevance to RDM	Practical Implications for RDM Training & Implementation
5	**Communities of Practice**	Learning occurs through participation in shared practices within a community.	RDM knowledge is often developed informally within disciplinary or institutional communities.	Creation of RDM communities (e.g. data steward networks, researcher forums) to support peer learning and knowledge exchange.
6	**Actor–Network Theory (ANT)**	Social and technical elements form networks in which agency is distributed.	Data, tools, standards, policies, and people all act as "actors" in RDM ecosystems.	Mapping RDM actor-networks to identify bottlenecks, dependencies, and leverage points for intervention.
7	**Legitimacy Theory**	Practices gain acceptance when they are perceived as appropriate and valuable.	RDM must be seen as legitimate research practice, not just administrative burden.	Framing RDM in terms of research quality, reproducibility, and scientific integrity to increase acceptance.
8	**Capability Maturity Model (CMM)**	Organizational capabilities develop through defined maturity stages.	RDM implementation progresses from ad hoc to optimized practices.	Use of maturity models to assess current RDM state, plan improvements, and measure progress over time.

RDM as a Socio-Technical System: Integrating People, Data, and Tools

Section	Description
Theoretical Foundation	**Science and Technology Studies (STS)**; **Actor–Network Theory (ANT)**
Rationale	Research Data Management is shaped by interactions between social actors (researchers, data stewards, institutions) and non-human actors (datasets, tools, repositories, policies). Understanding RDM as a socio-technical system enables more effective and sustainable adoption of RDM practices.
Target Group	PhD candidates, postdoctoral researchers, senior researchers, data stewards, research support staff
Competence Area	Socio-technical understanding of RDM; workflow integration; system awareness
Learning Outcomes	After completing the module, participants are able to:• explain RDM as a socio-technical system• identify key human and non-human actors in their RDM environment• analyze dependencies and bottlenecks in RDM workflows• adapt RDM practices to their own research context
Key Topics	• RDM beyond tools and policies• STS perspective on research practices• Actor–network mapping in RDM• Institutional and disciplinary RDM ecosystems
Learning Activities	• Short microlearning unit introducing STS and ANT concepts• Case-based discussion of a real research workflow• Hands-on exercise: mapping the RDM actor-network for participants' own projects• Group reflection and peer feedback
Training Methods	Blended learning; learning by doing; case-based learning; problem-based learning
Assessment Methods	• Submission of an RDM actor-network map• Short reflective statement on identified challenges and improvement actions
Evidence of Competence	• Completed actor-network diagram• Reflection demonstrating understanding of socio-technical dependencies
Duration & Format	2–3 hours (online or onsite workshop) + 20-minute self-paced microlearning
Tools & Resources	Whiteboard or online collaboration tools (e.g. Miro, Padlet); example RDM workflows; institutional RDM guidelines
Scalability & Sustainability	Module materials reusable across disciplines; adaptable by trained RDM trainers; suitable for train-the-trainer programmes
Expected Impact	Improved integration of RDM into everyday research workflows; increased awareness of systemic factors influencing RDM success

Many institutions benefit from a hybrid model, combining **Train the Trainer** for scalability**, Self-Development** for flexibility, and **Agents of Change** to support long-term cultural transformation.

Aspect	Train-the-Trainer	Self-Development	Agent of Change
Main idea	Train a small group who then train others	Individuals develop RDM skills independently	Empower key individuals to drive institutional change
Primary focus	Knowledge transfer and scalability	Personal learning and skill building	Cultural and organisational transformation
Typical actors	Trainers, librarians, data stewards	Researchers, support staff	Data stewards, research leaders, champions
Speed of implementation	Medium to fast	Slow to medium	Medium
Scalability	High	Low	Medium
Institutional impact	Moderate	Low	High
Sustainability	High if trainers are supported	Depends on motivation	High if leadership supports it
Resources needed	Training materials, coordination	Time and access to learning resources	Time, authority, and institutional backing
Strengths	Efficient, consistent messaging, scalable	Flexible, personalised learning	Long-term impact, embeds RDM in culture
Limitations	Risk of knowledge dilution	Uneven skill levels, limited reach	Requires strong mandate and support
Best use case	Rapid rollout of RDM training	Early or individual adoption	Strategic RDM implementation

Preparation of RDM training materials

Area	Principle	Description	Practical Application
Skills for EOSC	EOSC-aligned competencies	Training materials support the development of skills required for effective participation in the EOSC ecosystem.	Mapping materials to EOSC/Open Science competence frameworks; role-based learning paths.
FAIR by Design	FAIR learning materials	FAIR principles are applied to training materials themselves, not only taught as content.	Metadata for training units; open formats; PIDs; clear versioning and licensing (e.g. CC).
Modularity & Reuse	Modular design	Content is developed as reusable and recombinable learning units.	Microlearning modules; reusable templates; discipline-agnostic core content with extensions.
Openness	Open access & licensing	Training materials are openly available and legally reusable.	Publishing materials in open repositories; use of Creative Commons licences.
Practice Orientation	Learning by doing	Materials are grounded in real research workflows and tasks.	DMP exercises; dataset preparation; FAIRness assessment activities.
Disciplinary Sensitivity	Context-aware design	Materials reflect disciplinary diversity in data types, methods, and cultures.	Discipline-specific cases layered onto generic RDM principles.
Inclusiveness & Accessibility	Inclusive design	Materials are accessible to diverse audiences and learning needs.	Clear language; multiple formats; accessibility-aware design.
Quality & Evidence	Evidence-based content	Content reflects current policies, standards, and best practices.	Alignment with funder requirements; use of real-world cases.
Sustainability	Living training resources	Materials are maintained, updated, and improved over time.	Feedback loops; scheduled reviews; train-the-trainer model.

FAIR & EOSC Checklist for RDM Training Materials

Guiding principles for developing quality **Research Data Management** training **materials** aligned with EOSC (European Open Science Cloud) and FAIR-by-design standards.

EOSC Skills
Focus: EOSC-aligned competencies
- ✔ Mapped to EOSC competence frameworks
- ✔ Role-based learning pathways

FAIR Learning Objects
Focus: FAIR by design
- ✔ Clear learning objectives & metadata
- ✔ Open formats (e.g, PDF, CSV)
- ✔ Persistent identifiers (PIDs)
- ✔ Deposited in open repositories

Modular Units
Focus: Reusable, adaptable content
- ✔ Self-contained microlearning units
- ✔ Flexible, recombinable modules

Open & Licensed
Focus: Open availability
- ✔ Creative Commons licemces preferred
- ✔ Published in open repositories

Practice-Oriented
Focus: Real research tasks
- ✔ Aligned with real RDM workflows
- ✔ Exercises for drafting DMPs, preparing datasets, etc.

Discipline-Sensitive
Focus: Context-aware design
- ✔ Discipline-specific examples and cases
- ✔ Adaptable, inclusive approach

Accessible & Inclusive
Focus: Clear language & inclusive design
- ✔ Multiple formats, accessibility features

Living Resources
Focus: Sustainability and improvement.
- ✔ Regularly reviewed andupdated
- ✔ Feedback-driven improvements

- RDM training materials are developed to support **EOSC-relevant skills** and Open Science practices
- **FAIR principles apply to the materials themselves**, not only to the data they describe
- Modular, open, and reusable design enables **scalability and sustainability**
- Practice-oriented and discipline-sensitive content ensures **real-world applicability**
- Continuous review and train-the-trainer approaches support **long-term impact**

Well-designed RDM training materials are FAIR, open, reusable, and embedded in real research practice — enabling sustainable adoption of RDM across research communities.
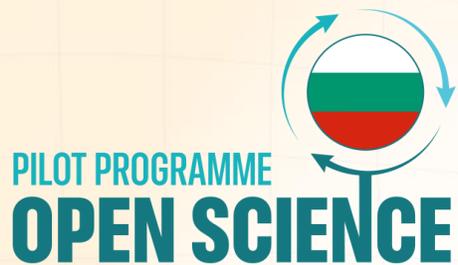
Magdalena Szuflita-Żurawska, Gdansk University of Technology | Head of the Scientific-Technical Information Services, Open Science Competence Center Leader; | Gdańsk Tech Rector's Plenipotentiary for Open Science

magdalena.szuflita@pg.edu.pl