17-04-2026

Deliverable GN5-2
D4.1 Data Management Guidelines

Contractual Date:	31-03-2026
Actual Date:	17-04-2026
Grant Agreement No.:	101194278
Work Package:	WP4
Task Item:	T4
Nature of Deliverable:	R
Dissemination Level:	PU
Lead Partner:	GÉANT
Document ID:	GN5-2-26-EB6250
Authors:	Mark van de Sanden (GÉANT); Dave Heyns (GÉANT); Helen Claire (Jisc); Magdalena Rzaca (GÉANT)

Abstract

This document provides reference material for members of the GÉANT community on the research lifecycle in terms of the curation and archival of research data and related artefacts. This context is useful and important for the many NRENs that provide services in support of their national research communities. This resource also provides deeper insight into European project outputs.

This deliverable is intended for a broad audience within the GÉANT community, including NREN staff, project support teams, repository and infrastructure operators, data stewards, and those working in technical, operational, project, policy, and support roles. It is designed for readers who require an introduction to the research data lifecycle, implementation of the FAIR data principles, and the wider European data-policy context.

Co-funded by the European Union

© GÉANT Association on behalf of the GN5-2 project. The research leading to these results has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101194278 (GN5-2).
Co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them.

Contents

Executive Summary 5
1 Introduction 7
1.1 How To Use This Document 7
2 European Strategy for Data 9
2.1 Laws and Policy Relevant to the European Strategy for Data 9
2.1.1 Data Act 9
2.1.2 Data Governance Act 11
2.1.3 Regulation on the Free Flow of Non-Personal Data 12
2.1.4 Open Data Directive and High-Value Datasets 14
2.1.5 Common European Data Spaces 16
2.2 Other Related EU Digital Legislation 17
2.2.1 AI Act 17
2.2.2 Cyber Resilience Act 19
2.2.3 Digital Services Act 21
2.2.4 Digital Markets Act 23
2.3 Summary 24
3 FAIR Data 26
3.1 FAIR Principles 26
3.1.1 FAIR Principles and Machine-Actionable Data 26
3.2 FAIR and the European Open Science Cloud 27
3.3 FAIR Cornerstone Technologies 29
3.3.1 Persistent Identifiers 30
3.3.2 Semantic Artefacts and Terminologies 31
3.3.3 Knowledge Graphs 34
3.4 Assessing FAIRness 36
3.4.1 RDA FAIR Data Maturity Model 36
3.4.2 FAIR Assessment Tools 39
3.5 FAIR Digital Objects 40
3.5.1 FAIR Digital Objects Forum 41
3.5.2 FAIR Signposting 42
3.5.3 FAIR Data Points 43
3.5.4 FAIR RO-Crates 44
4 Research Data Management 46
4.1 Research Data Management Best Practices 46
4.1.1 Before Research 47
4.1.2 During Research 49
4.1.3 After Research 52
4.2 Data Curation 57
4.2.1 CoreTrustSeal - Data Curation Levels 57

4.2.2 Data Curation Network - C-U-R-A-T-E-D steps model 58
4.3 Data Quality 58
4.3.1 Towards a Data Quality Framework for EOSC 59
4.3.2 EOSC EDEN Quality Assurance of Research Outputs 60
5 Training on Research Data Management Practices 63
5.1 Institutional Support for Research Data Management Training 63
5.2 National and Thematic Data Services and Competence Centres 64
5.3 Project Initiatives and International Organisations 64
5.4 Resources for Data Professionals 65
5.5 Quality Assurance and FAIR Training 65
5.6 The Role of NRENs and Research Infrastructures 66
6 Conclusions 67
Glossary 68
References 69

Figures

Figure 3.1: The seven steps of the FAIR Interoperability Framework with guiding questions 28
Figure 3.2: Organisational structure recommended for use by FAIR-IMPACT 33
Figure 3.3: Three main approaches to semantic artefact catalogue interoperability (FAIR-IMPACT) 33
Figure 3.4: SKG-IF Architecture and Data Model 35
Figure 3.5: EOSC Node research product catalogues: onboarding data sources (blue cylinders) and research products (red circles) from EOSC Nodes (light blue rectangles) 36
Figure 3.6: Basic concept of a FAIR Digital Object 40
Figure 3.7: DOIP used to create a single integrated FDO space 42
Figure 3.8: The use of typed web links to link resources from a landing page 42
Figure 3.9: FAIR Data Point high-level architecture 43
Figure 3.10: RO-Crates can link and be linked to research objects from many different sources using persistent identifiers like ORCIDs and DOIs) 44
Figure 4.1: Harmonised Research Data Lifecycle (RDL) Model created by the RDA working group on Mapping the Landscape of Digital Research Tools 46
Figure 4.2: Four quality aspects (i.e., science, product, stewardship and service) throughout a dataset lifecycle, including key stages and attributes associated with each 59
Figure 4.3: Reuse Fitness model 61

Tables

Table 2.1: High-level comparison of the Data Act and GDPR 10
Table 2.2: High-level comparison of the Data Governance Act and GDPR 12
Table 2.3: High-level comparison of the FFD Regulation and the GDPR 13
Table 2.4: High-level comparison of Open Data Framework / high-value datasets and GDPR 15
Table 2.5: High-level comparison of the AI Act and GDPR 18

Table 2.6: High-level comparison of the AI Act and GDPR 20
Table 2.7: High-level comparison of DSA and GDPR 22
Table 2.8: High-level comparison of the Digital Markets Act and the GDPR 23
Table 3.1: FAIR Data Maturity Model indicators 38
Table 3.2: Overview of FAIR assessment tools 40
Table 5.1: Useful resources for research data management 65

Executive Summary

The organisation and preservation of data is challenging for National Research and Education Networks (NRENs) as NREN staff, project support teams, repository and infrastructure operators, data stewards, and those working in technical, operational, project, policy, and support roles try to make sense of changing legislation, strict GDPR compliance and data security issues with limited resources.

As part of its objective to provide a joint-NREN approach to supporting Open Science, GN5-2 Work Package 4 Task 4 has developed these guidelines for the management of research data and related artefacts. It sets out the European legal context and provides examples of data management best practice in projects undertaken within the GÉANT community. This information provides a starting point for NRENs as they review, set out and update processes within their own organisations.

The report begins by outlining the European Strategy for Data and its related legislative instruments—including the Data Act, Data Governance Act, and Regulation on the Free Flow of Non-Personal Data—which collectively establish the regulatory environment for data sharing, access, interoperability, and sovereignty across Europe. It also highlights the Open Data Directive the increasing emphasis on secure, trustworthy, cross-border data flows.

Common European Data Spaces are a key implementation aspect of the European Strategy for Data, not only for data sharing, but also in the technical and organisational conditions needed to make such sharing trustworthy, interoperable, and workable in practice.

The wider EU digital regulatory framework includes other acts relevant to research infrastructures, digital services, connected products, and data-driven innovation – namely, the AI Act, Cyber Resilience Act, Digital Services Act, and Digital Markets Act.

Following on from the legislative overview, a substantial portion of the document is devoted to guidelines for improving the management and stewardship of research data by making it findable, accessible, interoperable, and reusable, based on EOSC use and examples from other Horizon Europe projects. It also describes the key technologies (i.e., persistent identifiers, semantic artefacts, and knowledge graphs) used to implement the FAIR principles and how the FAIRness of research data can be assessed.

The document then provides structured Research Data Management best-practice aligned with research data lifecycles. These include recommendations for planning and funding, data collection, processing, analysis, secure storage, publication, preservation, access control, and transformation. Emphasis is placed on using community standards, trusted digital repositories, rich metadata, reproducible workflows, data quality assurance, and appropriate handling of sensitive or confidential data—including through trusted research environments.

Further sections explore data curation frameworks, including the CoreTrustSeal curation levels and the Data Curation Network's CURATED model, offering concrete guidance for ensuring long term usability and trustworthiness of research outputs. The document also summarises European developments on data quality, highlighting the need for community standards, quality indicators, and systematic approaches for assessing contextual, technical, and metadata quality.

Finally, the report addresses the growing need for training in research data management, documenting related current initiatives across NRENs, national services, thematic infrastructures, and EU-funded projects. Given the

recognised skills gap in Europe, this section underscores the importance of coordinated training programmes, capacity building, and the development of FAIR, interoperable learning materials.

Overall, the deliverable provides reference material to equip the GÉANT community with the context, tools, and recommended best practices needed to support high quality, FAIR, and sustainable data management within their constituencies. Strengthening an understanding of research data management across NRENs and research organisations remains essential for maximising the long-term value and impact of European research.

1 Introduction

The European NREN community supports the region's research lifecycle in the most fundamental ways. The NRENs provide researchers and research communities network connectivity in a national, European, and global context. The community also provides the identity federation that underpins academic research and other related activities in Europe and beyond. In addition, the NRENs' security, storage, compute, and other bespoke industry sector services contribute towards the success of European research collaboration. It is therefore important that the NREN community has access to a comprehensive reference guide to the research lifecycle and related activities, particularly those in support of data curation and sharing. This document attempts to provide such a resource.

Research Data Management (RDM) Guidelines provide a shared framework for how research data should be planned, collected, processed, analysed, stored, preserved, published and shared across the research lifecycle. Their purpose is to support high quality, reproducible, and transparent science while ensuring researchers comply with funder policies, institutional standards, and broader European Open Science objectives. They also stress the importance of Data Management Plans as "living documents" that evolve throughout a project, supporting good practice and compliance with funder mandates such as those introduced under Horizon 2020's Open Research Data policies.

A driving force behind these guidelines is the need to align diverse institutional and national RDM practices across Europe, reducing confusion for researchers who move between organisations or collaborate internationally. Science Europe's *Practical Guide to the International Alignment of Research Data Management* [1] addresses this directly by offering common core requirements for Data Management Plans (DMPs), evaluation rubrics, and criteria for trustworthy repositories—all intended as minimum standards for organisations and disciplines to adapt (e.g., templates for organisations, researchers, and reviewers).

The European RDM landscape is also strongly shaped by the principles of findable, accessible, interoperable, and reusable (FAIR) data. These internationally recognised principles help ensure that research outputs can be effectively reused by both humans and machines, strengthening scientific integrity and long-term value of data. Institutions such as ETH Zürich highlight the role of RDM guidelines in anchoring data practices within scientific integrity frameworks, establishing responsibilities, and ensuring alignment with national and international regulations.

While the title of this deliverable is *Data Management Guidelines* it was also important to describe the context in which these guidelines need to be placed as the European Strategy for Data and related European Acts come into play. The adoption of Open Science and FAIR principles have also been driving forces within the research community during the last decade. While Section 2 (European Strategy for Data) and 3 (FAIR data) describe the context in which the RDM is placed, Section 4 provides an overview of RDM best practices. The RDM best practices are provided as recommendations built around the Research and Research Data Lifecycle. Adopting best practices, however, is not automatic. Researchers need to acquire the relevant skills and support, which is addressed in Section 5 Training on Research Data Management practices.

1.1 How To Use This Document

This guide is intended as a practical reference for NREN staff and members of the wider GÉANT community who support research projects, data stewardship, repositories, infrastructure, and compliance activities. It may be

used either as a general introduction to the European research data management landscape, or as a targeted reference by consulting the sections on FAIR data, research data management, curation, quality, and training according to the reader's role and project phase.

Audience

This document is intended for a broad audience within the GÉANT and NREN community, including NREN staff, project support teams, repository and infrastructure operators, data stewards, and those working in technical, operational, project, policy, and support roles. It is designed for readers who require a structured introduction to the research data lifecycle, implementation of the FAIR data principles, and the wider European data-policy context.

Disclaimer: Scope and Limitations

This document provides a high-level reference guide intended for the GÉANT and NREN community. It does not replace project-specific legal advice, funder requirements, institutional policies, ethics review processes, or the detailed operational instructions contained in project-level Data Management Plans.

2 European Strategy for Data

The European Strategy for Data, and associated Acts are relevant to NRENs due to the control and value of the research data generated by users and supported by NRENs' networks and services. Each country must develop its own policies and ensure both protection and harmonisation for data reuse. Data is vital to the research environment. As such, NRENs are now tasked with making data management trustworthy, interoperable and practical. Understanding the legislation behind data management is crucial for organisations to ensure compliance, mitigate risk, avoid penalties and build the trust of users.

The **European Strategy for Data** [2] is the European Union's overarching policy framework for creating a single market for data in which data can move freely and securely across sectors and member states, while ensuring appropriate protections for individuals and organisations and the retention of control over the data they generate. The Strategy's purpose is to increase the availability and reuse of data for innovation, research, public services, and economic growth, while upholding European values, rights, trust, and data sovereignty. The Strategy is implemented through legislation, governance mechanisms, and technical initiatives, most notably the **Data Governance Act**, the **Data Act**, the **Regulation on the free flow of non-personal data**, the EU's **Open Data Framework** including **high-value datasets**, and the development of **Common European Data Spaces**.

In summary [3], the European Strategy for Data (ESD) aims to:

- Increase the availability and reuse of data across the EU.
- Establish trustworthy governance rules for access to and sharing of data.
- Support innovation, including AI, through access to high-quality data and digital infrastructure.
- Strengthen the EU's competitiveness and strategic autonomy in the data economy.

The following sections provide a short overview of the main legislative pillars of the European Strategy for Data, related EU digital legislation, the Regulation on the free flow of non-personal data, and the framework on high-value datasets.

2.1 Laws and Policy Relevant to the European Strategy for Data

This section reviews the legislative and policy instruments that support the European Strategy for Data. The two main legislative pillars are the Data Act (DA) and the Data Governance Act (DGA). The wider EU framework includes the Regulation on the Free Flow of Non-Personal Data (FFD), the Open Data Framework, including high-value datasets, and the development of Common European Data Spaces. Together, these instruments and initiatives are intended to strengthen trust in data sharing, improve access to and use of data, support interoperability, reduce vendor lock-in, and promote the movement, accessibility, and reuse of data across the European Union. Accordingly, this section also considers how these instruments interact with the already-established General Data Protection Regulation (GDPR).

2.1.1 Data Act

The Data Act (Regulation (EU) 2023/2854) [4] harmonises the rules on the access to and use of data across the European Union. Its purpose is to make data generated by connected digital devices and related services more accessible and usable, improve fairness in the data economy, reduce barriers to switching between data-

processing services, and support interoperability. It is one of the two main legislative pillars of the European Strategy for Data. The Data Act was published on 22 December 2023, entered into force on 11 January 2024, and has applied since 12 September 2025.[1]

2.1.1.1 Main Features of the Data Act

The main aspects of the Data Act are:

- Greater access to data generated by connected devices and related services (i.e., apps).
- Fairer data-sharing rules in business-to-business contexts.
- Protection against unfair contractual terms.
- Interoperability requirements for easier transfer between cloud and other data-processing services.
- Public-sector access requirements to privately held data in certain situations of exceptional need.

2.1.1.2 Data Act and General Data Protection Regulation (GDPR)

The Data Act and the GDPR (Regulation 2016/679) [5] serve different but overlapping purposes. The GDPR governs the processing of personal data and protects the rights of individuals, whereas the Data Act regulates access to and use of data, including non-personal and mixed datasets, particularly in relation to digital devices and apps. The Data Act applies without prejudice to EU data protection law, including the GDPR. This means that where personal data is involved, organisations must still comply with GDPR requirements such as lawful basis, transparency, purpose limitation, and data minimisation.

Topic	Data Act	GDPR
Purpose	Fair access to and use of data	Protection of personal data
Scope	Mainly non-personal and mixed data, especially data from connected products and related services	Personal data only
Main logic	Access, sharing, portability, interoperability, competition	Lawfulness, fairness, transparency, minimisation, security
Main addressees	Data holders, manufacturers, service providers, cloud/data-processing providers	Controllers and processors
Relationship	Applies without prejudice to EU data protection law	Continues to govern personal-data processing

Table 2.1: High-level comparison of the Data Act and GDPR

Why this matters for NRENs

The Data Act is relevant to NRENs supporting connected services, cloud environments, data-sharing infrastructures, or research projects involving access to machine-generated or service-generated data.

[1] EU Regulations enter into force 20 days after publishing, meaning they have been formally and legally adopted by the EU. However, the date of application is when EU Regulations become mandatory.

Example

A research project uses data from connected laboratory devices stored on a cloud platform. The Data Act may be relevant to questions surrounding data access, sharing, and transfer, while the GDPR remains relevant if any of that data relates to identifiable individuals.

Practical Takeaway

For NRENs and research support environments, the Data Act may improve access to and reuse of certain types of data, especially in connected and digital service environments. However, where personal data is involved, the GDPR continues to apply and remains essential when assessing whether data can be lawfully shared or reused.

Practical Questions to Ask

- Does the dataset include personal data, non-personal data, or mixed data?
- Who is the data holder and who is requesting access?
- Does the request concern connected-device or service-generated data?
- Are GDPR obligations still triggered?
- Are there contractual confidentiality limitations?

2.1.2 Data Governance Act

The Data Governance Act (Regulation (EU) 2022/868) [6] is designed to increase trust in data sharing, strengthen mechanisms that increase data availability, and support the reuse of data across the European Union. It is a key legislative pillar of the European Strategy for Data and complements the Data Act's focus on mandatory access rights by outlining the governance conditions that make data sharing trustworthy and workable. The DGA entered into force on 23 June 2022 and has been applicable since 24 September 2023.

2.1.2.1 Main Features of the Data Governance Act

The main aspects of the DGA are:

- **Reuse of certain protected public-sector data**, including data affected by privacy, confidentiality, trade-secret, or intellectual-property restrictions.
- **Regulation of data intermediation services**, intended to create trustworthy neutral actors that facilitate data sharing.
- **A framework for data altruism**, allowing individuals and organisations to voluntarily make data available for research, healthcare, social welfare, and environmental protection.
- **Support for Common European Data Spaces**, intended to increase the amount of data available for access and reuse in trustworthy and secure environments.

2.1.2.2 Data Governance Act and GDPR

The Data Governance Act and the GDPR serve different purposes. The DGA focuses on trust, governance, and mechanisms for data sharing and reuse, whereas the GDPR governs the processing of personal data. The DGA does not override data protection law. Where personal data is involved, GDPR requirements continue to apply.

Topic	Data Governance Act	GDPR
Purpose	Create trustworthy frameworks for data sharing and reuse	Protect personal data
Scope	Reuse of certain protected public-sector data, data intermediation services, data altruism	Personal data only
Main logic	Trust, governance, neutrality, and sharing mechanisms	Lawfulness, fairness, transparency, minimisation, and security
Main addressees	Public-sector bodies, data intermediaries, data altruism organisations, other data-sharing actors	Controllers and processors
Relationship	Does not replace data protection law	Continues to govern personal-data processing

Table 2.2: High-level comparison of the Data Governance Act and GDPR

Why this matters for NRENs

The DGA is relevant to NRENs supporting research data infrastructures, trusted sharing environments, repository services, sectoral data spaces, or projects exploring controlled reuse of protected datasets.

Example

An NREN-supported research infrastructure wishes to facilitate controlled reuse of a protected public-sector dataset for scientific research through a trusted sharing environment or repository service. The Data Governance Act may be relevant here to the governance framework for reuse and intermediation, while the GDPR remains relevant wherever the dataset contains personal data.

Practical Takeaway

For NRENs and research support environments, the DGA is most relevant where trusted data-sharing frameworks, repository services, intermediaries, or controlled reuse models are being developed or supported. Note that the Data Act is mainly concerned with access to and use of data, whereas the Data Governance Act is mainly concerned with trust, governance, and mechanisms for data sharing and reuse.

Practical Questions to Ask

- Is the dataset public, protected, personal, or mixed?
- Is the intended activity sharing, reuse, or intermediation?
- Is a trusted intermediary or governance mechanism needed?
- Are GDPR, confidentiality, intellectual property, or trade-secret restrictions still triggered?
- Is data to be reused for research or another objective?

2.1.3 Regulation on the Free Flow of Non-Personal Data

The Regulation on the Free Flow of Non-Personal Data (Regulation (EU) 2018/1807, FFD Regulation) [7] establishes a framework to ensure that non-personal data can move freely across EU member states. Its purpose is to remove unjustified data-localisation requirements and support a more integrated internal market for cloud

and data-processing services. The FFD Regulation complements the GDPR by focusing on non-personal data, while mixed datasets remain subject to the relevant data-protection rules where personal data is involved. The FFD Regulation was adopted on 14 November 2018 and has applied since 28 May 2019.

2.1.3.1 Main Features of the FFD Regulation

The main features of the FFD Regulation are:

- Prohibition of unjustified data-localisation requirements, except where justified on grounds of public security.
- Cross-border access for competent authorities, even when data is stored in another Member State.
- Support for transfer between cloud and data-processing services, including through codes of conduct.
- Promotion of a more competitive internal market for data-processing services by reducing fragmentation across the EU.

2.1.3.2 FFD Regulation and GDPR

The FFD Regulation and the GDPR serve different but complementary purposes. The FFD Regulation applies to non-personal data and is intended to ensure its free movement within the EU, whereas the GDPR governs the processing of personal data. Where datasets are mixed, containing both personal and non-personal data, the GDPR continues to apply to the personal-data elements.

Topic	FFD Regulation	GDPR
Purpose	Ensure the free movement of non-personal data within the EU	Protect personal data
Scope	Non-personal data	Personal data only
Main logic	Free movement, removal of localisation barriers, internal market integration	Lawfulness, fairness, transparency, minimisation, and security
Main addressees	Member states, competent authorities, data-processing service users and providers	Controllers and processors
Relationship	Complements the GDPR for non-personal data	Continues to govern personal-data processing

Table 2.3: High-level comparison of the FFD Regulation and the GDPR

Why this Matters for NRENs

The FFD Regulation is relevant to NRENs supporting cross-border digital services, federated infrastructures, cloud environments, or research platforms processing large volumes of non-personal or mixed data.

Example

An NREN-supported research platform stores large volumes of machine-generated infrastructure data in a cloud service hosted in another Member State. Where that data is non-personal, the FFD Regulation supports cross-border storage and processing without unjustified localisation barriers, while the GDPR remains relevant if any part of the dataset includes personal data.

Making data portable or reusable does not in itself mean that it can be transferred or shared without further assessment; contractual, security, confidentiality, and data protection requirements may still apply.

Practical Takeaway

For NRENs and research support environments, the FFD Regulation is most relevant where cloud services, distributed infrastructures, and cross-border storage or processing arrangements are used for non-personal data. Its practical importance lies in supporting flexibility, reducing localisation barriers, and enabling more efficient use of digital infrastructure across the EU.

Practical Questions to Ask

- Is the dataset non-personal, personal, or mixed?
- Is there any national requirement affecting where the data must be stored or processed?
- Is the service cross-border or cloud-based?
- Do authorities need regulatory or audit access to the data?
- Are GDPR, confidentiality, security, or contractual restrictions also relevant?

2.1.4 Open Data Directive and High-Value Datasets

The Open Data Directive (ODD) creates the general EU framework for access to and reuse of public-sector information (the Open Data Framework). It is a Directive (as opposed to a Regulation) and so must be implemented by each Member State through national law.

By contrast, the rules on high-value datasets were specified in Commission Implementing Regulation (EU) 2023/138 [8] and adopted under Article 14(1) of the Open Data Directive. This means that the general open-data regime may vary somewhat by member state, due to the flexibility permitted in national transpositions of EU Directives, while the specific list of high-value datasets and the main publication and reuse conditions are set uniformly at the EU level through the Implementing Regulation. The Directive therefore establishes the general framework for the reuse of public-sector information, while the Implementing Regulation lays down the list of specific high-value datasets and the arrangements for their publication and reuse.

2.1.4.1 Open Data Framework

In 2003, the European Commission set up a legal framework for the reuse of public sector information (PSI) in many areas of activity (social, economic, geographical, business and educational information), This framework ensures fair, proportionate and non-discriminatory conditions for public sector information use. Determination of what information is public is addressed by member states, usually in freedom of information laws. The Framework is supported by the data.europa.eu portal and governed by the Data Governance Act and the Data Act.

2.1.4.2 High-Value Datasets

Under Commission Implementing Regulation (EU) 2023/138, public-sector bodies in all member states must make certain high-value datasets (HVDs) available for reuse. These datasets are considered to have particularly strong socio-economic value. Increasing their availability is intended to support innovation, transparency, public services, and cross-border data reuse. The Regulation identifies six thematic categories of HVDs:

- Geospatial data
- Earth observation and environment
- Meteorology

- Statistics
- Companies and company ownership
- Mobility

2.1.4.3 Arrangements for Publication and Re-Use

The Implementing Regulation requires that high-value datasets be made available:

- Free of charge
- In machine-readable formats
- Through application programming interfaces (APIs)
- By bulk download (where applicable)

HVDs must also be made available under conditions that support broad reuse and interoperability across the EU (although limited and temporary exemptions may apply in specific circumstances). This is intended to create a harmonised mandatory regime, specific to high-value data, to supplement the broader Open Data Framework created by the Directive.

2.1.4.4 Open Data Framework and the GDPR

The Open Data Framework promotes access to and reuse of public-sector information, but it does not remove legal constraints relating to personal data, confidentiality, intellectual property, or other protected interests. Where personal data is involved, the GDPR continues to apply. This means that openness and reuse under the Directive or the high-value datasets regime do not override data-protection obligations.

Topic	Open Data Framework / High-Value Datasets	GDPR
Purpose	Promote access to and reuse of public-sector information	Protect personal data
Scope	Public-sector information and designated high-value datasets	Personal data only
Main logic	Openness, reuse, interoperability, public value	Lawfulness, fairness, transparency, minimisation, and security
Main addressees	Public-sector bodies and reusers of public-sector information	Controllers and processors
Relationship	Does not remove data-protection obligations	Continues to govern personal-data processing

Table 2.4: High-level comparison of Open Data Framework / high-value datasets and GDPR

Why this Matters for NRENs

This framework is relevant to NRENs supporting data platforms, API-based services, repositories, or research environments that ingest, reuse, or build services on public-sector data.

Example

An NREN-supported service wishes to use openly available mobility or meteorological data published by a public-sector body through an API. In such a case, the Open Data Framework and the high-value datasets regime may

support reuse under harmonised conditions. However, if the dataset includes personal or otherwise protected information, additional legal safeguards remain relevant, including data-protection requirements where applicable.

Making data openly available does not in itself mean that all legal or practical barriers disappear; licensing, confidentiality, data quality, ethical, and data-protection considerations may still need to be assessed.

Practical Takeaway

For NRENs and research support environments, the Open Data Framework and high-value datasets are particularly relevant where public-sector data is reused through repositories, APIs, data services, or research infrastructures. Their practical importance lies in improving accessibility, interoperability, and reuse of data with high public and economic value, while still requiring attention to legal and governance constraints.

Practical Questions to Ask

- Is the dataset public-sector information or a designated high-value dataset?
- Is the data available under an open licence or another reuse condition?
- Is API or bulk access available?
- Does the dataset contain any personal, confidential, or otherwise protected information?
- Are quality, interoperability, attribution, or reuse conditions relevant?

2.1.5 Common European Data Spaces

Common European Data Spaces are a key implementation component of the European Strategy for Data. They are intended to enable trustworthy, interoperable, and secure data sharing within and across strategic domains, including health, open science, agriculture, manufacturing, energy, mobility, finance, public administration, skills, and the Green Deal, with additional sectoral data spaces such as tourism, media, language and cultural heritage [9].

Rather than being a single legislative act, data spaces are policy and infrastructure frameworks supported by the wider European Strategy for Data and the legal environment created by instruments such as the Data Governance Act and the Data Act.

2.1.5.1 Main Features of Common European Data Spaces

The main features specified by Common European Data Spaces are:

- Sector-specific and cross-sector data sharing.
- Interoperability across organisations and borders.
- Trusted governance and access mechanisms.
- Support for innovation, research, public services, and reuse.
- Alignment with European values, rights, and security requirements.

Why this Matters for NRENs

Common European Data Spaces are particularly relevant where NRENs support federated research environments, trusted data-sharing ecosystems, EOSC-related services, identity federation, secure connectivity, repositories, storage and compute services, and infrastructure for cross-border collaboration. Their relevance lies not only in data sharing itself, but also in the technical and organisational conditions needed to make such sharing trustworthy, interoperable, and workable in practice.

Example

An NREN-supported research environment participates in a cross-border health or research data-sharing ecosystem in which multiple institutions need secure connectivity, federated identity, interoperable access rules, and trusted infrastructure. A Common European Data Space is appropriate here as a framework for organising secure and trustworthy cross-border data sharing and reuse.

Creating a trusted data-sharing environment does not in itself remove legal, technical, ethical, or governance constraints; issues such as data protection, confidentiality, access control, interoperability, and security still need to be assessed.

Practical Takeaway

Common European Data Spaces are relevant to NRENs and research support environments because they depend on trusted digital infrastructure, secure connectivity, identity and access management, interoperability, and cross-border collaboration. NRENs may therefore play an important enabling role even where they are not directly responsible for the legal governance of the data space itself.

Practical Questions to Ask

- Is the activity part of a sectoral or cross-sector data-sharing ecosystem?
- Are trusted governance, access, and interoperability mechanisms needed?
- Is secure connectivity or federated identity required?
- Are cross-border legal, technical, or organisational issues likely to arise?
- What role does the NREN play: service and tool provider, infrastructure provider, federated access enabler, repository or platform supporter, or project partner?

2.2 Other Related EU Digital Legislation

In addition to the instruments directly linked to the European Strategy for Data, the wider EU digital regulatory framework includes other acts that are relevant to research infrastructures, digital services, connected products, and data-driven innovation – namely, the AI Act, Cyber Resilience Act, Digital Services Act, and Digital Markets Act. These do not primarily regulate data access and reuse, but they shape the broader legal environment in which data and digital services operate and are reviewed in the following subsections accordingly.

2.2.1 AI Act

The AI Act (Regulation (EU) 2024/1689) [10] establishes the EU's legal framework for artificial intelligence. It is the first comprehensive EU regulation specifically addressing AI and takes a risk-based approach. Its purpose is to promote trustworthy AI while protecting health, safety, and fundamental rights. Under the AI Act, an AI system is a machine-based system that operates with some degree of autonomy and uses inference to generate outputs—such as predictions, recommendations or decisions—that influence physical or virtual environments. The AI Act entered into force on 1 August 2024. Some provisions are already applicable, including the rules on prohibited AI practices and, from 2 August 2025, the rules on general-purpose AI (GPAI) models and some further obligations, including certain transparency obligations, are expected to apply from 2 August 2026.[2]

[2] The European Commission has proposed targeted simplification measures under the Digital Omnibus, including changing the timeline, but this remains a proposal and is not yet the final law.

2.2.1.1 Main Features of the AI Act

The main aspects of the AI Act are as follows:

- A four-tier, risk-based structure, covering:
  - Prohibited AI practices.
  - High-risk AI systems.
  - AI systems subject to transparency obligations.
  - Other AI systems presenting minimal or no specific regulatory risk.
- Strict requirements for high-risk AI systems, including obligations relating to risk management, data governance, technical documentation, logging, accuracy, robustness, cybersecurity, and human oversight.
- Transparency obligations for certain AI systems, such as where users need to be informed that they are interacting with AI or that content has been artificially generated or manipulated.
- Specific obligations for GPAI models, including additional duties for GPAI models with systemic risk.
- Governance and enforcement mechanisms involving the European AI Office, national competent authorities, and market surveillance authorities.

In simple terms, the AI Act works by assigning stricter obligations to AI systems considered higher risk, while imposing fewer or no specific obligations on lower-risk systems.

2.2.1.2 AI Act and the GDPR

The AI Act regulates AI systems through a risk-based framework focused on safety, transparency, governance, and fundamental-rights protection. The AI Act does not replace data protection law. Where an AI system involves the processing of personal data, GDPR requirements continue to apply alongside the AI Act.

Topic	AI Act	GDPR
Purpose	Regulate AI systems and promote trustworthy AI	Protect personal data
Scope	AI systems and certain GPAI models	Personal data only
Main logic	Risk-based regulation, safety, transparency, governance, oversight	Lawfulness, fairness, transparency, minimisation, and security
Main addressees	Providers, deployers, importers, distributors, product manufacturers and other relevant operators	Controllers and processors
Relationship	Adds AI-specific obligations	Continues to govern personal-data processing
Fines	Up to €35 million or 7% of worldwide annual turnover for the most serious infringements; lower tiers include €15 million or 3%, and €7.5 million or 1% for incorrect information	Up to €20 million or 4% of worldwide annual turnover for the most serious infringements; lower tier: up to €10 million or 2% of worldwide annual turnover for less serious infringements

Table 2.5: High-level comparison of the AI Act and GDPR

Why this Matters for NRENs

The AI Act is relevant to NRENs supporting AI-enabled research services, compute environments, data platforms, testing infrastructures, or collaborative projects involving the development or deployment of AI systems.

Example

An NREN-supported research project develops or deploys an AI system that analyses research data, including data that may relate to identifiable individuals. In such a case, the AI Act may be relevant to the classification of the system, transparency obligations, governance, and possible high-risk requirements, while the GDPR remains relevant wherever personal data is processed.

Using AI in research or service environments in line with the AI Act does not in itself mean that the system is exempt from legal or ethical requirements; questions of personal data, transparency, risk, accountability, and governance must still be assessed.

Practical Takeaway

For NRENs and research support environments, the AI Act is particularly relevant where AI tools, AI-enabled services, or research infrastructures support the development, testing, deployment, or use of AI systems. Its practical importance lies in introducing a structured framework for AI governance, transparency, and risk management, which may apply alongside the GDPR and other legal requirements.

Practical Questions to Ask

- Is the tool or service likely to qualify as an AI system under the AI Act?
- Could the system fall into a prohibited, high-risk, transparency, or GPAI category?
- Does the system process personal data and therefore also trigger GDPR obligations?
- Are documentation, logging, transparency, human oversight, or risk-management measures needed?
- Who is acting as the provider, deployer, or other relevant operator under the AI Act?

2.2.2 Cyber Resilience Act

The Cyber Resilience Act (CRA) (Regulation (EU) 2024/2847) [11] establishes cybersecurity requirements for products with digital elements, including software and connected devices, across their lifecycle. Its purpose is to improve the cybersecurity of digital products placed on the EU market by introducing requirements relating to secure design, vulnerability handling, and security support. The CRA entered into force on 10 December 2024. Reporting obligations apply from 11 September 2026, while the main obligations of the Act apply from 11 December 2027.

2.2.2.1 Main Features of the Cyber Resilience Act

The main measures enforced by the CRA are:

- Cybersecurity-by-design and by-default requirements for products with digital elements.
- Vulnerability handling obligations, including processes for identifying, documenting, and addressing vulnerabilities.
- Security update and support obligations during the product lifecycle.
- Incident and vulnerability reporting obligations for certain situations.
- Conformity assessment and compliance requirements before products are placed on the market.

2.2.2.2 Cyber Resilience Act and the GDPR

The CRA focuses on the cybersecurity of products with digital elements, whereas the GDPR governs the processing of personal data and protects the rights of natural persons in that context. The CRA does not replace data protection law. Where a digital product processes personal data, GDPR requirements continue to apply alongside the CRA.

Topic	Cyber Resilience Act	GDPR
Purpose	Improve the cybersecurity of products with digital elements	Protect personal data
Scope	Software, hardware, and connected products placed on the EU market	Personal data only
Main logic	Security-by-design, vulnerability handling, lifecycle cybersecurity	Lawfulness, fairness, transparency, minimisation, and security of personal data processing
Main addressees	Manufacturers, importers, distributors, and other economic operators	Controllers and processors
Relationship	Adds product cybersecurity obligations	Continues to govern personal-data processing
Fines / penalties	Up to €15 million or 2.5% of total worldwide annual turnover for the most serious infringements; lower tier: up to €10 million or 2%	Up to €20 million or 4% of worldwide annual turnover for the most serious infringements; lower tier: up to €10 million or 2%

Table 2.6: High-level comparison of the AI Act and GDPR

Why this Matters for NRENs

The CRA is relevant to NRENs supporting digital infrastructure, connected devices, software services, research platforms, or procurement processes involving products with digital elements.

Example

An NREN-supported research environment uses software and connected devices that collect, transmit, or process research-related data. In this case, the CRA may be relevant to the cybersecurity requirements of those products and their lifecycle support, while the GDPR remains relevant wherever personal data is processed.

Using secure digital products does not in itself guarantee compliance with all legal or organisational requirements; cybersecurity, procurement, contractual, operational, and data-protection considerations may still need to be assessed.

Practical Takeaway

For NRENs and research support environments, the CRA is particularly relevant where software, connected devices, digital tools, or infrastructure components are procured, deployed, or supported as part of research or service environments. Its practical importance lies in strengthening baseline cybersecurity expectations across the lifecycle of digital products.

Practical Questions to Ask

- Does the product qualify as a product with digital elements under the CRA?
- Who is the manufacturer, supplier, or other relevant economic operator?
- Are there obligations relating to secure design, vulnerability handling, or updates?
- Is the product used in a research or service environment involving personal data?
- Are procurement, lifecycle management, and incident-handling processes aligned with the CRA requirements?

2.2.3 Digital Services Act

The *Digital Services Act* (Regulation (EU) 2022/2065, Digital Services Act) [12] establishes EU-wide rules for intermediary services and online platforms. Its purpose is to make the online environment safer, more transparent, and more accountable, while protecting the fundamental rights of users. The DSA applies to a wide range of online services, including hosting services, online marketplaces, social media platforms, app stores, and certain travel or accommodation platforms. It entered into force on 16 November 2022 and has applied in full since 17 February 2024.

2.2.3.1 Main Features of the Digital Services Act

The main features of the DSA include:

- Safer online environments, including obligations concerning illegal content, goods, and services.
- Greater transparency, including clearer information on content moderation, advertising, and recommender systems.
- User protection and redress, including mechanisms to report illegal content and challenge platform decisions.
- Additional obligations for very large online platforms and search engines, especially in relation to systemic risks.
- Coordinated oversight and enforcement, involving national Digital Services Coordinators and the European Commission.

2.2.3.2 Digital Services Act and GDPR

The DSA focuses on the responsibilities, transparency, and accountability of online intermediary services and platforms. The DSA does not replace data protection law. Where personal data is involved, GDPR requirements continue to apply.

Topic	Digital Services Act	GDPR
Purpose	Make online intermediary services and platforms safer, more transparent, and more accountable	Protect personal data
Scope	Intermediary services and online platforms	Personal data only
Main logic	Platform accountability, transparency, user protection, systemic-risk mitigation	Lawfulness, fairness, transparency, minimisation, and security
Main addressees	Providers of intermediary services and online platforms	Controllers and processors
Relationship	Does not replace data protection law	Continues to govern personal-data processing
Fines / penalties	For very large platforms and search engines, up to 6% of annual worldwide turnover for certain infringements	Up to €20 million or 4% of worldwide annual turnover for the most serious infringements; lower tier: up to €10 million or 2%

Table 2.7: High-level comparison of DSA and GDPR

Why this Matters for NRENs

The DSA is relevant to NRENs supporting online collaboration environments, repository and hosting services, digital platforms, or other user-facing services that may fall within the category of intermediary services.

Example

An NREN-supported collaboration or repository platform allows users to upload and share research-related content. Here, the DSA may be relevant to reporting mechanisms, platform transparency, and user-protection obligations, while the GDPR remains relevant wherever the platform processes personal data.

Making content or data available through an online service does not in itself mean that it can be shared without restriction; legal, contractual, ethical, confidentiality, platform-governance, and data-protection requirements may still need to be assessed.

Practical Takeaway

For NRENs and research support environments, the DSA is most relevant where online platforms, hosted services, repositories, or other user-facing digital services are operated or supported. Its practical importance lies in setting clearer rules on platform governance, transparency, and user protection. Under the DSA, very large online platforms and search engines are those reaching more than 45 million average monthly active recipients in the EU.

Practical Questions to Ask

- Is the service acting as an intermediary or hosting platform?
- Does the service allow users to upload, share, or distribute content?
- Are there mechanisms for reporting illegal content or appealing decisions?
- Does the service process personal data and therefore also trigger GDPR obligations?

- Does the scale or function of the service create additional transparency or governance responsibilities?

2.2.4 Digital Markets Act

The Digital Markets Act (DMA) (Regulation (EU) 2022/1925) is designed to make digital markets in the European Union fairer and more contestable. It applies to large online platforms designated as gatekeepers, i.e., providers of core platform services that have a significant impact on the internal market and act as important gateways between business users and end users. The DMA entered into force on 1 November 2022 and has applied since 2 May 2023.

2.2.4.1 Main Features of the Digital Markets Act

The main features of the DMA are:

- Obligations for designated gatekeepers aimed at preventing unfair practices and improving contestability in digital markets.
- Restrictions on certain behaviours, such as self-preferencing and practices that lock users or business users into a gatekeeper ecosystem.
- Requirements supporting fairness and openness, including interoperability and access-related obligations in certain contexts.
- Strong enforcement powers for the European Commission, including investigations, fines, and periodic penalty payments.

2.2.4.2 Digital Markets Act and GDPR

The DMA focuses on market structure, fairness, and contestability in digital markets. The DMA does not replace data protection law. Where personal data is processed by gatekeepers or other actors, GDPR requirements continue to apply.

Topic	Digital Markets Act	GDPR
Purpose	Ensure fair and contestable digital markets	Protect personal data
Scope	Gatekeepers providing core platform services	Personal data only
Main logic	Market fairness, contestability, and control of gatekeeper power	Lawfulness, fairness, transparency, minimisation, and security
Main addressees	Designated gatekeepers	Controllers and processors
Relationship	Does not replace data protection law	Continues to govern personal-data processing
Fines / penalties	Up to 10% of total worldwide annual turnover; up to 20% for repeated infringements	Up to €20 million or 4% of worldwide annual turnover for the most serious infringements; lower tier: up to €10 million or 2%

Table 2.8: High-level comparison of the Digital Markets Act and the GDPR

Why this Matters for NRENs

The DMA is relevant where NRENs rely on, integrate with, or build services alongside major digital platforms that may be designated as gatekeepers.

Example

An NREN-supported service or research platform depends on access to digital ecosystems dominated by large platform providers. In such a case, the DMA may be relevant to questions of market access, interoperability, and platform fairness, while the GDPR remains relevant wherever personal data is processed.

Market openness or interoperability obligations do not in themselves remove legal, contractual, security, or data-protection constraints; these still need to be assessed in the specific context.

Practical Takeaway

For NRENs and research support environments, the DMA is most relevant as part of the broader digital market context. Its practical importance lies in addressing the power of large platform providers and promoting fairer and more open digital conditions, which may indirectly affect access, interoperability, and dependency on major platforms.

Practical Questions to Ask

- Does the service depend on a major platform ecosystem or core platform service?
- Could interoperability, access, or dependency on a gatekeeper affect the service?
- Are there market or contractual constraints linked to dominant platform providers?
- Does the service also process personal data and therefore trigger GDPR obligations?
- Could wider digital-market rules affect long-term platform strategy or procurement choices?

2.3 Summary

The **European Strategy for Data** is supported by a broader legal and policy framework aimed at improving the availability, accessibility, interoperability, and reuse of data across the European Union. The **Data Governance Act** and the **Data Act** are its two main legislative pillars. They are complemented by the **Regulation on the Free Flow of Non-Personal Data**, the **Open Data Framework**, including **high-value datasets**, and the development of **Common European Data Spaces**. Together, these measures support data sharing, trust, interoperability, and cross-border data use across sectors.

At the same time, the wider EU digital regulatory environment includes other acts that are also relevant in practice for NRENs and the wider GÉANT community. The **AI Act** focuses on the governance of AI systems through a risk-based framework. The **Cyber Resilience Act** focuses on cybersecurity requirements for products with digital elements. The **Digital Services Act** addresses the responsibilities and transparency of online intermediary services and platforms. The **Digital Markets Act** addresses fairness and contestability in digital markets, especially in relation to gatekeepers.

In simple terms, the main distinction may be expressed as follows:

- The **Data Act** covers access to and use of data.
- The **Data Governance Act** covers trust and governance for sharing.
- The **FFD Regulation** covers free movement of non-personal data.
- The **Open Data Framework / High-Value Datasets** cover reuse of public-sector information.

- The **AI Act** covers the governance of AI systems.
- The **Cyber Resilience Act** covers the cybersecurity of digital products.
- The **Digital Services Act** covers platform governance and transparency.
- The **Digital Markets Act** covers fairness in digital markets.

Across all of these instruments, one common point remains important for NRENs and research support environments: where **personal data** is involved, the **GDPR** continues to apply. Therefore, openness, reuse, access, interoperability, AI use, or platform operation do not remove the need to assess data protection, confidentiality, contractual, ethical, and security requirements in the specific context.

3 FAIR Data

The **FAIR data principles** [13] are guidelines to improve the management and stewardship of research data by making it **Findable, Accessible, Interoperable, and Reusable**. They are widely supported by organisations such as the European Commission and Research Data Alliance to promote open science and data sharing. The FAIR principles are also the cornerstone of the 'web of FAIR data and services' that the European Open Science Cloud (EOSC) seeks to create.

In summary, FAIR is not simply about making data 'open,' but about structuring and documenting data in a way that both humans and machines can efficiently find, access, integrate, and reuse.

This section provides a short overview of the FAIR principles, how EOSC uses the FAIR principles as a cornerstone to enable its web of FAIR data and services and how different Horizon Europe projects contribute to this. It also describes the key technologies (i.e., persistent identifiers, semantic artefacts, and knowledge graphs) used to implement the FAIR principles and how the FAIRness of research data can be assessed. In the last part of this section, the topic of FAIR Digital Objects (FDOs) is introduced along with emerging approaches for their implementation.

3.1 FAIR Principles

A 2014 European data science community meeting initially laid the groundwork for establishing guiding principles and practises to facilitate data discovery, access, interoperability, and reuse. Discussions coalesced around four elements that a modern data publishing environment should provide to support both manual and automated deposition, exploration, sharing, and use to support machines as well as humans. These became the FAIR data principles, as follows:

- Data should be **Findable**
- Data should be **Accessible**
- Data should be **Interoperable**
- Data should be **Reusable**

These FAIR principles are related, but technically independent of one another. Data providers and repositories implement these principles incrementally as the data ecosystem evolves towards increasing degrees of FAIR-ness.

3.1.1 FAIR Principles and Machine-Actionable Data

In 2016, *FAIR Guiding Principles for Scientific Data Management and Stewardship* was published in *Nature Scientific Data* [14]. The paper provides guidelines to improve the Findability, Accessibility, Interoperability, and Reuse (i.e., FAIRness) of digital assets, emphasising the importance of preparing data with a machine-first mindset – that is, focusing on the capacity of computational systems to find, access, interoperate, and reuse data with no or minimal human intervention, because humans increasingly rely on computational support to process data due to the increase in data volume, complexity, and creation speed. To this end, the authors offered the following guidelines:

- To be Findable:
  - (Meta)data is assigned a globally unique and persistent identifier.
  - Data is described with rich metadata (defined under 'To be Reusable' below).
  - Metadata clearly and explicitly includes the identifier of the data it describes.
  - (Meta)data is registered or indexed in a searchable resource.
- To be Accessible:
  - (Meta)data is retrievable by their identifier using a standardised communications protocol.
  - The protocol is open, free, and universally implementable.
  - The protocol allows for an authentication and authorisation procedure, where necessary.
  - Metadata is accessible, even when the data are no longer available.
- To be Interoperable:
  - (Meta)data uses a formal, accessible, shared, and broadly applicable language for knowledge representation.
  - (Meta)data uses vocabularies that follow FAIR principles.
  - (Meta)data includes qualified references to other (meta)data.
- To be Reusable:
  - (Meta)data is richly described with a plurality of accurate and relevant attributes.
  - (Meta)data is released with a clear and accessible data usage license.
  - (Meta)data is associated with detailed provenance.
  - (Meta)data meets domain-relevant community standards.

3.2 FAIR and the European Open Science Cloud

The objective of the European Open Science Cloud (EOSC) is to provide researchers and innovators in Europe with an open and trusted multi-disciplinary environment where they can publish, find, and reuse data, tools, and services for research and innovation. Through this environment, EOSC aims to mobilise, align and scale resources across Europe to accelerate open science, higher productivity and increased reproducibility and trust in research. To do this, EOSC emphasises the interoperability of data across and within disciplines, and the (re)use of data by interlinking it with the relevant tools and services, in line with the FAIR data principles [15].

The following subsections review a number of EC-funded projects that explore the relationship between EOSC and the FAIR principles to support their application to advance the reusability of resources distributed via EOSC.

FAIRsFAIR

FAIRsFAIR [16] contributes to EOSC by ensuring that research resources such as data, software and services can be maintained in line with the FAIR principles. This requires alignment and cooperation with existing organisations, long-term projects, and other initiatives involved in building EOSC. FAIRsFAIR thus engages institutional partners and the ESFRI research clusters, among others, to document and propagate FAIR policy and practice [17], gathering lessons learned in one domain and applying them to another, developing educational resources, and supporting the adoption of standardisation tools.

FAIRCORE4EOSC

The FAIRCORE4EOSC project [18] was launched in response to one of the priorities highlighted in the EOSC Strategic Research and Innovation Agenda (SRIA) [19] – namely, the establishment of the '*web of FAIR data and*

a Minimum Viable EOSC' by 2027, featuring the core components and functions that will enable EOSC to operate. Running from June 2022 to June 2025 and led by CSC in Finland, FAIRCORE4EOSC delivered nine EOSC-Core components to support a FAIR research life cycle, developing and testing these components through five user-centric case studies [20] in the following research areas:

- Climate change
- EOSC service providers & research data management (RDM) communities
- European integration of national-level services
- Mathematics
- Social sciences and humanities

The core components delivered by the project are unpacked in the following Section 3.3.

FAIR-IMPACT

To support the adoption and implementation of existing and emerging FAIR-related tools, methods and resources, the FAIR-IMPACT project [21] developed the FAIR Implementation Framework (FIF) [22] to assist stakeholders in assessing their current data management practices and develop realistic action and engagement plans to implement FAIR.

FIF underlines that, while the implementation of specific tools and methods is crucial for enabling FAIR, other factors impacting the overall success of FAIR-enabling efforts must be considered. To this end, FIF adapts the ACME-FAIR approach [23], placing equal emphasis on assessing capabilities for enabling FAIR and on active engagement to ensure uptake of FAIR-enabling practices. The diagram in Figure 3.1 below illustrates the seven components of FIF and provides a set of guiding questions for stakeholders to consider at each stage.

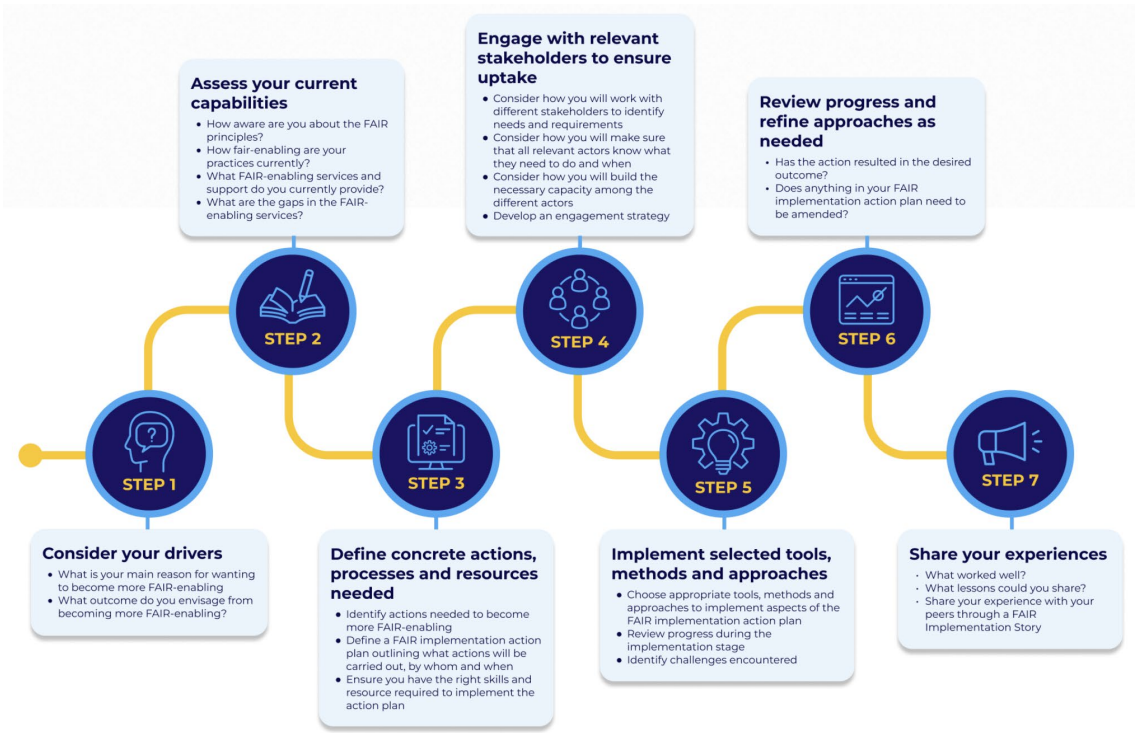

Figure 3.1: The seven steps of the FAIR Interoperability Framework with guiding questions

3.3 FAIR Cornerstone Technologies

As mentioned previously, EOSC considers the adoption of the FAIR data principles as a fundamental cornerstone in establishing its '*web of FAIR data and services*'. To achieve this vision during the EOSC implementation phase (2021–27), the FAIRCORE4EOSC project was initiated to develop the EOSC-Core components required for a FAIR EOSC. Leveraging existing technologies and services, the project developed nine, new EOSC-Core components intended to improve the discoverability and interoperability of an increased amount of research outputs [24].

The nine EOSC-Core components [25] delivered by FAIRCORE4EOSC are as follows:

- **Compliance Assessment Toolkit (CAT)**

  The Compliance Assessment Toolkit supports the EOSC policy on persistent identifiers (PIDs), which allow researchers to reliably identify, verify and locate data and are an integral aspect of FAIR. This is done via services to encode, record, and query compliance with the PID policy. To do so, a wide range of compliance requirements (TRUST, FAIR, PID Policy, Reproducibility, GDPR, Licences) will be evaluated as use cases for definition of a conceptual model. At the same time, vocabularies, concepts, and designs are intended to be reusable for other compliance needs: TRUST, FAIR, POSI, CARE, Data Commons.
- **Data Type Registry (DTR)**

  The Data Type Registry supports data interoperability and reuse in line with the FAIR principles. It offers a sustainable framework for the registration, discovery, and long-term management of structured, machine-readable data type definitions. Through the DTR, research communities can define and share standardised data structures that improve data quality, facilitate reuse, and support cross-disciplinary research.
- **Metadata Schema and Crosswalk Registry (MSCR)**

  The MSCR allows registered users and communities to create, register and version schemas and crosswalks with PIDs. The published content can be searched, browsed and downloaded without restrictions. The MSCR also provides an API to facilitate the transformation of data from one schema to another via registered crosswalks.
- **PID Graph**

  This component supports the harvesting of PID Graph metadata. The API service and data dumps were made available for the community to ingest and reuse the metadata seamlessly. In addition, the component focuses on the interoperability framework for graph data exchange.
- **PID Meta Resolver (PIDMR)**

  The PID Meta Resolver is a generalised resolver for mapping items into records. PIDMR aims to support researchers in their daily work to easily make use of the PIDs (resolution, metadata). At the time of writing, the PIDMR can resolve PIDs from around 50 different systems or providers. Providers can register their PID service and make their own PIDs resolvable via the PIDMR.
- **Research Activity Identifier service (RAiD)**

  The RAiD service provides persistent, unique and resolvable information for research projects. RAiD consists of a unique, resolvable DOI and a metadata record that connects organisations, people, inputs, and outputs to projects. It serves as a container for other PIDs (ORCIDs, RORs, DOIs, etc.), linking research components while tracking project changes over time.
- **Research Discovery Graph (RDGraph)**

  The EOSC RDGraph service delivers advanced discovery tools across EOSC resources and communities. RDGraph builds upon the EOSC catalogue's content, extending it with additional entities like RAiDs.
- **Research Software APIs and Connectors (RSAC)**

  The Research Software APIs and Connectors (RSAC) developed in FAIRCORE4EOSC enable research infrastructures to preserve, describe, and cite software by connecting directly to the Software Heritage archive. Using the CodeMeta metadata standard and Software Heritage identifiers (SWHIDs), these services ensure software artefacts can be reliably referenced and integrated into scholarly workflows.

RSAC has been implemented across a number of repositories (i.e., InvenioRDM (CERN), DataVerse (DANS)), publishers (i.e., Dagstuhl (LZI), Episciences (INRIA)), and aggregator technologies (i.e., swMATH (FIZ), OpenAIRE).

- **Software Heritage Mirror (SWHM)**

  The EOSC Software Heritage Mirror (SWHM) is a read-only copy of the Software Heritage universal source code archive, operated in agreement with (but independently of) the Software Heritage organisation.

The following subsections explore the cornerstone technologies that are fundamental to the above components' ability to support FAIR resources.

3.3.1 Persistent Identifiers

Persistent Identifiers (PIDs) are a cornerstone of the FAIR principles because they provide unique, machine-resolvable, and long-lasting references for data, researchers, and organisations, ensuring digital objects are consistently FAIR. By embedding persistent metadata, PIDs prevent link rot, enable automated discovery, and establish critical links between research, creators, and institutions. To support PID managers in developing EOSC-compliant PID policies, the FAIR-IMPACT project produced a set of guidelines. These are examined in the following section, along with a couple of PID selection tools to support researchers in selecting the most appropriate PID for a given use-case.

3.3.1.1 EOSC-Compliant PID Policy Guidelines

The FAIR-IMPACT project provides a list of 16 guidelines for creating a EOSC compliant PID policy for PID managers [26]. Guidance is offered both on formulating a new PID policy and evaluating an existing PID policy. These guidelines are intended for use with the EOSC-Core Compliance Assessment Toolkit to enable PID managers (among others) to assess a PID policy and make improvements if required:

1. **Select a PID Stack** that is globally unique and persistently resolvable.
2. **Manage persistence**

   Managers must develop policies and procedures to guarantee maintenance of the correct link between the identifier and the resolution target, and make sure the responsibilities are well defined in their agreements or contracts with the PID owner.
3. **Manage versions**

   Managers must have a clear policy on version management. The provisions of the policy depend on the purpose of referencing resources by the persistent identifier.
4. **Involve stakeholders**

   Managers should make time and resources available to participate in the governance structures of the PID Stacks that they use.
5. **Conform to a PID Stack checklist**

   Managers should confirm the degree to which PID Stacks support or conform to a number of important considerations, including:

   Certification and Compliance (e.g. adherence to standards)

   Continuity (e.g. long-term service availability)

   Sustainability (e.g. funding and operational models)

   Responsibilities (e.g. clearly defined roles for maintenance and governance)

   Value-added services (e.g. APIs, metadata enrichment).
6. **Select an appropriate scale**

Managers must consider the scale at which PIDs will be used. Two interrelated considerations are the scalability of the service, and the cost.

7. **Select appropriate identifier schema and structure**
8. **Consider resolution options**

   Managers should consider the type of resolution mechanism offered to users and owners when selecting a service and creating their own infrastructure. Examples of approaches for resolution options include via HTTP wrappers, a webpage, or an API.
9. **Maintain resolution integrity**

   Managers must maintain the link between the identifier and its resolution mechanism, and the object or concept being referenced.
10. **Manage metadata**

    Managers should manage metadata in alignment with community and disciplinary standards, and must maintain an authoritative version of the metadata.
11. **Consider implementation of machine-actionable extensions**
12. **Monitor resolution integrity**

    Managers should consider implementation of mechanisms (procedures) to verify the integrity of the resolution of the PID. Integrity verification includes two elements: link rot and content drift.
13. **Take sensitive metadata into consideration**

    Managers should consider implementation of practices to deal with sensitive metadata in cases where it is required.
14. **Consider periodic resolvability sampling**

    Managers who curate a large number of PID-referenced resources may consider random sampling to verify resolvability.
15. **Develop and implement sustainability and continuity mechanisms**
16. **Adopt a level for maturity and availability of services**

    Managers should adopt the EU Technology Readiness Level (TRL) classification 43 for services and web resources.

3.3.1.2 PID Selection Tools

Because of the many different types of PIDs, it can be a challenge to determine the right PID to reference information. To support researchers and data managers in the decision-making process, various PID selection tools have been developed. Two of these tools are discussed below:

- PID Selection Tool [27]: Developed by the PID4NFDI team [28] in Germany, the PID Selection Tool is designed for repository and infrastructure managers planning to integrate PIDs into an existing or new research service, repository, or tool. It supports the decision-making process by helping researchers to explore which PID system best fits a specific use case and integration goals.
- PID wijzer [29]: Developed at the National Library of the Netherlands in collaboration with the Dutch Digital Heritage Network, *PID wijzer* provides researchers with resources on important PID topics, and guides them through the first steps towards selecting a PID system. The guide focuses on the Dutch Digital Heritage sector, with libraries, archives and museums in mind, but the content is also applicable to research organisations.

3.3.2 Semantic Artefacts and Terminologies

A ‘semantic artefact’ in research refers to any structured, representational object – digital or conceptual – created to encode, store, or communicate meaning within a knowledge system. While an *artefact* can be any

human-created object, a *semantic artefact* specifically encodes meaning [30], allowing it to be processed computationally and consistently.

Semantic artefacts are typically defined as formalised representations of knowledge, such as an ontology, taxonomy, controlled vocabulary, knowledge graph, metadata schema, or any structured model that captures meaning and relationships. Research on semantic artefacts [31] emphasises how these objects allow information to be shared, reused, and interpreted across systems and contexts. Semantic artefacts can be collected into semantic artefact catalogues (SACs) to standardise definitions and improve interoperability across scientific and technical domains.

3.3.2.1 FAIR-by-Design Methodology

The FAIRsFAIR project work on FAIR Semantic Artefacts produced practical recommendations on how to make semantic artefacts FAIR. These include, for example, using persistent identifiers, providing rich and standardised metadata, ensuring clear licensing, and making artefacts accessible through interoperable formats and APIs. In addition, a service architecture for harmonising semantic artefact repositories' metadata was provided and the resulting harmonised metadata published in a FAIR Data Point [32].

3.3.2.2 Semantic Artefact Governance and Management

As EOSC strives to establish a unified, open environment for data and services, the governance of semantic artefacts is indispensable for seamless integration and interoperability across varied scientific domains. Following a structured methodology (see Figure 3.2), the FAIR-IMPACT team analysed practices, principles, and resources for semantic artefact governance and management, focusing on three types of initiatives:

- **Project-based initiatives** managing single-core semantic artefacts. Within this category, semantic artefacts are developed to address the need for standardised vocabulary in a target domain. These semantic artefacts enable semantic interoperability and support the creation of knowledge graphs by providing domain-specific concepts and properties, enhancing data integration between systems that deploy them.
- **Support-based initiatives** that assist research communities in developing and maintaining semantic artefacts. Individuals or groups of researchers independently build and maintain artefacts in a decentralised manner. Organisations such as INRAE Voc and NFDI4Biodiv operate within this category.
- **Harmonisation initiatives** that provide principles and guidelines, along with infrastructure, to harmonise the management of semantic artefacts *across* communities or organisations. Developers can use the upper-level ontologies, modularisation methods, and other standard workflows and practices established by these initiatives to ensure they develop interoperable semantic artefacts. The OBO Foundry is an example of this category, using semantic web technologies in biological and biomedical sciences. BASF also falls into this category but restricts its developed ontologies to internal organisational use [33].

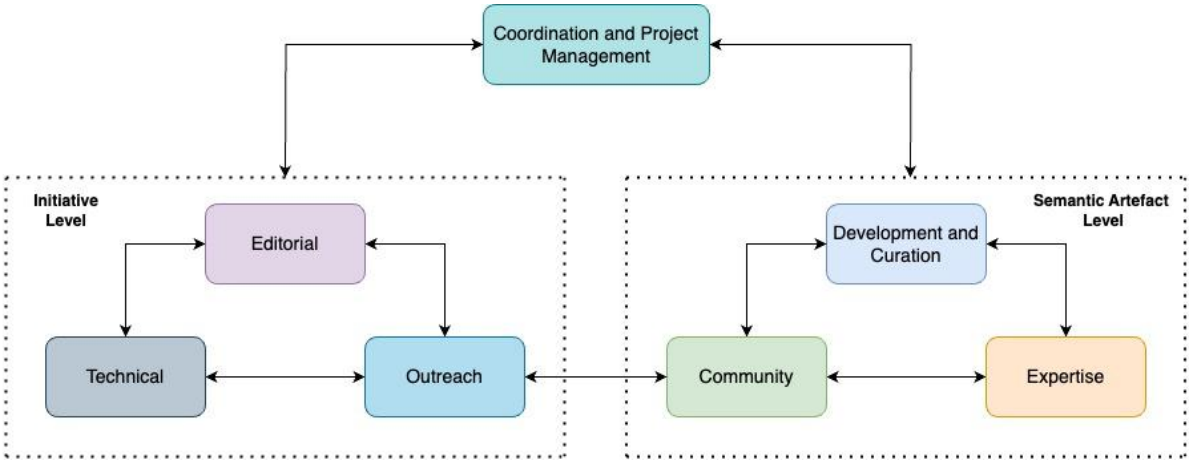

Figure 3.2: Organisational structure recommended for use by FAIR-IMPACT

3.3.2.3 Semantic Artefact Catalogues and Mappings

Semantic artefact catalogues (SACs) and ontology repositories are critical to EOSC and its aim of constructing a '*web of FAIR data and services*'. SACs serve as essential platforms for the distribution, sharing and reuse of semantic artefacts (e.g., ontologies, taxonomies, metadata schemas and standards). These catalogues also support compliance with the FAIR data principles, which are foundational to EOSC's mission of promoting open science and data sharing across diverse scientific disciplines.

The FAIR-IMPACT project developed a 'FAIR-by-design' methodology tailored to creating FAIR vocabularies and ontologies, with the potential to extend its application to other types of semantic artefacts. The proposed approach builds upon the Linked Open Terms (LOT) methodology [34], identifying three main approaches for SAC interoperability, each with its own advantages and disadvantages. These are summarised in the following diagram.

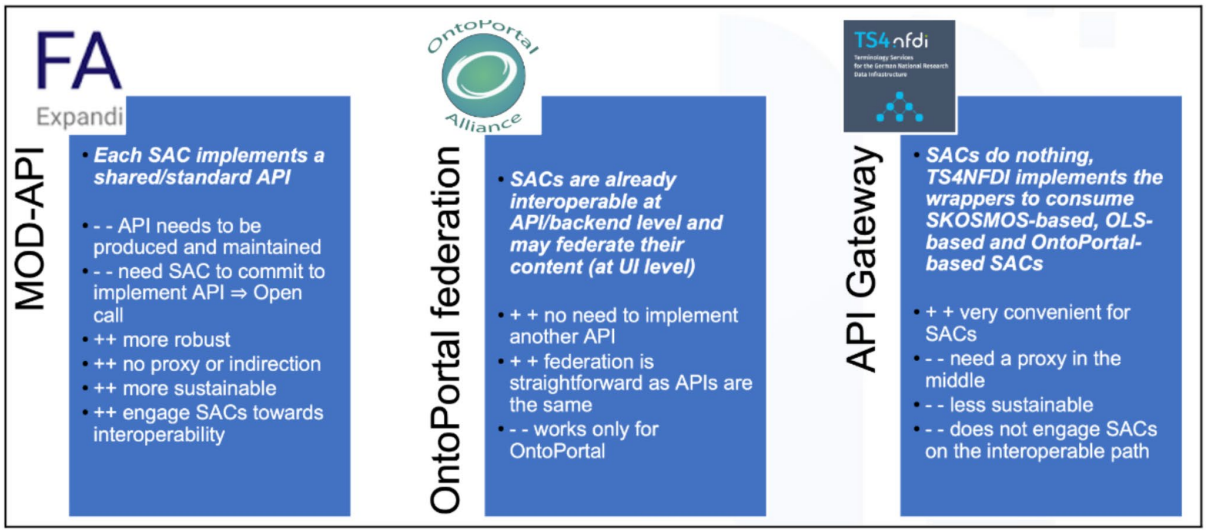

Figure 3.3: Three main approaches to semantic artefact catalogue interoperability (FAIR-IMPACT)

3.3.3 Knowledge Graphs

Knowledge graphs are not a new technology but have seen wider adoption in recent years. A knowledge graph is a form of knowledge base is a type of database that organizes information in the form of a network (a 'graph'), connecting related pieces of data. Knowledge graphs are often used to store interlinked descriptions of entities – objects, events, situations or abstract concepts – and to show how they are related, capturing the meaning behind those connections.

Knowledge graphs are used by organisations that maintain large structured data resources, including the Google search engine [35], Wikipedia [36] and OpenAIRE Graph [37]. The SURF Tech Trend Report [38] identifies knowledge graphs as one of the emerging technologies supporting the adoption of the FAIR principles. The EOSC EU Node's Resource Catalogue [39] is itself a knowledge graph making EOSC resources, such as services, tools, research data, publications, software, and other research outputs, available in a structured format. The following sections provide a short overview of important developments around knowledge graphs in the context of FAIR.

3.3.3.1 RDA Scientific Knowledge Graph Interoperability Framework

The Scientific Knowledge Graph – Interoperability Framework (SKG-IF) Working Group (WG) [40] aims to define a framework to enable a seamless exchange of information between diverse initiatives related to scientific knowledge graphs (SKGs) – a form of knowledge graph focused on academic content (e.g. repositories, databases and catalogues). The WG addresses the increasing need for interoperability among SKGs to support Open Science and FAIRness in research, and foster the transition to Open Research Information (ORI), in compliance with the 2024 Barcelona Declaration, which supports open and interoperable research information for FAIR and Open Science. [41].

In its final recommendations [42], the SKG-IF WG presents the core components of its SKG interoperability framework, including the data model (SKG-IF Data Model), API specifications, and extension guidelines. The framework facilitates the interoperable exchange of data on six core entities:

- **Research product**: Encompasses research literature, research data, research software, or other uniquely identified digital assets.
- **Agent**: Represents an individual or organisation involved in the creation, publication, dissemination, etc., of a research product.
- **Grant**: Describes funding awarded to an Agent by a funding body.
- **Venue**: The publishing 'gateway' used by an Agent to make their research products available.
- **Topic**: The scientific disciplines, subjects, and keywords related to a research product.
- **Data source**: A service or platform where a research product (its metadata and files) is stored, preserved, and made discoverable and accessible.

Figure 3.4 provides an architecture overview of the SKG-IF including the SKG-IF Data Model.

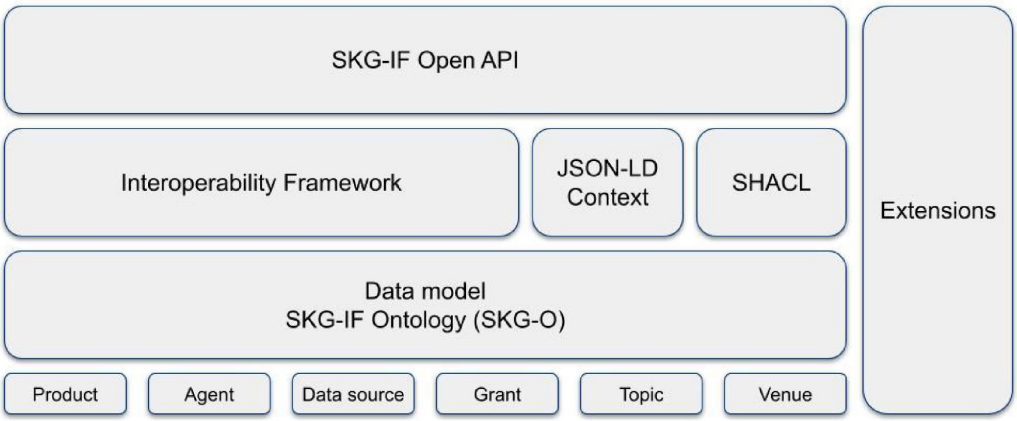

Figure 3.4: SKG-IF Architecture and Data Model

The OSTrails project [43] further develops the SKG-IF into an SKG Commons [44] to provide a common, reusable, extensible, and interoperable foundation for handling data exchange between SKG interfaces across tools and services that implement the SKG-IF.

3.3.3.2 Making Research Data Discoverable in the EOSC Federation

The aim of EOSC is to contribute to a '*web of FAIR data and services*' for science in Europe. The EOSC Federation is a network of interconnected, autonomous, and distributed nodes that federate data, tools, and services for research across Europe. The EOSC Federation is built around the broader network of interconnected EOSC Nodes (e.g. European, national and/or thematic) that operate under a common framework.

According to the EOSC Federation Handbook [45], resources can only be made available to the EOSC Federation by onboarding the resources into an EOSC Node and from that EOSC Node to the EOSC EU Node.

Note, the EOSC EU Node is the central, European-level node providing core services within the EOSC Federation. For researchers, this means that once data is onboarded to an EOSC Node, it can be discovered through the EOSC Federation, increasing its visibility and enabling access and reuse by other researchers across Europe.

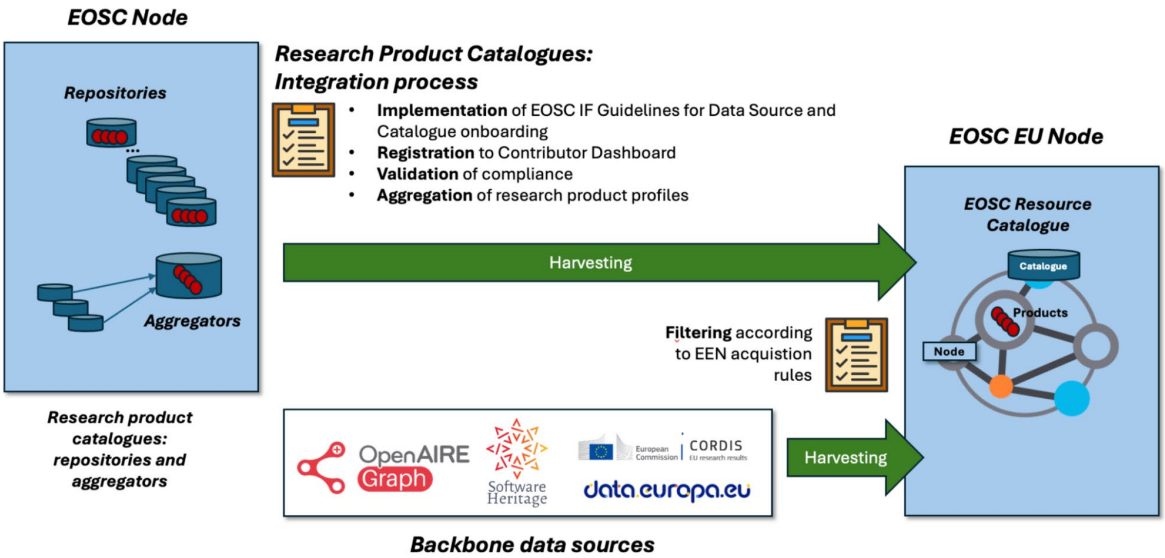

Figure 3.5: EOSC Node research product catalogues: onboarding data sources (blue cylinders) and research products (red circles) from EOSC Nodes (light blue rectangles)

Each EOSC Node maintains a resource catalogue, which is a structured collection of metadata describing its available research data, services, and tools. Figure 3.5 illustrates how an EOSC Node, responsible for the Node resource catalogues, will:

- Make their catalogue compliant with the EOSC Interoperability Guidelines for EOSC data sources.
- Register the catalogues in the EEN Resource Catalogue.
- Validate the catalogues to verify technical compliance with the Guidelines.

Like the EOSC Node Resource Catalogues, EOSC EU Node Resource Catalogue is a knowledge graph, a structured database that collects and links together detailed information about EOSC resources, such as services, tools, research data, publications, software, and other research outputs, as well as the organisations behind them. The EOSC EU Node Resource Catalogue will:

- Aggregate (harvest, harmonise, and index) research product metadata records from the catalogues on regular intervals.
- Filter the aggregated records based on a set of quality criteria, aiming at excluding records lacking "relevant" metadata information and non-scientific products (e.g., sales, explicit content).

The procedure on how to make the resources available to the EOSC Federation from an EOSC Node is described in more detail in the interoperability guideline on *Registration of EOSC Research Product Catalogues in the EOSC EU Node* [46].

3.4 Assessing FAIRness

To evaluate the rate of adoption of the FAIR principles by European researchers, and the FAIRness of shared data and resources, a set of assessment tools is required. The following two subsections give an overview of the RDA FAIR Data Maturity Model and a list of tools to assess the FAIRness of research data.

3.4.1 RDA FAIR Data Maturity Model

The Research Data Alliance (RDA) FAIR Data Maturity Model Working Group defined a Data Maturity Model [47] in 2019 to assess the proliferation of FAIR principles in research data, including a list of indicators to measure and evaluate the FAIRness of data. Different levels of priorities are assigned to each FAIRness indicator, reflecting the fact that some indicators are more important than others:

- **Essential (●●●):** These indicators address aspects of the utmost importance to achieve FAIRness under most circumstances, or, conversely, FAIRness would be impossible to achieve if the indicator were not satisfied.
- **Important (●●):** These indicators address an aspect that might not be of the utmost importance under specific circumstances, but its satisfaction would increase FAIRness.
- **Useful (●)**: These indicators address aspects that are *nice-to-have* but not necessarily indispensable in achieving FAIRness.

Error! Reference source not found. details the FAIR indicators and their importance:

FAIR	ID	Indicator	Priority	
F1	RDA-F1-01M	Metadata is identified by a persistent identifier	●●●	Essential
F1	RDA-F1-01D	Data is identified by a persistent identifier	●●●	Essential
F1	RDA-F1-02M	Metadata is identified by a globally unique identifier	●●●	Essential
F1	RDA-F1-02D	Data is identified by a globally unique identifier	●●●	Essential
F2	RDA-F2-01M	Rich metadata is provided to allow discovery	●●●	Essential
F3	RDA-F3-01M	Metadata includes the identifier for the data	●●●	Essential
F4	RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	●●●	Essential
A1	RDA-A1-01M	Metadata contains information to enable the user to get access to the data	●●	Important
A1	RDA-A1-02M	Metadata can be accessed manually (i.e. with human intervention)	●●●	Essential
A1	RDA-A1-02D	Data can be accessed manually (i.e. with human intervention)	●●●	Essential
A1	RDA-A1-03M	Metadata identifier resolves to a metadata record	●●●	Essential
A1	RDA-A1-03D	Data identifier resolves to a digital object	●●●	Essential
A1	RDA-A1-04M	Metadata is accessed through standardised protocol	●●●	Essential
A1	RDA-A1-04D	Data is accessible through standardised protocol	●●●	Essential
A1	RDA-A1-05D	Data can be accessed automatically (i.e. by a computer program)	●●	Important
A1.1	RDA-A1.1-01M	Metadata is accessible through a free access protocol	●●●	Essential
A1.1	RDA-A1.1-01D	Data is accessible through a free access protocol	●●	Important
A1.2	RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation	●	Useful
A2	RDA-A2-01M	Metadata is guaranteed to remain available after data is no longer available	●●●	Essential
I1	RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	●●	Important
I1	RDA-I1-01D	Data uses knowledge representation expressed in standardised format	●●	Important
I1	RDA-I1-02M	Metadata uses machine-understandable knowledge representation	●●	Important
I1	RDA-I1-02D	Data uses machine-understandable knowledge representation	●●	Important

FAIR	ID	Indicator	Priority	
I2	RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	●●	Important
I2	RDA-I2-01D	Data uses FAIR-compliant vocabularies	●	Useful
I3	RDA-I3-01M	Metadata includes references to other metadata	●●	Important
I3	RDA-I3-01D	Data includes references to other data	●	Useful
I3	RDA-I3-02M	Metadata includes references to other data	●	Useful
I3	RDA-I3-02D	Data includes qualified references to other data	●	Useful
I3	RDA-I3-03M	Metadata includes qualified references to other metadata	●●	Important
I3	RDA-I3-04M	Metadata include qualified references to other data	●	Useful
R1	RDA-R1-01M	Plurality of accurate and relevant attributes are provided to allow reuse	●●●	Essential
R1.1	RDA-R1.1-01M	Metadata includes information about the licence under which the data can be reused	●●●	Essential
R1.1	RDA-R1.1-02M	Metadata refers to a standard reuse licence	●●	Important
R1.1	RDA-R1.1-03M	Metadata refers to a machine-understandable reuse licence	●●	Important
R1.2	RDA-R1.2-01M	Metadata includes provenance information according to community-specific standards	●●	Important
R1.2	RDA-R1.2-02M	Metadata includes provenance information according to a cross-community language	●	Useful
R1.3	RDA-R1.3-01M	Metadata complies with a community standard	●●●	Essential
R1.3	RDA-R1.3-01D	Data complies with a community standard	●●●	Essential
R1.3	RDA-R1.3-02M	Metadata is expressed in compliance with a machine - understandable community standard	●●●	Essential
R1.3	RDA-R1.3-02D	Data is expressed in compliance with a machine-understandable community standard	●●	Important

Table 3.1: FAIR Data Maturity Model indicators

The work of the FAIR Data Maturity Model Working Group has strengthened the realisation that community involvement and input is crucial to identify and build the standards and approaches required to enable FAIR data in a range of disciplines. The FAIR Data Maturity Model represents a solid community-endorsed foundation on which further work can be built [48].

3.4.2 FAIR Assessment Tools

Alongside the FAIR Data Maturity Model, an array of tools has been developed by different initiatives to measure the FAIRness of datasets and data services. The following table provides a non-exhaustive list of tools for researchers to study and assess different aspects of data FAIRness.

Tool	Description
F-UJI [49]	F-UJI is a web service to programmatically assess the FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics [50]. It is an open-source tool implementing a REST API to assess the FAIRness of research data using PIDs. The tool outputs a score for each metric, as well as grouped scores and visualisations that present an overall picture of the FAIRness of the dataset, which can be downloaded for further implementation.
FAIR-Aware [51]	FAIR-Aware allows users to assess their knowledge of the FAIR principles and improve understanding of how data FAIRness can increase the potential value and impact of research data. This tool does *not* assess a specific object type; rather the knowledge of the people handling the objects.
O'FAIRe [52]	O'FAIRe (Ontology FAIRness Evaluator) is a metadata-based automatic FAIRness assessment methodology for ontologies and semantic artefacts against the 15 FAIR Principles. Assessment results are visualised and explained with user-friendly interfaces to help users improve the FAIRness of their resources.
FOOPS! [53]	FOOPS! (Ontology Pitfall Scanner for FAIR) is a tool for assessing the FAIRness of ontologies. The tool serves as a validator for researchers to assess whether a vocabulary (OWL or SKOS) conforms to best practices for publishing ontologies online.
FAIR Champion [54]	FAIR Champion is a general-purpose FAIR assessment tool intended for use by all communities and for all digital objects. Any test from any provider can be registered, so long as the test generates a metadata descriptor compliant with the FAIR Reference Model defined by OSTrails.
FAIROs [55]	FAIROs is a FAIR assessment tool for research objects. The tool uses external tools such as F-UJI, FOOPS! and RSFC to assess datasets, ontologies and software.
Research Software FAIRness Checks [56]	Research Software FAIRness Checks (RSFC) is a command-line interface to automatically evaluate the fairness of a GitHub or GitLab repository. Given a repository URL, RSFC will perform a series of 20 checks based on a list of Research Software Quality Indicators (RSQI).
FAIRsharing [57]	FAIRsharing is a curated, informative, and educational resource on data and metadata standards, databases, and data policies across all disciplines. FAIRsharing is a core integration component within a variety of research data management tools, including technical solutions for FAIR assessment and evaluation as well as data management plan (DMP) creation.
FAIR Assessment Authoring Tool [58]	This tool offers a complete framework for generating, registering, and authoring FAIR assessment components across different metadata repositories. The FAIR assessment metadata components available include FAIR Benchmark, FAIR Metric, FAIR Test and FAIR Benchmark Algorithm.
FAIR Validator [59]	The OpenAIRE Metadata Validator is a tool designed to evaluate metadata records for research products (publications, datasets, software, and other outputs) against the

Tool	Description
	OpenAIRE Guidelines, which define the minimum requirements for inclusion in the OpenAIRE Graph.
FAIR Data Self-Assessment Tool [60]	This researcher-friendly tool enables discovery of the FAIRness of research datasets based on questionnaires and provides practical tips to enhance FAIRness.
FAIR-Checker [61]	FAIR-Checker allows data FAIRness assessment and empowers data providers to enhance the quality of their digital and web resources. FAIR-Checker leverages semantic web technologies to check that metadata uses standards and recognised ontologies or controlled vocabularies.

Table 3.2: Overview of FAIR assessment tools

3.5 FAIR Digital Objects

The paper *Digital Objects as Drivers towards Convergence in Data Infrastructures* (2019) [62] defines Digital Objects (DOs) as meaningful entities in digital form that humans and machines need to be able to uniquely identify, such as data, metadata, software code, queries, configurations, etc. This also includes digital representations of physical entities such as persons or institutions, and abstract entities such as concepts and relations.

A DO is made up of some **content** (represented by a bit sequence stored in a repository), an **identifier**, and **properties** of different types. The core of the DO is a bit sequence encodes the content (data, metadata, software, etc.). The DO is described by metadata to enable access to and the correct interpretation of the DO's content. The DO has a PID that uniquely identifies it and its attributes point to the locations of the bit sequence, its metadata and other relevant properties, providing findability and accessibility of the DO. It should be noted that metadata descriptions are themselves DOs as they are associated with a PID. Figure 3.6 indicates the principles central to the concept of DOs.

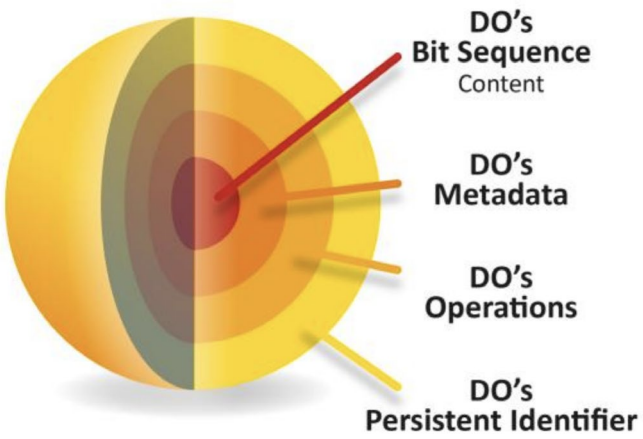

Figure 3.6: Basic concept of a FAIR Digital Object

While the FAIR Principles were first published in 2016, the concept of FAIR Digital Objects (FDOs) has not yet matured into a decisive standard like the TCP/IP standardised network protocol for internet access or HTTP(S) for accessing web resources. Four currently emerging concepts around FDOs – FDO Forum FAIR Digital objects, FAIR Signposting, FAIR Data Points, and FAIR RO-Crates – are discussed in the following subsections.

3.5.1 FAIR Digital Objects Forum

The FAIR Digital Object (FDO) Forum [63] is a bottom-up initiative that started in January 2021, holding workshops to bring together experts dealing with aspects of DOs in the Research Data Alliance (RDA) [64], DONA Foundation [65] and GO FAIR [66].

The approach advocated by the FDO Forum is built around the Digital Object Architecture (DOA) [67], using Handles/DOIs as PIDs and digital objects as its core data model forming the basis for the architecture. The DOA specifies three core components and two protocols:

- A *digital object* is defined as a sequence of bits, or a set of sequences of bits, incorporating a work or portion of a work or other information in which a party has rights or interests, or in which there is value, each of the sequences being structured in a way that is interpretable by one or more of the computational facilities, and having as an essential element an associated unique persistent identifier.
- DOA core components:
  - The *identifier/resolution system* enables:
    - The allotment of unique identifiers to data structured as digital objects, regardless of the location of the data or the technology used to serve the data.
    - The resolution of the identifiers to current state information about the corresponding digital object, e.g., its location(s), access & usage policies, timestamps, and/or public keys.

    The state information is stored in the form of a digital object; rapid resolution is enabled using the Digital Object Identifier Resolution Protocol (DO-IRP).
  - The *repository system* manages digital objects, including provision of access based on unique identifiers, and with integrated security. The use of identifiers in the access protocol abstracts away the details of the storage technologies from the clients, enabling a long-lived mechanism for depositing and accessing digital objects. Access to this system is enabled using the Digital Object Interface Protocol (DOIP).
  - The *registry system* is a specialised repository system intended to store *metadata* about digital objects rather than the digital information itself. Access to this system is enabled using the DOIP as well.
- DOA protocols:
  - The Digital Object Interface Protocol (DOIP) is a simple but powerful protocol for software applications (clients) to interact with services, which could be either the digital objects or the information systems that manage those digital objects.
  - The Digital Object Identifier Resolution Protocol (DO-IRP*)* is a rapid-resolution protocol for creating, updating, deleting, and resolving identifiers that are globally managed and allotted. Each identifier is associated with an identifier record that clients can resolve using this protocol.

DOIP is defined as a simple and universal protocol. The diagram on the left of Figure 3.7 illustrates DOIP's operation, allowing the connection of any client and server if the FDO model is supported. The right-hand diagram indicates how DOIP can be used to create single integrated FDO space, independently of the technology used by a given repository to organise its data.

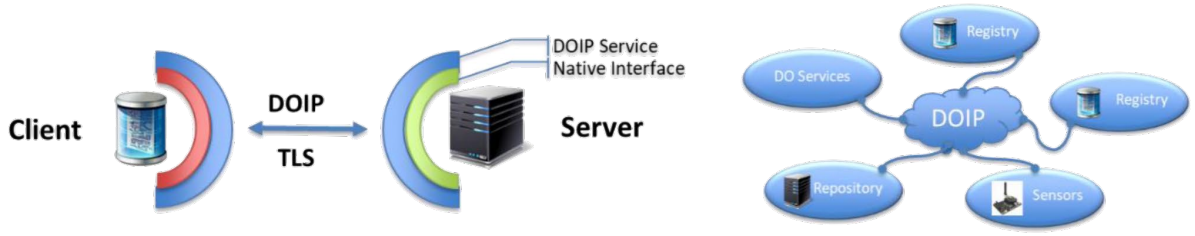

Figure 3.7: DOIP used to create a single integrated FDO space [68]

The FDO's Digital Object Architecture has been adopted by data communities including the Distributed System of Science Collections (DiSSCo) [69], a natural-science data community, and is gaining traction within the Handle identifier user community [70].

3.5.2 FAIR Signposting

Signposting [71] started as an initiative to improve the interoperability of scholarly repositories on the web, with the aim of providing a uniform approach to allow machines to interact with repositories. Signposting is not a formal standardisation effort; rather an accumulation of ideas from specialists in scholarly communication on the web, working on specifications to improve on the interoperability status quo.

The specifications provided under the Signposting umbrella are fully aligned with hypermedia (i.e., REST, HATEOAS) approaches to web interoperability, which focus on dynamic links, media types and machine-readable controls. These allow systems to automatically discover and access related data across repositories. By following the Signposting specifications, implementation should be more straightforward and will improve the FAIRness of data for machines.

The following diagram illustrates the relationship between FAIR Digital Objects and FAIR Signposting, sourced from a 2023 FAIR Digital Object Forum webinar [72].

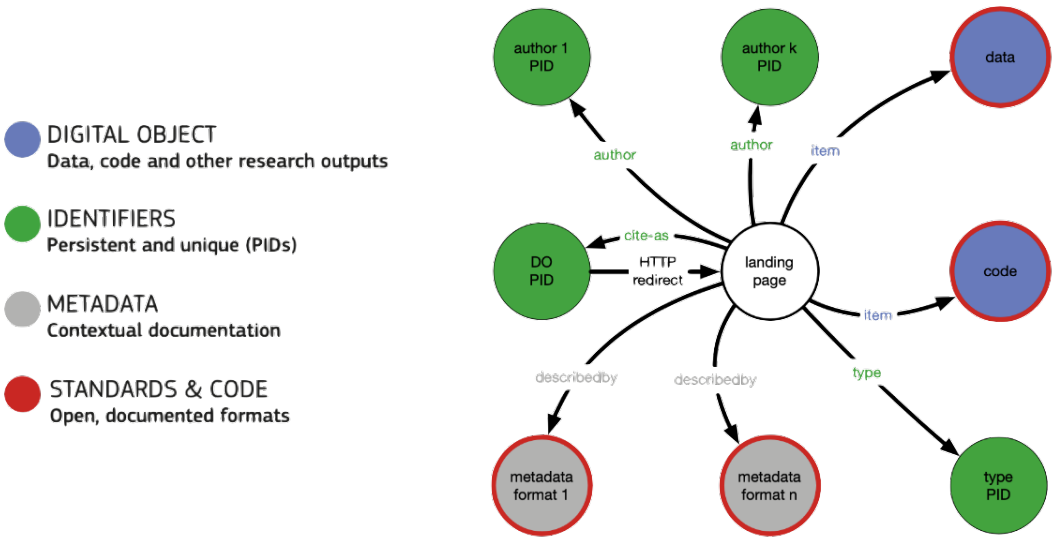

Figure 3.8: The use of typed web links to link resources from a landing page

In line with Signposting specifications, different types of identifiers are conveyed using different link types:

- "item" links are used to link to content resources.

- “Described by” links are used to link to metadata resources.
- “cite-as” to link to the PID of the digital object.
- “author” to link to identifiers of authors.
- “type” to link to URIs that convey what kind of digital object is involved.
- “license” to link to license URIs.

The net result of providing these typed links is that when a web crawler lands on the PID of the digital object or directly on its landing page, it can immediately build a picture of the complete object by observing these typed links and can follow the links to gather further information.

FAIR Signposting has received broad support within the scholarly repository community [73] and in popular repository technologies including CKAN [74], Dataverse [75], DSpace [76], and InvenioRDM [77], to name a few.

3.5.3 FAIR Data Points

The FAIR Data Point (FDP) concept [78] concerns presenting (meta)data on the web using existing standards. FDPs store *information about datasets*, i.e., *metadata*. FDPs advance the FAIRness of data by providing the metadata needed for **F**indability and **R**eusability, along with a uniform open way of **A**ccessing the data. FDPs also address the **I**nteroperability of the metadata stored, but leave the Interoperability aspects of the data itself to the data provider.

The most important component of FDPs is the definition of the FAIR Data Point API specification[79], which is completely based on semantic metadata standards (mostly Data Catalog (DCAT) [80] and Dublin Core [81], combined with the Linked Data Platform (LDP) [82]), and based on the REST philosophy [83]. This combination targets the highest possible technical interoperability in combination with simple implementation. Other components include a metadata registration service implementing the API specification, allowing maintainers to define and update metadata; and the client of the API, a web frontend to add, edit and query information in the metadata registration service. Figure 3.9 provides a high-level overview of the FAIR Data Point architecture.

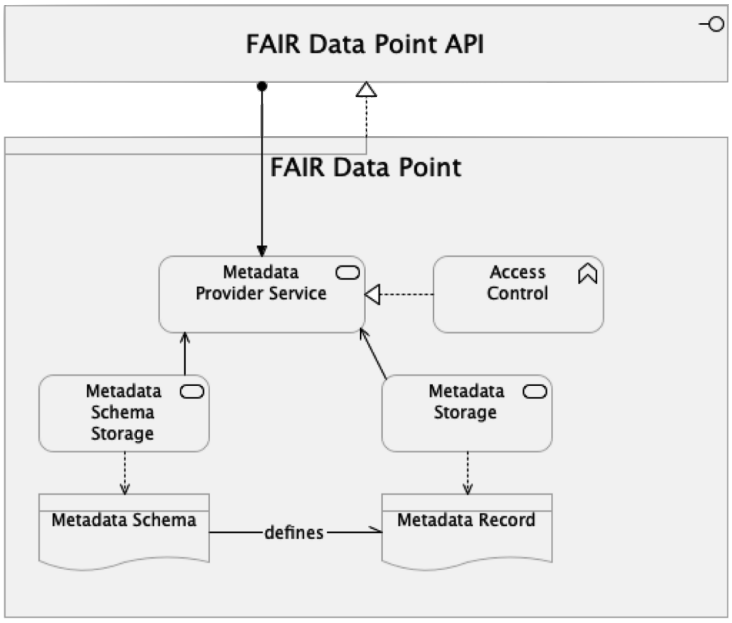

Figure 3.9: FAIR Data Point high-level architecture [84]

The FAIR Data Point Index homepage [85] offers an overview of active FAIR Data Points.

3.5.4 FAIR RO-Crates

An RO-Crate [86] is an integrated view allowing the user to see an entire *research object* – the methods, data, output and outcomes of a project or piece of work. Linking all this together enables the sharing of research outputs with their context as a coherent whole. RO-Crates link data and metadata irrespective of their storage locations, meaning that the user can find the data from a paper, can find its authors from the data, and so on. For example, an RO-Crate contains an author's name alongside their ORCID, which in turn is connected to their affiliations, their funding, and their other publications.

An RO-Crate is a structured package for sharing research data. It can:

- Contain files and link to large datasets stored online.
- Record an entire analysis, including inputs, outputs, and tools used.
- Connect related publications and datasets.
- Preserve datasets for long-term archiving.
- Combine multiple uses to fit different needs.

A general overview of the RO-Crate architecture is shown in the following figure.

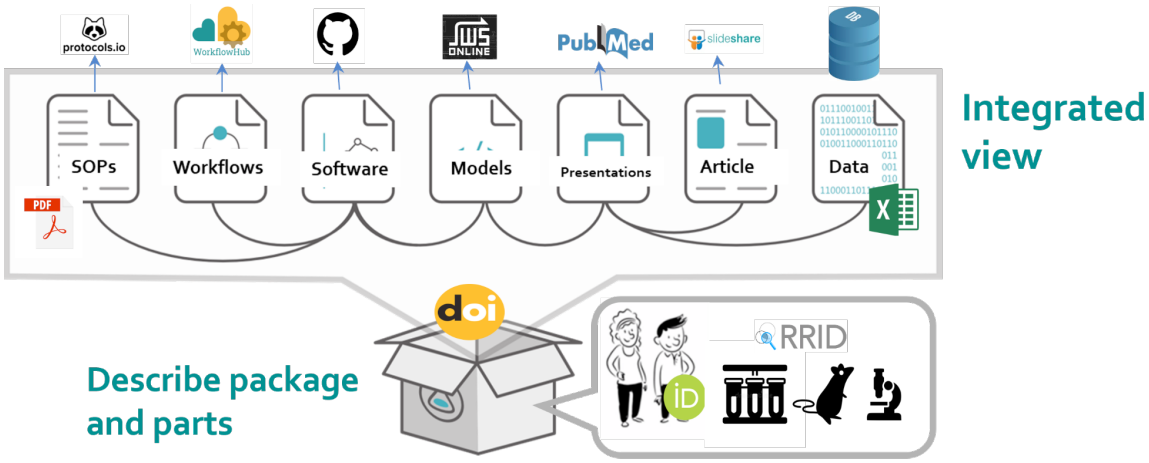

Figure 3.10: RO-Crates can link and be linked to research objects from many different sources using persistent identifiers like ORCIDs and DOIs (image credit: Goble, C. 2024 [87])

RO-Crates improve the FAIRness of research data as follows:

- **Findable**: RO-Crates can be published to repositories like Zenodo and given a DOI, just like any other data. The linked metadata helps consumers of the RO-Crate to find related work, people, and organisations.
- **Accessible**: RO-Crates are compatible with FAIR Signposting, enabling the metadata to be retrieved automatically via the DOI or other persistent identifier.
- **Interoperable**: RO-Crates are based on standard, widely used web technologies (including JSON-LD and schema.org [88]), making it usable and interoperable across platforms. It is flexible enough to reference any kind of external (meta)data, whether that external data uses RO-Crate itself or not.
- **Reusable**: RO-Crates are designed to capture detailed provenance. For example, an RO-Crate could include a full representation of the individual steps executed as part of an analysis, including the inputs, outputs, computational environment, and the researchers involved.

When RO-Crates are combined with FAIR Signposting, it results into a practical implementation of the FAIR Digital Object Framework [89]. RO-Crates have been adopted in fields such as bioinformatics, digital humanities, and regulatory sciences, and a growing number of tools offer native RO-Crate support [90].

4 Research Data Management

Science Europe [91], an association representing major public organisations that fund or perform high-quality, novel research in Europe, defines Research Data Management (RDM) as follows:

> *Research data management refers to the handling of research data (collection, organisation, storage, and documentation) during and after a research activity. Good data management helps ensure that researchers share their data in a FAIR way (findable, accessible, interoperable, and reuseable). Research organisations increasingly require their researchers to develop a data management plan to ensure that all aspects are considered from the start of an activity on.* [92]

Good data management practices are the cornerstones of the scientific process to achieve trust in the results of research. This section provides an overview of RDM best practices throughout the research and research data lifecycles, and addresses the topics of data curation (managing the use of data from its point of creation to ensure it is available for discovery and reuse in the future) and data quality.

4.1 Research Data Management Best Practices

This section provides an overview of RDM best practices. This is not intended as an exhaustive list, but a series of references on best practices structured around (1) the research lifecycle, defined as the activities performed before, during and after a research project, and (2) the different stages of the research data lifecycle (RDL), as defined by the RDA WG on Mapping the Landscape of Digital Research Tools [93], illustrated in the following diagram.

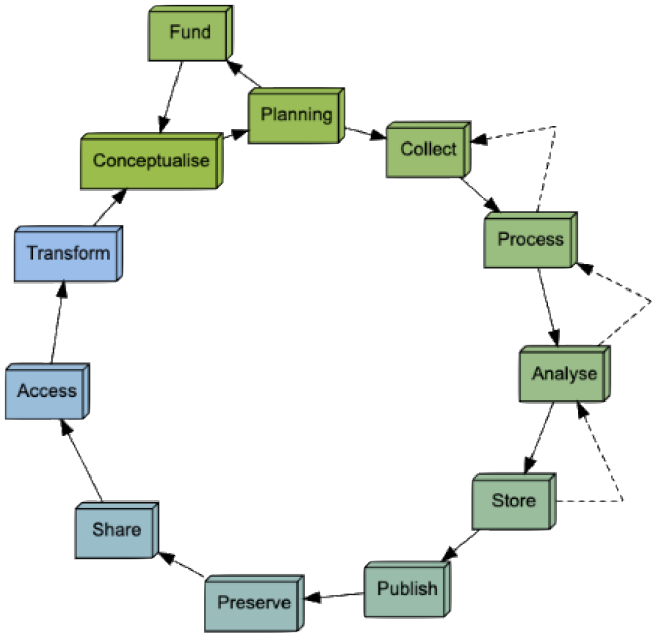

Figure 4.1: Harmonised Research Data Lifecycle (RDL) Model created by the RDA working group on Mapping the Landscape of Digital Research Tools [94]

The stages of the RDL, shown in the preceding figure, are grouped into the phases of the broader research lifecycle as follows:

- **Before Research**: Conceptualise, Fund, Planning
- **During Research**: Collect, Process, Analyse, Store
- **After Research**: Publish, Preserve, Share, Transform

Each stage of the RDL is introduced with a short definition in italics from the aforementioned deliverable of the RDA working group on Mapping the Landscape of Digital Research Tools [95].

4.1.1 Before Research

This section provides best practices and guidelines for the stages of the research data lifecycle that are normally planned before starting the active research phase: Conceptualise, Planning and Fund.

4.1.1.1 Conceptualise

Definition: *To formulate the initial research idea or hypothesis, and define the scope of the research project and the data component/requirements of that project.*

Best Practice Recommendations

Conceptualisation is the starting point of a research project where the initial ideas are identified, defined and recorded. Important questions such as the Five Ws (*Who, What, When, Where and Why*) about a research idea or hypothesis are asked to determine the scope and context of the research project. Information produced during the conceptualisation phase is needed to communicate the ideas with fellow researchers at all phases of the research lifecycle:

- *Before* the start of a research project, to invite fellow researchers to contribute and to request funding supporting the research activity.
- *During* a research project, to assess the research results fit with the hypothesis or not, and if not, to understand why not.
- *After* a research project, for researchers who want to reuse the research output and need to understand the origin and background of the research project and research idea.

Therefore, all the information gathered during the conceptualisation phase should be referenced when the research output is published and shared.

4.1.1.2 Fund

Definition: *To identify and acquire financial resources to support the research project, including data collection, management, analysis, sharing, publishing and preservation.*

Best Practice Recommendations

Funding agencies will have their own guidelines on requesting funding for and reporting on research projects. Generally, funding agencies request to be informed about any research output generated through the project and for funding sources to be acknowledged.

The following aspects come into play during the funding phase:

- **Storage costs**: Estimating the effort and resources required for a research project is challenging and affected by many uncertainties. In the RDM context, calculating storage costs for data during and after a project is a complex task because it depends heavily on the technical characteristics of a dataset and on how the data is used during processing and analysis (see Section 4.1.2.4).

  In general, researchers are not data or storage experts and lack experience of sizing data storage infrastructures and/or calculating storage costs. Best practice to produce a realistic estimate for storage costs in a funding request is to collaborate with the storage service provider intended to be used in the research project. The service provider can use specific questions to assess the researcher's storage resource requirements.
- **Tracking research activities**: To support researchers in keeping track of all the information used and produced during a research project, researchers can use Research Activity IDentifiers (RAiDs). A RAiD is a unique PID developed by the Australian Research Data Commons (ARDC) [96] to identify research projects and activities and allow access by research communities worldwide.

  RAiDs create a single place for storing and retrieving project information using a RAiD name (identifier) and its associated metadata record. Each RAiD metadata record stores project or activity information for long-term discovery, resolution, access, sharing, reporting and other uses by the global research community. While RAiD use is not yet a mandatory requirement of funding agencies, it is advised as a best-practice way of tracking the information used and produced during a research project. In Europe, SURF (NL) is the leading NREN in advocating and offering RAiDs [97].

4.1.1.3 Planning

Definition: *To establish a structured strategic framework for management of the research project, outlining aims, objectives, methodologies, and resources required for data collection, management and analysis. DMPs should be established for this phase of the lifecycle.*

Best Practice Recommendations

A Data Management Plan (DMP) is a document that explains how research data will be collected, organised, stored, and shared during and after a project. Its purpose is to ensure data is well managed from the start of a project, and properly prepared for long-term preservation and reuse.

Producing a DMP requires researchers to think about the data and storage resources needed during and after the project. It is advised for researchers to engage potential storage service providers in the early stages of project planning to develop storage resource requirements both *during* and *after* the project – the latter aspect is sometimes forgotten. Compared to computing resources for processing and analysis, which are temporary needs, storing data requires a permanent allocation of storage resources and needs constant attention, from either the researcher or the storage service provider.

DMPs are thus an essential tool for researchers, funders and service providers. Many funding agencies, including the EC for European projects, request a DMP before or at the start of a project. Some agencies may even stipulate the use of certain DMP services, such as the following:

- OpenAIRE ARGOS [98]: ARGOS is an open extensible service that simplifies the management, validation, monitoring and maintenance of plans. It allows actors (researchers, managers, supervisors, etc.) to create actionable DMPs that can be freely exchanged among infrastructures to carry out specific aspects of the data management process in accordance with the DMP.
- DMPonline [99]: DMPonline helps researchers to create, review, and share DMPs that meet institutional and funder requirements. DMPonline is offered as a web service and made available as open source by the Data Curation Centre (DCC) as DMP Roadmap [100]. Because the software is open source, many organisations run their own DMPonline instance, such as DMP Tuuli [101] operated by CSC and Belnet's DMPonline [102].

At the time of writing, a number of international initiatives are working on standardising machine-actionable DMPs:

- **RDA's Common Standard for Machine-actionable Data Management Plans** [103]: The RDA's DMP Common Standards WG aims to create a standardised machine-actionable DMP format. This enables the exchange of information between systems acting on behalf of stakeholders involved in the research lifecycle (researchers, funders, repository managers, data stewards, etc.). It also allows the automation of typical data management tasks, thus reducing researchers' workloads. To this end, the WG has developed an application profile for the automatic exchange, integration, and validation of information provided in DMPs [104].
- **OSTrails** [105]: An EU-funded project under Horizon Europe, OSTrails aims to transform data management by establishing interoperable knowledge graphs and deliver modular FAIR tests, ultimately providing a comprehensive solution to drive FAIR research practices forward.

For GÉANT project-specific data management requirements, NRENs should also consult the applicable DMP and related project rules for the relevant GÉANT project documentation [106].

4.1.2 During Research

This section provides the guidelines for the stages of the research data lifecycle that are normally planned during the active research phase: Collect, Process, Analyse and Store.

4.1.2.1 Collect

Definition: *To use predefined procedures, methodologies and instruments to acquire and store data that is reliable, fit for purpose and of sufficient quality to test the research hypothesis.*

Best Practice Recommendations

After initial discovery of datasets to be explored in the active research process, the data needs to be collected and stored on the researcher's chosen service. This may entail making the data available to other researchers or across multiple computing platforms for processing and analysis. Data may be collected either from existing repository services or generated as part of the research process. In cases where existing datasets are reused, the data is typically transferred from a repository service to the storage service made available to researchers. The data collection stage involves the following considerations:

- **Access**: Depending on the sensitivity of the data, researchers follow different procedures:
  - **Open data**: Data is openly accessible and researchers can collect data via direct download from the repository service where it is stored.
  - **Embargoed data**: Embargo means that access to data is closed for a period before being made openly accessible. Data can be embargoed because it needs to be reviewed before publication, or because a researcher is planning follow-up research on the same dataset. Researchers seeking to use embargoed data need to request access from the data owner. Data repository services normally have capabilities to support these requests.
  - **Restricted and/or closed data**: In case of sensitive and/or confidential data, specific processes and procedures apply for requesting and getting access to datasets. This can include researchers submitting data access requests to a data access body. For example, in the European Health Data Space, each participating country establishes a Health Data Access Body (HDAB) as a governance structure for the assessing and approving secondary use of health data. (See also Section 4.1.3.4, Access.)
- **Data transfer considerations**:

- **Size does matter**: When planning data transfers, researchers should consider the size of the dataset to be transferred. Specific tools exist for transferring large datasets (both in data volume and in number of files) versus smaller data volumes or only a few files.
- **Less is more**: When many files need to be transferred, researchers should opt for parallel transfers (i.e., multiple files being transferred at the same time, in parallel). Parallel transfers can be more efficient than sequential file transfers, but there is an optimal number of parallel transfers for maximal efficiency. This depends on the configuration of the technical infrastructure and storage services used, and the network between them.

  Researchers can make use of network performance testing tools such as iperf [107] to optimise parallel transfers. Researchers are encouraged to collaborate with system administrators to test data transfer performance over specific network protocols (e.g. SCP/SFTP, GridFTP, S3).

- **Data transfer tools**: As above, multiple tools exist for data transfers of different sizes:
  - **FileSender** [108]: A well-known tool within the NREN community, FileSender is a web-based application for sending large files quickly and securely. It can handle files of any size and uses extensive security protocols to ensure that files only reach their intended recipients. FileSender has been adopted across research institutes, healthcare organisations and educational facilities.
  - **SSH**: Secure Shell (SSH) is a very popular tool set for providing secure access to Unix-/Linux-based systems. The SSH toolset provides multiple utilities to transfer data, including SCP [109] and SFTP [110] for tens of files, and Rsync [111] for any larger number, including full directory trees.
  - **Rclone** [112] is a command-line program to manage files on cloud storage, supporting many different protocols [113]. Its primary use case is similar to Rsync, managing large data transfers of many files, including full directory tree transfers.
  - **File Transfer Service (FTS)** [114]: An open source tool developed for managing large-scale data transfers within the World LHC Computing Grid (WLCG) [115]. FTS supports both third-party and parallel data transfers through WebDav, HTTPS, GridFTP, and domain-specific protocols such as xRootD [116] and SRM [117]. Best suited for transferring petabyte-scale datasets, FTS can easily be integrated into the orchestrated data flows of research communities.

In general, best practice is to use network protocols that encrypt data in transit, which is supported by all the protocols mentioned above. However, researchers should note that in certain cases, such as gigabyte-scale files, data encryption can cause performance bottlenecks from the storage services having to encrypt and decrypt the files. Some data transfer tools offer the option to encrypt only the control path (e.g., authentication) and leave the data path unencrypted. This allows researchers to securely authenticate themselves on the storage services, and benefit from faster data transfers by not encrypting all data. This option should only be used for non-sensitive and/or non-confidential data.

4.1.2.2 Process

Definition: *To make new and existing data analysis-ready. This may involve standardised pre-processing, cleaning, reformatting, structuring, filtering, and performing quality control checks on data. It may involve the creation and definition of metadata for use during analysis, such as acquiring provenance from instruments and tools used during data collection.*

Best Practice Recommendations

By the processing phase, most data will already have been generated. Consider the following best practices:

- **Folder/data organisation**: Organising research files/data in a logical and standardised structure has advantages both for researcher's daily work and when sharing data with others. Additionally, RO-Crates, supported by many services, can be used to structure datasets in line with the FAIR principles (see Section 3.5.4).

- **File naming**: Use file names that reflect the content of the file, facilitating the searching, discovering and understanding of the data.
- **Version control**: Version and/or status information can be used in file names to help track changes in each iteration of a document.
- **File formats**: Self-defined files formats are strongly discouraged for FAIRness. Researchers are encouraged to use file formats widely adopted within the research community to ensure usability, accessibility and sustainability (see, for example, the preferred formats from DANS, the Dutch national centre of expertise for research data [118]). Other popular file formats within the scientific domain include NetCDF [119] and HDF5 [120]. The advantage of using these file formats is that many tools are available to read and interpret the content.
- **Data provenance**: Datasets are reliable when the processes used to create them are reproducible and analysable for defects. This is done by tracking provenance information in detail throughout the whole process of generating, transforming and analysing datasets. Workflow systems make it easier to track the research process and to maintain a provenance trail.
- **Workflows**: Many research communities have developed workflow and workflow systems to support their researchers. Popular workflow systems are Galaxy [121][122], CERN's Reana [123], and the WorkflowHub [124], a registry for describing, sharing and publishing scientific computation workflows. Researchers are advised to use *scientific* workflows to process research data.
- **Sensitive and/or confidential data**: When working with sensitive and/or confidential data, data must be processed in a Trusted Research Environment (TRE), a secure digital system that holds and provides access to sensitive data for approved researchers [125]. Some NRENs offer access to TREs, such as SURF's Secure ANalysis Environment (SANE) [126] and the ePouta virtual private cloud from CSC [127]. At the time of writing, three Horizon Europe projects are currently developing trusted environments for sensitive data management in EOSC (EOSC-ENTRUST [128], TITAN [129] and SIESTA [130]).

4.1.2.3 Analyse

Definition: *To derive insights, knowledge, and understanding from processed data. Data analysis involves iterative exploration and interpretation of experimental or computational results, often utilising mathematical models and formulae to investigate relationships between experimental variables. Distinct data analysis techniques and methodologies are applied according to the data type (quantitative vs qualitative).*

Best Practice Recommendations

During the analysis phase, researchers will produce research outputs that will be used for the conclusions on the scientific findings. Reproducibility of findings is a major principle underpinning the scientific method. For the findings of a study to be reproducible, the results obtained by an experiment, observational study, or statistical analysis should be able to be achieved again with a high degree of reliability when the study is replicated. To support reproducibility, researchers need to maintain research outputs in a structured way.

4.1.2.4 Store

Definition: *To record data using technological media appropriate for processing and analysis whilst maintaining data integrity and security.*

Best Practice Recommendations

Many different types of storage media are available to store digital data, this ranges from removable and portable storage devices, local drives on laptops, in cloud storage, object stores and/or storage resources made available on HPC, HTC, Cloud computing services. Which storage media to choose depends on characteristics of the data and how it is being used during processing and analysis. Therefore, it is not easy to provide general

guidelines to which storage media to store and manage research data. The following aspects need to be considered:

- **Storage media**: Avoid using removable media, such as USB flash drives, memory cards, external hard drives and smartphones. Removable media can get lost, which presents a security risk. Many organisations offer access to an *Enterprise File Sync & Share* [131] cloud storage service to store and manage personal and, depending on the size, also research data. Through the syncing capabilities, data can be easily made available across different systems (e.g. mobile, personal computer, VMs, containers) and through the sharing capabilities data can easily be shared with fellow researchers and colleagues.
- **Data location**: Due to the current geopolitical situation, data location has become a very important topic. Researchers need to be aware of *where* data is stored and *who* is providing the storage service. The preference is that data is stored on services within the European domain and offered by a European native cloud provider, which is 100% EU-owned, operated and governed by EU law.
- **Size of a dataset**: As storage capacity is always a limited resource, it is important to ensure user access to sufficient storage capacity to store the dataset. While this sounds obvious, it can be an issue when dealing with large datasets, for example those with TBs or PBs amounts of data. Plan for sufficient storage capacity during the funding and planning phase.
- **Storage performance**: Application performance strongly depends on how fast data can be read from storage media. To optimise application performance, select storage media matching the expected application performance. Storage performance also strongly depends on the characteristics data (e.g. large files versus small files and/or streaming access versus random access)
- **Large files versus small files**: If possible, organise data in large files (>1GB), storage media is more efficient in reading and writing large files.
- **Streaming access versus random access**: In general, streaming access functions better as random access. If an application is randomly accessing data, data needs to be stored on storage media with high input/output per Second (IOPs). If an application is streaming data, for example during data transfers, it is more important to use storage media with high bandwidth.
- **Pseudonymisation**: is a data management and de-identification procedure by which personally identifiable information fields within a data record are replaced by one or more artificial identifiers. For example. researchers can use the OpenAIRE [132] Amnesia [133] tool to anonymise research data.
- **Encryption**: Wikipedia defines encryption (more specifically, encoding) as the process of transforming information in a way that, ideally, only authorised parties can decode. Encryption can be done at many different levels. Storage services can make use of encrypted media through which all data stored on the storage service is automatically encrypted. This prevents reading of data when storage media is deprecated or gets stolen. Researchers can independently encrypt data. In this case, research data is secured, including from the system administrators having administrative access to the storage service. And of course, researchers always need to use encrypted network protocols to transfer data to and from storage services.

4.1.3 After Research

This section provides guidelines for the stages of the Research Data Lifecycle that are normally planned after the active research phase: Publish, Preserve, Share and Transform.

4.1.3.1 Publish

Definition: *To release research data in published form for use by others with appropriate metadata for citation (including a unique persistent identifier) based on FAIR principles.*

Best Practice Recommendations

The Discussion Paper [134] from the EOSC Association LTDP TF elaborates on different aspects that come into play, when selecting a repository for publishing a dataset.

- Disciplinary vs Generic data repositories
  The discussion paper from the EOSC Association Long-Term Data Preservation Task Force (LTDP-TF) elaborates on the differences between disciplinary and generic data repositories, citing that many funders may recommend Trusted Digital Repositories (TDR) with disciplinary expertise (e.g. Science Europe [135]). Defining and agreeing on disciplinary expertise, however, is not trivial, and will become more challenging as repositories and data services claim to be interdisciplinary, multi-disciplinary or cross-disciplinary.

- Trust versus Trustworthiness
  Trust (confidence and belief in the reliability or truth of something) is offered, accepted, sought and earned between parties. It is subjective and may be right or wrong. When applied to data, it is the confidence that the data is accurate, ready to use, and of a certain level of quality. **Trustworthiness** in the context of a TDR is about demonstrating that practices meet a set of standard requirements through an assessment process, for example, through one of the TDR standards (i.e. CoreTrustSeal [136], Nestor Seal [137] or ISO16363 [138]). What sets a TDR apart is the provision of organisational and technical infrastructures, and security designed to ensure 'preservation' for its designated community of users. Preservation is a promise and a process to ensure that risks to the continued access, usability and understandability of digital objects are minimised.

OpenAIRE [139] provides the following guidelines for finding a trustworthy digital repository [140]:

- Use a disciplinary repository if there is one.
- Alternatively, use an institutional repository, if you have one where the data will also be available for the long term.
- Use the catch-all (generic) repository, for example Zenodo [141], maintained by CERN [142], or B2SHARE [143] offered by EUDAT [144].
- Or search in a global registry - re3data [145] or FAIRsharing [146] - for a suitable repository (they provide several filtering options).

4.1.3.2 Preserve

Definition: *To ensure the safety, integrity, and accessibility of data for as long as necessary so that data is as FAIR as possible. Data preservation is more than data storage and backup, since data can be stored and backed up without being preserved. Preservation should include curation activities such as data cleaning, validation, assigning preservation metadata, assigning representation information, and ensuring acceptable data structures and file formats. At a minimum, data and associated metadata should be published in a trustworthy digital repository and clearly cited in the accompanying journal article unless this is not possible (e.g. due to the privacy or safety concerns).*

Note: Preservation includes the active processes required to ensure that the data remains accessible, readable, and usable over time; archiving refers to storing data for long-term retention.

Best Practice Recommendations

- EOSC Association Long Term Data Preservation Task Force (LTDP-TF) *Final Report & Recommendations* [147].

The EOSC Association Long Term Preservation Task Force (LTP-TF) ran between 2021-2023. The TF report provides an overview of outputs, the vision for optimal preservation of FAIR Digital Objects within the context of EOSC, the conclusions from the consultation process, and the final set of key recommendations. Preservation aims to generate assets for research with long- term value. Therefore, digital objects must be retained, stored, findable, accessible, interoperable and reusable over time. Deciding which digital objects should and which should not be preserved must be done via a transparent appraisal processes, any change to levels of care must also be transparent and justified. This includes criteria set at the point of deposit in a repository, initial curation actions, and measures to ensure long term preservation.

- EOSC EDEN Report on Identification of Core Preservation Processes [148]

  Core Preservation Processes (CPP) are identified as those processes that, in principle, every trustworthy, digital, long-term archive (TDA) should have in place. The initial CPP framework was adapted from the Digital Preservation Coalition's Core Requirements [149] and refined through systematic comparison with institutional policies, professional expertise, and selected standards. The CPPs are designed to be organisation- and system-agnostic, setting good-practice baseline expectations rather than prescribing an idealised model. This ensures they are flexible, widely applicable, and able to support different institutional strategies and capacities. The list of CPPs is established by a group of digital preservation practitioners. The CPP writing group identified 30 processes, including: checking the integrity of files and managing data corruption, generating checksums and validating the Integrity files, identifying and validating file formats, metadata extraction, ingest and management, rights and risk management, and more. For each of the processes, a detailed process description is provided with descriptive steps for implementation.

4.1.3.3 Share

Definition: *To make data available and accessible to humans and/or machines. Data may be shared with project collaborators or published to share it with the wider research community and society at large. Data sharing is not limited to open data or public data, and can be done during various stages of the research data lifecycle. At a minimum, data and associated metadata should be published in a trustworthy digital repository and clearly cited in the accompanying journal article.*

Best Practice Recommendations

In general, Research Performing Organisations develop and provide their own guidelines to share and deposit research data into a repository, examples can be found at the University of Ghent [150], University of Utrecht [151] and DANS [152]. Researchers need to prepare the data for sharing, meaning that fellow researchers and/or other users of the data need to be able to read, consume and understand the data technically, semantically and contextually. The prepare data for sharing the following aspects come into play:

- Find a suitable repository for publishing the dataset, see Section 4.1.3.1 Publish, and follow the guidelines for depositing.
- Prepare dataset for sharing:
  - *Folder/data organisation:* Organise files/data in a logical structure, using a standard way of organising research files has indisputable advantages, both for daily work and when sharing data with colleagues or others
  - *File naming:* Use file names that reflect the content of the file, facilitating searching, discovering and understanding of the data.
  - *Version control:* Version and/or status information can be used in file names.
  - *File formats:* Prevent using self-defined files formats, use file formats that are used within the research community and/or recommended the repository, they will offer the best long-term guarantees in terms of usability, accessibility and sustainability, see for example the preferred

formats from DANS [153]. If these recommendations are not available, it is recommended to use popular file formats used within the scientific domain, for example NetCDF [154] or HDF5 [155]. The advantage of using these file formats is that many tools are available to read and interpret the content.

- *Licensing data:* (See Section 4.1.3.4 Access)
- *Persistent identifiers:* Use persistent identifiers where possible to reference information, for example, use digital object identifiers (DOIs) for publications [156], ORCIDs for researchers [157], RoRs for Research Organisations [158], Grant Identifiers to reference funding applications [159] and RAIDs for projects and/or research activities [160].
- *Metadata:* Provide rich metadata to describe the dataset and the data. Metadata to describe the dataset is normally provided during the depositing process. Depending on the file formats used, additional metadata needs to be provided to semantically interpret the structure and content of a file. Providing this level of information will significantly improve the reusability of the data.
- *Terminologies and Data Types:* Also to describe content and to type data values, terminology services and data type registries can be used, for example the Terminology service [161] as offered by the Leibniz Information Centre for Science and Technology and University Library [162] and/or the Data Type Registry [163] offered by the ePIC PID Consortium [164]. The TS4NFDI service [165], developed in the National Research Data Infrastructure (NFDI) [166], provides an overview of the thematic and generic terminology service providers.
- *Reference related resources and information:* This will provide additional context to understand the dataset in relation to a research activity, when referencing additional resources, use PIDs.
- *Sensitive and/or confidential data*: Because of the nature of the data, sharing of sensitive and/or confidential data is always a challenge. Data with a high-level of sensitivity is normally maintained within Trusted Research Environments (TREs) and is only accessible by approved researchers. To receive access, researchers need to go through specific processes and procedures to request and gain access to the data. While direct access to the data is not available, sensitive and/or confidential data can be made FAIR by publishing the metadata describing the dataset in a repository to assign PID and to make the dataset discoverable. The metadata should also include a procedure on how to request access. In this way researchers can discover the dataset and understand the process and procedure to receive access, see also Section 4.1.3.4 Access.

4.1.3.4 Access

Definition: *To control and manage data access by designated users and reusers. This may be in the form of publicly available published information. Necessary access control and authentication methods are applied.*

Best Practice Recommendations

One of the main differences between Open Access and FAIR is that Open Access needs to be made accessible , while in the case of Open Science and FAIR, the mantra is *as open as possible, as closed as necessary* [167]. This is an important difference because in specific research domains, data cannot be made open accessible if data contains personal information and is therefore sensitive and/or it needs to be treated with confidentiality. Metadata describing a dataset is treated as open to improve discoverability.

- *Licensing data:* When making research data publicly available, it is important to let potential users know in advance what they are allowed to do with the data, licensing is an effective way to communicate such permissions. It is a good practice to apply a standard and open license for open research data, as it ensures legal interoperability and the widest possible reuse. To guide researchers selecting an appropriate license, EUDAT [168] developed the *license selector tool* [169].

- *Embargo period:* An embargo is a request by a researcher to delay the publication of their dataset until a specified time [170]. Embargoes (or embargo periods) are most commonly used when researchers want to publish datasets, but are unable to due to reasons such as data sensitivity, impending publication plans or industry/funder agreements.

- *Sensitive and/or confidential data*: Data is published under restricted access, with the dataset findable via its metadata. To receive access, researchers need to follow specific procedures to request and to be granted access. Research communities dealing with sensitive and/or confidential data have developed solutions to manage the requesting and granting process, for example the Resource Entitlement Management System (REMS) tool [171] that has been developed by CSC [172] and is used by the ELIXIR research community [173] that brings life science resources from across Europe together.

- *Access control:* In case of restricted and/or closed access, researchers normally need to authenticate themselves to get access to a service. This does not automatically mean that a user has access to the dataset he/she wants access to. This access needs to be granted through the access control system of the service. How this access is granted strongly depends on service and technology.

4.1.3.5 Transform

Definition: *To create new data from the original, for example: (i) by migration into a different format; (ii) by creating a subset, by selection or query, to create newly derived results, for publication; or, iii) combining or appending with other data.*

Best Practice Recommendations

The *Long-Term Data Preservation Task Force (LTDP TF)* of the *EOSC Association* defined long-term data/digital preservation in the *EOSC Preservation: Overview Discussion Paper* [174] as an activity that guarantees continued access to digital materials, or at least to the information contained in them, indefinitely. Note, in this context, indefinitely means a "long-term preservation intention without a predefined end date." Examples include national archives, museums, libraries, national data archives, and research data repositories, such as CERN and ESA. This permanent preservation cannot be guaranteed in absolute terms, as it is what the technological capabilities make possible.

Medium-term data preservation is the continued access to digital materials beyond changes in technology for a defined period but not indefinitely. (For comparison, the lifespan of standard storage media is three to five years.)

Data needs to be actively maintained, migrated and curated over time, see Section 4.1.3.2 Preserve. The topic of Transform focuses on one aspect of digital preservation: ensuring that digital content can be read from digital storage and through available software tools. Everyone responsible for maintaining IT infrastructure is aware of technical changes in hardware and software, and aware of the lack of backwards compatibility. It is recommended to:

- **Store data on current and actively supported storage media**. While, for example, a tape cartridge has a durability for 15 to 30 years of archival storage [175], it is not guaranteed that a manufacturer will maintain the availability of the tape drives, which could have an impact on reading the data. Therefore, administrators managing data for long-term preservation should plan regular data migrations.
- Similarly for files, **ensure that tools are available to read and interpret the structure and content of files**. This needs to be actively monitored over time. If tools and or file formats are planned to be deprecated data curators need to plan a file migration (EOSC EDEN CPP-014 [176])

4.2 Data Curation

Data curation is one of the most important processes used to maintain and preserve data throughout the data life cycle. The Dutch Data Curation Network of the LCRDM Task Group [177] drafted "Report on the state of the art of data curation in the Netherlands and the feasibility of creating a dedicated Dutch Data Curation Network" [178]. Here data curation is defined as:

> *The activity of managing the use of data from its point of creation to ensure it is available for discovery and reuse in the future. Examples of data curation range from adding, verifying and improving metadata to checking if files open as expected and recording who did what with the dataset in a repository. Researchers, research support staff and repository staff carry out this kind of activity, in different phases of the research data life cycle.*

While data curation is one of the main processes to preserve data over time, it is implemented at different levels and in different ways. The following two, sub-sections provide guidelines on data curation and preservation levels, as defined by the CoreTrustSeal organisation and the standardised CURATED steps checklist from the Dutch Data Curation Network [179].

4.2.1 CoreTrustSeal - Data Curation Levels

CoreTrustSeal [180] is an international, community based, non-governmental, and non-profit organisation promoting sustainable and trustworthy data infrastructures. CoreTrustSeal is best known for its data repository certification programme [181] and the CoreTrustSeal Requirements [182] for trustworthy data repositories.

In addition to the certification program, CoreTrustSeal aims to align data curation and preservation practices for data repositories and data managers. To align these practices, the CoreTrustSeal Board developed a discussion paper to develop guidelines for data curation and preservation levels. At the time of this writing, the position paper has evolved to Version 3 [183].

The CoreTrustSeal position paper identifies four levels:

- **Z. Level Zero.** Content distributed as deposited. Unattended deposit-storage-access.

  Data content and supporting metadata is stored for a given period, or indefinitely. This may include multiple copies and monitoring of bitstreams for integrity. Data content and supporting metadata are distributed to users exactly as they are provided by depositors. Beyond these measures, there are no checks of deposit compliance, no initial curation or active long-term preservation.
- **D. Deposit Compliance**

  Data content and supporting metadata deposited are checked for compliance with defined criteria, e.g. data formats, metadata elements, and compliance with legal and ethical norms. Digital objects that do not meet these criteria may be rejected, or moved forwards to initial curation if provided by the repository.
- **C. Initial Curation**

  The digital objects are curated by the repository to meet defined criteria, which may exceed those defined for Deposit Compliance. This initial curation for access and use may include, e.g., the correction or enhancement of metadata and/or data content, or the creation of dissemination formats.
- **A. Active preservation**

  In addition to D and/or C above the repository takes long-term responsibility for ensuring that the data and metadata can be understood and rendered as required by the designated community for reuse. The preservation actions can be aimed at logical-technical, semantic, or quality aspects of the (meta)data, for example, in response to the threat of technological obsolescence, to accommodate changing needs of the Designated Community, or in response to other considerations such as security or legal concerns.

Logical-technical measures include updating hard- and software environments, archival and dissemination formats of digital objects, and metadata.

Semantic measures include updating the content of metadata elements and other semantic artefacts such as controlled vocabularies and ontologies if necessary. It may include responsibility for editing the structure and content of deposited data.

From the perspective of the CoreTrustSeal, Levels D and C alone are not in scope for CoreTrustSeal certification as they do not entail active long-term preservation and hence do not provide a long-term perspective beyond bit storage.

4.2.2 Data Curation Network - C-U-R-A-T-E-D steps model

The Data Curation Network (DCN) [184] is a membership organisation of institutional and non-profit data repositories whose vision is to advance open research by making data more ethical, reusable, and understandable. DCN has a mission to empower researchers to publish high quality data in an ethical and FAIR way, collaboratively advance the art and science of data curation by creating, adopting, and openly sharing best practices, and supporting thoughtful, innovative, and inclusive data curation training and professional development opportunities.

To support researchers and data curators to prepare datasets for sharing, DCN developed a standardised set of C-U-R-A-T-E-D steps and checklists [185] to ensure that all datasets submitted to the Network receive consistent treatment. The C-U-R-A-T-E-D steps are defined as follows:

- **C -** Check files/code and read documentation (risk mitigation, file inventory, appraisal/selection)
- **U -** Understand the data (or try to), if not… (run files/code, QA/QC issues, readmes)
- **R -** Request missing information or changes (tracking provenance of any changes and why)
- **A -** Augment metadata for findability (DOIs, metadata standards, discoverability)
- **T -** Transform file formats for reuse (data preservation, conversion tools, data visualisation)
- **E -** Evaluate for FAIRness (transparent usage licenses, responsibility standards, metrics for tracking use)
- **D** - Document all curation activities throughout the process

To support the C-U-R-A-T-E-D steps, DCN developed for each of the steps a checklist with essential tasks to perform [186].

The Dutch Data Curation Network also used the C-U-R-A-T-E-D steps model to analyse the data curation practices within the Netherlands, as reported in the *Report on the state of the art of data curation in the Netherlands and the feasibility of creating a dedicated Dutch Data Curation Network.*

4.3 Data Quality

Reproducibility is one of the cornerstones of the scientific process. To be able to reproduce a scientific result a researcher must be able to understand and trust the scientific result and how the scientific process has been conducted. And to be able to trust the scientific result you need quality data.

To ensure data quality, data producers, data managers and data curators need to assess the quality of the data throughout the full research data lifecycle, when a researcher collects a dataset from a data sources for reuse, when new data and research output is generated during processing and analysis, when prepared for preservation, publication, sharing and reuse.

Next to ensure trust in the scientific process, data quality becomes one of the key aspects, in the current age of Artificial Intelligence (AI). In AI there are popular sayings about the importance of quality data: "Garbage In, Garbage Out" or "AI is only as good as Its data"[187]. Preparing data for the use in AI will provide a big push towards improving data quality.

Recent reports have featured data quality as a key topic:

- EOSC Association FAIR Metrics and Data Quality task force subgroup on Data Quality report *Towards a data quality framework for EOSC* [188]
- EOSC EDEN Quality assurance of research outputs [189]

In the next two subsections the key data quality related aspects of these deliverables will be summarised.

4.3.1 Towards a Data Quality Framework for EOSC

During 2021–23, the EOSC Association operated 13 task forces on many EOSC related topics and one of these task forces focused on *FAIR Metrics and Data Quality* [190]. The subgroup on Data Quality produced the report "Towards a data quality framework for EOSC" [191], describing *the basic concepts to build a solid basis for a mutual understanding of data quality in a multidisciplinary environment*.

Data quality is defined *as "*the degree to which a set of inherent characteristics of data fulfils requirements" *(ISO 8000* [192]*)* and a schematic diagram [193], first published in the *Data Science Journal 21*, is used to relate the dataset quality aspects of science, product, stewardship and services to the different phases of the research data lifecycle.

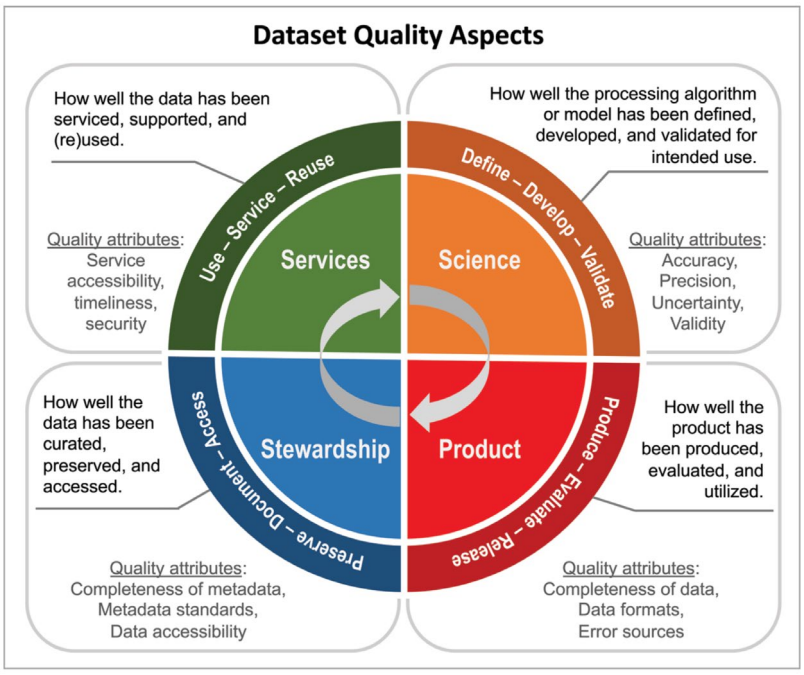

Figure 4.2: Four quality aspects (i.e., science, product, stewardship and service) throughout a dataset lifecycle, including key stages and attributes associated with each

Data quality dimensions can be categorised in many ways. Figure 4.2 categorises quality dimensions using four quality aspects and the phases of the research data lifecycle, indicating important quality attributes for each

aspect. The report also describes intrinsic quality, contextual quality, and representational and accessibility quality.

- **Intrinsic quality** implies that data has essential quality independent of other factors.
- **Contextual quality** highlights the need for quality to be considered within the context of the task at hand. Data must be relevant, timely, complete, and appropriate to add value.
- **Representational and accessibility quality** emphasises the importance of computer systems that store and provide access to data; that is, the system must present information in such a way that it is interpretable, easy to understand, easy to manipulate, and is represented concisely and consistently.

The main recommendations related to data quality include:

- Data quality assessment needs standards to check data against. Unfortunately, not all communities have agreed on standards, therefore EOSC should assist and push each community to agree on community standards to guarantee the FAIR exchange of research data.
- Data in EOSC needs to be presented with enough information for the user to understand how to read and correctly interpret the dataset, what restrictions are in place to use it, and what processes participate in its production. EOSC should ensure that the dataset is structured and documented in a way that can be (re)used and understood. Quality assessments in EOSC should not be concerned with checking the soundness of the data content.
- Errors found by the curators or users need to be rectified by the data producer/provider. If this is not possible, errors need to be documented. Improving data quality as close to the source (i.e., producer or provider) as possible is highly recommended.
- User engagement is necessary to understand the user requirements (needs, expectations, etc.); it may or may not be part of a quality management function. Determining and evaluating stakeholder needs is not a one-time requirement but a continuous and collaborative part of the service delivery process.
- It is recommended to develop a proof-of-concept quality function performing basic quality assessments tailored to EOSC needs (e.g., data reliability and usability). These assessments can also support rewarding research teams most committed to providing FAIR datasets.
- Data quality is a concern for all stakeholders, detailed further in this document. The quality assessments must be a multi-actor process between the data provider, EOSC, and users, potentially extended to other actors in the long run.
- A number of requirements valid for all datasets in EOSC (and beyond) and specific aspects of a maturity matrix gauging the maturity of a community when dealing with quality have been defined.

The FAIR Metrics and Data Quality task force ended in 2023, but the work is continued in the FAIR Assessment and Alignment Opportunity Area [194].

4.3.2 EOSC EDEN Quality Assurance of Research Outputs

EOSC EDEN [195] is a Horizon Europe project [196] that is developing a framework to identify what data are candidates for long-term preservation based on use, benefit, and quality and a model for re-appraisal points along data lifecycle and test usability.

One of the core activities of the EOSC EDEN project is to develop a framework to identify data that is a candidate for long-term preservation based on data quality. Although the EOSC EDEN project started in January 2025 and the delivery of the framework is planned by the end of the project in December 2027, its work provides a good example.

The following information is retrieved from two EOSC EDEN deliverables:

- *D2.1 - EOSC EDEN Specifications and Architecture (Iteration 1)* [197]
- *D3.1 - Report on Discipline Requirements and Needs* [198]

EOSC EDEN has structured the analysis around the Reuse Fitness model (Figure 4.3), which evaluates long-term digital preservation (LTDP) and digital object quality (DOQ) practices through five key components: contextual data quality, technical data quality, metadata quality, trustworthy digital archive capabilities, and preservation objectives/designated community needs.

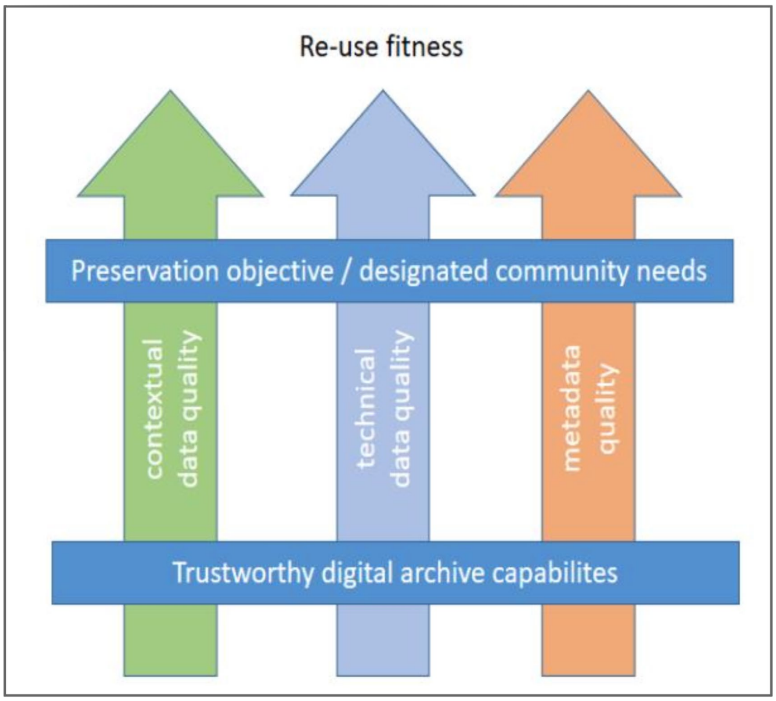

Figure 4.3: Reuse Fitness model

The five key components of the Re-use Fitness model are defined as:

- **Contextual data quality** addresses aspects of quality that should be verified by curators on the basis of best practice and standards. The objectives of such quality assurance is largely focused on authenticity and integrity of objects submitted for publication and archiving, it focuses on applicability and alignment with the policies of the repository, completeness of the deposit, intellectual property and ownership, authenticity and originality, provenance metadata on the creation of the objects supporting reuse, *legal* and *contractual aspects* and the *fitness for use*.
- **Technical data quality** refers to such properties of a research object (such as a dataset) that can be established and verified via automated or algorithmic methods. Such methods can be standardised and provided by services. There are two main considerations for determining and evaluating such properties:
    - *Metadata Confirmation*: To enhance or otherwise supplement the scope of metadata and its accuracy (metadata completeness). As examples: a file may be presented as a given MIME type, but it is prudent to verify that the file is indeed such a type; spatial coverage can be accurately determined directly from file data, or one may want to compute the checksum of a file or collection of files for integrity management.
    - *Data Assessment*: Computing properties of the dataset to inform on aspects such as precision and variance, presence of outliers, or nature of the data distribution. Such (mostly statistical)

properties can be used to identify errors and anomalies in the data. It can also assist with the identification of fake or fraudulent data.

- **Metadata quality** is an important factor in long-term data preservation and reuse. High-quality metadata ensures that archived datasets remain findable, accessible, interoperable, and reusable (FAIR) over time. The FAIR principles focus on rich, machine-readable metadata as the basis for discovery and reuse of digital objects. In practice, this means repositories must collect sufficient descriptive information in a standardised, machine-actionable way aligned with community needs. Poor metadata quality can cause even well-preserved data to become undiscoverable, thus unusable. In a preservation context for metadata quality, the following aspects need to be considered: completeness, schema compliance, machine-actionability, multiple language, fitness for reuse and FAIR metadata assessment.
- **Trustworthy digital archive capabilities** address topics such as: Does a repository apply clear and documented deposit criteria? Does it support versioning? Are datasets machine-readable and/or accessible through an API? Does a repository apply backup and retention policies? Is it compliant with a specific framework/policies (e.g. OAIS [199], CoreTrustSeal [200], FAIR [201], CARE [202], TRUST [203])? Are datasets machine-readable and/or API accessible? Does a repository have a plan on how it is funded? Does it have an exit strategy if long-term funding fails? and Does a designated community offer a choice of well-established repositories? The capabilities a Trustworthy digital archive adequately should undertake are defined as core preservation processes, see Section 4.1.3.2 Preserve.
- **Preservation objectives/designated community needs**: Preservation objectives are a specific, actionable and measurable way in which an object is to be used. The designated community is a group of users, now or in the future, who can understand and use the preserved objects. While the designated community describes the who (the user), the preservation objectives describe how the object is used. Preservation objectives/designated community needs addresses topics as if there is a need for generalist or specialist curators, the level of curation that is applied by a repository, reassessment schedule, if deposits are reassessed over time, and if the original and/or access to a copy of the material are retained.

As previously mentioned, the EOSC EDEN project will further develop a quality assurance framework, including metrics indicating reuse fitness for digital objects.

5 Training on Research Data Management Practices

The 2016 report *Realising the European Open Science Cloud* [204] highlighted an alarming shortage of data expertise in the EU and a pressing requirement with regards to the data expertise needed to support the aims of EOSC was apparent.

As a result of this and the various acts, policies and mandates around sharing data described earlier in this document, many projects and initiatives have emerged to support researchers to acquire relevant data skills and receive the support they need. Training is now provided by a range of different providers including:

- Research institutions providing direct support for their researchers.
- National data services providing general guidance on data skills as well as use of their own repositories.
- Thematic data services providing guidance on discipline-specific data management, as well as their own repositories.
- National and thematic competence centres.
- Projects funded by Horizon Europe and other programmes to support skills development.
- International data-focussed organisations.

There is now a wide range of resources available for training in research data management (RDM) practices and this section will highlight some starting points for those NRENs considering running their own training, or looking for resources for themselves.

It should be noted that this section focuses specifically on research data management, data produced as part of the research process, and not data collected more broadly in organisational operations and service delivery. Also note that several of the resources referenced below relate to open science, as data management is an integral part of this process.

5.1 Institutional Support for Research Data Management Training

The increased need for data management skills has led to the emergence of data stewardship as a profession supporting researchers within RPOs. Skills4EOSC [205] identifies two different types of data stewards:

- **Coordinator data stewards:** Coordinator Data Stewards act as a 'centralised knowledge and communication hub' for researchers. They advise and train on policy, guidelines, data management plans, institutional infrastructure and tools to implement FAIR and CARE principles across the organisation.
- **Embedded data stewards:** Embedded Data Stewards serve research teams, faculties, departments, sections of organisations directly involved in producing research outputs, supporting them to plan and implement FAIR and CARE principles, meeting needs of researchers as they arise, and working with others to ensure outputs are preserved and reusable in the long term.

Data stewards may have different names as there is no consistency between institutions around terminology, or the location of these roles within the institution. A selection includes Data Steward, Data Librarian, Research Data Management Specialist, Research Data Manager or Data Curator.

The additional work required for researchers to meet RDM requirements can be seen as a burden by researchers and the strength of institutional data stewards is in providing local context and personalised, ongoing support. Universities within one's own country may therefore be a good starting point to find resources on RDM training. Some examples of institutional resources include the University of Edinburgh's *Guide to Research Data Management* [206] and the RDM support pages at the University of Utrecht [207] and University of Vienna [208].

5.2 National and Thematic Data Services and Competence Centres

For any training and support, making delivery as relevant as possible to the audience's local contexts is important.

National contexts vary in terms of local systems, infrastructure and policy and legal requirements. National data services may offer guidance and training on their own collections and tools, as well as offering general advice on data management. The *UK Data Service's Learning Hub* [209] and *Research Data Netherlands* [210] are examples.

Disciplinary differences in types of data can affect how training is approached, with subject knowledge and expertise increasing relevance to researchers. Thematic research infrastructures and services provide training in data management and FAIR in their own disciplines, such as the ELIXIR ERIC [211] distributed infrastructure for life *sci*ences data and Consortium of European Social Science Data Archives (CESSDA) ERIC [212].

Centres of support for training and skills development are increasingly referred to as 'competence centres' and may be national, thematic, or both. The Skills4EOSC and OSCARs projects described in Section 5.3 have done work on the coordination of competence centres across the EOSC ecosystem.

5.3 Project Initiatives and International Organisations

In addition to institutional support for RDM, much support is provided by national and thematic competence centres, project initiatives and international organisations. This document has already highlighted many useful resources from various projects, but this section highlights those focused specifically on training and education for RDM.

Resource	Description
Data Curation Network [213]	The DCN has partnered with many members of the curation community to develop resources and share knowledge of data curation practices and tools.
Research Data Alliance [214]	These workshops have been carefully developed to help data professionals grow in step with the rapidly evolving research data landscape.
CODATA [215]	The International Science Committee's Committee on Data, CODATA provides resources including the Terms4FAIRskills [216] project and the Data Science Summer Schools in collaboration with the RDA [217].
The Carpentries [218]	The Carpentries teaches foundational coding and data science skills to researchers worldwide. It has Train the Trainer programmes and materials on Data Carpentry and Software Carpentry.
Skills4EOSC [219]	The Skills4EOSC project ran from 2022-25 developed a competence centre network and common methodologies, activities and training resources to unify the current

Resource	Description
	training landscape into a collaborative and reliable ecosystem for FAIR and open science.
OSCARs [220]	This project aims to consolidate approaches to EOSC across five discipline science clusters, including developing competence centres.
EOSC Academy [221]	EOSC Gravity will establish a dedicated EOSC Academy, offering structured learning opportunities through national and project-based tracks. This will help build a skilled community capable of driving EOSC-related activities within institutions and countries. This is aimed at prospective nodes and services in the EOSC Federation.

Table 5.1: Useful resources for research data management

5.4 Resources for Data Professionals

As data stewardship is emerging as a profession, resources for those wishing to develop skills in these roles are increasing. The Skills4EOSC Minimal Viable Skills Set [222] for Data Stewards provides a starting point for development and they have also developed a more detailed curriculum for data stewards [223], which contains eight sections:

- Training skills
- Research data management
- Research software management
- Policy and governance
- Usage rights and licences
- Ethics
- Personal data and GDPR
- Transversal or soft skills

The Skills4EOSC minimal viable skillset and curricula are increasingly being adapted to create localised versions, for example being used as the foundation for the Sonrai Irish Data Steward Network's Data microcredential, *Introduction to Data Stewardship* [224] and the Research Data Netherlands' *Competency Framework for Data Professionals* [225]. While not common, some institutions offer data stewardship certificates, for example, the University of Vienna [226].

There are also a number of informal, online courses available specifically for those interested in data stewardship roles, such as "Essentials 4 Data Support" [227] and the EOSC Synergy / FAIRsFAIR's "Data Steward Training" course [228].

5.5 Quality Assurance and FAIR Training

There are many factors to consider when delivering training in RDM, including key decisions such as audience, level and delivery mode. There are 'Train the Trainer' courses some free and some fee-paying, which cover pedagogic good practice specifically for this topic. The RDA runs a *Train the Trainer workshop on RDM* [229] and OpenAIRE launched an *RDM Train the Trainer Bootcamp* in 2025 [230].

Ensuring creation of *good quality training* is important and the Skills4EOSC *Quality Assurance Certification Framework* provides detailed guidance on this [231], along with a quick *Quality Compass checklist* [232]. The

EOSC Synergy Online Training Handbook is a useful resource with practical planning templates [233]. Skills4EOSC also provides recommendations for the set-up of *Open Science and Research Data Management thematic training* [234].

Training materials should be made as reusable as possible by applying the FAIR principles. This aims to ensure materials can be discovered and reused more easily, saving time and reducing duplication of effort. If you are producing training materials, guidelines can be found in the Skills4EOSC *FAIR by Design Methodology* [235] and Elixir's *FAIR training handbook* [236].

5.6 The Role of NRENs and Research Infrastructures

As described throughout this section, there is already a wide range of training materials and services available to support research data management. The level of support provided by NRENs and research infrastructures will depend on the remit and strategies of each individual organisation and to what extent data services and RDM training are in scope.

For example, EUDAT and CSC have a strong focus on data, OpenAIRE on Open Science, and have training offers on these topics.

- EUDAT: https://eudat.eu/eudat-training
- OpenAIRE: https://openplato.eu/
- CSC: https://csc.fi/en/trainings/

In many cases it may be appropriate to be only a broker, to signpost to training, or collaborate with partners to deliver training. NRENs and research infrastructures should consider their audience and aims carefully and explore what other relevant provisions already exist and this document provides some starting points.

6 Conclusions

This document has outlined the key concepts, frameworks, and best practices underpinning effective Research Data Management (RDM) within the GÉANT community. By situating RDM within the wider context of the European Strategy for Data, FAIR principles, and key technologies such as persistent identifiers, semantic artefacts, and knowledge graphs, it highlights the evolving expectations placed on researchers, data stewards, and service providers. The overview of the research and research data lifecycles provides a structured reference for planning, managing, preserving, and sharing data in ways that support transparency, reproducibility, and long-term value.

European initiatives — including EOSC, FAIRsFAIR, FAIRCORE4EOSC, FAIR IMPACT, EOSC EDEN, and OSTrails — continue to strengthen the technical and policy landscape for FAIR and trustworthy data. These efforts emphasise the need for high quality metadata, interoperable standards, robust preservation processes, and machine actionable tooling that can scale with scientific demand. The document also underscores the growing importance of training and capacity building, reflecting the recognised skills gap across Europe in data stewardship and research data management practices.

Overall, the recommendations presented here aim to support NRENs and the broader research community in adopting consistent, sustainable approaches to data management. Continued collaboration, alignment with European frameworks, and investment in skills and infrastructure will be essential to enabling a resilient Web of FAIR data and services and to maximising the impact of publicly funded research in the years ahead.

Glossary

AI	Artificial Intelligence
API	Application Programming Interface
CAT	Compliance Assessment Toolkit
CPP	Core Preservation Process
CRA	Cyber Resilience Act
DA	Data Act
DGA	Data Governance Act
DiSSCo	Distributed System of Science Collections
DMA	Digital Markets Act
DMP	Data Management Plan
DO	Digital Object
DO-IRP	Digital Object Identifier Resolution Protocol
DOA	Digital Object Architecture
DOIP	Digital Object Interface Protocol
DTR	Data Type Registry
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, and Reusable
FDO	FAIR Digital Object
FDP	FAIR Data Point
FFD	Free Flow of non-personal Data
GDPR	General Data Protection Regulation
GPAI	General Purpose AI
HVD	High-Value Dataset
IOPS	Input/Output per Second
LTDP-TF	Long-Term Data Preservation Task Force
MSCR	Metadata Schema and Crosswalk Registry
NREN	National Research and Education Network
OOD	Open Data Directive
PID	Proportional Integral Derivative
PIDMR	PID Meta Resolver
RAiD	Research Activity iDentifier service
RDGraph	Research Discovery Graph
RDA	Research Data Alliance
RDM	Research Data Management
REMS	Resource Entitlement Management System
RO-Crate	Research Object Crate
RSAC	Research Software APIs and Connectors
SAC	Semantic Artefact Catalogue
SKG-IF	Scientific Knowledge Graph – Interoperability Framework
SSH	Secure Shell
SWHN	Software Heritage Mirror
TDA	Trustworthy Digital Archive
TDR	Trusted Digital Repository
TLS	Transport Layer Security
TRE	Trusted Research Environment
WG	Working Group

References

[1] https://zenodo.org/records/4915862

[2] European Strategy for Data https://digital-strategy.ec.europa.eu/en/policies/strategy-data

[3] European Strategy for Data fact page https://digital-strategy.ec.europa.eu/en/factpages/data

[4] EU Data Act https://digital-strategy.ec.europa.eu/en/policies/data-act

[5] EU General Data Protection Regulation https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

[6] EU Data Governance Act https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

[7] EU Regulation on the Free Flow of Non-Personal Data https://digital-strategy.ec.europa.eu/en/policies/non-personal-data

[8] High-Value Datasets – Commission Implementing Regulation (EU) 2023/138 https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj/eng

[9] Common European data spaces | Shaping Europe's digital future https://digital-strategy.ec.europa.eu/en/policies/data-spaces

[10] EU AI Act https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

[11] EU Cyber Resilience Act https://eur-lex.europa.eu/eli/reg/2024/2847/oj/eng

[12] EU Digital Services Act https://digital-strategy.ec.europa.eu/en/policies/digital-services-act

[13] FAIR data – guiding principles https://force11.org/info/guiding-principles-for-findable-accessible-interoperable-and-re-usable-data-publishing-version-b1-0/

[14] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[15] EOSC implementation https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science/european-open-science-cloud-eosc_en#eosc-implementation

[16] FAIRsFAIR https://www.fairsfair.eu/

[17] FAIRsFAIR in EOSC https://www.fairsfair.eu/fairsfair-eosc

[18] FAIRCORE4EOSC https://faircore4eosc.eu/

[19] Horizon Europe Co-programmed Partnership for the European Open Science Cloud (EOSC). (2024). Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC) (v1.3, 2024) (1.3). Zenodo. https://doi.org/10.5281/zenodo.17582648

[20] FAIRCORE4EOSC case studies https://faircore4eosc.eu/case-studies

[21] FAIR-IMPACT https://fair-impact.eu/

[22] FAIR Implementation Framework https://fair-impact.eu/fair-implementation-framework

[23] ACME-FAIR: a Guide for Research Performing Organisations (RPO) https://www.fairsfair.eu/acme-fair-guide-rpo

[24] Core Components Supporting a FAIR EOSC https://cordis.europa.eu/project/id/101057264

[25] EOSC-Core components https://faircore4eosc.eu/eosc-core-components;

[26] van Horik, R., & Hugo, W. (2024). D3.3 - Guidelines for creating a user tailored EOSC Compliant PID Policy. Zenodo. https://doi.org/10.5281/zenodo.11354246

[27] PID Selection Tool https://pid.services.base4nfdi.de/pidtool/

[28] PID4NDFI https://pid.services.base4nfdi.de/

[29] PID wijzer https://www.pidwijzer.nl/en/pid_results/new

[30] Dr. Martin Siefkes https://www.siefkes.de/

[31] Oscar Corcho et al, "A maturity model for catalogues of semantic artefacts", https://arxiv.org/abs/2305.06746

[32] Gonzalez Beltran, A. N., & Wilson, A. (2024). M4.3 - Specification of semantic artefact description. Zenodo. https://doi.org/10.5281/zenodo.10725304

[33] Semantic Artefacts Governance and Management https://fair-impact.eu/semantic-artefacts-governance-and-management

[34] Linked Open Terms https://lot.linkeddata.es/

[35] Google Knowledge Graph https://en.wikipedia.org/wiki/Knowledge_Graph_(Google)

[36] HPCwire: Scaling the Knowledge Graph Behind Wikipedia https://www.hpcwire.com/bigdatawire/2025/07/10/scaling-the-knowledge-graph-behind-wikipedia/

[37] OpenAIRE Graph https://catalogue.openaire.eu/service/openaire.graph/overview

[38] SURF Tech Trends 2025: Data Management https://www.surf.nl/files/2025-10/sf_ttr26_en_datamanagement.pdf

[39] EOSC EU Node Resource Catalogue https://open-science-cloud.ec.europa.eu/news/eosc-eu-node-resource-catalogue

[40] Scientific Knowledge Graphs – Interoperability Framework (SKG-IF) Working Group https://www.rd-alliance.org/groups/scientific-knowledge-graphs-interoperability-framework-skg-if-wg/

[41] Barcelona Declaration on Open Research Information https://barcelona-declaration.org/

[42] Mannocci, A., Peroni, S., & RDA Scientific Knowledge Graphs Interoperability Framework (SKG-IF) WG Members. (2026). Scientific Knowledge Graphs Interoperability Framework (SKG-IF): Enabling Data Exchange for SKGs (1.1). Zenodo. https://zenodo.org/records/18713833

[43] OSTrails https://ostrails.eu/

[44] SKG Commons https://docs.ostrails.eu/en/latest/commons/skg/skg-commons.html

[45] EOSC Association. (2026). EOSC Federation Handbook (2.0). EOSC Association. https://doi.org/10.5281/zenodo.18454649

[46] Manghi, P., Athanasiou, S., Karmas, T., & Szegedi, P. (2025). Registration of EOSC Research Product Catalogues in the EOSC EU Node (3.1). Zenodo. https://doi.org/10.5281/zenodo.17513272

[47] FAIR Data Maturity Model Working Group. (2020). FAIR Data Maturity Model. Specification and Guidelines (1.0). Zenodo. https://doi.org/10.15497/rda00050

[48] The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments https://datascience.codata.org/articles/dsj-2020-041#B4

[49] F-UJI https://www.f-uji.net/

[50] Devaraju, A., Huber, R., Mokrane, M., Herterich, P., Cepinskas, L., de Vries, J., L'Hours, H., Davidson, J., & Angus White. (2022). FAIRsFAIR Data Object Assessment Metrics (0.5). Zenodo. https://doi.org/10.5281/zenodo.6461229

[51] FAIR-Aware https://fairaware.dans.knaw.nl/

[52] O'FAIRe https://github.com/agroportal/fairness

[53] FOOPS! https://w3id.org/foops

[54] FAIR Champion https://github.com/OSTrails/FAIR-Champion

[55] FAIROs https://oeg-upm.github.io/FAIR-Research-Object/ro/

[56] Research Software FAIRness Checks (RSFC) https://github.com/oeg-upm/rsfc

[57] FAIRsharing https://fairsharing.org/

[58] FAIR Assessment Authoring Tool https://ostrails-fair.fair-wizard.com/

[59] OpenAIRE Metadata Validator https://code-repo.d4science.org/MaDgIK/openaire-ostrails-api/

[60] FAIR Data Self-Assessment Tool https://ardc.edu.au/resource/fair-data-self-assessment-tool/

[61] FAIR-Checker https://fair-checker.france-bioinformatique.fr/

[62] Wittenburg, P. (2019). Digital Objects as Drivers towards Convergence in Data Infrastructures [Data set]. B2SHARE v2. https://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11

[63] FAIR Digital Object (FDO) Forum https://fairdo.org/

[64] Research Data Alliance (RDA) https://www.rd-alliance.org/

[65] DONA Foundation https://www.dona.net

[66] GO FAIR https://www.go-fair.org/

[67] Digital Object Architecture https://www.dona.net/digitalobjectarchitecture

[68] Wittenburg, P., Anders, I., Blanchi, C., Buurman, M., Goble, C., Grieb, J., Hardisty, A., Islam, S., Jejkal, T., Kálmán, T., Kirkpatrick, C., Lannom, L., Lauer, T., Manepalli, G., Peters-von Gehlen, K., Pfeil, A., Quick, R., van de Sanden, M., Schwardmann, U., … Wieder, P. (2022). FAIR Digital Object Demonstrators 2021 (final). Zenodo. https://doi.org/10.5281/zenodo.5872645

[69] Distributed System of Science Collections (DiSSCo) https://www.dissco.eu/

[70] Handle Identifier System https://www.dona.net/handle-system

[71] Signposting the Scholarly Web https://signposting.org/

[72] Van de Sompel, H. (2023, May 27). FAIR Digital Objects and FAIR Signposting. https://doi.org/10.5281/zenodo.7977333

[73] Signposting adopters https://signposting.org/adopters/

[74] CKAN https://ckan.org/

[75] Dataverse https://dataverse.org/

[76] DSpace https://dspace.org/

[77] InvenioRDM https://inveniosoftware.org/

[78] FAIR Data Point https://www.fairdatapoint.org/

[79] FAIR Data Point API specification https://specs.fairdatapoint.org/

[80] DCAT https://www.w3.org/TR/vocab-dcat-2/

[81] Dublin Core https://dublincore.org/

[82] Linked Data Platform (LDP) https://www.w3.org/TR/ldp/

[83] Representational State Transfer | Wikipedia https://en.wikipedia.org/wiki/Representational_state_transfer

[84] Luiz Olavo Bonino da Silva Santos, Kees Burger, Rajaram Kaliyaperumal, Mark D. Wilkinson; FAIR Data Point: A FAIR-oriented approach for metadata publication. Data Intelligence 2022; https://direct.mit.edu/dint/article/5/1/163/112599/FAIR-Data-Point-A-FAIR-Oriented-Approach-for

[85] FAIR Data Point Index homepage https://home.fairdatapoint.org/

[86] FAIR RO-Crate https://www.researchobject.org/ro-crate/about_ro_crate

[87] Goble, C. (2024). FAIR Digital Research Objects: Metadata Journeys. University of Auckland Seminar, Auckland. Zenodo. https://doi.org/10.5281/zenodo.10710142

[88] Schema.org http://schema.org/

[89] FAIR Digital Object Framework https://fairdigitalobjectframework.org/

[90] RO-Crate Tools https://www.researchobject.org/ro-crate/tools

[91] Science Europe https://scienceeurope.org/

[92] Science Europe – Research Data Management https://scienceeurope.org/our-priorities/open-science/research-data-management/

[93] RDA Mapping the landscape of digital research tools WG https://www.rd-alliance.org/groups/rda-ofr-mapping-landscape-digital-research-tools-wg/activity

[94] Deliverable 1: The creation of a harmonised research data lifecycle (RDL) model and crosswalk to existing models https://www.rd-alliance.org/wp-content/uploads/2024/09/D1_The-creation-of-a-harmonised-research-data-lifecycle-RDL-model-and-crosswalk-to-existing-models-.pdf

[95] Deliverable 1 – see [92]

[96] Australian Research Data Commons (ARDC) https://ardc.edu.au/

[97] SURF: Research Activity Identifier (RAiD) https://www.surf.nl/en/themes/open-science/open-research-information/research-activity-identifier-raid

[98] OpenAIRE ARGOS https://argos.openaire.eu

[99] DCC DMPonline https://dmponline.dcc.ac.uk/

[100] https://github.com/DMPRoadmap/roadmap

[101] CSC's DMP Tuuli https://dmptuuli.fi/

[102] Belnet's DMPonline instance https://dmponline.be/

[103] RDA DMP Common Standards WG: Common Standards for Machine-actionable Data Management Plans https://www.rd-alliance.org/groups/dmp-common-standards-wg/outputs/?output=94576

[104] Miksa, T., Walk, P., & Neish, P. (2020). RDA DMP Common Standard for Machine-actionable Data Management Plans. Zenodo. https://doi.org/10.15497/rda00039

[105] OSTrails https://ostrails.eu/

[106] Deliverable GN5-2 D1.1 Data Management Plan GN5-2_D1-1_Data-Management-Plan.pdf [available to project participants only]

[107] iperf https://software.es.net/iperf/

[108] FileSender https://filesender.org/

[109] Linux manual page – scp https://man7.org/linux/man-pages/man1/scp.1.html

[110] Linux manual page – sftp https://man7.org/linux/man-pages/man1/sftp.1.html

[111] Linux manual page – srsync https://man7.org/linux/man-pages/man1/rsync-ssl.1.html

[112] rclone https://rclone.org/

[113] rclone documentation https://rclone.org/docs/

[114] CERN File Transfer Service https://fts.web.cern.ch/fts/

[115] World LHC Computing Grid (WLCG) https://wlcg.web.cern.ch/

[116] xroot protocol https://xrootd.org/

[117] SRM https://sdm.lbl.gov/srm-wg/

[118] DANS research data file formats https://dans.knaw.nl/en/file-formats/

[119] NetCDF https://www.unidata.ucar.edu/software/netcdf

[120] HDF5 https://www.hdfgroup.org/solutions/hdf5/

[121] Galaxy Project https://galaxyproject.org/

[122] European Galaxy server https://usegalaxy.eu/

[123] Reana https://www.reana.io/

[124] Workflow Hub https://workflowhub.eu/

[125] DARE UK – What is a Trusted Research Environment? https://dareuk.org.uk/wp-content/uploads/2023/07/DARE_UK_Public_Dialogue-What_is_a_trusted_research_environment.pdf

[126] SANE: secure environment for analysing sensitive data www.surf.nl/en/themes/research-infrastructure/sane-secure-environment-for-analysing-sensitive-data
[127] CSC's ePouta virtual private cloud https://research.csc.fi/service/epouta/
[128] EOSC-ENTRUST https://eosc-entrust.eu/
[129] TITAN https://titan-eosc.eu/
[130] SIESTA https://eosc-siesta.eu/
[131] https://en.wikipedia.org/wiki/Enterprise_file_synchronization_and_sharing
[132] https://www.openaire.eu/
[133] https://amnesia.openaire.eu/
[134] Andreu, T., Anglada, L., Antos, D., Bähr, T., Brzeźniak, M., Burgi, P.-Y., Cavet, C., Celjak, D., Crépé-Renaudin, S., De Loof, C., Dillo, I., Dubois, O., Fernandes, R., Forsström, P.-L., Ganis, G., Gibney, E., Holl, A., L'Hours, H., Lamers, D., … Wyns, R. (2023). EOSC Preservation: Overview Discussion Paper. Zenodo. https://doi.org/10.5281/zenodo.7516259
[135] Science Europe. (2021). Practical Guide to the International Alignment of Research Data Management - Extended Edition. https://doi.org/10.5281/zenodo.4915862
[136] https://www.coretrustseal.org/
[137] https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html
[138] http://www.iso16363.org/
[139] https://www.openaire.eu/
[140] https://www.openaire.eu/find-trustworthy-data-repository
[141] https://zenodo.org/
[142] https://home.cern/
[143] https://b2share.eudat.eu/
[144] https://eudat.eu/
[145] https://www.re3data.org/
[146] https://fairsharing.org/
[147] Anglada, L., Antos, D., Bähr, T., Burgi, P.-Y., Celjak, D., De Loof, C., Dillo, I., Dubois, O., Fernandes, R., Forsström, P.-L., Ganis, G., Holl, A., Lamers, D., Lammert, A., L'Hours, H., Majdanski, M., Nief, J.-Y., Piroi, F., Presser, K., … Wyns, R. (2024). LTDP-TF Final Report & Recommendations. Zenodo. https://doi.org/10.5281/zenodo.10820893
[148] EOSC EDEN T1.2, Lindlar, M., Caron, B., Benauer, M., Kylander, J., Dekeyser, K., Addis, M., Levlin, M., Laukkanen, M., Lehtonen, J., Burger, F., Koho, T., Schwab, F., Molloy, L., & Zhang, F. (2025). EOSC EDEN M1.1 – Report on Identification of Core Preservation Processes. Zenodo. https://doi.org/10.5281/zenodo.16992452
[149] https://www.dpconline.org/digipres/implement-digipres/procurement-toolkit/procurement-toolkit-core-requirements
[150] https://www.ugent.be/en/research/openscience/datamanagement/after-research/sharing.htm
[151] Research Data Management Support. (2025). Utrecht University Data Publication Guidelines (1.0). Utrecht University. https://doi.org/10.5281/zenodo.15149066
[152] https://dans.knaw.nl/en/depositing-data-manual/before-depositing_ds/
[153] https://dans.knaw.nl/en/file-formats/
[154] https://www.unidata.ucar.edu/software/netcdf
[155] https://www.hdfgroup.org/solutions/hdf5/
[156] https://www.doi.org/
[157] https://orcid.org/
[158] https://ror.org/
[159] https://www.crossref.org/documentation/research-nexus/grants/
[160] https://www.raid.org/
[161] https://terminology.tib.eu/ts
[162] https://www.tib.eu/en/
[163] https://dtr-pit.pidconsortium.net
[164] https://www.pidconsortium.net/

[165] https://terminology.services.base4nfdi.de/
[166] https://www.nfdi.de/?lang=en
[167] https://rea.ec.europa.eu/open-science_en
[168] https://eudat.eu/
[169] https://ufal.github.io/public-license-selector/
[170] https://library.nd.edu.au/researchers/rdm/publication#s-lib-ctab-21794921-4
[171] https://www.elixir-finland.org/en/aai-rems-2/
[172] https://csc.fi/
[173] https://elixir-europe.org/
[174] Andreu, T., Anglada, L., Antos, D., Bähr, T., Brzeźniak, M., Burgi, P.-Y., Cavet, C., Celjak, D., Crépé-Renaudin, S., De Loof, C., Dillo, I., Dubois, O., Fernandes, R., Forsström, P.-L., Ganis, G., Gibney, E., Holl, A., L'Hours, H., Lamers, D., … Wyns, R. (2023). EOSC Preservation: Overview Discussion Paper. Zenodo. https://doi.org/10.5281/zenodo.7516259
[175] https://en.wikipedia.org/wiki/Linear_Tape-Open
[176] https://zenodo.org/records/16992452/files/EOSC-EDEN_CPP-014_File_Migration.pdf
[177] https://lcrdm.nl/
[178] Slouwerhof, I., Jetten, M., Elsenga, C., Groot, N. de ., Grootveld, M., Karvovskaya, L., Ras, M., Smaele, M. de ., Visscher, R., Berg, B. van . den ., & Verheul, I. (2019). Dutch Data Curation Network. Report on the state of the art of data curation in the Netherlands and the feasibility of creating a dedicated Dutch Data Curation Network /Een Nederlands netwerk voor datacuratie. Rapportage over de stand van zaken rond datacuratie in Nederland en de haalbaarheid van een speciaal datacuratienetwerk. (Final Version, December 2019). Zenodo. https://doi.org/10.5281/zenodo.3564861
[179] https://datacuration.network/
[180] https://www.coretrustseal.org/
[181] https://www.coretrustseal.org/apply/
[182] https://www.coretrustseal.org/why-certification/requirements/
[183] CoreTrustSeal Standards and Certification Board. (2024). Curation & Preservation Levels: CoreTrustSeal Position Paper. Zenodo. https://doi.org/10.5281/zenodo.11476980
[184] https://datacuration.network/
[185] https://datacuration.network/outputs/workflows/
[186] “The DCN CURATE(D) Steps." z.umn.edu/curate
[187] https://saifr.ai/blog/garbage-in-garbage-out-why-data-quality-is-critical-to-ai
[188] Lacagnina, C., David, R., Nikiforova, A., Kuusniemi, M. E., Cappiello, C., Biehlmaier, O., Wright, L., Schubert, C., Bertino, A., Thiemann, H., & Dennis, R. (2023). TOWARDS A DATA QUALITY FRAMEWORK FOR EOSC (1.0.0). Zenodo. https://doi.org/10.5281/zenodo.7515816
[189] https://zenodo.org/records/17984753
[190] https://eosc.eu/advisory-groups/fair-metrics-and-data-quality
[191] Lacagnina, C., David, R., Nikiforova, A., Kuusniemi, M. E., Cappiello, C., Biehlmaier, O., Wright, L., Schubert, C., Bertino, A., Thiemann, H., & Dennis, R. (2023). TOWARDS A DATA QUALITY FRAMEWORK FOR EOSC (1.0.0). Zenodo. https://doi.org/10.5281/zenodo.7515816
[192] https://www.iso.org/obp/ui/#iso:std:iso:8000:-2:ed-2:v1:en:term:3.4.8
[193] Peng, G, Lacagnina, C, Downs, RR, Ganske, A, Ramapriyan, HK, Ivánová, I, Wyborn, L, Jones, D, Bastin, L, Shie, C-L and Moroni, DF. 2022. Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets. Data Science Journal, 21: 8, pp. 1–20. DOI: https://doi.org/10.5334/dsj-2022-008
[194] https://eosc.eu/opportunity-area-exp/oa3-fair-assessment-and-alignment
[195] https://eden-fidelis.eu/
[196] https://cordis.europa.eu/project/id/101188015
[197] Middelbos, W., Addis, M., Hugo, W., Steinhoff, W., Snyder, K. P., Wijngaarden, M., Presser, K., Cieśla, M., Le Meur, J.-Y., Koho, T., Kylander, J., Laukkanen, M., Lehtonen, J., Levlin, M., Dekeyser, K., Huber, R., Weiser, J., Høier, S., Caron, B., … Klemetsen, T. (2025). D2.1 - EOSC EDEN Specifications and Architecture (Iteration 1) (V1.0). Zenodo. https://doi.org/10.5281/zenodo.17232536

[198] Andreassen, H. N., Flügel, A.-L., Klemetsen, T., Smart, K., Benauer, M., Cannizzaro, G., Ciesla, M., Conzett, P., Huber, R., Høier, S., Laik, A., Lammert, A., Le Meur, J.-Y., L'Hours, H., Lindlar, M., Mehl, F., Mendes de Farias, T., Middlebos, W., Parkes, O., … Märkälä, A. (2025). EOSC EDEN D3.1 - Report on Discipline Requirements and Needs (V1.0). Zenodo. https://doi.org/10.5281/zenodo.15789261
[199] http://www.oais.info/
[200] https://www.coretrustseal.org/
[201] https://force11.org/info/the-fair-data-principles/
[202] https://www.gida-global.org/care
[203] https://www.nature.com/articles/s41597-020-0486-7
[204] https://digital-strategy.ec.europa.eu/en/library/first-report-high-level-expert-group-european-open-science-cloud
[205] Hansen, K. K., White, A., Horton, L., Koteska, B., Rauste, P., & van Leersum, N. (2025). Data Steward: Minimum Viable Skills Profile. Zenodo. https://doi.org/10.5281/zenodo.16760455
[206] https://library.ed.ac.uk/sites/default/files/2024-08/A%20Guide%20to%20Research%20Data%20Management%20at%20Edinburgh%202019.pdf
[207] https://www.uu.nl/en/research/research-data-management
[208] https://rdm.univie.ac.at/
[209] https://ukdataservice.ac.uk/learning-hub/
[210] https://researchdata.nl/en/home/
[211] https://elixir-europe.org/platforms/training
[212] https://www.cessda.eu/Training
[213] Data Curation Network https://datacuration.network/education-resources/
[214] Research Data Alliance https://www.rd-alliance.org/regions-rda-europe/rda-europe-training-workshops/
[215] CODATA https://codata.org/initiatives/data-skills/
[216] https://codata.org/initiatives/data-skills/wg-terms4fairskills/
[217] https://codata.org/initiatives/data-skills/research-data-science-summer-schools/
[218] The Carpentries https://codata.org/initiatives/data-skills/
[219] Skills4EOSC https://www.skills4eosc.eu/resources/deliverables-milestones
[220] OSCARs https://oscars-project.eu/oscars-objectives
[221] EOSC Academy https://eosc.eu/eosc-gravity
[222] https://zenodo.org/records/16992361
[223] https://skills4eosc-dscurriculum.github.io/DataSteward-Training-Curriculum/latest/
[224] https://datastewards.ie/microcredential/
[225] https://researchdata.nl/en/about-training/competency-framework/
[226] https://rdm.univie.ac.at/data-stewards-at-the-university/become-a-data-steward/
[227] https://researchdata.nl/en/essentials-4-data-support/
[228] https://moodle.learn.eosc-synergy.eu/course/view.php?id=132&section=0#tabs-tree-start
[229] https://www.eventbrite.be/e/rda-europe-train-the-trainer-workshop-on-research-data-management-2026-tickets-1978259002931?aff=oddtdtcreator
[230] https://www.openaire.eu/community/blogs/empowering-the-next-generation-of-data-stewards-join-the-fair-rdm-bootcamp
[231] https://www.skills4eosc.eu/resources/quality-assurance-certification-framework
[232] https://www.skills4eosc.eu/resources/skills4eosc-quality-compass
[233] https://moodle.learn.eosc-synergy.eu/course/view.php?id=15
[234] https://zenodo.org/records/16992361
[235] https://www.skills4eosc.eu/resources/fair-by-design-methodology
[236] https://elixir-europe-training.github.io/ELIXIR-TrP-FAIR-training-handbook/