Project Title	FIDELIS: Establishing A European Network of Trustworthy Digital Repositories
Project Acronym	FIDELIS
Grant Agreement No.	101188078
Start Date of Project	2025-01-01
Duration of Project	36 months
Project Website	https://eden-fidelis.eu/

Digital repository capabilities and characteristics mapping report

Work Package	WP5 - Building common understanding of TDRs and a harmonised matrix of repository capabilities and characteristics
Lead Author (Org)	Philipp Conzett (UiT The Arctic University of Norway, 0000-0002-6754-7911)
Contributing Author(s) (Org)	Severine Duvaud (SIB Swiss Institute of Bioinformatics, 0000-0001-7892-9678) Thomas Jouneau (INRAE/Université de Lorraine, 0000-0001-5986-8128) Joel Kallio (Tampere University, Finnish Social Science Data Archive, 0009-0003-5076-9985) Terje Klemetsen (UiT The Arctic University of Norway, 0000-0002-2024-1798) Jonas Recker (GESIS - Leibniz Institute for the Social Sciences, 0000-0001-9562-3339) Pavel Straňák (Charles University, 0000-0002-6895-8536) Sanni Tujunen (Tampere University, Finnish Social Science Data Archive, 0009-0003-9076-8397) Mari Kleemola (Tampere University, Finnish Social Science Data Archive, 0000-0001-8855-5075) Henna Kaartinen (Tampere University, Finnish Social Science Data Archive, 0009-0001-2148-1822)
Date	2025-10-29
Version	V1.0
DOI	https://doi.org/10.5281/zenodo.17471910

Funded by the European Union

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Author	Notes
0.1	2025.02.19	Philipp Conzett (UiT)	First draft of spreadsheet for resource mapping
0.2	2025.06.25	Philipp Conzett (UiT), Severine Duvaud (SIB Swiss Institute of Bioinformatics), Terje Klemetsen (UiT), Sanni Tujunen (TAU-FSD), Joel Kallio (TAU-FSD), Jonas Recker (GESIS)	First resources added to spreadsheet
0.3	2025.07.09	Philipp Conzett (UiT)	Aligned spreadsheet with TTRAM v1.0
0.4	2025.05.28	Philipp Conzett (UiT)	Outline and ToC of report
0.5	2025.10.17	Philipp Conzett (UiT), Severine Duvaud (SIB Swiss Institute of Bioinformatics), Terje Klemetsen (UiT), Sanni Tujunen (TAU-FSD), Joel Kallio (TAU-FSD), Jonas Recker (GESIS)	First draft of Results section
0.6	2025.10.20	Philipp Conzett (UiT)	Revised Results section and created first complete draft of report.
0.7	2025.10.26	Severine Duvaud (SIB), Terje Klemetsen (UiT), Sanni Tujunen (TAU-FSD), Joel Kallio (TAU-FSD), Jonas Recker (GESIS), Thomas Jouneau (INRAE)	Reviewed complete draft. Revised Agri-Food sections, added methodology

Funded by the European Union

0.8	2025.10.26	Philipp Conzett (UiT)	Reviewed revised version and created final draft
0.9	2025.10.28	Mari Kleemola (TAU-FSD)	WPL review
1.0	2025.10.29	Philipp Conzett (UiT)	Final version

Funded by the European Union

Table of Contents

Funded by the European Union

Funded by the European Union

TERMINOLOGY

Terminology/Acronym	Description
AF	Activity or Function drawn from the reference list of Activities and Functions in TTRAM
API	Application Programming Interface
CARE	Principles for Indigenous Data Governance: Collective benefit, Authority to control, Responsibility, Ethics
CMDI	Component MetaData Infrastructure – a metadata framework used in CLARIN
CTS	CoreTrustSeal – a certification for trustworthy data repositories
DMP	Data Management Plan
EOSC	European Open Science Cloud – a pan-European initiative to provide researchers with access to FAIR data and services
FAIR	Findable, Accessible, Interoperable, Reusable – guiding principles for scientific data management and stewardship
GDPR	General Data Protection Regulation
LoRCaP	Levels of Retention, Curation and Preservation
OAIS	Open Archival Information System – a reference model for digital preservation
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting – a protocol for harvesting metadata from repositories
PID	Persistent Identifier – a long-lasting reference to a digital object, such as a DOI or Handle
TDR	Trustworthy Digital Repositories
TRUST	Transparency, Responsibility, User focus, Sustainability, Technology – principles for trustworthy data repositories
TTRAM	Transparent Trustworthy Repository Attributes Matrix – a reference model developed in FIDELIS to better understand repositories and align repository capabilities
Zenodo	An open-access repository for research outputs, often used for publishing deliverables in EU projects

Funded by the European Union

Executive Summary

This report presents one of three main outcomes of Task 5.2 (T5.2) “Preparing for Federation” within the FIDELIS project, which aims to establish a European Network of Trustworthy Digital Repositories (TDRs) aligned with the European Open Science Cloud (EOSC). The task addresses the challenge of supporting digital repositories in improving their practices and achieving greater interoperability and alignment with EOSC requirements.

To this end, T5.2 conducted a comprehensive mapping of repository capabilities and characteristics across five scientific communities – Agri-food, Biomedical Sciences, Climate Science, Linguistics, and Social Sciences – through desk research and analysis of existing standards, best practices, solutions, landscape analyses, etc. The mapping was guided by the Transparent Trustworthy Repository Attributes Matrix (TTRAM), a reference model developed in Task 5.1, which defines 30 repository Activities and Functions (AFs) across thematic areas such as digital object management, organisational infrastructure, technology, and security.

The report identifies commonalities and specificities in repository practices, highlights gaps in current capabilities, and provides insights into areas where FIDELIS can offer targeted support. These findings will inform the development of recommendations and tools to facilitate the federation of repositories and enhance their trustworthiness, discoverability, and alignment with EOSC.

Funded by the European Union

1. Introduction

This report presents a key outcome of Task 5.2 (T5.2): Preparing for Federation, within the FIDELIS project. The overarching goal of T5.2 is to chart the landscape of European digital repositories and gather insights on how FIDELIS can best support their efforts to improve, align, and federate with the European Open Science Cloud (EOSC) as Trustworthy Digital Repositories (TDRs).

The work in T5.2 is structured into two main components:

1. Community Survey: Conducted jointly with Task 8.1, this survey collected input from repository stakeholders regarding their current practices, challenges, and needs. Topics included technical and organisational standards, existing federation models, and best practices for managing digital objects. The survey also explored expectations for how the FIDELIS network could provide support. The results are published on Zenodo (Jouneau et al., 2025).

2. Desk Research and Mapping: This present report documents a desk-based mapping of current landscape analyses, best practices, standards, catalogues, registries, and other solutions used and needed by repositories. These are collectively referred to as “repository capabilities and characteristics”.

The mapping exercise is guided by the FIDELIS Transparent Trustworthy Repository Attributes Matrix (TTRAM), developed in Task 5.1. TTRAM serves as a reference model for alignment and cooperation across repositories in the FIDELIS network and EOSC. It enables a detailed examination of the artefacts that support repository practices and provide evidence for assessment, such as standards, semantic artefacts, policies, and procedures. The first version of TTRAM is available on Zenodo (L’Hours et al., 2025a, 2025b, 2025c).

TTRAM defines 30 repository activities and functions (AFs), grouped into four main areas:

- Context
  - AF01 Identification & Contact
  - AF02 Mission & Scope
- Digital Object Management
  - AF03 Conceive, Create, Collect
  - AF04 Deposit & Appraisal
  - AF05 Curation, Quality & Compliance
  - AF06 Discovery & Identification
  - AF07 Access
  - AF08 Reuse
  - AF09 Workflows
  - AF10 Preservation
  - AF11 Provenance and Authenticity
  - AF12 Support
- Organisational Infrastructure

Funded by the European Union

    - AF13 Governance
    - AF14 Policy & Standards
    - AF15 Rights
    - AF16 Resources
    - AF17 People & Expertise
    - AF18 Third Party Dependencies
    - AF19 Continuity of Service
    - AF20 External Engagement
    - AF21 Release & Publishing
    - AF22 Interoperability
    - AF23 Legal & Ethical
    - AF24 Criteria, Assessment, Improvement
    - AF25 Analysis & Impact
    - AF26 Training
    - AF27 Research & Development (R&D)
- Technology
    - AF28 Storage & Integrity
    - AF29 Technical Infrastructure
- Security
    - AF30 Security

This report is organised as follows:

- Section 2 outlines the methodology used for mapping repository capabilities and characteristics.
- Section 3 presents the results, structured according to the 30 TTRAM AFs and grouped by domain-agnostic and domain-specific resources. Domain-specific resources are further categorised by five research communities: Agri-food, Biomedical Sciences, Climate Science, Linguistics, and Social Sciences.
- Section 4 discusses key aspects of the reviewed and mapped repository capabilities and characteristics presented in Section 3.
- Section 5 summarises the main takeaways from the mapping.

Funded by the European Union

2. Methodology

This chapter outlines the methodology employed in Task 5.2 of the FIDELIS project to map repository capabilities and characteristics of digital repositories. The methodology was executed through a series of coordinated actions, partly in an iterative manner:

1. Resource Identification: Members of Task 5.2 identified potentially relevant resources (e.g., best practice descriptions, policies, guidelines, landscape analyses, etc.) within the six scientific/research communities acting as Network beacons in FIDELIS: Agri-food, Biomedical Sciences, Climate Science, Linguistics, Physical Sciences, and Social Sciences. Each community is represented by project partner organisations as indicated in brackets below:

- Agri-food (INRAE)
- Biomedical Sciences / ELIXIR (SIB)
- Climate Science (DKRZ)
- Linguistics / CLARIN (CU, UiT)
- Physical Sciences (ESRF)
- Social Sciences / CESSDA (GESIS, TAU-FSD)

Due to capacity limitations within Task 5.2, the Physical Sciences community was not included in this mapping exercise.

2. Resource Collection: Identified resources were added to the joint EOSC EDEN and FIDELIS Zotero group library under the sub-folder T5.2_Landscape_Analysis.

3. Mapping Framework: A Google spreadsheet was designed to capture repository capabilities and characteristics. Each row represents a reviewed resource, and columns were structured to reflect metadata, repository types, subject classifications, and alignment with the 30 TTRAM Activities/Functions (AFs), as outlined in the following row and column descriptions:

- **Rows 1–3** contain heading information that defines the structure of the spreadsheet. Starting with **row 4**, each row represents a reviewed resource, capturing relevant information aligned with the columns described below.
- **Row 4** is dedicated to the FIDELIS TTRAM, which serves as the foundational framework for the mapping exercise. It includes detailed information across all 30 repository Activities/Functions (AFs) defined in TTRAM.
- **Column A (“ID”)** assigns a unique identifier to each reviewed resource.
- **Column B (“Report”)** indicates whether the resource is included in the present report (“Yes”) or not (“No”).
- **Columns C–F** form the **Reference Information** section, capturing:
    - **C**: Author(s)
    - **D**: Year of publication
    - **E**: Title of the resource
    - **F**: Zotero URL linking to the resource entry in the Zotero group library

Funded by the European Union

- **Columns G–L** represent the **Resource Type**, identifying whether the resource is a:
  - Standard
  - Recommendation
  - Mode of federation
  - Best practice
  - Solution
  - Other (free text)
- **Columns M–Q** indicate the **Repository Type**, distinguishing between:
  - National
  - Institutional
  - Disciplinary
  - Generalist
  - Other (free text
- **Column R** identifies the relevant **Network beacon community** (e.g., Agri-food, Biomedical Sciences, Climate Science, Linguistics, Physical Sciences, Social Sciences, Generic, Other discipline). Note on coverage: While the mapping aimed to include resources from all six Network beacon communities identified in the FIDELIS project, resources from the Physical Sciences community were not included in this iteration due to capacity limitations within Task 5.2.
- **Columns S–W** reflect the **DFG Classification of Subject Areas** used by re3data, including:
  - Humanities and Social Sciences
  - Life Sciences
  - Natural Sciences
  - Engineering Sciences
  - Other (free text)
- **Columns X–BA** correspond to the 30 **FIDELIS TTRAM Activities/Functions (AF01–AF30)**. Each cell may contain either:
  - An “x” or similar marker indicating coverage
  - Relevant excerpt(s) from the resource
- **Column BB** captures any repository capabilities or characteristics not addressed by TTRAM.
- **Column BC** includes additional comments or observations related to the reviewed resource.

4. Data Entry and Annotation: Task participants populated the spreadsheet with relevant information, including excerpts and indicators of alignment with TTRAM AFs. The level of detail varied across resources, with some entries containing full excerpts and others marked with indicators.

Note: No mapping between ISO16363 and the TTRAM Activities/Functions has been undertaken as the levels of granularity of both frameworks is on opposite ends of the spectrum. The level of detail of ISO16363 makes exact matching difficult and does not seem immediately beneficial at the current stage.

Funded by the European Union

An Excel version of the Google spreadsheet is available on Zenodo (Conzett et al., 2025).

5. Report Drafting: A Google document was created to draft the report, with initial guidance provided for Section 3. Participants transferred information to the draft document either from the spreadsheet or directly from the source documents.

Note on the use of Artificial Intelligence (AI):

For the following resources:

- Ala-Lahti et al., 2022
- Alaterä et al., 2022
- Behnke et al., 2020
- Durinx et al., 2017
- Field et al., 2011
- Karsch-Mizrachi et al., 2025
- Lin et al., 2024
- Rehm et al., 2021
- Sansone & Rocca-Serra, 2016
- Wilkinson et al., 2016
- Yilmaz et al., 2011

Each article was downloaded and thoroughly reviewed, with particular attention given to sections describing activities and functions. The annotated document was then uploaded to Microsoft Copilot, accompanied by the following prompt:

"The TTRAM activity and function XX is defined as 'XXX'. Is it covered in the article?"

If the AI's interpretation aligned with the author's annotations, a statement was added to the report. In cases of discrepancy, the author revisited the original text to determine the most accurate interpretation and updated the report accordingly. In this process, AI served as a safeguard against misinterpretation or overinterpretation, ensuring consistency and reliability in the analysis.

For the following resources, where time constraints were met :

- Caracciolo et al., 2020
- Harper et al., 2018
- Lokers et al., 2018
- Sen et al., 2020
- Drakos et al., 2015
- Top et al., 2022
- Devare et al., 2022
- Sestak et al., 2023
- Marrano et al., 2025
- Ali B & P. Dahlhaus, 2022

Funded by the European Union

the adapted process was reversed. Each article was downloaded, then uploaded to ChatGPT 5 with the following prompt:

“For each topic, give a passage of no more than 50 words from each of the articles that deal with the topic, with a short summary of the authors' position on that topic”. Every TTRAM activity and function was then given, one after the other. A lot of calibrations and corrections of the prompts were necessary before a sensible result could be obtained.

The outputs were thoroughly reviewed and assessed against the articles, and subsequent corrections, sorting and pruning were performed on them.

6. Report Enhancement and Completion: The draft report was converted to a Word document and transferred to the institutional SharePoint environment of the main author. Resource descriptions in Section 3 were aligned, Sections 1, 2, 4, and 5 were added, and missing citations and references were added.

Note on the use of AI:

After a complete draft of the report was created as described above, the draft document was uploaded to Microsoft Copilot, accompanied by the following prompt:

“Please review Section 1 of the uploaded EU project milestone report draft and suggest improvements to its structure, clarity, and logical flow. Make sure the content and context are clearly communicated to a non-expert reader.”

The answer was thoroughly reviewed, adjusted as needed, and then implemented in Section 1 of the draft report.

The process was repeated for Sections 2, 3, 4, and 5.

7. Finalisation: Following the completion of the full draft, the involved Task 5.2 participants, along with other contributors from the FIDELIS project, reviewed the document and provided comments and suggestions for improvement. These inputs were carefully considered and incorporated where appropriate.

The main author of the report then conducted a final review of all feedback, resolved outstanding issues, and completed the finalisation of the report.

Funded by the European Union

3. Results

This chapter presents the findings from the desk research-based mapping of repository capabilities and characteristics conducted as part of Task 5.2 in the FIDELIS project. The analysis is structured according to the FIDELIS Transparent Trustworthy Repository Attributes Matrix (TTRAM), which defines 30 repository Activities and Functions (AFs). These AFs are grouped into five thematic areas:

- **Context** (Section 3.1)
- **Digital Object Management** (Section 3.2)
- **Organisational Infrastructure** (Section 3.3)
- **Technology** (Section 3.4)
- **Security** (Section 3.5)

Within each thematic area, the AFs are addressed in ascending numerical order, beginning with AF01. For each AF, the reviewed resources are categorised into two groups:

- **Domain-agnostic resources**, which apply broadly across disciplines
- **Domain-specific resources**, which are further organised by five research communities: Agri-food, Biomedical Sciences, Climate Science, Linguistics, and Social Sciences.

3.1.Context

Context is about the repository's mission, scope, and designated community, establishing the foundational understanding of its responsibilities, boundaries, and the types of digital objects, depositors, and users it serves.

3.1.1. IDENTIFICATION & CONTACT (AF01)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Identification & Contact as follows: "Names, abbreviations and contact details for the organisation to provide reference information and context for the rest of the activities and functions" (L'Hours et al., 2025c, p. 10) and suggests the following items for transparent information:

- Organisation Name
- Organisation Alternative Names
- Organisation Identifier(s)
- Catalogue identifiers and locations
- Contact details

The A|F Identification & Contact is covered by the following elements of the **CoreTrustSeal** requirement R0. Background Information & Context (CoreTrustSeal Standards and Certification Board, 2022, pp. 7–8):

- Re3data Identifier
- Repository type

Funded by the European Union

    - Generalist repository
    - Specialist repository
- Overview. Provide a short overview of key characteristics of the repository, reflecting the repository type selected. This should include information about the scope and size of data collections, data types and formats. Further contextual information may also be added.
- Designated Community. A clear definition of the Designated Community demonstrates that the applicant understands the scope, knowledge base, and methodologies—including preferred software/formats—of the group(s) of users at whom the curation and preservation measures are primarily targeted.

The **FAIRsFAIR project D2.3** deliverable report states that the repository itself should have a PID and be listed in registries of repositories with machine-readable and interpretable metadata about the repository (Behnke et al., 2020).

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF01 Identification & Contact is addressed in the following Agri-Food resources:

Harper et al., 2018: Repositories should list the people responsible and provide direct contact methods to avoid anonymous barriers to communication; showing staff improves accountability and user support. "Make it obvious who is doing the work. Many online resources and databases do not mention the people on their teams and only provide an anonymous contact form. Individuals working on a resources or database should be named on the website." And also : “Every GGB database should have a feedback form”.

Drakos et al. (agINFRA), 2015: Repositories should be registered in agricultural aggregators/catalogues and institutions should expose researcher/institution profiles so repositories are discoverable and their maintaining organisations clear. "2) Repository registration in an agricultural aggregator (like the CIARD RING: http://ring.ciard.net) 3) Institutions’ researchers’ profiles and/or Institutions’ profiles need to be publicly available and accessibly by the agri-food community (using tools like registration in the AgriVIVO: http://www.agrivivo.net or the future AgriProfiles)"

Biomedical Sciences

The TTRAM Activity/Function AF01 Identification & Contact is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: Each ELIXIR Core Data Resource (CDR) is associated with a host organization or consortium. The URL used to access the resource must be provided as well as the type of database (Deposition database or knowledge base). A short description with the scientific purpose, coverage and comprehensiveness of the resource is also requested.

Funded by the European Union

Lin et al., 2024: The article emphasizes the importance of clear repository metadata such as the repository name, the existence of persistent identifiers, the organizational affiliation as well as contact information for support or inquiries.

Climate Science

The Activity/Function (A|F) Identification & Contact is not addressed by any of the reviewed resources from Climate Science.

Linguistics

For **CLARIN** B centres, the following requirements apply (Wittenburg, Van Uytvanck, Zastrow, Straňák, et al., 2023):

- 2.c Visibility of connection to CLARIN. Requirement: Each centre needs to refer to CLARIN in a visible way on its website.
- 2.f Registration in the Centre Registry. Requirement: Each centre should be registered in the Centre Registry. [...] Centre Registry information is automatically transmitted to re3data.org – a registry of research repositories.

Social Sciences

The CESSDA Brief (Kleemola et al., 2025) recommends that data repositories should register in relevant registries and include repository identifiers on websites. Examples mentioned include re3data, OpenDOAR, Fairsharing, and Research Organization Registry (ROR).

3.1.2. MISSION & SCOPE (AF02)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function Mission & Scope follows: “Defining the purpose (or mission) of the organisation and the boundaries of activities for which the repository is responsible. Where activities involve partners or outsourced services, the organisation retains ultimate responsibility for the activities, functions and outcomes. Provides reference information and context for the rest of the activities and functions.” (L’Hours et al., 2025c, p. 11) and suggests the following items for transparent information:

- Organisation Mission Statement
- The Levels of Retention, Curation and Preservation (LoRCAP) offered by the repository
- Repository Type
- Designated Community
- Permitted/Restricted Object, Depositor and User Characteristics
- Geographic and linguistic coverage
- Whether a repository is FAIR-enabling or endorses the TRUST or CARE Principles

The A|F Mission & Scope is covered by **CoreTrustSeal** requirement Mission & Scope (R01) requiring certified repositories to have an explicit mission to provide access to and preserve digital objects (CoreTrustSeal Standards and Certification Board, 2022, p. 11).

Funded by the European Union

The **nestor Seal** addresses aspects of this A|F in “C2 Responsibility for preservation” and “C3 Designated communities”. C2 requires that the archive “responsibility for the long-term preservation of the information objects on the basis of legal requirements or its own objectives”. C3 specifies the need to define the archive’s Designated Community, including (nestor Certification Working Group, 2025).

In the **OAIS Reference Model**, an Open Archival Information System (OAIS) is defined as an organization which has “accepted the responsibility to preserve information and make it available for a Designated Community [...]. The system meets a set of mandatory responsibilities that allows an OAIS Archive to be distinguished from other uses of the term ‘archive’” (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 1, p. 13). The mandatory responsibilities (Section 3.2) have additional implications for what the mission of an OAIS must entail. This includes

- defining the community for whose understanding and use the information objects will be preserved,
- ensuring that the preserved objects are “Independently Understandable” to this community,
- the implementation of “documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the Archive”, and
- making the preserved information objects available to the defined community (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 3, p. 1).

The **EOSC Federation Handbook** does not address the A|F Mission & Scope explicitly, but provides the following high-level guidelines for FAIR data repositories (EOSC Association, 2025, pp. 32–33):

> *EOSC Node data repositories should align with the FAIR principles and aim to implement them as rigorously as possible. [...] Data repositories should implement the EOSC Guidelines for Research Data Sources. The repositories must be onboarded as a service in one of the EOSC Nodes [...]. [...] In the cases where the EOSC Node repositories are not community standards, they should link to community standard repositories if they exist, to increase their impact and reduce proliferation of data repositories.*

The **Desirable Characteristics of Data Repositories for Federally Funded Research** outlines its mission as to improve consistency across U.S. Federal departments and agencies in guiding researchers on selecting repositories for Federally funded research data, aiming to ensure data is Findable, Accessible, Interoperable, and Reusable (FAIR) while integrating privacy and security. Its scope defines repositories' responsibilities in managing and sharing data from Federally funded research, including specific considerations for human data, to enhance public access and the overall return on R&D investments (White House Office of Science and Technology Policy, 2022).

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. Trusted repositories have specific provisions in place and offer

Funded by the European Union

explicit information online about their policies, which define their services (e.g. acquisition, access, security of content, long-term sustainability of service including funding, etc) (European Commission, 2024).

The **Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements** (Jahn et al., 2023) states that there is no single standardised system for organisations or individuals to codify and express commitment to and endorsement of specific repositories. For example, endorsement can be seen to happen through use, so that repositories popular among researchers can be considered endorsed by their research community.

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) states on repository coverage that the higher-level subject areas/disciplines the repository covers, as well as cross-disciplinary domains, such as the types of data, technology and study.

The **Research Data College Working Group** has put forward exclusion criteria for selecting a trustworthy **subject-specific** repository for self-depositing data. This methodology was devised and applied to the French national list of recommended trust repositories: https://recherche.data.gouv.fr/en/repositories). When it comes to the mission and scope, the Research Data College Working Group rule out "subject-specific repositories that limit data deposits to certain scientific communities where only scientists affiliated to the institution hosting the repository are authorised to deposit" (Lamotte et al., 2024, p. 11). In addition the Research Data College Working Group recommends that subject-specific repositories provide transparent information about the following items (Lamotte et al., 2024, pp. 12–13):

- Disciplinary field. Where possible, the disciplinary field employs the nomenclature used by HAL, which proposes a corollary tailored to the human and social sciences, which is not always the case with other existing nomenclatures. Furthermore, it offers up to three different levels of granularity, which makes it possible to provide better descriptions of the selected repositories.
- Data accepted. Here it is a question of describing the type of data accepted by the repository, ensuring that the terminology specific to each discipline is employed in order to make the depositor's choice easier (e.g. NMR spectra, 3D structures of biological molecules, TEI-encoded corpus, etc.).
- Volume limit. This information can prove important for researchers from disciplines that generate substantial volumes of data. It also helps to anticipate the expected cost if the repository requires a financial contribution above a certain volume of data.

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) implicitly recognizes organizational scope and mission by stressing that the relevance and priority of its indicators must be adjusted according to the specific requirements and practices of the disciplinary community. This community focus translates into explicit essential indicators requiring that both metadata and data comply with domain-relevant

Funded by the European Union

community standards (R1.3), thereby setting the necessary technical boundaries for FAIRness within that defined scope.

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), all 14 respondents, among which 11 repositories and 3 organisations, had a mission statement. A closer examination of the statements revealed differences in the content and length, although there were similarities between them, particularly amongst the certified repositories.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) report presents a first iteration of guidelines that will help to expose relevant information at the organisational and object level to facilitate discovery, provide context, and support interoperability between repositories, registries, and other related stakeholders.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) discusses the mission of repository and how some are not in scope of CTS certification e.g. due to not having long-term preservation mission. In all cases, repositories need to be clear on their responsibilities and priorities. For repositories with a long-term preservation mission, CoreTrustSeal is the core certification. If, for example, data findability and accessibility are important, FAIR evaluation is useful. If security aspects are important, there are ISO standards (like ISO 2700141) for security, and if IT service management is essential, FitSM42 defines a baseline of IT service management effectiveness.

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) requires repositories to clearly define their operational boundaries by providing public documentation outlining the scope of accepted resources. Essential to the mission, the repository must also clearly indicate the organization responsible for management and governance, and maintain a publicly available policy detailing the fate of resources if operations cease.

Domain-specific resources

Agri-food

No relevant mapping to the TTRAM Activity/Function AF02 Mission & Scope was found in the gathered Agri-Food resources.

Biomedical Sciences

The TTRAM Activity/Function AF02 Mission & Scope is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: Having a clear mission and scope is a defining feature of an ELIXIR Core Data Resource (CDR), as emphasized in the article. This is reflected in several key evaluation criteria:

Funded by the European Union

- First, the scientific focus and quality of a resource is assessed by determining whether it functions primarily as a deposition database, accepting data submissions, or as a knowledge base, which adds value through integration, annotation, or interpretation.
- Second, the resource must provide a scope statement that outlines its scientific coverage and comprehensiveness. This includes specifying whether the resource addresses all species or a subset, particular families, or outputs from specific experimental methods. It also involves positioning the resource in relation to other similar data resources, highlighting its unique contributions or complementary role.
- Finally, the community and potential usage are considered by estimating the size of the global user community that could benefit from the resource. This helps gauge the resource's relevance and impact within the life sciences domain.
- Together, these criteria ensure that each ELIXIR CDR has a well-defined mission and scope, providing essential context for its activities and supporting its long-term sustainability and strategic value.

Lin et al., 2024: In the article, the mission of data repositories is to provide core services such as ingestion (intake) of data, data management, preservation, archival storage, administration, access. It emphasizes the importance of clearly articulating the repository's mission, including the type of data it accepts, the target user community, and the goals of the repository (e.g., long-term preservation, open access, domain-specific reuse). Besides, the article supports transparency through the TRUST principles and the provision of information such as:

- **Types of data repositories**
    - **Domain-specific repositories:** These repositories store data of a specific type (e.g., protein structure, nucleotide sequence, clinical data) or discipline (e.g., cancer, neurology). They often form a nexus of resources for their research communities interested in these specialized data.
    - **Generalist repositories:** These repositories store data of multiple types and disciplines, accepting data regardless of its type, format, content, disciplinary focus, or research institution affiliation. NIH has established agreements with several generalist repositories under the NIH Generalist Repository Ecosystem Initiative (GREI)[16].
    - **Project-specific data repositories:** These repositories store domain-specific data generated from a project or collaboration (e.g., NIH All of Us[17]) and enable data sharing and reuse by making the project-specific data available for reuse by other projects or researchers. This is not to be confused with a project data coordinating center (DCC), which facilitates the project collaboration, curation, and data analysis but does not serve as a repository as data are not widely available for reuse by other researchers. Note that a DCC may also later facilitate submission of the project data to a data repository.

Funded by the European Union

    - **Institutional repositories:** These repositories store data primarily created by members of an institution or a group of institutions, such as principal investigators (PIs), postdocs, and students. This category addresses the needs of the institution's staff and may serve to collect data from one-to-many projects, and, depending on the institutional mission, may function as a domain-specific or generalist repository.

- **Data type**: Categories of digital assets that are managed by a repository.

Barrett et al., 2012:

- BioSample Database: Stores submitter-supplied metadata about biological materials (e.g., cell lines, tissue biopsies, organisms, environmental isolates).
- BioProject Database: Provides an organizational framework for accessing information about research projects.

Karsch-Mizrachi et al., 2025: The article clearly states that INSDC captures, preserves, and presents globally comprehensive nucleic acid sequence information as part of the permanent scientific record and as a forum for sharing amongst the broad scientific community.

Climate Science

The Activity/Function (A|F) Research & Development (R&D) is not addressed by any of the reviewed resources from Climate Science.

Linguistics

For **CLARIN** B and E centres, the following requirements apply when it comes to the AF Mission & Scope (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023, p. 3):

- Centres need to offer useful services to the CLARIN community
- Each centre needs to make clear statements about their policy of offering data and services
- Centres need to employ activities to relate their role in CLARIN to the research community in order to guarantee a research based status of the infrastructure and allow researchers to embed their services in their daily research work
- Centres that are offering infrastructure type of services (E) need to specify their services for CLARIN and the terms of giving service

For **CLARIN** B centres, the following requirements apply (Wittenburg, Van Uytvanck, Zastrow, Straňák, et al., 2023):

- 2.a Description of the repository's context, mission and scope. Requirement: The centre's repository context, mission and scope needs to be clearly defined.
- 2.e Details about resources and services provided (updated). Requirement: Each centre needs to make explicit statements about CLARIN compliant resources and services available at the centre.

Social Sciences

Funded by the European Union

The CESSDA Data Management Expert Guide (CESSDA Training Team, 2022, Chapter 7. Discover) outlines the core mission of trusted social science domain repositories: integrating data into the research lifecycle to ensure it is published, shared, discovered, and reused in line with the FAIR principles. Moreover, they:

- archive and preserve data;
- offer and manage access to the data;
- provide complex services focused on data reuse for research, teaching and learning;
- check data quality and compliance;
- improve data interoperability, e.g. by accompanying data with rich standardised metadata;
- maintain data catalogues;
- seek to add new data to their collections;
- develop training for data producers and data users.

3.2. Digital Object Management

Digital Object Management covers the lifecycle of digital objects from creation and deposit through appraisal, curation, access, reuse, and preservation, emphasizing transparency in practices and alignment with levels of retention, curation and preservation (LoRCAP).

3.2.1. CONCEIVE, CREATE, COLLECT (AF03)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Conceive, Create, Collect as follows: “The stages of the research data lifecycle, before data enters a repository. Working with data or metadata creators and owners can help improve quality and highlight the benefits of curation, preservation and reuse.” and suggests for transparent information “any information or guidance provided to researchers about how to manage their digital objects’ data and metadata during the pre-repository phase” (L’Hours et al., 2025c, p. 12) including:

- Funding Information
- Project Information
- Data Management Plan (DMP)

The phase immediately preceding Ingest into the repository is covered in the **PAIMAS** (Consultative Committee for Space Data Systems (CCSDS), 2004) and **PAIS** (Consultative Committee of Space Data Systems (CCSDS), 2014) standards. PAIMAS defines roles and the different sub-phases of the pre-Ingest phase, beginning with the first contact between data producer and archive. PAIS provides guidance for preparing the information objects to be submitted to and ingested by the archive during the “formal definition phase” as defined in PAIMAS.

The **Desirable Characteristics of Data Repositories for Federally Funded Research** requiring Federally funded researchers to develop Data Management Plans (DMPs) that specify how data will be managed and shared before it enters a repository. Federal agencies provide guidance to help

Funded by the European Union

researchers select appropriate repositories (White House Office of Science and Technology Policy, 2022).

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) addresses the pre-repository phase by being designed for use during the development of Research Data Management Plans, allowing data producers and project managers to specify the expected level of FAIRness their resources should achieve before data and metadata are produced.

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), most respondents (10 out of 14) offered support for the initial conceptualisation of research projects, access to data, and the collection and/or creation of research (meta)data.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) highlight that data citations provide credit for the data producer and are among the biggest motivators for researchers to publish their data.

Within AF03 Conceive, Create, Collect, the **Data Curation Network** defines the following Curation Activity (Johnston et al., 2016): "Conversion (Analog): In effort to increase the usability of a data set, the information is transferred into digital file formats (e.g., analog data keyed into a database). Note: digital conversion is also used to convert "fixed" data (e.g., PDF formats) into machine-readable formats."

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF03 Conceive, Create, Collect was only found to be addressed in the following Agri-Food resource:

Harper et al., 2018: AgBioData recommends that repositories advise data creators on management plans, metadata completeness, and FAIR readiness before data submission. "Proper data management is a critical aspect of research and publication… Data are the lifeblood of research, and their value often do not end with the original study, as they can be reused for further investigation if properly handled.". And also: "Provide training for researchers on responsible data management. Tutorials covering all aspects of data management, including file formats, the collection and publishing of high value metadata along with data, interacting with GGB databases, how to attach a license to your data, how to ensure your data stays with your publication and more will be useful training tools."

Biomedical Sciences

The TTRAM Activity/Function AF03 Conceive, Create, Collect is addressed in the following Biomedical Sciences resources:

Field et al., 2011: GSC emphasizes standards for describing and capturing rich contextual information, exemplified by the MIxS checklists, ensures the maximised usefulness, quality, and

Funded by the European Union

quantity of data for public collections of genomes, metagenomes, and marker gene sequences before they enter a repository.

Yilmaz et al., 2011: The MIxS standard forms mandatory guidelines for collecting and reporting comprehensive contextual data at the time of sample acquisition and sequencing, effectively serving as an "electronic laboratory notebook".

Barrett et al., 2012: The BioProject and BioSample databases and Submission Portal represent a proactive approach by NCBI to organize and integrate data across interdisciplinary resources and to obtain a rich set of contextual metadata from data producers.

Karsch-Mizrachi et al., 2025: The article describes how INSDC engages with data creators and submitters before data enters the repository by setting expectations for metadata quality, submission standards, and validation processes. Besides, INSDC has set up working groups to increase consistency and alignment with other standards organizations such as the Genomic Standards Consortium, the Public Health Alliance for Genomic Epidemiology, and the Global Alliance for Genomics and Health, across eleven categories of data and metadata.

Rehm et al., 2021: The article emphasizes collaboration with data creators through its Driver Projects and Work Streams, which help shape standards and policies from real-world needs. GA4GH develops data models for clinical and phenotypic data (e.g., Phenopackets, Pedigree) and standardized file formats (e.g., SAM, BAM, CRAM, VCF/BCF, VRS, VA), for consistent quality, reusability, and effective data sharing before data formally enters a repository.

Climate Science

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that data providers are responsible for an initial contact phase where they describe their project, data characteristics, and publication needs to the DKRZ Data Management (DM). Following this, they must undertake thorough data preparation, which includes ensuring files are in accepted open-source formats (like NetCDF with CF conventions), meet specific file header standards, adhere to file size and compression guidelines, and are consistently labeled and organized into appropriate Dataset directories.

Linguistics

The Activity/Function (A|F) Conceive, Create, Collect is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

CESSDA's Data Management Expert Guide supports researchers in the social sciences in caring for their data throughout the research data lifecycle in accordance with community standards (CESSDA Training Team, 2022).

Funded by the European Union

The combination of informed consent, anonymization, and access control enables the sharing of personal data, so attention should be paid to these issues when collecting and processing data. (CESSDA Training Team, 2022, Chapter 5. Protect). Thorough and systematic documentation of research data is essential for ensuring it can be published, discovered, cited, and reused. Clear and comprehensive metadata enhances the overall quality and usability of the data. (CESSDA Training Team, 2022, Chapter 2. Organise and Document).

Pre-ingest ensures data can be properly preserved, easing the workload for curators. While high value data should ideally be shared via trusted archives in a structured and well-documented format, researchers often face barriers like limited time, resources, or incentives. This leads to incomplete or unsuitable data submissions. Pre-ingest helps identify and resolve issues with data and metadata quality, completeness, and format early on. It also fosters collaboration with researchers, promoting a stronger culture of data sharing and good data management practices. (CESSDA Training Team, 2025a, Chapter 3.1 Pre-Ingest Basics).

3.2.2. DEPOSIT & APPRAISAL (AF04)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Deposit & Appraisal as follows: “Accepting custody of digital objects from depositors, transferring responsibility to the repository. It may also include appraising offered or requested deposits to ensure they meet established criteria for acceptance.”, and suggests for transparent information “[d]ocumentation of compliance criteria applied at the point of deposits, whether automated or manually applied and the degree to which they are required or optional” (L’Hours et al., 2025c, p. 13), including

- Collections Development and Appraisal Policy
- Deposit Compliance Criteria
- Deposit Procedures
- Acceptable/Preferred File Formats List
- ReAppraisal Plan, Data Management Plan
- Deposit licence (see also AF15 “Rights”).

The suggested CoreTrustSeal Levels of Retention, Curation and Preservation (**LoRCaP**) address the deposit and appraisal aspect with level “D. Deposit Compliance”. On this level, “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” (CoreTrustSeal Standards and Certification Board, 2024) . Metadata required on object- and repository-level to demonstrate how the LoRCaP are implemented are proposed in L’Hours et al., 2024.

The A|F Deposit & Appraisal is covered by **CoreTrustSeal**’s Deposit & Appraisal (R08), which requires certifying repositories to accept data and metadata based on defined criteria to ensure relevance and understandability for users. To document the fulfilment of R08, CoreTrustseal asks applicants to

Funded by the European Union

provide references to a number of items (CoreTrustSeal Standards and Certification Board, 2022, pp. 18–19):

- Any documented deposit process that includes steps to ensure that data and metadata are sufficient for long-term preservation.
- A collection development policy or procedures to guide the selection of digital objects.
- Criteria for prioritisation and any different curation-levels or preservation levels defined during appraisal.
- The approach to digital objects that do not fall within the mission/collection profile.
- Procedures to determine that the metadata required to interpret and use the digital objects are provided.
- Any automated assessment of metadata adherence to relevant schemas.
- The repository approach if metadata provided is insufficient for long-term preservation.
- A list of preferred formats.
- Checks in place to ensure that depositors adhere to the preferred formats.
- The approach towards digital objects that are deposited in non-preferred formats.
- The transfer of custody and responsibility during the handover from the depositor to the repository.

The **nestor Seal** addresses this A|F in criteria "C1 Selection of information objects and their representations" and "C14: Integrity: Ingest interface". C1 requests that "[c]riteria have been defined for the selection of information objects and their representations in the digital archive". C14 states the need for an "interface for ingesting the representations in a way which retains their integrity. The interface contains all of the functions and processes aimed at transferring the submission information packages from the producers, transforming them into archival information packages and incorporating them into the digital archive" (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** Deposit and Appraisal processes are described as part of the Ingest and Administration functional entities (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4.2.3.3, 4.3.3.6). The Ingest functional entity receives deposits in the form of a Submission Information Package (SIP) from data producers, performs technical quality assurance of the transfer process and prepares the information object for archiving (see CURATION, QUALITY & COMPLIANCE (AF05)). These actions are complemented by the Administration functional entity which checks the contents of the SIP against the submission agreement to "verify that the quality of the data meets the requirements of the Archive" and that adequate metadata is present "to ensure that the Content Information is understandable and independently usable to the Designated Community" (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4, p. 14).

Desirable Characteristics of Data Repositories for Federally Funded Research (White House Office of Science and Technology Policy, 2022) provides Clear Use Guidance to ensure repository datasets are accompanied by documentation describing terms of dataset access and use (e.g., reuse licenses and need for approval by a data use committee). The repository ensures datasets must be assigned unique persistent identifiers (PIDs,) accompanied by metadata to enable discovery, reuse, and

Funded by the European Union

citation of datasets, using schemas that are appropriate to, and ideally widely used across, the communities that the repository serves. The policy expects repositories to provide or facilitate expert curation and quality assurance to improve data accuracy and integrity, and implement mechanisms to record the provenance of submitted datasets and metadata.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. Trusted repositories have specific provisions in place and offer explicit information online about their policies, which define their services (e.g. acquisition, access, security of content, long-term sustainability of service including funding, etc) (European Commission, 2024).

The **Repository Features to Help Researchers: An invitation to a dialogue** includes Data Deposition Condition, which is expected to clarify who can deposit data into the repository (Cannon et al., 2021).

The following exclusion criterion from the **Research Data College Working Group** applies to the A|F Deposit & Appraisal (Lamotte et al., 2024, p. 10): “No moderation of data deposits. The following should be excluded: repositories that do not practise data moderation (human or automated) designed to ensure a minimum quality of the metadata, which makes it possible to avoid transferring data that is incomplete or poorly described.”

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) establishes the objective requirements for appraisal by defining Essential indicators that must be met for a resource to be considered FAIR, notably requiring that both metadata and data comply with domain-relevant community standards (R1.3) and that rich metadata is provided for discovery (F2). This framework is explicitly intended to assist in defining the expected level of FAIRness for resources during the development of Research Data Management Plans (DMPs), thereby pre-defining the compliance criteria required for acceptance during the deposit and appraisal phase.

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), 11 out of 14 respondents provide support at Deposit stage, defined as the transfer of data from the active research phase into longer term storage, curation or preservation systems, including any appraisal and selection processes or standards.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) propose that exposing repository metadata related to certification requirements allows Validation Authorities (like CoreTrustSeal) to harvest this information, thereby simplifying the assessment process for trustworthiness, reducing administrative overhead, and supporting both machine and human validation of compliance.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that repositories are encouraged to define clear deposit

Funded by the European Union

policies and appraisal criteria. Some repositories outsource or insource these functions, and certification requires clarity on these processes.

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) states that repositories must publicly outline the scope of resources accepted and ensure the agreement with the depositor grants the necessary rights to copy, transform, and store items to fulfil preservation responsibilities. To facilitate the transfer of custody, the repository should support mediated submission using standardized protocols such as SWORD and record the resource's checksum upon submission.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) states that data citations should be collected during the acceptance phase through self-report by authors as they deposit or update data records. To ensure consistency and compliance during deposit, repositories are recommended to utilize specific DataCite metadata fields such as relatedIdentifier and relationType to establish the necessary citation connections between the deposited dataset and the citing scholarly output.

The **Core Preservation Process CPP-029 Ingest** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA performs all operations necessary to transform an SIP into an AIP.” (EOSC EDEN T1.2 et al., 2025).

Within AF04 Deposit & Appraisal, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

- “Deposit agreement: The certification by the data author (or depositor) that the data conform to all policies and conditions (e.g., do not violate any legal restrictions placed on the data) and are fit for deposit into the repository. A deposit agreement may also include rights transfer to the repository for ongoing stewardship.”
- “Selection: The result of a successful appraisal. The data are determined appropriate for acceptance and ingest into the repository according to local collection policy and practice.”

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF04 Deposit & Appraisal is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData stresses transparent deposit policies and metadata standards to ensure quality and FAIR compliance. “Repositories must develop and communicate submission procedures, metadata requirements, and preferred data formats to ensure consistent data quality and usability.”

Caracciolo et al., 2020: The Agrisemantics group highlights appraisal through metadata validation and semantic consistency checks. “Repositories should evaluate submitted datasets for compliance with community-agreed metadata vocabularies and semantic standards before acceptance.”

Funded by the European Union

Sen et al., 2020: WheatIS applies coordinated deposit procedures ensuring that partner repositories meet metadata and quality thresholds. “The WheatIS portal harvests metadata from distributed data repositories that have adopted the common metadata standards defined by the working group to ensure consistent access and quality of information.”

Marrano et al., 2025: Marrano et al. require repositories to publish deposit and appraisal documentation, including file format and compliance policies. “Repositories must provide open documentation detailing deposit workflows, accepted formats, and appraisal criteria, distinguishing between required and optional elements.”

Biomedical Sciences

The TTRAM Activity/Function AF04 Deposit & Appraisal is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: In ELIXIR CDRs, Deposition databases are evaluated based on the scientific quality of their content, which includes Appraisal of incoming data to ensure it meets standards. While not explicitly described as a “custody transfer,” the article implies that ELIXIR CDRs take responsibility for the long-term stewardship of the data they host, which aligns with the concept of accepting custody.

Yilmaz et al., 2011: The MIxS standard serves as the established criteria for acceptance for digital objects submitted to public databases like GenBank, EBI-ENA, and the Sequence Read Archive. These repositories validate submitted information for MIxS compliance upon submission, which enables them to accept custody of digital objects that meet these standards.

Barrett et al., 2012: The BioProject database organizes metadata for research projects [...] and provides a central portal to access the data once it is deposited into an archival database.

Karsch-Mizrachi et al., 2025: The article describes how the INSDC accepts custody of nucleotide sequence data through structured submission systems managed by its member organizations. These systems include mechanisms for data validation and quality control, ensuring that submitted data meets established criteria before being incorporated into the repository. The Implementation Committee (IC) plays a central role in defining and enforcing these acceptance standards.

Lin et al., 2024: The article does not detail formal appraisal workflows (e.g., peer review of datasets, rejection criteria) nor any existing decision-making process for accepting or rejecting deposits. However, it confirms that repositories accept submissions and take custody of data from depositors and assumes responsibility for its preservation and access, an implicit transfer of responsibility from depositor to repository.

Climate Science

The METAFOR project: preserving data through metadata standards for climate models and simulations implies the use of metadata collection for CMIP5 (Callaghan et al., 2010).

Funded by the European Union

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) describes the CMIP5 framework to address deposit and appraisal by mandating the use of the CF-NetCDF format for output data and relying on the CMIP5 Questionnaire for metadata submission. This metadata entry tool appraises deposits by enforcing Controlled Vocabulary constraints, validating against the Common Information Model (CIM) XSD, and checking for deeper coherency using Schematron validation to ensure compliance with established criteria prior to data publication.

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC accepts custody of digital objects from depositors through a defined submission process, where data management first appraises offered deposits to ensure they align with the WDCC's scope and quality requirements. This appraisal involves quality control, checking for adherence to open-source data formats, file sizes, and consistent data organization, alongside both technical quality assurance by WDCC and scientific quality assurance by the data provider to ensure long-term usability and citability of the data.

Linguistics

CLARIN centres offering deposition services are encouraged to register their file format recommendations in the CLARIN Standards Information System (SIS), which aggregates and visualizes the list of recommendations for data deposition formats (CLARIN Standards and Interoperability Committee, n.d.). The SIS list distinguishes between three levels of recommendation, which should be viewed as relative to the profile of the given centre, along the following rules of thumb:

- “recommended” should be interpreted as meaning that the centre in question will in most cases be able to process the data without much manipulation and that it is likely that the data will be preserved long-term in that format (the specifics are up to that centre);
- “acceptable” should be interpreted as meaning that the centre may need to spend some time and resources on the up-conversion of the data, and that the data may be preserved in one of the recommended formats instead;
- “discouraged” should be understood as indicating that the centre may find it problematic to up-convert the data.

Social Sciences

The CESSDA Data Archiving Guide Chapter 2.1 presents some typical appraisal criteria that make data acceptable for a given archive. These criteria are often presented in the collection development policy or acquisition policy of the archive and may include, for example:

- Purpose: research, teaching, replication
- Level of curation – short or long term
- Alignment with other internal and external policies, e.g., research priorities of funders
- Flexibility to deal with exceptions, such as data with a high risk of loss
- Thematic coverage

(CESSDA Training Team, 2025a, Chapter 2.1 Data Collection and Data Acquisition Policies).

Funded by the European Union

Before ingesting the data, it should be ensured that the data file contains sufficient metadata and/or accompanying documentation to make the contents understandable and to support future reuse. (CESSDA Training Team, 2025a, Chapter 3.4.3 Review of documentation).

3.2.3. CURATION, QUALITY & COMPLIANCE (AF05)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Curation, Quality & Compliance as follows: "Ensuring digital objects reach a defined level of quality and standards compliance before they are made available for reuse" and suggests for transparent information "[d]ocumentation of any initial curation steps taken after deposit to meet defined criteria for access and reuse and potentially to enable preservation, including any curation steps like conversions to file formats and any additions to metadata." (L'Hours et al., 2025c, pp. 13–14), including

- Quality and Standards statements
- Standard operating procedures (SoP, Quality criteria)
- Compliance and Quality assurance policies and procedures

The suggested CoreTrustSeal Levels of Retention, Curation and Preservation (**LoRCaP**) address activities relevant to this A|F with Levels D. Deposit Compliance (see Deposit & Appraisal above) and C. Initial Curation. Initial curation entails curation of digital objects "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" (L'Hours et al., 2024, p. 4). Note that the LoRCaP are expected to replace the current CoreTrustSeal Level of Curation categories. Metadata required on object- and repository-level to demonstrate how the LoRCaP are implemented are proposed in L'Hours et al., 2024.

CoreTrustSeal addresses the A|F Curation, Quality & Compliance in two places. First, in requirement R0. Background Information & Context, CoreTrustSeal asks applicant to select which of the following level(s) of curation they apply to their digital assets (CoreTrustSeal Standards and Certification Board, 2022, p. 9):

- A. Content distributed as deposited;
- B. Basic curation – e.g. brief checking, addition of basic metadata or documentation;
- C. Enhanced curation – e.g. conversion to new formats during ingest, enhancement of documentation and metadata;
- D. Data-level curation – as in C above, but with additional editing of deposited data.

Second, in requirement Quality Assurance (R10), CoreTrustSeal requires repositories to address technical quality and standards compliance, and ensure that sufficient information is available for end users to make quality-related evaluations. CoreTrustSeal asks applicants to provide references to the following items (CoreTrustSeal Standards and Certification Board, 2022, pp. 20–21):

- The approach to data and metadata quality taken by the repository including variations for different curation-levels.

Funded by the European Union

- The standards that data, metadata and documentation must comply with to be acceptable for preservation and access. Whether these are general external standards, internally developed standards or specific to a community of practice.
- The quality control checks in place ensure the completeness and understandability of data and metadata.
- The approach to resolving issues e.g. whether the digital objects are returned to the depositor for rectification, fixed by the repository, noted by quality flags, and/or included in the accompanying metadata.
- The approach to managing changes to expected standards (e.g. new or updated data formats of metadata schemas) in response to changes in the technical environment or to changes in the needs of the Designated Community.
- Any links provided to other digital objects' data and metadata e.g. related digital objects, publications, or the use of controlled vocabularies and ontologies.

In the **nestor Seal** this A|F is addressed in the criteria C21-C25, which require that "the digital archive has issued specifications" for SIPs and AIPs as well as for the transformation of a SIP to an AIP and an AIP to a DIP. In addition, "C1 Selection of information objects and their representations" as well as "C20 Technical authority" are relevant here. (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** initial curation actions are performed in the Ingest functional entity. In particular, the "Generate AIP function" creates an Archival Information Package (AIP) from the SIP in accordance with archive standards. This may comprise the following actions:

- file format conversions,
- determining Transformational Information Properties[1] (sometimes also referred to as significant properties),
- gathering adequate Representation Information, including from the Producer, and creating the Descriptive Information that completes the AIP,
- reorganizing the content information stored in the SIPs. (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 4/7).

The **EOSC Federation Handbook** partially addresses the A|F Curation, Quality & Compliance, when requiring that "[d]ata repositories in the EOSC Federation should measure the FAIRness of their data through FAIR metrics and they should be able to demonstrate compliance to the FAIR principles by implementing at least the Findable and Accessible guidelines" (EOSC Association, 2025, p. 32).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** highlights that repositories should provide or facilitate expert curation and quality assurance to enhance the

[1] "An Information Property the preservation of the value of which is regarded as being necessary but not sufficient to verify that any NonReversible Transformation has adequately preserved information content. This could be important as contributing to evidence about Authenticity. Such an Information Property is dependent upon specific Representation Information, including Semantic Representation Information, to denote how it is encoded and what it means." (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 1/17).

Funded by the European Union

accuracy and integrity of datasets and metadata, ultimately making Federally funded research data FAIR to the fullest extent possible (White House Office of Science and Technology Policy, 2022).

The **Recommendations Consultation. EOSC-A Long Term Data Preservation Task Force** (Andreu et al., 2023) formulates the Curation & Preservation Levels from Z. (Level Zero), D. (Deposit Compliance) C. (Initial Curation) B. (Logical-Technical Preservation) and A. (Conceptual preservation) for understanding and reuse. Data services, including repositories should specify all the levels of care they apply to objects within their collection, including through repository and digital object registry metadata. Different types of data services benefit from being transparent on their current level of storage, curation and preservation practice, as this increases trust by the user and funders alike.

The **EU Annotated Grant Agreement** states repositories have mechanisms or provisions for expert curation and quality assurance for the accuracy and integrity of datasets and metadata, as well as procedures to liaise with depositors where issues are detected They ensure that contents are accompanied by metadata sufficiently detailed and of sufficiently high quality to enable discovery, reuse and citation and contain information about provenance and licensing. Their metadata is machine-actionable and standardized (e.g. Dublin Core, Data Cite, etc) preferably using common non-proprietary formats and following the standards of the respective community the repository serves, where applicable (European Commission, 2024).

The **Update of the Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements** states that the lack of a public policy for preservation, curation and security of the contents is the most frequent reason, followed by not adhering to a specific metadata standard (Lazzeri, 2024).

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) states on data curation that review and annotation of the data performed by the repository (e.g. via a data submission tool that enforces some curation, or by its curation team). Does the repository curate its holdings?

Findings from **D1.3 Recommendations for a FAIR EOSC - White Paper of the FAIR-IMPACT Synchronisation Force** (Grootveld, 2025) emphasize that building an inclusive and collaborative network of Trustworthy Digital Repositories (TDRs) is essential to share expertise and help repositories improve their trustworthiness, which includes achieving or preparing for certified status, often involving curation processes. Develop semantic interoperability through FAIR Semantic Artefacts, via adoption of policies, guidelines and mappings. Furthermore, achieving high data quality (fitness for purpose) is important for daily research and its value, even though quality is distinct from the assessment of FAIRness.

The **O'FAIRe makes you an offer: Metadata-based Automatic FAIRness Assessment for Ontologies and Semantic Resources** (Amdouni et al., 2022) details a methodology and tool, O'FAIRe, for automatically assessing the level of FAIRness—a key standard of quality and compliance—for ontologies and semantic resources based largely on their metadata descriptions as managed by repositories.

Funded by the European Union

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) addresses curation, quality and compliance by setting Essential indicators for Reusability (R1), notably requiring that metadata and data comply with domain-relevant community standards (R1.3-01M, R1.3-01D) and that metadata is expressed in compliance with a machine-understandable community standard (R1.3-02M). Furthermore, achieving a high level of FAIRness relies on providing a plurality of accurate and relevant attributes (R1-01M) and including clear license information (R1.1-01M), both classified as Essential indicators necessary for enabling reuse.

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), most of the 14 respondents reported performing a combination of curation levels, with only five respondents selecting one level. The most common choice was enhanced curation by converting data to new formats and enhancing documentation, followed closely by basic curation by briefly checking the data and adding basic metadata. Six respondents reported providing data level curation by further editing deposited data for accuracy. A further two organisations indicated that they perform no curation for some data in their collection, alongside higher levels of curation for other parts of their collection.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.1, Contextual Data Quality, that it is strongly recommended to provide standardised, accessible, and comprehensive methods for applying documentation to datasets, intended to exist alongside standardized metadata schema.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) recommend that repositories express the levels of care offered and received by digital objects, which should detail the levels of curation and preservation in place and how these might change over time. Furthermore, standardizing the exposure of assessment results using the Data Quality Vocabulary (DQV) is recommended, which can embed FAIR assessment outcomes within the metadata of assessed datasets via DCAT, including crucial information such as the test date and the testing tool used.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) introduce several indicators that address aspects related to quality and standards, particularly under the Reusable (R) principle. Specifically, it is considered Essential (R1.3-01M, R1.3-01D) that both metadata and data comply with domain-relevant community standards, which directly supports the goal of ensuring defined levels of quality and compliance prior to making objects available for reuse.

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) emphasizes that the repository must undertake a lightweight review of basic metadata upon resource submission, and enhance it if necessary. Repositories must also provide documentation or a policy that outlines the curation processes applied to both the resources and their metadata to ensure defined criteria for access and reuse are met.

Funded by the European Union

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) recommend that repositories store citation links using specific DataCite metadata fields (such as relatedIdentifier and relationType) to establish a standardized relationship between the dataset and the citing scholarly output. Furthermore, to increase transparency and trust in the quality of the citation information, repositories must expose the provenance (source) for every asserted data citation listed on the dataset landing page.

The **Core Preservation Process CPP-019 Data Quality Assessment** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA evaluates and re-evaluates the data quality of Information Objects." (EOSC EDEN T1.2 et al., 2025).

Within AF05 Curation, Quality & Compliance, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

- "Arrangement and Description: The re-organization of files (e.g., new folder directory structure) in a dataset that may also involve the creation of new file names, file descriptions, and the recording of technical metadata inherent to the files (e.g., date last modified)."
- "Code review: Run and validate computer code (e.g., look for missing files and/or errors) in order to find mistakes overlooked in the initial development phase, improving the overall quality of software."
- "Data Cleaning: A process used to improve data quality by detecting and correcting (or removing) defects & errors in data."
- "Deidentification: Redacting or removing personally identifiable or protected information (e.g., sensitive geographic locations) from a dataset prior to sharing with third-parties."
- "File renaming: To rename files in a dataset, often to standardize and/or reflect important metadata."
- "Indexing: Verify all metadata provided by the author and crosswalk to descriptive and administrative metadata compliant with a standard format for repository interoperability."
- "Interoperability: Formatting the data using a disciplinary standard for better integration with other datasets and/or systems."
- "Quality Assurance: Ensure that all documentation and metadata are comprehensive and complete. Example actions might include: open and run the data files; inspect the contents in order to validate, clean, and/or enhance data for future use; look for missing documentation about codes used, the significance of "null" and "blank" values, or unclear acronyms."
- "Restructure: Organize and/or reformate poorly structured data files to clarify their meaning and importance."

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF05 Curation, Quality & Compliance is addressed in the following Agri-Food resources:

Funded by the European Union

Harper et al., 2018: AgBioData recommends that repositories define and publish data-quality and compliance policies to ensure community standards are met. “Developing data standards and practices to facilitate data curation, integration and reuse is essential. Standards will ensure data quality, facilitate interoperability and reduce redundant work.”

Sen et al., 2020: WheatIS requires repositories to apply common quality criteria and ensure compliance with agreed data and metadata standards across distributed nodes.

Top et al., 2022 : One of the cases in the fourth part of the article illustrates the need for a specific curation and data alignment process : “Selected data goes through a dedicated data curation procedure. This is a crucial step to enable use of the data beyond its original purpose of collection. In this procedure metadata is checked and completed using all information available in the data files, supporting documents, publications in data and scientific journals, etc. “

Biomedical Sciences

The TTRAM Activity/Function AF05 Curation, Quality & Compliance is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017:

- Compliance: FAIR is a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. These indicators will be used to demonstrate that ELIXIR Core Data Resources (CDR) are compatible with the FAIR data principles. For instance, a CDR is expected to provide persistent and unique identifiers to enable usability and community-recognised standards for metadata and data to enable interoperability.
- Level of quality: scientific quality is one of the key indicators, especially for knowledge bases which add substantial value through expert curation, annotation of metadata. The curation effort and outputs linked to a resource are an important measure of the quality of ELIXIR CDRs. This is why the number of FTE corresponding to Curators position to support adherence to metadata requirements or support for extraction of information from the scientific literature is one of the indicators.

Lin et al., 2024: In the article, curation, described as the process of employing various standards and best practices to transform data into meaningful organized, structured, and computable forms is one of the 6 common characteristics of data repositories. To ensure transparency, users of a data repository should understand the operational aspects of a repository, such as data validation and curation procedures.

Field et al., 2011: The GSC developed the MIxS standard, including checklists (MIGS/MIMS/MIMARKS) that require core information upon data submission and publication to ensure compliance and richer entries.

Funded by the European Union

Yilmaz et al., 2011: The MIxS standard is crucial in ensuring sequence data reaches a defined level of quality and standards compliance before it is made available for reuse by major data providers, including partners of the International Nucleotide Sequence Database Collaboration.

Barrett et al., 2012: The article describes mechanisms that support metadata quality and standardization, but it does not fully implement the level of curation, quality assurance, and compliance enforcement envisioned by FIDELIS TTRAM. The responsibility for metadata quality largely remains with the submitters.

Karsch-Mizrachi et al., 2025: The article explains how the INSDC ensures that nucleotide sequence data meets defined quality standards and compliance requirements before it is made available for reuse. This is achieved through the implementation of validation rules, metadata checklists, and minimal standards that are applied during the submission process. These efforts reflect a clear commitment to ensuring that digital objects are curated and meet quality and compliance benchmarks before being integrated into the repository and made accessible to the global scientific community.

Climate Science

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) states that Curation, Quality and Compliance is actively managed through the CMIP5 Questionnaire, a metadata entry tool that enforces Controlled Vocabulary constraints, validates against the CIM XSD, and performs Schematron-based validation to check the deeper coherency of collected parameters, ensuring digital objects and their documentation adhere to defined standards before publication.

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) explains a multi-stage process involving a pre-publication quality control phase, which encompasses both Technical Quality Assurance (TQA) performed by WDCC and Scientific Quality Assurance (SQA) overseen by the data provider, verifying adherence to open source file formats, CF Metadata Conventions, and WDCC Minimal Standards for File Headers. Post-publication, curation activities maintain long-term usability and citability by enabling the addition of references, management of "open time series" for evolving datasets, and the creation of new versions or handling of errata with explanations.

The **NetCDF Climate and Forecast (CF) Metadata Conventions** (Eaton et al., 2024) state that the CF conventions ensure requirements for self-describing metadata that is both human-readable and easily parsable by programs. This framework defines precise attributes for variables, units, coordinate systems, and data representation to facilitate unambiguous interpretation and interoperability.

Linguistics

The Activity/Function (A|F) CURATION, QUALITY & COMPLIANCE is not addressed by any of the reviewed resources from Linguistics.

Funded by the European Union

Social Sciences

(Kleemola et al., 2025) recommends that data repositories should:

- set standards for preservation and other repository operations by for example contributing to defining TDRs and by building consensus on curation and preservation levels.
- leverage domain-specific expertise in long-term data preservation, advocating for domain-specific services such as metadata creation and data curation and advocating the benefits of long-term data preservation for future research and society.

A recommended best practice is for the repository to have a clearly defined data curation policy that outlines how data is maintained and how its value is enhanced to support reuse and long-term preservation. (CESSDA Training Team, 2025a, Chapter 4.4 Quality assurance of data and documentation material).

The CESSDA Data Management Expert Guide addresses the limitations of the DDI metadata standard, which is recommended for social science research. It has limitations in "describing biases caused by data mining interfaces of social media platforms, in data availability and formats, explanations about code and scripts used in collection, cleaning and analysis etc." These can be described only as free-text comments outside the structured standard. (CESSDA Training Team, 2022, Chapter 2. Organise and Document).

The Guide also emphasizes the value of trusted CESSDA repositories as they ensure long-term access and data quality through expert guidance. Experts help improve metadata completeness, advise on suitable file formats for long-term preservation and review the quality of data. (CESSDA Training Team, 2022, Chapter 6. Archive&Publish).

3.2.4. DISCOVERY & IDENTIFICATION (AF06)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Discovery & Identification as follows: "Applying persistent identifiers and descriptive metadata to digital objects to support resource discovery. Providing discovery systems and making metadata available for harvesting." and suggests for transparent information "[p]ersistent identifiers used (ideally chosen from a controlled vocabulary) for objects, organisations, researchers, software etc, and information about whether all objects are persistently identified. Information about identification at a more granular level (files within objects, questions within surveys, variables within statistics). Links to resource discovery systems that the repository provides or third parties that index metadata harvested from the repository collection." (L'Hours et al., 2025c, p. 14) including

- Persistent Identifier system(s)
- PID Policy
- Discovery metadata and documentation

Funded by the European Union

- Resource discovery catalogues
- Harvesting protocols (e.g. OAI-PMH)

The A|F Discovery & Identification is covered by the **CoreTrustSeal** requirement with the identical name, requiring certifying repositories to enable users to discover the digital objects and refer to them in a persistent way through proper citation. The CoreTrustSeal extended guidance asks applicant to provide references to the following items (CoreTrustSeal Standards and Certification Board, 2022, pp. 22–23):

- The search facilities offered by the repository.
- The standards that a searchable metadata catalogue complies with.
- The approach to ensuring that identifiers are unique and persistent.
- Machine harvesting of the metadata.
- Repository, or repository data and metadata, inclusion in disciplinary or generic registries of resources.
- Recommended data citations.

The **nestor Seal** criteria relevant to this A|F are “C16 Integrity: User Interface”, “C27 Identification”, and “C28 Descriptive Metadata”. C16 requires the archive to provide an interface “which allows users and the digital archive administration to check and maintain the integrity of the representations”. C27 deals with the use of unique and persistent internal and external identifiers. In addition, “C4 Access” mentions the need for “appropriate search possibilities” (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** the Data Management functional entity manages the metadata required for data discovery - i.e. the descriptive information generated during Ingest - and will return responses to queries from the Access functional entity (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4.2.3.5).

The **EOSC Federation Handbook** addresses parts of the A|F Discovery & Identification in several of its sections (EOSC Association, 2025):

- **5.2.1 Research publications:** The EOSC Federation will automatically reference publications through OpenAIRE and be searchable via the EU Node resource catalogue.
- **5.2.2 Research data sources:** Linking between data repositories and databases is strongly encouraged and is fundamental to creating the Web of Data.
- **5.2.2.1 FAIR Data repositories:** The repositories must be [...] be findable in the EOSC Federation catalogue and searchable through a custom search interface. Data repositories should implement a search function which is optimised to find data from the scientific domain they serve. [...] Data must be citable via a Persistent IDentifier (PID) according to the “Guidelines for creating a user tailored EOSC Compliant PID Policy” and the Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC). [...] Data repositories must ensure that Metadata are harvestable via the standard protocols OAI-PMH (https://www.openarchives.org/pmh/).

Funded by the European Union

The **Desirable Characteristics of Data Repositories for Federally Funded Research** emphasizing that repositories should assign unique, citable PIDs (like DOIs) to datasets and ensure they are accompanied by descriptive metadata using appropriate schemas to enable discovery, reuse, and citation (White House Office of Science and Technology Policy, 2022).

The **EU Annotated Grant Agreement** states repositories assign persistent unique identifiers to contents (e.g. DOIs, handles, etc), such that the contents (publications, data and other research outputs) are unequivocally referenced and thus citable. They ensure that contents are accompanied by metadata sufficiently detailed and of sufficiently high quality to enable discovery, reuse and citation and contain information about provenance and licensing. Their metadata is machine-actionable and standardized (e.g. Dublin Core, Data Cite, etc) preferably using common non-proprietary formats and following the standards of the respective community the repository serves, where applicable. For an overview/comparison of EC GA metadata requirements, see EC GA Metadata Requirements (European Commission, 2024).

The **Update of the Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements** show also in this case that the main reasons for not meeting the essential characteristics for trusted repositories are the lack of a licence field in the metadata, along with missing adherence to a specific standard for metadata. Furthermore, grant information is often not offered through separate metadata fields, resulting in a lack of machine-actionable, interoperable, and standardized metadata, as requested in the MGA (Lazzeri, 2024).

The **Recommendations for Services in a FAIR data ecosystem** (Bangert et al., 2019) states that services supporting FAIR data should offer or make use of the following components:

A. PID services for a wide range of objects, such as publications, researchers, data sets and organisations. Emerging PID types (e.g. for instruments) should be monitored and used when they are mature.
B. Domain-specific ontologies, as domain-specific requirements have to be taken into account.
C. Human and machine-readable standards to make datasets findable, reusable and interoperable (licences as one particular example of standards needed for machine readability).
D. If applicable, metadata that complies with appropriate (domain) standards should be generated and captured automatically (for e.g by instruments) .

The **Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements** (Jahn et al., 2023) states that for fulfilment of the Horizon Europe MGA metadata access and licensing criteria, we consider the repository to be required to provide metadata related to each digital object via a public domain dedication such as CC0, Public Domain or equivalent. However, as it is further shown in the Analysis section many repositories do not readily provide this information in a clear human or machine-readable manner.

Funded by the European Union

The **Practical Guide to the International Alignment of Research Data Management** (Science Europe, 2021) states that:

1. Provision of Persistent and Unique Identifiers (PIDs):
    a. Allow data discovery and identification
    b. Enable searching, citing, and retrieval of data
    c. Provide support for data versioning.
2. 2. Metadata:
    a. Enable finding of data
    b. Enable referencing to related relevant information, such as other data and publications
    c. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data
    d. Use metadata standards that are broadly accepted (by the scientific community)
    e. Ensure that metadata are machine-retrievable.

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) states on data and metadata standards that the community-defined standards the repository implements to enable the representation of data and/or metadata in a consistent, machine readable form (e.g. via models, formats, schemas, vocabularies, ontologies). These standards facilitate the discovery and interpretation of data and/or metadata. Which data and metadata standards (if any) has the repository implemented? Persistent Identifiers for Data. Globally unique and Persistent IDentifiers (PIDs). Does the repository assign PIDs to its holdings? If so, which PID schema has been implemented? Citation to related publications. A mechanism to link datasets to related articles or pre-prints. Does the repository enable data to article linking? At what stage of data deposition is article information required?

The **Current State and Future Directions for Open Repositories in Europe** (Shearer et al., 2023) states that despite the fact that a repository supports certain metadata schemas and PIDs does not always equate to the collections having high quality metadata. While most repositories do support standardised and granular metadata schemas, they often rely on the author to fill in the metadata fields. Since authors may not be aware of the standards, this often leads to lower quality metadata records. While most repositories do undertake basic metadata curation and checking, this may not be sufficient to optimize discovery and reuse of repository resources. There are opportunities to improve the quality of metadata - either through data curation activities at the repository, or by introducing machine extraction of metadata information - but this may require greater commitment in terms of staff and technical resources at the repository.

The exclusion criteria from the **Research Data College Working Group** rule out subject-specific repositories that do not assign long-term identifiers (Lamotte et al., 2024, p. 10).

The **O'FAIRe makes you an offer: Metadata-based Automatic FAIRness Assessment for Ontologies and Semantic Resource's** (Amdouni et al., 2022) analysis of 149 semantic resources showed generally high scores for Findability (F), with an average normalized score of 70, confirming that most

Funded by the European Union

ontologies utilize URIs as local/primary identifiers (65%) and that AgroPortal provides discovery systems, making metadata explicitly available. Specifically regarding identification, 44% of resources provide an external/secondary identifier, but significant improvement is needed as only 14% provide a version-specific URI.

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) establishes that the Findability (F) area is entirely comprised of Essential indicators focusing on these concepts, requiring that both metadata and data must be identified by a persistent and globally unique identifier (F1). Furthermore, it is essential that rich metadata is provided to allow discovery (F2).

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), all the 11 surveyed repositories use one or more metadata schema. Persistent identifiers were used by the majority of respondents. All 11 repositories provided guidance to depositors regarding required data and metadata formats. Nine repositories reported that their metadata was available through resource discovery services such as B2FIND, CLARIN VLO, or CESSDA Data Catalogue. Just over half of these repositories also had an OAIPMH endpoint. Finally, most of the respondents supported data citation by providing a sample citation or citation instructions, some of them via a machine actionable sample citations for individual datasets.

The discovery and identification activity and function are covered in the **FAIRsFAIR project D2.3** report (Behnke et al., 2020) via the mention of community standards and ontologies from public registries that must be reused by the repository, which must also provide metadata in different formats to be harvested by different search engines in RDF or JSON-LD. For each Digital Object's page, the HTTP header should contain technical metadata about the Digital Objects. Besides, the repository should provide a search interface or be linked to aggregating services that enable findability of Digital Objects. The metadata can be provided at the level of files, variables, attributes, individual cells. The granularity is decided by the repository.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.3, Metadata Quality, that flexible, discipline-agnostic metadata standards (e.g. DCAT) are needed to facilitate interdisciplinary research, data discovery, and data package transfer between repositories. Recommendations and workflows are needed to increase granularity for PIDs – for example, at file-, sample-, and instrument-level. In section 4.3.4, Trustworthy Digital Archive Capabilities, explains that while most infrastructures have considered machine readability and have API access to their data, open data and metadata harvesting protocols should be standardized with improved documentation such as through use of OAI-PMH, or other well-established suitable protocols.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) recommend using established standards like DCAT and schema.org to describe repositories and their digital objects, which fosters interoperability and supports machine-actionability for harvesters and discovery services. Furthermore, the guidelines emphasize that repositories must expose Persistent IDs (PIDs) with their

Funded by the European Union

(meta)data and ensure they are resolvable and associated with metadata to improve findability and reuse.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that persistent identifiers, metadata management, and cataloguing are discussed as essential for discovery. Certified repositories provide more comprehensive discovery information.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) strongly aligns with metadata availability, establishing multiple indicators under the Findable (F) principle as Essential, including requiring that metadata (RDA-F1-01M) and data (RDA-F1-01D) are identified by persistent and globally unique identifiers, that rich metadata is provided for discovery (RDA-F2-01M), that metadata includes the identifier for the data (RDA-F3-01M), and that metadata is offered in a way that allows harvesting and indexing (RDA-F4-01M).

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) states that to achieve Discovery and Identification, repositories must assign persistent identifiers (PIDs) that point to the resource's landing page and also support PIDs for related entities like authors and funders. Effective discovery also requires the application of basic Dublin Core and granular metadata elements using controlled vocabularies, and the support of metadata harvesting using OAI-PMH so that records can be indexed by external academic discovery services.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) recommend that repositories submit data citations they collect to DataCite (see DataCite's documentation for contributing data citations to DataCite) so that these connections can be aggregated and discoverable by others in the community. It further recommend that repositories use the metadata fields recommended by GREI to collect data citations, specifically, those listed below to establish the relationship between the dataset and the citing object:

- 12. relatedIdentifier
- 12.a relatedIdentifierType
- 12.b relationType
- 12.f resourceTypeGeneral

In order to encourage researchers and other parties to cite datasets they use, data repositories should provide a citation template on the landing page of the dataset. Repositories should expose the citations for individual datasets on the landing page for the dataset record. To increase transparency and trust in the information displayed for data citations, we recommend that repositories indicate the provenance for the data citation, i.e. the source for the asserted citation. The citation information listed at the repository should include: The number of citations to the dataset; The list of citing objects and their associated identifiers, or a link to where such a list of citing objects can be accessed; A breakdown of the citations according to their provenance, in text

Funded by the European Union

format - where the repository harvests data citations from external sources. The repository should also document, as part of the information pages for contributors and users, information on the mechanisms by which the repository collects and stores data citation, including the sources it employs. It recommends that all data repositories contribute their data citations to DataCite so that those citations can be integrated into the Data Citations Corpus.

In the **D4.5 Report on Completed FAIR Data Standard Adoption and Certifications of Data Repositories in the Region** (Alaterä et al., 2022), it is recommended to ensure that data and metadata resources are assigned a unique and persistent identifier, such as a globally unique ID (GUID) like a URL or DOI. As a best practice, it is further advised to provide (meta)data with a Global Unique Persistent Resolvable Identifier (GUPRI). Additionally, metadata should be structured and made available in formats such as RDF or JSON-LD. The report also recommends separating metadata from data, ensuring that metadata explicitly includes the identifier of the related data. To enhance findability, resources should be indexed in registered, searchable repositories or machine-actionable search engines. The (meta) data must be associated with community / domain standards and based upon agreed vocabularies and semantic models. An explicit pointer to the templates, schemas, ontologies, vocabularies, variables must be provided. Beyond these FAIR-aligned recommendations, the authors tested and validated several generic best practices, including:

- Promoting the use of schema.org and other widely adopted metadata standards, embedded or linked in machine-actionable linked data formats (e.g., JSON-LD, RDF).
- Signposting generic metadata as typed links in HTTP headers, as recommended by signposting.org.
- Expressing data licenses in machine-actionable formats, for example by using standard Creative Commons licenses.

The **OpenAIRE** Guidelines for Data Archives 2.0 (OpenAIRE, 2022) provides “instruction for data archive managers to expose their metadata in a way that is compatible with the OpenAIRE infrastructure. [...] By implementing the OpenAIRE Guidelines, data archive managers are facilitating the creation of enhanced publications and building the stepping-stones for a linked data infrastructure for research. Metadata: OpenAIRE builds on the DataCite Metadata Schema v3.1 by making some of the otherwise optional DataCite properties mandatory, as well as enforcing specific encoding schemes on the values of some DataCite properties; Use of OAI-PMH.”

The **Core Preservation Process CPP-005 Identifier Management** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “Identifiers are assigned to Objects, Information packages and/or Metadata, and managed along to their life cycle.” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-024 Enabling Discovery** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA provides catalogue services to its

Funded by the European Union

consumers to help them identify Objects that they may be interested in." (EOSC EDEN T1.2 et al., 2025).

Within AF06 Discovery & Identification, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

- "Metadata: Information about a data set that is structured (often in machine-readable format) for purposes of search and retrieval. Metadata elements may include basic information (e.g. title, author, date created, etc.) and/or specific elements inherent to datasets (e.g., spatial coverage, time periods)."
- "Persistent Identifier: A URL (or Uniform Resource Locator) that is monitored by an authority to ensure a stable web location for consistent citation and long-term discoverability. Provides redirection when necessary. E.g., a Digital Object Identifier or DOI."
- "Data Citation: Display of a recommended bibliographic citation for a dataset to enable appropriate attribution by third-party users in order to formally incorporate data reuse as part of the scholarly ecosystem."
- "Discovery Services: Services that incorporate machine-based search and retrieval functionality that help users identify what data exist, where the data are located, and how can they be accessed (e.g., full-text indexing or web optimization)."
- "Full-Text Indexing: Enhance the data for discovery purposes by generating search-engine-optimized formats of the text inherent to the data."
- "Metadata Brokerage: Active dissemination of a data set's metadata to search and discovery services (e.g., article databases, catalogs, web-based indexes) for federated search and discovery."

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF06 Discovery & Identification is addressed in the following Agri-Food resources:

Caracciolo et al., 2020: The Agrisemantics group instructs repositories to ensure persistent identifiers and machine-readable metadata to make agricultural data discoverable.

Harper et al., 2018: Recommends that repositories apply common indexation technologies such as Apache SolR to facilitate discovery; and persistent identifiers and standardized metadata be applied across genomic databases to enable discovery and reuse.: **"**Make your database discoverable by indexing and providing a search engine. This will expose your data through a standard interface and allows creation of a federated database system.**"**

Top et al., 2022: AgroDataCube demonstrates repository best practice by assigning DOIs and linking catalogue entries to discovery systems through persistent and accessible web resources. "A DOI

Funded by the European Union

(needed for 'F' and 'A' of FAIR) is available, which refers to a digital catalogue entry at the Wageningen University & Research library and directs the reader to the associated web page of AgroDataCube."

Drakos et al., 2015: agINFRA recommends that repositories interconnect through extended metadata and adopt European standards for discovery and identification.

Sen et al., 2020: Discoverability is at the heart of WheatIS, which directs repositories to provide standardized metadata for global discovery and federated search across distributed nodes. "The mission of the WheatIS EWG was to create an informational infrastructure, establish data standards, and build a single portal that allows search, retrieval, and display of globally distributed wheat data sets that are indexed in standard data formats at servers around the world."

Biomedical Sciences

The TTRAM Activity/Function AF06 Discovery & Identification is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017:

- Persistent identifiers: ELIXIR CDRs are expected to use stable and resolvable identifiers (e.g., DOIs, accession numbers) for their digital objects to support long-term access and citation, which is a key aspect of discovery and identification.
- Metadata: The article emphasizes the importance of high-quality metadata, which must be rich, standardized, and machine-readable.

Lin et al., 2024: In the article, the importance of the use of rich metadata to support dataset discovery, reuse, and citation is highlighted. Besides, according to authors, datasets and metadata should follow a common, non-proprietary format based on community standards to ensure interoperability and long-term accessibility.

Sansone & Rocca-Serra, 2016: The article addresses the importance of FAIR principles and use of metadata standards for achieving discoverability, citation, credit as well as interoperability. Indeed, it strongly supports and elaborates on the Discovery & Identification function by:

- Advocating for persistent identifiers.
- Promoting rich, structured metadata.
- Highlighting tools and initiatives that enable metadata harvesting and indexing.
- Connecting these efforts to the broader goal of making digital objects FAIR.

Yilmaz et al., 2011: MIxS standardization addresses the previous issue of sparsely annotated sequences in databases that made it impossible to reliably retrieve data based on environmental or geographic origins, thus significantly enhancing the ability to discover, integrate, and analyze natural genetic diversity.

Funded by the European Union

Barrett et al., 2012: Each BioProject and BioSample submission is assigned a unique accession number, which serves as a persistent identifier to link metadata with corresponding experimental data across NCBI's archival databases. The system captures rich descriptive metadata for both projects and samples, and encourages the use of controlled vocabularies and attribute dictionaries to ensure consistency and improve metadata quality. These metadata records are indexed within NCBI's Entrez search system, enabling users to perform both free-text and fielded queries, utilize advanced search features, and navigate across related databases like SRA, GenBank, and GEO. Additionally, metadata are made available for bulk download via FTP in XML and tab-delimited formats, and can be accessed programmatically through Entrez Utilities.

Karsch-Mizrachi et al., 2025: The article explains that the INSDC applies persistent identifiers to all data records in the form of globally unique accessions, which are essential for data citation and discovery. These accessions are widely recognized and used by publishers to support reproducibility and traceability in scientific research. The article also discusses the importance of descriptive metadata, particularly contextual information such as sample type, geolocation, and collection date, which enhances the discoverability and utility of the data. Additionally, the INSDC provides public access to its databases through public search and retrieval tools.

Wilkinson et al., 2016: The article addresses the Findability (F) as a core principle, requiring that (meta)data are assigned globally unique and persistent identifiers (F1) and are richly described with metadata (F2) that explicitly includes the data's identifier (F3). This ensures (meta)data are registered or indexed in searchable resources (F4).

Rehm et al., 2021: The article describes the use of persistent identifiers and metadata standards to support data discovery and interoperability. It also outlines APIs like Beacon and Data Connect for genomics data discovery and searching across federated repositories.

Climate Science

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) explains that the METAFOR project (Callaghan et al., 2010) established a crucial standardized Controlled Vocabulary (CV) and metadata structure, integrated within the Common Information Model (CIM), to support Discovery and Identification of climate models and simulations, ensuring consistency and comparability essential for end-users accessing CMIP5 data. This systematic approach employs persistent identifiers defined in the CMIP5 Data Reference Syntax (DRS) and utilizes an information pipeline that converts the CV into an OWL ontology and broadcasts CIM-compliant documents as "atom feeds," enabling resource discovery via tools like Earth System Grid Federation (ESGF) gateway interfaces and faceted browsing.

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC assign persistent identifiers (DOIs and PIDs) to archived data and metadata, and require comprehensive descriptive metadata entered through a graphical user interface called MetaXA.

The **NetCDF Climate and Forecast (CF) Metadata Conventions** (Eaton et al., 2024) state that the CF conventions enhance digital object discovery and identification by ensuring datasets are

Funded by the European Union

self-describing through standardized, machine-parsable metadata. This design's use of persistent identifiers (e.g., DOIs for the document itself and LSIDs for data elements) directly supports automated discovery systems.

Linguistics

The **Tromsø recommendations for citation of research data in linguistics** (Andreassen et al., 2019) provide guidance on how to cite data from linguistics and language research and are further elaborated in the handbook chapter **"Guidance for Citing Linguistic Data"** (Conzett & De Smedt, 2022). The Tromsø recommendations suggest two templates for citation of datasets (Conzett & De Smedt, 2022, p. 146):

1. The template for a minimal bibliographic reference to a dataset has the following elements: **Author, Date, Title, Publisher, Locator.**
2. The template for an expanded bibliographic reference to a dataset including conditional elements (i.e., required in certain cases depending on resource characteristics) is as follows: **Author**, Other Attribution (Roles), **Date, Title, Publisher, Locator,** Version, Date accessed.

Elements rendered in bold are part of the minimal template, in other words, they are always required, while elements rendered in italics are considered to be conditional. Conditional elements are elements whose presence is conditioned by either the characteristics of the resource (e.g., references to versioned datasets should include the version number), or on subfield-specific traditions (e.g., in language documentation, it is common to acknowledge the contributions of language consultants by name).

Repositories and other resource providers are advised to provide metadata conformant to the following (Conzett & De Smedt, 2022, p. 153):

- At minimum, the metadata should include the elements in the minimal templates recommended herein.
- Metadata should preferably be structured according to a standard format (e.g., component MetaData Infrastructure, RIS) so that information from it can be extracted by programs.
- Metadata should be available freely, without cost or restrictions, even if the data themselves have restrictions.
- The metadata should allow persistent reference to the data set, and to the metadata themselves, to avoid link rot. This implies that PIDs should be assigned to the data and to the metadata.
- Data repositories and other resource providers should provide metadata in machine-readable as well as human-readable form and should preferably also generate ready-made citations, both in formats for export to reference managers and in textual format.

The **CLARIN** Data Citation Guidelines version 1.0 (Matthiesen & Leonardič, 2025) differ from the Tromsø recommendations on the following items:

Funded by the European Union

- "Labelling different versions is simply a matter of convention (rather than a necessary condition for identification) given that each new version of a dataset should be assigned a unique PID."
- "For resources that are in CLARIN repositories, URLs other than those containing PIDs are not acceptable since they might break."
- "Note that the use of PIDs makes mentioning "accessed at <date>" redundant. We therefore do not recommend it, in contrast to other recommendations."

For **CLARIN** B and E centres, the following requirements and recommendations apply when it comes to the AF DISCOVERY & IDENTIFICATION (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023, p. 3):

- Centres need to offer component based metadata (CMDI) that make use of elements from accepted registries such as the CCR in accordance with the CLARIN agreements, i.e. metadata needs to be harvestable via OAI PMH.
- Centres need to associate PIDs records according to the CLARIN agreements with their objects and add them to the metadata record.
- Centres are advised to participate in the Federated Content Search with their collections by providing an SRU/CQL Endpoint. This content search is especially suitable for textual transcriptions and resources.

CLARIN Type C centres are expected to serve metadata via the OAI-PMH protocol (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023, p. 2).

For CLARIN B centres, the following requirements apply (Wittenburg, Van Uytvanck, Zastrow, Straňák, et al., 2023):

- 3.c Licenses on data and metadata. Requirement: Data and metadata are licensed.
- 6.a Metadata harvestable via OAI-PMH to the VLO. Requirement: Computer access to the repository: Metadata harvesting to VLO should work - see https://vlo.clarin.eu/data/.
- 6.b CMDI metadata validation and use of CMDI profiles with CCR ConceptLinks. Requirement: The metadata should be CMDI-compliant (see https://www.clarin.eu/cmdi).
- 6.c Metadata PID and references to resources. Requirement:
    - 1) State if the harvested CMDI files contain a PID in the MdSelfLink header field;
    - 2) State if the harvested CMDI files refer to web-accessible files or a landing page with a ResourceProxy.
- 7. Persistent Identifiers. Requirement: Centres need to associate PIDs (handles or DOIs) with their metadata records. These PIDs should be suitable for both human and machine interpretation, taking into account the HTTP-accept header. Individual files (e.g. a text, zip or sound file) can be referred to with either the handle of the describing metadata record in combination with a part identifier or with another handle.
- 8. Federated Content Search (optional). Requirement: Centres can choose to participate in the Federated Content Search with their collections by providing an SRU/CQL Endpoint.

Funded by the European Union

Social Sciences

The CESSDA Data Access Policy (Bolton, 2022) emphasizes the importance of discovery and identification by requiring that metadata be openly available to anyone (Principle 1), aligned with FAIR principles (Principle 4), harvestable using controlled vocabularies (Principle 5), include persistent identifiers and citation guidance (Principle 6), and clearly state access conditions (Principle 7).

The CESSDA Metadata Model (CMM) serves as a metadata schema for social science data. It contains metadata elements, their definitions and information on other requirements, such as repeatability. The model is intended to enhance the discoverability and clarity of data for users, as well as to support interoperability between CESSDA service providers. It is built from the viewpoint of quantitative (social science) data and is based on the DDI Lifecycle 3.2 metadata standard. (Akdeniz & Moilanen, 2023).

(Kleemola et al., 2025) suggest that data repositories should extend the use of PIDs to cover datasets, studies, authors, data contributors, funders, and study components while standardising machine-readable metadata across repositories.

The CESSDA Data Archiving Guide Chapter 2.4 (CESSDA Training Team, 2025a, Chapter 2.4 Data Access Policies) states that having a persistent identifier (PID) for data is a key element in enabling data access - and thereby supports discovery.

Staff should ensure that metadata is created according to FAIR principles, making research data more accessible and reusable. Data descriptions follow international standards, such as those defined by the Data Documentation Initiative (DDI) and CESSDA. To support consistency and interoperability across datasets and catalogues, controlled vocabularies should be used for key metadata fields. Using the controlled vocabularies should be mandatory for all published datasets. (CESSDA Training Team, 2025a, Chapter 4.3.2 Vocabularies (CMM, DDI, etc.)).

The article recommends that repositories should provide rich citation-related metadata in machine-actionable data format and citation suggestions that include at least the six core citation components, including persistent identifier (PID) and version information. PIDs should be available also to embargoed datasets. General metadata containers should describe relationships between the dataset and other entities, using machine-actionable PIDs (ORCID, ROR ID, DOI). The document recommends that citation requests could be included in metadata and repository entries. (Bornatici et al., 2025, pp. 6–8).

3.2.5. ACCESS (AF07)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Access as follows: “Determining the appropriate method of access—such as direct download or use within a secure remote environment—and facilitating user access accordingly. Access criteria are based on the characteristics of both the digital objects and the users, in alignment with rights management

Funded by the European Union

policies (see AF15 "Rights").” (L’Hours et al., 2025c, p. 15) and suggests the following for transparent information:

- Access routes & methods
- Access Management Policy
- Documentation of access routes and methods, including machine harvesting
- Licence and terms of access (see also AF15 “Rights”)

In the **OAIS Reference Model** access is handled by the Access functional entity. It maintains a search interface for the data users, e.g. a catalog, and creates Dissemination Information Packages (DIPs) to provide archive content to the users (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4.2.3.8).

In the **nestor Seal**, criterion “C4 Access” is relevant to this A|F. It states: “The digital archive ensures that authorised users in the designated communities can access the representations. This includes appropriate search possibilities. The digital archive openly declares its conditions of use and any costs which may arise, listing these in a transparent manner”. In addition, C16 Integrity: User Interface has some relevance to this A|F (nestor Certification Working Group, 2025).

The **EOSC Federation Handbook** addresses parts of the A|F Access in two of its sections (EOSC Association, 2025):

- **5.2.2.1 FAIR Data repositories:** Data repositories must have a data policy that governs the access to the data which conforms with the general data policy of the EOSC Federation.
- **5.2.5 Research Services:** All Research Services must be accessible either anonymously e.g. for open data repositories, or through an EOSC compliant AAI (cf. Chapter 4) in accordance with documented User Access Policies (see Chapter 6).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** policy emphasizes that repositories should provide broad, equitable, and maximally open access to datasets and metadata free of charge, ensuring clear documentation of access and use terms. For human data, repositories must implement specific procedures like tiered access, user credentialing, download control, and review of access requests to protect privacy and ensure fidelity (White House Office of Science and Technology Policy, 2022).

The **Practical Guide to the International Alignment of Research Data Management** (Science Europe, 2021) states that:

- Data access and usage licences:
    - Enable access to data under well-specified conditions
    - Ensure data authenticity and integrity
    - Enable retrieval of data
    - Provide information about licensing and permissions (in ideally machine-readable form)
    - Ensure confidentiality and respect rights of data subjects and creators.

Funded by the European Union

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) states on data access conditions that data access mechanisms and terms to define access at repository and/or dataset level. Does the repository explicitly state data access conditions on dataset landing pages?

The **REPOSITORIES: Key Infrastructure For Maintaining European Research Excellence** (Shearer et al., 2025) states regarding powered by AI-ready infrastructure that adopting machine-readable metadata, linked data practices, and text/data mining capabilities to ensure repository resources remain accessible and valuable in an AI-driven research ecosystem.

Findings from **D1.3 Recommendations for a FAIR EOSC - White Paper of the FAIR-IMPACT Synchronisation Force** (Grootveld, 2025) emphasize that facilitating secure data sharing and access to data requires defining a harmonized operational and legal framework, and recommend advocating for Creative Commons licenses as the default option to enhance the legal and organizational interoperability that governs access criteria.

The **Research Data College Working Group** recommends that subject-specific repositories provide transparent information about whether "the repository offers the possibility of combining the deposit with an embargo" [...], as "[s]ome research teams may be keen to delay open-access publication of their datasets, specifying a specific embargo period" (Lamotte et al., 2024, p. 12).

The **O'FAIRe makes you an offer: Metadata-based Automatic FAIRness Assessment for Ontologies and Semantic Resources** (Amdouni et al., 2022) found that the Accessibility (A) score for semantic resources in AgroPortal was high, with an average normalized score of 90, due to all ontologies being accessible via the open, free, and universal HTTP protocol (A1.1), and the repository (AgroPortal) systematically supporting versioning (A2) and authentication and authorization (A1.2) for access control.

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) addresses appropriate user access through numerous Essential indicators in the Accessible (A) area, requiring that both metadata and data be retrievable using standardized and free access protocols (A1.1-01M, A1-04M, A1-04D) and that manual access is possible (A1-02M, A1-02D). Although the most secure access methods (like those supporting authentication and authorization) are only classified as a Useful indicator (A1.2-01D), it is Important that the metadata contains the necessary information regarding access conditions and required actions for the user to retrieve the data (A1-01M).

The **FAIRsFAIR project D2.3** report (Behnke et al., 2020) recommends different access policies for different versions of the data.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.3, Metadata Quality, that standardized time recommendations for file restrictions, access control or embargoes are needed in order to strike a balance between FAIR requirements for data, publication requirements and researcher concerns surrounding open data sharing. A further need is

Funded by the European Union

for repositories to ensure the existence of clear channels of communication to discuss restricted or embargoed data (e.g. author and/or repository contact information that is not time-sensitive).

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) propose that guidance should be incorporated directly into Data Management Planning (DMP) tools to support users in making informed choices about data deposit concerning access, along with storage, curation, and preservation, from the earliest stages of their research. Repositories are further advised to be transparent about the levels of care offered and received by digital objects, which should include details about the levels of retention, curation, and preservation in place and how they might change, impacting eventual access.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that access management is addressed through licensing, user support, and access portals. The report emphasizes the importance of transparency in access policies and the role of certification in improving access services.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) strongly aligns with appropriate method of access by making several access-related indicators essential, requiring that data and metadata identifiers resolve to the digital object or metadata record (RDA-A1-03M, RDA-A1-03D), that they are accessible through standardized protocols (RDA-A1-04M, RDA-A1-04D), and that metadata contains information necessary for the user to gain access (RDA-A1-01M, an Important indicator). Essential indicators also mandate that metadata is accessible through a free access protocol (RDA-A1.1-01M).

In the **D4.5 Report on Completed FAIR Data Standard Adoption and Certifications of Data Repositories in the Region** (Alaterä et al., 2022), repositories are required to ensure that data can be retrieved and properly resolved via an open and free protocol, by testing the resolution of the data's GUID. Furthermore, any discovered GUID must support authentication and authorization mechanisms within its resolution protocol and comply with an explicit data access policy. Metadata must also include a clearly identified, machine-readable persistence policy to ensure long-term accessibility and reliability.

The **Core Preservation Process CPP-025 Enabling Access** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA gives access to its Information Objects to authorised internal users or end users.” (EOSC EDEN T1.2 et al., 2025).

Within AF07 Access, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

- “Embargo: To restrict or mediate access to a data set, usually for a set period of time. In some cases an embargo may be used to protect not only access, but any knowledge that the data exist.”
- “File download: Allow access to the data materials by authorized third parties.”

Funded by the European Union

- "Restricted Access: In order to maintain the privacy of research subjects without losing integral components of the data, some data access will be protected and/or mediated to individuals that meet predefined criteria."

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF07 Access is addressed in the following Agri-Food resources:

Top et al., 2022: AgroDataCube provides a concrete repository model for transparent access documentation, combining DOI-based catalogue links, API-based retrieval, and license information. "The web page of AgroDataCube also provides information about the used license and the access token needed for data retrieval."

Drakos et al., 2015: agINFRA requires repositories to define user access methods through interoperable infrastructures and to validate access routes that enable collaboration. "Deploy tools (exchange standards, software, methodologies) for collaboration between European institutions in data-intensive research; validate the approach by enabling users to interact with each other and the data."

Sen et al., 2020: WheatIS directs repositories to establish standardized access methods enabling users to query distributed data sets securely and efficiently.

Caracciolo et al., 2020: The Agrisemantics group recommends that repositories ensure access through machine-readable services and interoperable web standards. "Providing explicit and machine-readable description of data makes it possible to programmatically integrate and reuse data."

Biomedical Sciences

The TTRAM Activity/Function AF07 Access is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: One of the key evaluation indicators for ELIXIR Core Data Resources (CDRs) is accessibility. ELIXIR CDRs must be freely and openly accessible to the scientific community, including direct access to data via downloads, APIs, or web interfaces. ELIXIR CDRs must support FAIR principles, among which accessibility.Data and metadata should be retrievable using standardized protocols. However, this methodology does not explicitly discuss secure remote environments or controlled access for sensitive data. It focuses primarily on open access to public data resources.

Lin et al., 2024: The article showcases the various access models to data repositories depending on the nature of the data and user requirements. These include:

- Open access, where data is freely available without registration or fees.

Funded by the European Union

- Registration-required access, which allows use after logging in, helping repositories track usage.
- Controlled access, typically for sensitive human data, requiring identity verification and research justification, often reviewed by a committee.
- Data enclaves, a subset of controlled access, where data cannot be downloaded and must be used within secure environments.
- Pay-to-access or donation-based models, which may support sustainability but can conflict with principles of equitable access.
- Closed access, where data is restricted, often for proprietary or commercial reasons.

These models reflect how repositories balance accessibility, security, and rights management in alignment with ethical and regulatory frameworks.

Sansone & Rocca-Serra, 2016: Addresses optimal interoperability access and use of data and other digital objects is completely automated, and accessible to both human and machine. Mentions the BioSharing/FairSharing (https://fairsharing.org)

Barrett et al., 2012: The article describes multiple access mechanisms:

- Direct download of metadata and data via FTP in XML and tab-delimited formats.
- Programmatic access through Entrez Utilities.
- Interactive search and retrieval via the Entrez system and BioProject/BioSample portals.
- Submission of data requires authentication through NIH or NCBI accounts, ensuring controlled access for depositors.
- Clinical samples with privacy concerns are excluded from BioSample and handled via dbGaP, which supports controlled access.

Karsch-Mizrachi et al., 2025: The article explains that INSDC provides free and unrestricted access to its nucleotide sequence data through publicly available databases and retrieval tools. The collaboration is built on the principle of open science, and all member organizations commit to making their data resources accessible without barriers. While the article does not go into detail about secure remote environments or differentiated access methods based on user characteristics, it does acknowledge that controlled-access human data is managed separately by each member and is not shared across the INSDC.

Wilkinson et al., 2016: In the article, access is defined as a core principle (A) for scholarly digital objects, asserting that (meta)data must be retrievable by their globally unique and persistent identifiers using standardized, open protocols, which can include authentication and authorization procedures when required. This ensures that both humans and computational agents can effectively

Funded by the European Union

access the data and metadata, with metadata remaining accessible even if the data itself is no longer available or is subject to restrictions, guided by clear usage licenses.

Rehm et al., 2021: access is facilitated through technical standards and policy frameworks that enable secure and responsible sharing of genomic and health data, supporting both direct data retrieval and analysis within secure federated environments. This include standards like the Data Repository Service API, htsget, and Crypt4GH to manage data access, while using GA4GH Passports and the Data Use Ontology to enforce access criteria based on data sensitivity and user authorization, ensuring alignment with rights management policies including Machine Readable Consent Guidance.

Climate Science

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) states that securing remote environments based on digital object and user characteristics, relies on the Earth System Grid Federation (ESGF) infrastructure for data publication and download in the climate modeling community. Crucially, the controlled vocabulary (CV) developed by the METAFOR project (Callaghan et al., 2010) provides standardized, high-level metadata essential for guiding end-users through data discovery, interpretation, and comparison by supporting tools like ESGF gateway interfaces and faceted browsing.

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC ensures data accessibility by offering DataCite DOI publication for archived items, which resolve to a WDCC-hosted landing page serving as an entry point for data access. The access criteria are primarily determined by the selected Creative Commons Licenses.

Linguistics

For CLARIN B centres, the following requirements apply (Wittenburg, Van Uytvanck, Zastrow, Straňák, et al., 2023):

- 6.d User access to the repository. Requirement: Each centre should set up a repository (a web-accessible server that offers human access to language resources/services and their metadata). Specify how human access is enabled.
- 3.a The centre should offer data access/sharing for users from other CLARIN ERIC countries.

Social Sciences

The CESSDA Data Archiving Guide covers the A|F Access in chapter 2.4 Data Access Policies. It describes three levels of access that depend on the type of data and the agreements with data depositors. The levels are:

- Open access to data: Data holdings can be accessed by users at any time and by any means.
- Restricted data access:

Funded by the European Union

1. *Standard access*: Data holdings are fully anonymized and available for scientific purposes, upon registration.
2. *Special conditions access*: User may gain access to data following the data depositor's requirements (e.g. written approval).
3. *Access under special licence*: Data that require users to complete a detailed special licence form describing in detail the terms of using that data.

- Controlled data access: The archive reserves the right to permit access to data only through a physical or virtual secure environment or after special training.

(CESSDA Training Team, 2025a, Chapter 2.4 Data Access Policies).

3.2.6. REUSE (AF08)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Reuse as follows: "Making and keeping digital objects usable and understandable depends on deposit, curation, and preservation activities. Facilitating reuse depends on understanding the needs and expectations of users, including—but not limited to—identified 'designated communities', through ongoing external engagement (see AF20 "External Engagement"). Repositories may mediate reuse through secure remote access systems, safe rooms, or other tools that keep the digital objects fully or partially under the service provider's control."). Repositories may mediate reuse through secure remote access systems, safe rooms, or other tools that keep the digital objects fully or partially under the service provider's control." and suggests for transparent information "References to artefacts including file formats, metadata and ontologies (potentially already mentioned under Deposit or Curation) that are specifically designed to enable reuse. Designated Community definitions, digital object models." (L'Hours et al., 2025c, p. 15) including

- Terms and Conditions for Use
- Designated Community Consultation information

CoreTrustSeal addresses the A/F Reuse through their requirement R13 with the same name. R13 requires repositories to "ensure that data and metadata continue to be understood and used effectively into the future despite changes in technology and the Designated Community's knowledge base" and evaluates "the measures taken to ensure that data and metadata are reusable" by asking applicants to provide evidence including references to the following items(CoreTrustSeal Standards and Certification Board, 2022, pp. 23–24):

- The ways in which the repository engages with their Designated Community of users to identify their needs.
- The data formats, metadata schemas, controlled vocabularies and ontologies used to support reuse, and how these meet the community needs.
- The metadata and documentation provided at the point of access to support understandability and reuse appropriate to the Designated Community. This may include information specific to data type, e.g. manuals, calibration records, photos, protocols.

Funded by the European Union

- Measures to ensure that data and metadata remain understandable.
- Management of changes to data, metadata, documentation or other information that supports reuse.
- Responses to this Requirement should focus on engagement with the Designated Community, identification of their needs and specifying how their needs are met.

In the **nestor Seal** criteria “C5 Interpretability” and “C13 Significant” properties are particularly relevant to this A|F. C5 requires that the “digital archive has defined measures to ensure the long-term interpretability of at least one of the representations, thereby meeting a basic precondition for appropriate use now and in the future. This includes the interpretability of both content data and metadata”. C13 states: “The digital archive identifies and documents which of the transferred representations' properties are significant for preservation of the information objects”. In addition, “C26 Dissemination information packages”, “C28 Descriptive metadata”, “C29 Structural metadata”, “C30 Technical metadata” have some relevance for this A|F (nestor Certification Working Group, 2025).

According to the **OAIS Reference Model** an Open Archival Information System has to ensure that the digital objects it preserves remain “Independently Understandable”[2] to its Designated Community (i.e. the defined community it primarily serves). In an OAIS the main types of information relevant for reuse are Representation Information and Preservation Description Information.

Structure Representation Information is any information about “the arrangement of and the organization of the parts or elements of the Data Object” (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 1/16), including information about how to render the bits of which a digital object is composed (e.g. by providing the necessary software). Semantic Representation Information “further describes the meaning of the Data Object, and its parts or elements” (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 1/16). An example for such Semantic Representation Information are column names in a spreadsheet. Preservation Description Information entails Provenance Information, Context Information, Reference Information, Fixity Information, and Access Rights Information (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 1/14). Provenance and Context Information can help potential users to better understand the preserved object semantically. To maintain the understandability of digital objects, an archive has to monitor the skills and needs of its Designated Community and update the Representation Information as needed (see Consultative Committee for Space Data Systems (CCSDS), 2024, p. 3/5).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** (White House Office of Science and Technology Policy, 2022) state that the repository must ensure datasets are accompanied by documentation describing terms of dataset access and use (e.g., reuse licenses and need for approval by a data use committee). The repository must ensure datasets are accompanied

[2] “Independently Understandable: A characteristic of information that is sufficiently complete to allow it to be understood by the Designated Community, as exemplified by the associated Preservation Objectives, without having to resort to special resources not widely available, including named individuals.” p.1–12

Funded by the European Union

by metadata that describe terms of reuse and provide the ability to measure attribution, citation, and reuse of data (e.g., through assignment of adequate and openly accessible metadata and unique PIDs).

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. They ensure that contents are accompanied by metadata sufficiently detailed and of sufficiently high quality to enable discovery, reuse and citation and contain information about provenance and licensing (European Commission, 2024).

The **Update of the Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements** show also in this case that the main reasons for not meeting the essential characteristics for trusted repositories are the lack of a licence field in the metadata, along with missing adherence to a specific standard for metadata (Lazzeri, 2024).

The **Practical Guide to the International Alignment of Research Data Management** (Science Europe, 2021) states that:

- Data access and usage licences:
    - Enable access to data under well-specified conditions
    - Ensure data authenticity and integrity
    - Enable retrieval of data
    - Provide information about licensing and permissions (in ideally machine-readable form)
    - Ensure confidentiality and respect rights of data subjects and creators.

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) states on data and metadata Standards. The community-defined standards the repository implements to enable the representation of data and/or metadata in a consistent, machine readable form (e.g. via models, formats, schemas, vocabularies, ontologies). These standards facilitate the discovery and interpretation of data and/or metadata. Which data and metadata standards (if any) has the repository implemented? Data Contact Information. Contact info (for the person, depositor, producer or owner, ideally with ORCID; or organization) of the data. Does the repository show data depositor contact information on dataset landing pages? Data Reuse Condition. License or terms of use for reuse of existing data in the repository; these can be the same for every dataset in the repository, or vary from dataset to dataset. What are the range of data reuse conditions which may be applied to the data holdings?

The **O'FAIRe makes you an offer: Metadata-based Automatic FAIRness Assessment for Ontologies and Semantic Resources** (Amdouni et al., 2022) study on 149 semantic resources in AgroPortal found that the average normalized score for Reusability (R) was significantly low, standing at 38, reflecting challenges in making digital objects usable and understandable over time.

Funded by the European Union

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) addresses reuse by defining five Essential indicators for Reusability (R1), primarily requiring that a plurality of accurate and relevant attributes are provided (R1-01M) and that metadata includes clear information about the license under which the data can be reused (R1.1-01M). Furthermore, ensuring understandability and usability (R1) is achieved through the Essential requirement that metadata complies with a machine-understandable community standard (R1.3-02M).

The **FAIRsFAIR project D2.3** report (Behnke et al., 2020) covers the Reuse activity and function via the use of permanent identifiers (PID) as well as machine-readable licenses, standard formats, and a strong emphasis on APIs.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) recommend, to enhance this reuse, that repositories express the history of FAIR assessment outcomes for an object over time to provide insights into improvements in FAIRness. Furthermore, since an object's FAIR score is heavily influenced by repository practices, repositories can report a "baseline" FAIR score at the organizational level to inform potential depositors about the achievable FAIR outcomes for objects deposited with them.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that certification supports data reuse by ensuring data is well curated, documented, and accessible. CoreTrustSeal certification encourages the adoption of FAIR principles to enhance reusability.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) directly addresses reusability (R) by mandating several Essential indicators: that a plurality of accurate and relevant attributes are provided to allow reuse (RDA-R1-01M), that metadata includes information about the license under which the data can be reused (RDA-R1.1-01M), and that both metadata and data comply with domain-relevant community standards (RDA-R1.3-01M, RDA-R1.3-01D).

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) requires repositories to stipulate reuse conditions by including licensing information or terms of use in the metadata record. Effective reuse also necessitates that resources are stored in machine-readable, non-proprietary formats and that landing pages include machine-readable citation metadata.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) conclude that data citations are among the biggest motivators for researchers to publish their data, which is essential for making objects available and usable. Furthermore, these citations increase the rigor and reproducibility of research by enabling data users to clearly document the source of the data they employ.

Funded by the European Union

In the **D4.5 Report on Completed FAIR Data Standard Adoption and Certifications of Data Repositories in the Region** (Alaterä et al., 2022), it is recommended to ensure resuse of digital objects by including to the metadata an explicit pointer to a licence or to use existing schemas that include licence terms.

Within AF08 Reuse, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

- "Documentation: Information describing any necessary information to use and understand the data. Documentation may be structured (e.g., a code book) or unstructured (e.g., a plain text "Readme" file)."
- "Contextualize: Use metadata to link the data set to related publications, dissertations, and/or projects that provide added context to how the data were generated and why."
- "Data Visualization: The presentation of pictorial and/or graphical representations of a data set used to identify patterns, detect errors, and/or demonstrate the extent of a data set to third party users."

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF08 Reuse is addressed in the following Agri-Food resources:

Caracciolo et al., 2020: The Agrisemantics group instructs repositories to support reuse by publishing semantic artefacts, adopting shared ontologies, and providing machine-actionable metadata.

Harper et al., 2018: AgBioData recommends that repositories design systems and metadata to enable ongoing data reuse across biological databases. "Proper data management is a critical aspect of research and publication… Data are the lifeblood of research, and their value often do not end with the original study, as they can be reused for further investigation if properly handled."

Sen et al., 2020: WheatIS recommends that repositories structure metadata and develop tools that permit reuse of global wheat datasets by a designated research community.

Top et al., 2022: AgroDataCube demonstrates repository-level reuse facilitation through standardized formats and web-accessible APIs that allow controlled, repeatable use of data. "This web page contains the information on how data can be retrieved from the repository, using standard web technologies and formats such as HTTP requests, REST and (Geo)JSON ('I', 'R')."

Biomedical Sciences

The TTRAM Activity/Function AF08 Reuse is addressed in the following Biomedical Sciences resources:

Funded by the European Union

Durinx et al., 2017: The article explicitly states that ELIXIR Core Data Resources must support the FAIR principles, including Reusability. The data must be well-described, richly annotated, and linked to relevant metadata and standards. The scientific quality indicator includes curation practices that enhance the usability of data. The article also includes community indicators, such as size and diversity of the user base, engagement with scientific communities, and responsiveness to user needs. However, The article does not mention secure remote access systems. It focuses on open access and interoperability.

Lin et al., 2024: In this article, repositories are expected to promote broad and responsible reuse of data and metadata, ensuring proper attribution and citation. To support this, they should adopt the FAIR principles—especially the "Reusable" aspect, which requires that data and metadata be well-described, licensed clearly, include detailed provenance, and conform to relevant community standards.

Sansone & Rocca-Serra, 2016: Mentioned in relation to the ISA framework (https://elixir-europe.org/intranet/forms/rir-selection/sid433)

Field et al., 2011: GSC highlights that maximizing usefulness depends on community-based standardization activities. This necessitates enriching public databases with their crucial biological context through the adoption of shared standards like MIxS upon submission and publication, supported by compliance and curation efforts, and stored within robust, coherent electronic systems.

Yilmaz et al., 2011: GSC developed MIxS for reuse purposes to ensure comprehensive contextual data is deposited at the time of submission, enhancing the quality, accessibility, and utility of sequence data.

Barrett et al., 2012: The article provides a strong foundation for facilitating reuse through metadata quality, interoperability, and community engagement.

Karsch-Mizrachi et al., 2025: Open access is a core tenet of INSDC and is highly aligned with the FAIR principles for data reuse. No restrictions are placed on the use of data made available through INSDC databases.

Wilkinson et al., 2016: in the article, it is emphasized that digital objects are made reusable by being richly described with accurate and relevant attributes, associated with detailed provenance, and released with clear usage licenses, while also meeting domain-relevant community standards.

Rehm et al., 2021: reuse is achieved through standardized data models (e.g., Phenopackets, Pedigree, VRS, VA), common file formats (e.g., SAM, BAM, CRAM, VCF/BCF), and APIs (e.g., htsget, TES, WES).

Climate Science

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) states that metadata and CVs [Controlled Vocabularies] are designed for reuse in future projects. This CV imposes consistency and comparability across descriptions of climate models

Funded by the European Union

and simulations, providing the crucial information necessary for end-users to interpret and compare results.

Linguistics

The Activity/Function (A|F) REUSE is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The CESSDA Data Access Policy (Bolton, 2022) states that all publicly funded data should be available to anyone regardless of status, nation, or type of use, unless restricted by agreed access conditions (Principle 2).

Kleemola et al. (2025, p. 23) recommend that data repositories publicly commit to community principles such as FAIR, TRUST and CARE on websites and research outputs to build credibility and confidence with stakeholders.

The CESSDA Data Management Expert Guide addresses the A|F Reuse in Chapter 2: Organise & Document. The guide provides instructions and examples on data management for data depositors to help ensure that digital objects are understandable and usable by organising and documenting data. The main practices covered include:

- Designing a clear data file structure
- Naming files consistently and organising data files in a well-structured and unambiguous folder structure
- Documenting data systematically to make data publishable, discoverable, citable and reusable according to common metadata standards
- Adapting a Data Management Plan (DMP)

(CESSDA Training Team, 2022, Chapter 2. Organise and Document)

The article promotes reuse by recommending proper citation practices and metadata standards that support reproducibility. It also recommends including licensing and reuse conditions in metadata. (Bornatici et al., 2025, pp. 7–8).

3.2.7. WORKFLOWS (AF09)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Workflows as follows: “Processes and activities for managing, changing and maintaining digital objects and their associated data and metadata in line with legal, ethical, policy, and other standards and operating procedures (see AF23 “Legal & Ethical” and AF14 “Policy & Standards Management”).” and suggests for transparent information “Standard operating procedures, workflow/business process diagrams and

Funded by the European Union

documentation related to how the repository approaches digital object management."(L'Hours et al., 2025c, p. 16)

The A|F Workflows is covered by R11 Workflows in **CoreTrustSeal** which requires the digital object management of repositories to take place "according to defined workflows from deposit to access" and elaborates on these workflows as follows (CoreTrustSeal Standards and Certification Board, 2022, p. 22):

> *Workflows may be specified in a mixture of standard operating procedures, business process descriptions and diagrams that guide normal practice and provide mechanisms for handling exceptions.*
>
> *The response statement and evidence should include references to the following items:*
>
> - *Workflows/business process descriptions covering the curation levels performed.*
> - *How workflows are adjusted for different types of data and metadata.*
> - *Decision handling within the workflows.*
> - *Change management of workflows.*
> - *Ability to track workflow execution, with mechanisms to handle exceptions.*

The functional model of the **OAIS Reference Model** is a generalized model of workflows relevant to performing digital preservation tasks in an OAIS. It can therefore be considered as a blueprint for specifying and documenting workflows in a given repository (CCSDS 2024, Chapter 4.2).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** (White House Office of Science and Technology Policy, 2022) state that the repository must support authentication of data submitters, has technical capabilities that facilitate associating submitter PIDs with those assigned to their deposited digital objects, such as datasets. It further states that the repository must provide broad, equitable, and maximally open access to datasets and their metadata free of charge in a timely manner after submission, consistent with legal and policy requirements related to maintaining privacy and confidentiality, Tribal and national data sovereignty, and protection of sensitive data.

The **Recommendations Consultation. EOSC-A Long Term Data Preservation Task Force** (Andreu et al., 2023) formulates FAIR-enabling practices to be undertaken by all data services should be defined. That FAIR-enabling practices undertaken by data services should be made transparent to users and funders to increase trust in services. Where responsibility is distributed, accountability should remain clear, including accountability for (meta)data loss or destruction. Digital object management outcomes, including preservation, should be integrated into a roles and responsibilities framework that integrates all actors and actions. The roles and responsibilities framework should be aligned with clear process models that meet the needs of different stakeholder communities. Different roles should use a living data management plan as a key artefact for periodic audit, review and revision.

Funded by the European Union

The **EU Annotated Grant Agreement** states repositories have mechanisms or provisions for expert curation and quality assurance for the accuracy and integrity of datasets and metadata, as well as procedures to liaise with depositors where issues are detected (European Commission, 2024).

The **D1.3 Recommendations for a FAIR EOSC - White Paper of the FAIR-IMPACT Synchronisation Force** (Grootveld, 2025) states that processes for managing, changing, and maintaining digital objects in line with legal, ethical, and policy standards, should be made more transparent in repositories to increase their overall trustworthiness. Furthermore, to develop semantic interoperability, transparent workflows and governance mechanisms are necessary to support the alignment and mapping of Semantic Artefacts (SAs).

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) approaches workflows primarily through its requirement for compliance with community standards across all stages, ensuring that the workflows producing and managing the data/metadata adhere to agreed-upon frameworks. Furthermore, the model requires transparency of information about the provenance and history of the data/metadata.

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), ten out of 14 respondents provided legal and ethical information, eight had metadata and data versioning policies in place, and seven provided descriptions of service processes or workflows.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** report (Kleemola et al., 2022) describes how repositories documented and improved their workflows during the certification process, including ingest, curation, and preservation workflows.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) itself is not a workflow description but acts as a tool to normalize assessment, which can be integrated into institutional workflows by specifying the expected level of FAIRness that resources must achieve within Research Data Management Plans (DMPs) before data is produced. Furthermore, the FDMM includes indicators that implicitly require established data management processes, such as the Essential indicator RDA-A2-01M, which verifies that metadata is guaranteed to remain available even after data is gone, an assurance based on the repository's documented lifecycle information.

The **Core Preservation Process CPP-013 Object Management Reporting** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA delivers reporting to enable the effective management of Objects. This must include a wide range of functional, operational, and statistical reports and analytics.” (EOSC EDEN T1.2 et al., 2025).

Within AF09 Workflows, the **Data Curation Network** defines the following Curation Activity (Johnston et al., 2016): “Curation Log: A written record of any changes made to the data during the curation process and by whom. File is often preserved as part of the overall record.”

Funded by the European Union

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF09 Workflows was not found to be addressed in the gathered Agri-Food resources.

Biomedical Sciences

The TTRAM Activity/Function AF09 Workflows is addressed in the following Biomedical Sciences resources:

Yilmaz et al., 2011: MIxS standard defines workflows and activities for the consistent acquisition, validation, and maintenance of metadata during submission to public databases.

Barrett et al., 2012: The article outlines submission and metadata workflows that support structured data management and user interaction. However, it does not fully address the broader scope of repository-level workflows related to legal, ethical, and policy compliance

Karsch-Mizrachi et al., 2025: The article outlines the processes and governance structures that manage the lifecycle of nucleotide sequence data within INSDC. These workflows include submission, validation, metadata enrichment, and public dissemination, all of which are guided by formalized policies and standards.

Climate Science

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC manages digital objects and their associated data and metadata through a comprehensive workflow that includes initial contact, submission agreement, data preparation, metadata submission, data submission, quality control, DOI/PID assignment, and continuous curation.

The **NetCDF Climate and Forecast (CF) Metadata Conventions** (Eaton et al., 2024) state that the CF framework includes an audit trail for modifications through the "history" attribute and a strong commitment to backward compatibility for sustained data utility and long-term maintenance.

Linguistics

The Activity/Function (A|F) WORKFLOWS is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The CESSDA Data Management Expert Guide and Data Archiving Guide give a broad overview of the workflows that need to be carried out and documented throughout the lifespan of research data (CESSDA Training Team, 2022, 2025a).

3.2.8. PRESERVATION (AF10)

Funded by the European Union

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Preservation as follows: “Monitoring both the technology landscape (see AF29 "Technical Infrastructure") and the user community (see AF20 "External Engagement") for any changes that could impact the usability or understanding (see AF08 "Reuse") of digital objects. When necessary, preservation actions are taken—such as migrating data formats, updating metadata (e.g., revising ontologies), or emulating the environment in which the digital object is used. These actions help ensure the long-term viability and continued accessibility of the digital object.” and suggests for transparent information “[p]reservation policies, strategies, plans and procedures that define how long artefacts will be preserved for, how potential preservation actions (emulation, file format updates, updates to semantic artefacts such as controlled vocabularies and ontologies) are approved and implemented. Preservation interacts with re-appraisal criteria set during Deposit & Appraisal, the changing Technical Infrastructure Landscape over time, and how the ReUse needs of the community are monitored.” (L’Hours et al., 2025c, pp. 16–17) The FIDELIS TTRAM also notes that “[s]pecialist repositories need specific skills in data formats, research methodologies, metadata, and ontologies related to disciplines or content types” and that “[u]ser characteristics influence community watch, object characteristics inform technology watch.” (L’Hours et al., 2025c, p. 17)

The proposed CoreTrustSeal Levels of Retention, Curation and Preservation (**LoRCaP**) define requirements for Active Preservation as follows: “In addition to [D. Deposit Compliance and/or C. Initial Curation] 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” (L’Hours et al., 2024, p. 4). Metadata required on object- and repository-level to demonstrate how the LoRCaP are implemented are proposed in L’Hours et al., 2024.

In **CoreTrustSeal**, the A|F Preservation is dealt with in requirement 9 (RO9), called Preservation plan, which requires repositories to “assume[...] responsibility for long-term preservation and manage[...] this function in a planned and documented way” and lists the following issues to be addressed by the repository (CoreTrustSeal Standards and Certification Board, 2022, pp. 19–20):

- The documented approach to preservation, including whether this involves format migration, emulation, etc. Ensuring bit level integrity is vital but not sufficient for preservation Ensuring bit level integrity is vital but not sufficient for preservation.
- File formats and metadata schemas for long term preservation.
- How the level of responsibility for the preservation of each item is defined.
- Plans related to future migrations or similar measures to address the threat of obsolescence.
- Actions relevant to preservation specified in documentation, including custody transfer, submission information criteria, and preservation information metadata.

Funded by the European Union

- Measures to ensure these actions are taken.
- Any minimum stated retention and/or preservation periods.
- How often the digital objects are re-appraised and the possible outcomes of reappraisal.
- The repository approach to deleting/removing data and metadata from collection/holdings including the impact on persistent identifiers.

In the **nestor Seal** criteria "C11 Preservation measures", "C13 Significant properties" and "C24 Interpretability of the archival information" are relevant to this A|F. C11 states that the archive "should conduct strategic planning as a means of preserving the digital objects entrusted to it [...]. Long-term planning should be based on the monitoring of legal and social changes, the demands and expectations of the designated communities and all technical changes relevant for the sustained preservation and appropriate use of the information objects in the form of their representations". C13 focuses on the properties "significant for preservation of the information objects" and C24 addresses "technical preservation measures [...] undertaken to ensure the interpretability of the archival information packages" (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** core preservation activities are performed by the Preservation Planning functional entity and the Administration functional entity. Preservation Planning (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4.2.3.7) includes a "Preservation Watch" function which compiles all information necessary to make informed preservation decisions. It receives input from a number of internal and external sources, including the "Monitor Designated Community" and the "Monitor Technology" functions. The former ensures that the archive continues to understand the needs and skills of the users who are the primary target group of the preserved digital objects while the latter guarantees that the archive has up to date information about developments in the technological landscape which may affect the renderability of the digital objects it holds. In addition, Preservation Planning is tasked with developing preservation strategies and standards, performing risk assessments, and developing preservation plans among other things. Administration in turn is tasked with actually performing any updates to be made to the archived objects (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4.2.3.6). Chapter 5 of the OAIS Reference Model addresses "Preservation Perspectives", namely, "various practices that have been, or might be, used to preserve digital information and to preserve access services to digital information" (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 5/1), including different types of migration.

The **Desirable Characteristics of Data Repositories for Federally Funded Research** (White House Office of Science and Technology Policy, 2022) address long-term organizational sustainability when the repository has a plan for long-term management of data, including maintaining integrity, authenticity, and availability of datasets; has contingency plans to ensure data are available and maintained during and after unforeseen events.

The report **FAIR Forever? Long Term Data Preservation Roles and Responsibilities** (Currie & Kilbride, 2021) states that research repositories should urgently prioritise and adapt workplans to include quality improvement mechanisms where these do not already exist, including DPC Rapid Assessment

Funded by the European Union

Model, establishing thereby a strategic framework to achieve baseline certification for primary preservation services, or identifying preservation pathways for data. For Research Repositories there is a medium priority, identifying costs of action versus inaction with respect to high value, critically endangered content.

The **Recommendations Consultation. EOSC-A Long Term Data Preservation Task Force** (Andreu et al., 2023) formulates that effective storage, including multi-copy redundancy and integrity measures, is necessary but not sufficient for preservation. Minimum criteria for acceptable storage practices in different scenarios should be defined as a foundation for all levels of retention, curation and preservation services. Different types of data services benefit from being transparent on their current level of storage, curation and preservation practice, as this increases trust by the user and funders alike. Data services, including repositories should specify all the levels of care they apply to objects within their collection, including through repository and digital object registry metadata. Digital objects should include metadata that specify their level of care and the timeframes or criteria for reappraisal of the level of care. Unique preservation functions and activities should be defined alongside functions and activities that apply to all (meta)data services. Preservation roles must include monitoring the changing needs of communities at the point of reuse. This community watch must be aware of the knowledge base, methodologies and technologies of the user communities. Preservation roles must include monitoring the changing nature of available technologies for the deposit, storage, curation, discovery, access and reuse of data and metadata. This technology watch must continue to meet the needs identified through community watch and be proactive as well as reactive. Managers must integrate preservation planning into operational management including staffing, funding, service development and procurement.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. They further facilitate mid- and long-term preservation of the deposited material (European Commission, 2024).

The **Update of the Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements** states that the lack of a public policy for preservation, curation and security of the contents is the most frequent reason, followed by not adhering to a specific metadata standard (Lazzeri, 2024).

The **Practical Guide to the International Alignment of Research Data Management** (Science Europe, 2021) states that:

1. Preservation:
    a. Ensure persistence of metadata and data
    b. Be transparent about mission, scope, preservation policies, and plans (including governance, financial sustainability, retention period, and continuity plan).

Funded by the European Union

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) states on data preservation policy that details how the preservation of the data is ensured. Does the repository provide information on its data preservation policies?

The **Research Data College Working Group** recommends "prioritising maintained repositories with a data-preservation period of at least five years, following the practices put in place by Recherche Data Gouv" (Lamotte et al., 2024, p. 10).

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), respondents were asked whether their organisation employed any of the common digital preservation strategies e.g., format and schema migration and/or emulation. All three of the non-repository service providers selected 'Can't say or N/A' which was expected, as they do not provide long-term digital preservation. Three of the repositories selected 'Can't say or N/A'; seven of the repositories selected 'format and schema migration' and one repository selected 'both'.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.5, Preservation Objective/Designated Community Needs, that although all disciplines indicate that expert curators are utilized across all disciplines, variability in the level of curation requires that a standard level of acceptable curation that enables long-term digital preservation (LTDP) and digital object quality (DOQ) must be promoted. Across all disciplines, standardized and comprehensive strategic reappraisal and reassessment schedules and guidelines must be established and implemented. The existing tendency to follow long-standing practices or the simple statement that data will be preserved "forever" is inadequate and comprehensive strategies must be put in place. Regular, standard reappraisals and reassessment checks, that include methodologies for enhancement (e.g. file format, metadata enrichment), are required to ensure LDP and DOQ for both historical and legacy data, as well as relevant standards. The relevance of older data and standards is not sufficiently assessed across many disciplines, but can be highly relevant for state-of-the-art science that includes, for example, field recordings that capture linguistic variables, sampling of targeted organisms, or time-series climate data. Older data can be vital to understand temporal changes in many disciplines, and the rapid evolution of technology presents new challenges for both evaluating and supporting historical and legacy data plus standards. Concerning identifying needs in the designated community (community watch), developing a targeted, inclusive, and systematic approach to community engagement should be considered within the disciplines, preferably as a joint effort between repositories to reduce duplication of work.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) recommend that repositories express the levels of care offered and received by digital objects, detailing what levels of retention, curation, and preservation are in place and how these might change over time. Additionally, the guidelines conclude that digital objects not undergoing active preservation to address evolving technology (e.g., file format migration) or community changes (e.g., ontology updates) risk a deteriorating FAIR score over time.

Funded by the European Union

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that long-term preservation is a core requirement of TDRs. The report discusses OAIS-based responsibilities and the need for documented preservation policies. Some repositories lacked explicit preservation missions, which limited their eligibility for certification.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) emphasize persistence and preservation of metadata, notably through the Essential indicator RDA-A2-01M, which mandates that metadata is guaranteed to remain available even after the data itself is no longer accessible, a requirement evaluated based on the organization's documented life cycle information. Furthermore, the FDMM's focus on Active preservation (A) within the Levels of Retention, Curation and Preservation (LoRCAP) variable directly supports this function by confirming that active preservation is defined as the long-term responsibility for ensuring that data and metadata can be understood and rendered for reuse by the designated community.

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) states that Preservation requires repositories to establish a digital preservation plan that outlines the duration of resource management, identifies responsible roles, and documents procedures for preserving different resource formats to ensure long-term usability. To enable necessary preservation actions, repositories must secure the rights to copy, transform, and store items from the depositor, and maintain a business continuity plan detailing procedures for managing disruptions like cyber-attacks or natural disasters.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) require a standardized approach to tracking data usage, mandating the consistent use of DataCite metadata fields to store citation relationships. Furthermore, the recommendations ensure future access to this contextual information by requiring repositories to expose the provenance (source) of every asserted citation and by encouraging contribution to centralized resources like the Data Citation Corpus.

The **Core Preservation Process CPP-012 Risk Mitigation** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA enables the design, development and management of plans for mitigating identified preservation risks." (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-014 File Migration** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA supports batch modifications of previously ingested Files to prevent a preservation- or access-related risk." (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-015 Emulation and Rendering Tools** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA enables the rendering of Objects via the application of emulation and/or other specialist tools." (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-017 Disposal** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA enables the managed disposal of

Funded by the European Union

Information Packages and permits the retention and maintenance of Metadata even when the content of the Information Package has been removed from the TDA." (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-022 Significant Properties Definition** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA defines significant properties for sets of Information Objects (i.e., properties that it commits to preserve over the long term through preservation actions and rendition)." (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-023 Risk Definition and Extraction** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA monitors technology evolution, identifies risks and defines detection methods for these risks." (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-026 File Format Normalisation** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA performs operations on the data Objects prior to ingest in order to comply with its format requirements." (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-028 Creation of Derivatives** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA generates derivative copies to address the specific needs of its Designated Community." (EOSC EDEN T1.2 et al., 2025).

Within AF10 Preservation, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

- "File Format Transformations: Transform files into open, non-proprietary file formats that broaden the potential for long-term reuse and ensure that additional preservation actions might be taken in the future. Note: Retention of the original file formats may be necessary if data transfer is not perfect."
- "File Inventory or Manifest: The data files are inspected periodically and the number, file types (extensions), and file sizes of the data are understood and documented. Any missing, duplicate, or corrupt (e.g., unable to open) files are discovered."
- "Software Registry: Maintain copies of modern and obsolete versions of software (and any relevant code libraries) so that data may be opened/used overtime."
- "Transcoding: With audio and video files, detect technical metadata (min resolution, audio/video codec) and encode files in ways that optimize reuse and long-term preservation actions. (E.g, Convert QuickTime files to MPEG4)."
- "Emulation: Provide legacy system configurations in modern equipment in order to ensure long-term usability of data. (E.g., arcade games emulated on modern web-browsers."
- "Migration: Monitor and anticipate file format obsolescence and, as needed, transform obsolete file formats to new formats as standards and use dictate."

Domain-specific resources

Agri-food

Funded by the European Union

The TTRAM Activity/Function AF10 Preservation was not found to be addressed directly into the gathered Agri-Food resources. However, the following statements can be mentioned:

Harper et al., 2018: AgBioData recommends that repositories develop long-term preservation and sustainability plans, including metadata and software curation, to prevent data loss. “Long-term sustainability and maintenance of databases is critical to ensure continued access to data and tools; mechanisms must be put in place to guarantee that curated data and metadata remain available to the community.” ; and also positions metadata as crucial to “long-term management of persistence and relevance, adaptation to changes in technologies, support of old data formats or conversion into new and continued integration with new data.”

Caracciolo et al., 2020: The Agrisemantics group instructs repositories to maintain and update semantic artefacts and controlled vocabularies to preserve the interpretability of data. “Providing explicit and machine-readable description of data makes it possible to programmatically integrate and reuse data; keeping semantic resources updated ensures long-term interpretability.”

Biomedical Sciences

The TTRAM Activity/Function AF10 Preservation is addressed in the following Biomedical Sciences resources:

Lin et al., 2024: In the article, Preservation, defined as *“the extent to which the data repositories invest resources in archiving data for long-term use, including adapting to evolving user needs, changes in storage technology, and changes in media formats”* is considered as one of the six major characteristics of the different types of data repositories. Besides, preservation is considered a mandatory responsibility as long as the project or institutional repository is relevant to its user base and mission. Therefore, the article encourages the adoption of principles such as TRUST and more specifically the Transparency and Sustainability practices.

Field et al., 2011: GSC proactively monitors the evolving technology landscape, such as the democratization of sequencing access, and sociological changes within the scientific community. This vigilance ensures the continuous development and harmonization of shared standards.

Karsch-Mizrachi et al., 2025: The Membership Arrangement described in the article states that INSDC Member institutions are independent government or non-profit organizations that manage nucleotide sequence databases that capture, preserve, and present comprehensive nucleotide sequence information and annotations to preserve the scientific record and enable broad sharing of such data. Additionally, INSDC has identified several areas of data management that are currently undergoing evolution or emerging as areas in which new development is needed to meet the needs of data submitters and users or to ensure the sustainability of sequence repositories.

Funded by the European Union

Wilkinson et al., 2016: Preservation is achieved through the Reusability (R) principle, which mandates that digital objects are richly described with accurate attributes, associated with detailed provenance, released with clear usage licenses, and meet domain-relevant community standards.

Rehm et al., 2021: in the article, recommendations for data repositories are made: set standards for preservation and other repository operations, contribute to defining TDRs and building consensus on curation and preservation levels. Leverage domain-specific expertise in long-term data preservation: Advocate for domain-specific services, such as data curation and metadata creation, and the benefits of long-term data preservation for future research and society.

Climate Science

The METAFOR project: preserving data through metadata standards for climate models and simulations (Callaghan et al., 2010) defines a core goal for the digital preservation of climate models.

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) describes the ensured preservation by recognizing that the core scientific standard, the Controlled Vocabulary (CV), must evolve, improve, and be reused in future projects, thus requiring its long-term preservation. This long-term viability is addressed by planning to establish an international governance committee under IS-ENES2 to manage the evolution and preservation of the CV. Additionally, the system planned future work, such as developing a tool to convert CV XML back to mindmap format, to ensure the CV XML remains the primary preservation and governance artifact.

The **NetCDF Climate and Forecast (CF) Metadata Conventions** (Eaton et al., 2024) state that to preserve usability and understanding regardless of the technology landscape, the CF data model is designed to be independent of specific encoding formats, providing a framework that anticipates potential changes in underlying data storage technologies like NetCDF.

Linguistics

The Activity/Function (A|F) Preservation is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

Kleemola et al. (2025, p. 23) recommends that data repositories should:

- set standards for preservation and other repository operations by for example contributing to defining TDRs and by building consensus on curation and preservation levels.
- leverage domain-specific expertise in long-term data preservation, advocating for domain-specific services such as metadata creation and data curation and advocating the benefits of long-term data preservation for future research and society.

FAIR principles improve data accessibility, but without sustained access to data, they don't ensure long-term preservation. Researchers must trust that archives will preserve their data and provide

Funded by the European Union

stable access. To earn this trust, data archives need to be able to convince researchers that 1) they will receive support, data curation, and preservation to a satisfactory degree, and 2) that the repository services will be stable and consistent over time. (CESSDA Training Team, 2025a, Chapter 5.4 Trustworthy data archives).

The CESSDA Data Archiving Guide Chapter 2.2 describes preservation as “the heart of a repository” and presents that a digital preservation policy is part of the preservation environment. Its purpose is to define the rules, the responsibilities, the roles and monitoring systems for managing data within a data archive. The policies include information about:

- an organizational framework (e.g. mission, acquisition, access and use of data)
- address challenges such as legal issues and technological change
- establish principles like guidelines for data preservation, cooperation, transparency, and trustworthiness
- define the roles and responsibilities required for data preservation

(CESSDA Training Team, 2025a, Chapter 2.2 Preservation Policies).

3.2.9. PROVENANCE & AUTHENTICITY (AF11)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Provenance & Authenticity is described as follows: “Providing information about digital objects that details their history (provenance) and ensures their authenticity. Provenance and authenticity information can be captured at the point of deposit, generated during curation and preservation processes, and shared with data or metadata users.” and suggests for transparent information “Information about the creation, creators and history of the object requested at deposit. How the repository records information about changes to digital objects and what information is communicated externally (e.g. to end users).” (L’Hours et al., 2025c, p. 17)

This A|F is addressed by **CoreTrustSeal** requirement R07, with the (near) identical name, Provenance and authenticity, which states that “[t]he repository guarantees the authenticity of the digital objects and provides provenance information”, which means that “[a]ny intentional changes to data and metadata should be documented, including the rationale and originator of the change”, and that [a]uthenticity covers reliability and provenance, including the relationship between the deposited digital objects and those provided at the point of access” (CoreTrustSeal Standards and Certification Board, 2022, pp. 17–18). The response statement from the repository should address the following items (CoreTrustSeal Standards and Certification Board, 2022, p. 18):

- The repository approach to changing and versioning data and metadata. How the approach and records of changes are communicated to data depositors and users.
- The provenance information and audit trails recorded for data and metadata processing and versioning.

Funded by the European Union

- How the repository compares the essential properties of different versions of the same file.
- Identification checks for depositors.

In the **nestor Seal**, criteria C17-C19, dealing with Authenticity in the context of Ingest (C17), Preservation measures (C18), and Use (C19) are relevant to this A|F. In addition "C31 Logging the preservation measures" and "C32 Administrative metadata" are of interest here (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** one of the mandatory responsibilities of an OAIS is to provide access to the preserved digital objects "as copies of, or as traceable to, the original submitted Content Information with evidence supporting its Authenticity" (p. 3/1). An important means to fulfilling this responsibility is provenance information, which is part of the Preservation Description Information. It "documents the history of the Content Data Object. This tells the origin or source of the Content Data Object, its Information Properties to be preserved (Transformational Information Properties), any changes that may have taken place since it was originated, and who has had custody of it since it was originated, providing an audit trail for the Content Data Object" (p. 4/35). This means that all OAIS functions interacting with a given digital object are required to log these interactions.

The **Desirable Characteristics of Data Repositories for Federally Funded Research** (White House Office of Science and Technology Policy, 2022) addresses provenance when the repository has mechanisms in place to record the origin, chain of custody, version control, and any other modifications to submitted datasets and metadata.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. They ensure that contents are accompanied by metadata sufficiently detailed and of sufficiently high quality to enable discovery, reuse and citation and contain information about provenance and licensing. Their metadata is machine-actionable and standardized (e.g. Dublin Core, Data Cite, etc) preferably using common non-proprietary formats and following the standards of the respective community the repository serves, where applicable (European Commission, 2024).

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) states on data versioning that mechanism and process to make and track edits to a dataset after deposition. Does the repository enable modifications to published data (e.g., to correct it or append additional information)? Is there a process to distinguish, link and access all public versions of the data?

The **O'FAIRe makes you an offer: Metadata-based Automatic FAIRness Assessment for Ontologies and Semantic Resources** (Amdouni et al., 2022) methodology addresses Provenance and Authenticity through the Reusability (R) sub-principle R1.2, finding that semantic resources generally struggle to meet detailed provenance requirements. Specifically, the analysis showed that while many ontologies provide information about the actors involved (77%) and general provenance (87%),

Funded by the European Union

only a small number document accrual methods and policies (4%) or describe the ontology rationale using information like competency questions (6%).

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) addresses provenance and authenticity through the Important indicator RDA-R1.2-01M, which requires that metadata includes provenance information according to community-specific standards. Additionally, providing provenance information using a more general, cross-community language is recognized as a useful indicator (RDA-R1.2-02M).

In the **FAIRsFAIR project D2.3** report (Behnke et al., 2020), it is recommended to gather provenance metadata on digital objects and files upon upload.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.4, Trustworthy Digital Archive Capabilities, that versioning needs better support towards appropriate logging within a repository, and in particular the transfer of version histories and logs when dataset packages are transferred to other repositories. This is reflected in the level of versioning by infrastructures and inconsistent knowledge by researchers concerning discipline specific practices. Versioning of data, metadata and the service framework of the repositories are the most common solutions encountered.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) recommend that repositories expose the history of FAIR assessment outcomes for an object over time to provide insight into improvements in FAIRness. Furthermore, to ensure authenticity and provenance of assessment results, the guidelines advise using the Data Quality Vocabulary (DQV) to embed results—including crucial details like the test date, metric used, and name and software version of the testing tool—within the metadata of assessed datasets.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that repositories must ensure data authenticity and provenance. Certification requires evidence of procedures that maintain data integrity and traceability, including version control and audit trails.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) directly address provenance under the Reusable (R) principle, defining the indicator RDA-R1.2-01M as Important, which requires metadata to include provenance information according to community-specific standards. Additionally, the FDMM includes the Useful indicator RDA-R1.2-02M, which requires that metadata provides provenance information using a cross-community language (such as PROV-O).

In the **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022), Provenance and Authenticity are addressed in the COAR Community Framework by requiring the repository to collect basic preservation metadata, including provenance, date of upload, and file format. To ensure authenticity, the repository must also record the checksum when a resource is submitted or modified.

Funded by the European Union

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) conclude that repositories should increase transparency and trust in data citation information by indicating the provenance (source) for every asserted data citation listed on the dataset landing page. Furthermore, repositories are advised to document, as part of their information pages for users and contributors, the mechanisms and external sources employed to collect and store data citations.

In the **D4.5 Report on Completed FAIR Data Standard Adoption and Certifications of Data Repositories in the Region** (Alaterä et al., 2022), repositories are required to ensure that metadata is structured and describes provenance, for example using RDF or embedded JSON. Furthermore, metadata should include an explicit reference to a schema that provides machine-readable provenance information.

The **Core Preservation Process CPP-021 AIP Versioning** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA creates successive versions of the Information Objects on its own or on the producer initiative." (EOSC EDEN T1.2 et al., 2025).

Within AF11 Provenance and Authenticity, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

- "Authentication: The process of confirming the identity of a person, generally the depositor, who is contributing data to the data repository. (e.g., password authentication or authorization via digital signature). Used for tracking provenance of the data files."
- "Chain of custody: Intentional recording of provenance metadata of the files (e.g., metadata about who created the file, when it was last edited, etc.) in order to preserve file authenticity when data are transferred to third-parties."
- "Contact Information: Keep up-to-date contact information for the data authors and/or the contact persons in order to facilitate connection with third-party users. Often involves managing ephemeral information that will change over time."
- "Versioning: Provide mechanisms to ingest new versions of the data overtime that includes metadata describing the version history and any changes made for each version."

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF11 Provenance and Authenticity is addressed in the following Agri-Food resources:

Caracciolo et al., 2020: The Agrisemantics group directs repositories to use metadata, semantic frameworks and persistent identifiers to record provenance and ensure traceability of digital objects.

Sen et al., 2020: WheatIS highlights repository responsibilities in documenting metadata about the creation and origin of wheat datasets, ensuring provenance in distributed systems.

Funded by the European Union

Drakos et al., 2015: agINFRA instructs repositories to guarantee provenance by harmonizing metadata across data sources and ensuring consistent identification of datasets.

Top et al., 2022: AgroDataCube demonstrates repository-level provenance through persistent identifiers and catalog entries that document data sources and retrieval history. "A DOI (needed for 'F' and 'A' of FAIR) is available, which refers to a digital catalogue entry at the Wageningen University & Research library and directs the reader to the associated web page of AgroDataCube."

Biomedical Sciences

The TTRAM Activity/Function AF11 Provenance and Authenticity is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: An important aspect of service quality for ELIXIR Core Data Resources is the provision of links to documentation of provenance. This involves ensuring that the resource connects its data to relevant scientific literature, thereby establishing the origin and biological context of the facts it presents.

Lin et al., 2024: In the article, provenance, defined as a way to *"record the origin, chain of custody, and modifications to data/metadata"* is considered as desirable characteristics for data repositories. Therefore, the article encourages the adoption of principles such as FAIR and more specifically Reusability.

Barrett et al., 2012: The article supports basic provenance and authenticity through accessioning, submitter attribution, and object linking. However, it lacks formal mechanisms for tracking data history, verifying authenticity, or presenting detailed provenance metadata to users.

Karsch-Mizrachi et al., 2025: The article discusses the inclusion of rich contextual metadata, such as sample type, collection date and geolocation, which is now mandatory for new submissions. This metadata contributes to the provenance of the data by documenting when and where the sample was collected and under what conditions. Furthermore, the INSDC expresses a commitment to improving data attribution, recognizing the contributions of those involved in sample acquisition, sequencing, and submission. These practices collectively ensure that the authenticity and provenance of digital objects are preserved and made available to users, supporting transparency and trust in the repository's contents.

Wilkinson et al., 2016: Provenance and Authenticity is directly addressed by making detailed provenance (R1.2) a core component of the Reusability (R) principle. This underscores that scholarly digital objects are associated with their history and accurate attributes.

Climate Science

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) describes the act of ensuring authenticity, the system employs a Conformance concept to track how a simulation adheres to its experiment requirements, and utilizes a validation

Funded by the European Union

pipeline, including Schematron-based validation, to guarantee internal coherency between the many descriptive parameters within the harvested metadata.

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that information regarding provenance is managed through versioning in WDCC.

The **NetCDF Climate and Forecast (CF) Metadata Conventions** (Eaton et al., 2024) state that provenance is achieved through metadata attributes such as "history", which offers an audit trail for modifications to the original data, and "institution", "source", and "references", which detail the origin and production methods, thereby directly supporting the provenance and authenticity of the digital objects. Additionally, metadata for data quantization aims to provide provenance information so users can reproduce data transformations and understand how the data might differ from its original unquantized state.

Linguistics

The Activity/Function (A|F) Provenance and Authenticity is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The CESSDA Data Archiving Guide covers the A|F Provenance & Authenticity in Chapter 4.5 Updates and versioning. The guide strongly advises repositories to track the provenance and history of all files they receive and store. It is crucial to identify the current version of the data, and the guide presents structured versioning practices - such as changelog folders and version- or date-based file naming - to ensure that data is clear and reproducible. (CESSDA Training Team, 2025a, Chapter 4.5 Updates and versioning)

Repositories are recommended to include data versioning and changelogs, and ensure access to metadata about all published versions of datasets, which contribute to documenting the history and authenticity of data. (Bornatici et al., 2025, p. 7).

3.2.10. SUPPORT (AF12)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Support is described as follows: “Providing guidance and responding to requests from depositors and users around deposit and appraisal, discovery, access and re-use of data and metadata.”, suggests for transparent information “[l]ink to primary location where the repository provides supporting information or offers supporting services to depositors, users and others.” and notes that “[s]pecialist repositories need specific skills in data formats, research methodologies, metadata, and ontologies related to disciplines or content types, to deliver relevant support.” (L’Hours et al., 2025c, pp. 17–18)

Funded by the European Union

The A|F Support is not explicitly addressed by **CoreTrustSeal**, but has affinities with their requirement R06 Expertise and Guidance, which is discussed in section PEOPLE & EXPERTISE (AF17) below.

In the **OAIS Reference Model**, interaction with producers is managed by the Administration functional entity ("Customer Service", see Chapter 4, p. 14). The function "Coordinate Access Activities" in the Access functional entity "provides assistance to OAIS Consumers including providing status of orders and other Consumer support activities in response to an assistance request via the Deliver Response function." (Chapter 4, p. 19). Beyond defining these functions, the OAIS Reference Model does not state any requirements for how such support functions should be implemented.

PAIMAS provides detailed suggestions as to which information should be exchanged and which points clarified between the producer and the archive before Ingest (Consultative Committee for Space Data Systems (CCSDS), 2004, Chapter 4).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** (White House Office of Science and Technology Policy, 2022) address the topic of clear use guidance when the repository ensures datasets are accompanied by documentation describing terms of dataset access and use (e.g., reuse licenses and need for approval by a data use committee). Furthermore, broad and measured reuse is when the repository ensures datasets are accompanied by metadata that describe terms of reuse and provides the ability to measure attribution, citation, and reuse of data (e.g., through assignment of adequate and openly accessible metadata and unique PIDs).

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) states on user support that provides support to users during or after submission. Does the repository have a contact point (e.g. helpdesk email or contact form) to assist data depositors and data users?

In the **FAIRsFAIR project D2.3** report (Behnke et al., 2020), a strong emphasis is made on the APIs for which active support must be provided to users as documentation or training taught by the repository's staff. Besides, technical support for predefined file formats must be provided, as well.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.3, Metadata Quality, that direct communication between the repository and the depositor needs to be available, with increased communication about the implementation of, clarification/explanation of, and compliance with metadata standards. The communication or explanation needs can be expressed as both monitory, which includes repositories having one-on-one contact with depositors, and automated, with respect to improving workflows, standardizing metadata annotation and facilitating completeness. 4.3.5 Preservation Objective/Designated Community Needs: Researchers wish to deposit data efficiently and in line with good practice, and having access to an active support desk and/or a contact point inside the repository, where someone takes the time to answer, is strongly wanted. Interaction between repository personnel and depositor cannot be neglected, even when elaborated repository user guides exist. For the researcher, who often deposits data at a very irregular frequency, having someone with trustworthy digital repositories competencies who can give advice and answer

Funded by the European Union

questions, is highlighted as very important for trust and user experience. For the repository, this human contact point aids monitoring the needs in the community. Common repository personnel-directed FAQs within disciplines may improve and facilitate live guidance, as well as reducing time spent formulating advice and answers.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) propose incorporating guidance directly into Data Management Planning (DMP) tools to help users make informed choices about data deposit concerning access, storage, curation, and preservation from the earliest stage of their research. Furthermore, transparency about the levels of care offered by repositories (retention, curation, and preservation) supports mutual trust, which is a critical precursor to trusted relationships between actors such as object depositors and users.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) provides one-on-one trust support work, webinars, and documentation templates. The repositories involved found the support process useful. Peer support and expert feedback were crucial for helping repositories navigate certification. Managing trust in the future entails peer collaboration through support programmes and networks, while ensuring solid resources for the upkeep of these collaborative efforts. Future endeavours to manage trust should make use of the existing and planned networks of trustworthy repositories that can share both expertise and responsibility, while recognising the need for more enduring sources of funding for managing trust sustainably. Some individual ERICs, such as CESSDA and CLARIN, provide targeted support for their members on issues related to trust and seeking compliance

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) conclude that repositories should provide guidance and information to users and contributors regarding the mechanisms used to manage data citations. Specifically, this guidance should be documented as part of the information pages for contributors and users, detailing how the repository collects and stores data citations, including the external sources it employs.

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF12 Support is addressed in the following Agri-Food resource:

Harper et al., 2018: AgBioData explicitly recommends that repositories provide personal and responsive support to users and contributors to build trust and maintain communication. (See AF01 : Identification & Contact).

Biomedical Sciences

The TTRAM Activity/Function AF12 Support is addressed in the following Biomedical Sciences resources:

Funded by the European Union

Durinx et al., 2017: A key component of quality of service for ELIXIR Core Data Resources is the provision of effective customer support and user engagement mechanisms. This includes operating a helpdesk to assist users with queries and technical issues, thereby ensuring accessibility and responsiveness. In addition, resources are expected to actively seek and incorporate user feedback into service design and improvement processes, fostering a user-driven approach to development. Finally, the provision of training activities—such as workshops, tutorials, or documentation—demonstrates a commitment to empowering users and enhancing their ability to effectively engage with the resource.

Lin et al., 2024: In the article, User diversity is defined as a way to "accommodate a diverse audience, offering resources and support for users across multiple disciplines and skill levels" is considered as one of the six major characteristics of the different types of data repositories. Therefore, the article encourages the adoption of principles such as TRUST and more specifically the User focus practices.

Field et al., 2011: GSC provides support for digital object management through its compliance working group to help depositors adhere to the MIxS standard. Additionally, its developer's working group provides technical assistance for implementing GSC standards in software and database projects.

Barrett et al., 2012: Minimal support - BioProject and BioSample [...] submissions are supported by a web-based Submission Portal that guides users through a series of forms for input of rich metadata describing their projects and samples.

Wilkinson et al., 2016: The FAIR principles provides intrinsic guidance and facilitates autonomous discovery, access, and reuse for both human and computational agents.

Rehm et al., 2021: GA4GH provides guidance for data creation, deposit, discovery, access, and reuse across the genomic data lifecycle. Additionally, user support is achieved via the GA4GH Starter Kit and through community engagement initiatives such as the Genomics in Health Implementation Forum (GHIF) to foster understanding and share best practices among depositors and users.

Karsch-Mizrachi et al., 2025: The article does not describe a formal helpdesk or user support system. However, it demonstrates that INSDC provides structured guidance and documentation to facilitate deposit, appraisal, discovery, access, and reuse of data and metadata.

Climate Science

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC offers comprehensive support and guidance to depositors for the entire data publication process, encompassing initial data appraisal, preparation (e.g., format, structure), metadata submission, quality control, and assignment of persistent identifiers (DOIs/PIDs).

Linguistics

The Activity/Function (A|F) Support is not addressed by any of the reviewed resources from Linguistics.

Funded by the European Union

Social Sciences

When promoting data sharing, archives should highlight benefits such as:

- merits for data sharing (research visibility and increased recognition through higher citation rate);
- data from publicly funded research should be published and reused;
- long-term preservation of their own research data;
- opportunity to inspire further research and hands-on learning for students;
- contribution to quality, transparency and accountability of research;
- compliance to funders', institutional, or journals' data policies;
- address researchers' fear of misinterpretation, misuse, and fear of losing control over data;
- address practical constraints that researchers perceive as hindrances to proper data management.

Generalist repositories usually allow quicker data publication, which is why it may be beneficial to emphasize the benefits of using trusted domain-specific repositories in the social sciences, as they offer added value in terms of data quality, curation and preservation processes. This ensures data is more discoverable, reusable, and less prone to misinterpretation. (CESSDA Training Team, 2025a, Chapter 3.2.2 Advocate for data sharing).

3.3.Organisational Infrastructure

Organisational Infrastructure encompasses governance, policy development, resource allocation, staffing, partnerships, and continuity planning, ensuring the repository's operational sustainability and alignment with its mission.

3.3.1. GOVERNANCE (AF13)

Domain-agnostic resources

The **FIDELIS TTRAM** the repository Activity/Function (A|F) Governance is described as follows: "The organisational hierarchies and processes by which the entity's mission is managed and executed." and suggests for transparent information "[m]anagement and Decision documentation, organograms, management roles, role of host organisation, funders etc." (L'Hours et al., 2025c, p. 19)

The A|F Governance is dealt with in **CoreTrustSeal** requirement R05 Governance & Resources, which requires repositories to have "adequate funding and sufficient numbers of staff managed through a clear system of governance to effectively carry out the mission" (CoreTrustSeal Standards and Certification Board, 2022, p. 16). For the Governance part of this requirement, CoreTrustSeal asks for applicants to provide evidence through "descriptions and diagrams of governance bodies, groups and hierarchies" (CoreTrustSeal Standards and Certification Board, 2022, p. 16).

The **nestor Seal** states in C10 Organisation and processes: "The organisational structure should be appropriate for the objectives, tasks and processes of the digital archive. The structural and procedural organisation should be defined. The responsibilities should be established. The digital

Funded by the European Union

archive is incorporated at the appropriate point in the schedule of responsibilities" (nestor Certification Working Group, 2025).

The Functional Model of the **OAIS Reference Model** details the organization, processes and workflows of an Open Archival Information System (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4.2).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** policy (White House Office of Science and Technology Policy, 2022) emphasizes that a repository's "Organizational Infrastructure" must include Long-term Organizational Sustainability through comprehensive plans for data management, integrity, and authenticity, as well as Risk Management capabilities with documented safeguards for sensitive data.

The **Recommendations Consultation. EOSC-A Long Term Data Preservation Task Force** (Andreu et al., 2023) formulates that transparency about and analysis of current roles and responsibilities associated with data services functions and activities are necessary inputs into financial calculations related to salaries and funding streams. Identify and support staffing costs for preservation specific roles and responsibilities. Any decision to store, retain and then undertake initial curation will imply costs. Any additional costs needed to deliver active preservation are uniquely preservation costs.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term (European Commission, 2024).

The **REPOSITORIES: Key Infrastructure For Maintaining European Research Excellence** (Shearer et al., 2025) states that aligned with institutional research policies. Integrate repository management into institutional data policies, funding allocation, and strategic planning.

Repositories and Beyond: Analysis of Survey for SSHOC Organisations (Ala-Lahti et al., 2022) does not cover this AF – or at least not explicitly. However, the fact that most of the respondents had a business continuity plan as well as a preservation policy, procedure or plan in place, legal and ethical information, rights information and/or licences, metadata and data versioning policies, and descriptions of service processes or workflows suggest strong governance.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) conclude that, in the context of trust and transparency, repositories should be transparent about granular characteristics such as their governance structure. Furthermore, if a mechanism for information exchange, such as the one proposed by the guidelines, is used to align a network of Trustworthy Digital Repositories (TDRs), it will be the decision of the Network governance or decision-making body to set any expectations around the sharing or content of specific metadata fields.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that modern repositories are often developed through

Funded by the European Union

partnerships and with a wide range of insource and outsource options. Trust between these different data service actors is essential for the data ecosystem, and future assessment and certification options and networks should cover data service providers and the issue of enabling FAIR data. Governance structures were evaluated as part of the certification process. Repositories were encouraged to clarify roles, responsibilities, and decision-making processes, especially in complex partnership models.

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) states that adhering to good practices regarding governance, a repository must clearly indicate the organization responsible for its management and the nature of its governance. Additionally, successful management and execution of the entity's mission require a publicly available policy stating what will happen to resources if operations cease.

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF13 Governance was not found to be addressed in the gathered Agri-Food resources.

Biomedical Sciences

The TTRAM Activity/Function AF13 Governance is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: A strong legal, funding, and governance infrastructure is essential for the long-term sustainability of ELIXIR Core Data Resources. One key element is the presence of an independent, international Scientific Advisory Board, which provides expert guidance, ensures scientific relevance, and supports strategic decision-making. In addition, resources must demonstrate sustainable support and funding, including evidence of past and ongoing commitments from host institutions and other funding bodies. These elements collectively ensure that the resource is not only scientifically robust but also institutionally and financially resilient.

Field et al., 2011: GSC comprises a board, several standing committees, and working groups, with its mission managed and executed through face-to-face meetings, the formation of working groups, and the development of consensus products/standards.

Yilmaz et al., 2011: MIxS standard is engaged by a network of experts from sequencing centers, major resource maintainers, and individual investigators. This governance includes structured processes such as community surveys for standard development, ensuring adoption by key data providers.

Karsch-Mizrachi et al., 2025: The article describes the formal organizational structure of INSDC, which includes the establishment of two key committees: the Executive Committee (EC), responsible for strategic direction, and the Implementation Committee (IC), responsible for operational policies and procedures. This governance model was formalized through the signing of the Founders

Funded by the European Union

Arrangement in 2023, which codifies the collaboration among the founding members and outlines their mutual responsibilities. The article also discusses the Membership Arrangement, which sets expectations for all members regarding data stewardship, access, and compliance. These structures and agreements demonstrate how the INSDC manages and executes its mission through defined hierarchies and processes.

Rehm et al., 2021: GA4GH is a structured alliance with eight Work Streams (WSs) involving foundational groups for regulation, ethics, and data security, and technical groups that develop standards and policy frameworks driven by the needs of twenty four real-world genomic data initiatives (Driver Projects). Products are proposed to the Steering Committee, undergo review by specialized Work Streams and a Product Review Committee.

Climate Science

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) states, for long-term management and execution of the CV as a standard, the governance structure includes the intent to establish an international governance committee under the auspices of the IS-ENES2 project to manage the CV's evolution and preservation.

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC is hosted by the DKRZ and maintained by its data management department, which is responsible for collecting, archiving, and disseminating Earth System data.

Linguistics

The Activity/Function (A|F) Governance is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The CESSDA Resource Directory (CESSDA, 2025) contains links to resources on the topic of "Organisation: Organisational Structure" that are relevant to the topic of Governance (https://www.cessda.eu/Resource-Directory?tree=17,22). The CESSDA Resource Directory (RD) lists resources with the intention to "help to build sustainable and mature data archives and support the development of new services and features within existing data archives. Information on relevant documents, training materials, tools and support services are collected, selected and reviewed, making the RD a curated inventory of existing resources" (CESSDA, 2025).

3.3.2. POLICY & STANDARDS (AF14)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Policy & Standards is described as follows: "The adoption and development of policies, standards and other criteria to guide practice, aligned with governance structures (see AF13 "Governance") and operational workflows (see AF09 "Workflows"), and in compliance with the applicable legal and ethical framework (see AF23 "Legal

Funded by the European Union

and Ethical").” and suggests for transparent information “[p]rocedures for adopting and developing policies and standards, lists of policies and standards.” (L’Hours et al., 2025c, pp. 19–20)

In the **OAIS Reference Model**, it is a function of the Administration functional entity to “Establish Standards and Policies” based on input from other functions and functional entities (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 4/13). Among the standards and policies mentioned are “format standards, documentation standards and the procedures to be followed during the Ingest process” and “approved standards and Preservation Objectives [...] for example specifying goals relating to Transformational Information Properties and usability of the Content Information” (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 4/13).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** policy (White House Office of Science and Technology Policy, 2022) emphasizes that a repository's organizational infrastructure must include documented policies for data retention, provide clear use guidance detailing access and reuse terms, and implement risk management capabilities that comply with applicable confidentiality and monitoring requirements.

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) suggests that repositories adopt community-defined standards to ensure the consistent and machine-readable representation of data and metadata. These standards—encompassing models, formats, schemas, vocabularies, and ontologies—play a critical role in facilitating the discovery, accessibility, and interpretation of data and metadata. Considerations include:

- Data and Metadata Standards: What data and metadata standards, if any, has the repository implemented to enhance interoperability and usability?
- Persistent Identifiers (PIDs): Does the repository assign globally unique and persistent identifiers to its holdings? If so, which PID schema (e.g., DOI, Handle, ARK) is utilized to ensure long-term accessibility and traceability?
- Citation and Linking to Related Publications: Does the repository provide mechanisms to link datasets to related articles, preprints, or other publications? At what stage of the data deposition process is information about related articles required?
- By adhering to these standards and practices, the repository aims to support researchers in managing, sharing, and citing their data effectively, while fostering a robust ecosystem for data discovery and reuse.

The **O'FAIRe makes you an offer: Metadata-based Automatic FAIRness Assessment for Ontologies and Semantic Resources** (Amdouni et al., 2022) focuses on technical compliance with the FAIR principles for semantic resources, it indirectly addresses the operationalization of standards by noting that high average scores for Findability (F) (70) and Accessibility (A) (90) reflect the success of repository policies in adopting standards like URIs and the HTTP protocol for access.

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) approaches policy and standards activities by making Essential the requirement that both metadata and data comply with domain-relevant community

Funded by the European Union

standards (R1.3-01M, R1.3-01D) and that metadata adheres to a machine-understandable community standard (R1.3-02M), ensuring that policy outcomes are directly reflected in the technical quality and usability of the digital objects. These indicators establish a minimum compliance baseline, requiring repositories to adopt standards that facilitate the reuse of their resources by the designated community.

In the **FAIRsFAIR project D2.3** report (Behnke et al., 2020), explicit data policies (like versioning and dynamic data) and PID policies are recommended, together with global PID for each digital object or file. Besides, an explicit data deletion policy with the roles and responsibilities in human and machine interoperable ways must be provided to users. A tombstone procedure must be put in place, as well in case the repository ceases to operate.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) recommend that transparent information exposed by repositories should take account of, and map to, existing standards and criteria, such as DCAT, schema.org, and DataCite, to minimize divergence and maximize interoperability. Furthermore, specific information consumers (like F-UJI) can design their own requirements or standards regarding the content of information exposed, which can then be communicated to their audience.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** report (Kleemola et al., 2022) promotes the adoption of CoreTrustSeal, OAIS, ISO 16363, and TRUST principles. It also discusses the need for community-agreed definitions and minimum standards for data services. Feedback on the CoreTrustSeal certification pointed to the need to clarify its guidance and to define in even more detail the necessity to include evidence and documentation rather than spell out the procedures in the assessment. The CoreTrustSeal expectations are appropriate for SSH repositories seeking TDR status and no clashes between CoreTrustSeal and more specific SSH requirements were found. There is no apparent demand for anything more detailed and extensive than CoreTrustSeal at the moment.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) strongly emphasize compliance with standards, establishing several Essential indicators that require metadata and data to comply with domain-relevant community standards (RDA-R1.3-01M, RDA-R1.3-01D) and to use standardized protocols for access (RDA-A1-04M, RDA-A1-04D), thus providing specific criteria for assessing a repository's adherence to standards in practice. Furthermore, the FDMM itself serves as a standardizing framework, providing a common set of indicators and priorities intended to normalize assessment of FAIRness, which informs policy adoption and compliance efforts for data providers, publishers, project managers, and funding agencies.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) mandate the use of DataCite metadata fields (relatedIdentifier, relationType, etc.) to store standardized citation relationships. This policy is aligned with broader community criteria, drawing upon the principles

Funded by the European Union

outlined in the FORCE11 Data Citation Principles to ensure transparent and consistent handling of data citations across the ecosystem.

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF14 Policy & Standards is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData calls for repositories to adopt, document, and harmonize data standards and policies across agricultural genomics databases. “Developing data standards and practices to facilitate data curation, integration and reuse is essential. Standards will ensure data quality, facilitate interoperability and reduce redundant work.”

Drakos et al., 2015: agINFRA directs repositories to implement shared policies and standards to ensure interoperability across agricultural data infrastructures. “Improve interoperability between existing e-infrastructures; successfully interconnect agricultural data repositories through extended metadata; advanced implementation and adoption of European standards and specifications.”

Biomedical Sciences

The TTRAM Activity/Function AF14 Policy & Standards is addressed in the following Biomedical Sciences resources:

Field et al., 2011: GSC serves as a key open-membership organizational infrastructure, established to drive community-based standardization activities aimed at enhancing the usefulness of 'omics information. Its core mission involves implementing new genomic standards, such MIxS which includes checklists for genomes, metagenomes, and marker gene sequences, and harmonizing information collection and analysis efforts.

Karsch-Mizrachi et al., 2025: The article details how the INSDC has formalized its collaboration through the Founders Arrangement and the Membership Arrangement, which establish shared principles, responsibilities, and expectations for all members. These documents guide operational practices and are aligned with the governance structure, particularly through the roles of the Implementation Committee, which is responsible for establishing INSDC policies and procedures for building and maintaining nucleotide sequence databases. The article also describes the development of minimal standards for data and metadata, which are being harmonized with external standards organizations such as the Genomic Standards Consortium and the Global Alliance for Genomics and Health. These efforts ensure that INSDC’s policies and standards are not only internally consistent but also compliant with broader scientific, ethical, and technical frameworks, supporting interoperability and responsible data stewardship.

Climate Science

Funded by the European Union

The METAFOR project: preserving data through metadata standards for climate models and simulations (Callaghan et al., 2010) introduces the Common Information Model (CIM) that acts as a policy/standard for metadata.

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) explains the organizational infrastructure adopted and enforced key existing criteria, such as the CF convention for data format and metadata constraints imposed by CMIP5 tables, while centering its conceptual framework on the Common Information Model (CIM). A Controlled Vocabulary (CV), was established through wide scientific consultation to impose consistency and comparability in model descriptions, with plans to establish an international governance committee to manage its long-term evolution and preservation.

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC operates as a Core Trust Seal-certified FAIR repository under the guidance of DKRZ's Data Management, establishes an organizational infrastructure with clear policies and standards to ensure long-term data accessibility and re-usability.

The **NetCDF Climate and Forecast (CF) Metadata Conventions** (Eaton et al., 2024) state that the CF conventions establish a community-driven and publicly licensed standard for climate and forecast metadata. The standard's development and adoption are guided by design principles that prioritize practicality for producers and users, consistency, error minimization, and crucial backward compatibility for long-term data utility.

Linguistics

The Activity/Function (A|F) Policy & Standards is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The CESSDA Data Archiving Guide outlines a set of core policies covering five key areas in Chapter 2 - Policies of Data Archives:

1. Data Collection and Acquisition – These policies define how data is selected and acquired
2. Preservation – These policies' purpose is to define the rules, the responsibilities, the roles and monitoring systems for managing data within a data archive.
3. Curation - Data repositories should implement clear curation policies to ensure proper management and long-term preservation of datasets. These policies should include provisions for assigning PIDs (e.g. DOIs, Handles, or URNs) and defining access conditions for each digital object.
4. Data Access – These policies define the procedures that an archive must implement when providing access to data.
5. Data Dissemination – These policies outline the principles and conditions under which data are made available to users.

Funded by the European Union

In addition, policies can support certification processes, such as applying for CoreTrustSeal, or may be required to outline certain standards, such as ISO 27001. (CESSDA Training Team, 2025a, Chapter 2: Policies of Data Archives).

Internal policies aligned with FAIR principles are essential for an organisation's long-term success. Sharing these policies with users and designated communities can help build trust in the archive. (CESSDA Training Team, 2025a, Chapter 5.2 Enable FAIR in your organisation).

3.3.3. RIGHTS (AF15)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Rights is described as follows: "The management of permissions, prohibitions, and obligations governing how digital objects are accessed and used by individuals and organisations, both within and outside the repository." and suggests for transparent information "Rights Policy, Licences for deposit, access and reuse, non-compliance policy, sufficient rights to permit LoRCAP, rights related to succession plans." (L'Hours et al., 2025c, p. 20)

CoreTrustSeal addresses the A|F Rights in requirement R02 Rights Management, according to which "[t]he repository maintains all applicable rights and monitors compliance", which means that the repository "manages, and communicates to relevant stakeholders, all rights (permissions, prohibitions, obligations) covering data and metadata deposit, storage, preservation, access, and use" and has in place the "system, methods and artefacts (e.g. licenses, agreements, terms and conditions, and related policies and procedures)" necessary for rights management(CoreTrustSeal Standards and Certification Board, 2022, p. 12). The applicants need to provide evidence addressing the following items:

- The overall rights management approach to deposited files, data and metadata.
- The rights to copy, transform, and store digital objects for preservation, as well as provide access to them[.]
- Conditions of use (e.g. intellectual property rights, distribution, intended use, protection of sensitive data, etc.).
- Deposit and access agreements or licenses.
- How rights metadata is managed for humans (e.g. license documents/files) or machines.
- Monitoring of compliance at deposit, during curation/preservation, and during access and reuse. Describe any circumstances where compliance monitoring is not possible.
- Measures in place if non-compliance is detected.

The **OAIS Reference Model** defines Access Rights Information as "[t]he information that identifies the access restrictions pertaining to the Content Data Object, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the Submission Agreement, related to both preservation (by the OAIS) and final usage (by the Consumer). It also includes the specifications for the application of rights enforcement measures." (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 1/9). It is stored in the

Funded by the European Union

Preservation Description Information (PDI), a specific Information Object included in Archival Information Packages. PDI "will support the trust in, the access to and context of the Content Data Object over an indefinite period of time [...]. The PDI must include information that is necessary to adequately preserve the particular Content Data Object with which it is associated. It is specifically focused on describing the past and present states of the Content Data Object, ensuring it is uniquely identifiable, and ensuring it has not been unknowingly altered" (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4.3.2.4.3).

The **nestor Seal** addresses Rights in the following criteria: C6 Legal and contractual basis, focusing on the relation to the producers, as well as C7 Legal conformity, which addresses the need to ensure continued "conformity with relevant regulations concerning the ingest, archiving and use of digital objects" (nestor Certification Working Group, 2025).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** policy (White House Office of Science and Technology Policy, 2022) addresses rights by stating that repositories must provide Clear Use Guidance, ensuring datasets are accompanied by documentation describing terms of dataset access and use (such as reuse licenses and the need for approval by a data use committee). Furthermore, repositories must provide Broad and Measured Reuse, ensuring datasets include metadata that describe terms of reuse.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. They ensure that contents are accompanied by metadata sufficiently detailed and of sufficiently high quality to enable discovery, reuse and citation and contain information about provenance and licensing (European Commission, 2024).

The **Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements** (Jahn et al., 2023) states that for the fulfilment of the Horizon Europe MGA license criteria we consider the repository to be required to provide a field in the metadata in which the licence of the deposited item can be indicated.

The **Practical Guide to the International Alignment of Research Data Management** (Science Europe, 2021) states that:

1. Data access and usage licences:
    a. Enable access to data under well-specified conditions
    b. Ensure data authenticity and integrity
    c. Enable retrieval of data
    d. Provide information about licensing and permissions (in ideally machine-readable form)
    e. Ensure confidentiality and respect rights of data subjects and creators.

The exclusion criteria from the **Research Data College Working Group** rule out subject-specific repositories that practice the transfer of rights, as "[t]he intellectual property practices of some

Funded by the European Union

publishers do not guarantee unrestricted access and unrestricted reuse of data deposited in the repositories that they develop and recommend" (Lamotte et al., 2024, p. 10).

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) addresses rights directly in the Reusable (R) area by making it an Essential indicator (RDA-R1.1-01M) that metadata includes information about the licence under which the data can be reused. Furthermore, the model prioritizes the adoption of a standard reuse licence (RDA-R1.1-02M) and ensuring the licence information is expressed in a machine-understandable way (RDA-R1.1-03M), both considered Important indicators for enabling automated rights management and reuse.

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), ten out of 14 respondents offered rights information and/or licences.

The **FAIRsFAIR project D2.3** report (Behnke et al., 2020) acknowledges that a variety of access restrictions exists, but doesn't go into details.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.3, Metadata Quality, that there need to be clear recommendations for non-CC licences when necessary. This includes the recommendation of a standard software license that is free, open, and interoperable such as the EU Commission-recommended EUPL. Furthermore, guidelines must be established for how to select licenses that facilitate optimization and accessibility for sensitive, person-identifying, and/or Indigenous data. Discipline-, geographic-, and organization-specific licences are commonly in use, and would benefit from semantic indexing and cataloguing to enhance both license understandability and data reusability.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) emphasize that transparent information regarding "Rights" is critical, specifically mentioning that the deposit and reuse licenses for an object could be included as characteristics of a specific digital object resulting from a repository activity/function. Furthermore, managing rights involves ensuring that Access criteria are aligned with rights management policies, which determines the appropriate method of access for users.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) strongly align with rights, such as the management of permissions, prohibitions, and obligations by making it an Essential indicator (RDA-R1.1-01M) that metadata includes information about the license under which the data can be reused. Furthermore, the FDMM includes Important indicators requiring that the metadata refers to a standard reuse licence (RDA-R1.1-02M) and that the reuse license is machine-understandable (RDA-R1.1-03M), facilitating automated management of these permissions and obligations.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) encourage practices that promote open data sharing and citation, noting that data citations increase the rigor and reproducibility of research and act as a major motivator for researchers to publish their data. Additionally, the guidance relates to the rights of data creators by emphasizing that data citations

Funded by the European Union

provide credit for the data producer, recognizing the individuals or organizations that shared the data.

The **Core Preservation Process CPP-020 Rights Management** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA manages rights related to the Information Objects, both for agents inside its scope (access, migrate, etc.) and for end users (access, reuse, etc.)” (EOSC EDEN T1.2 et al., 2025).

Within AF15 Rights, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

- “Rights Management: The process of tracking and managing ownership and copyright inherent to a data set as well as monitoring conditions and policies for access and reuse (e.g., licenses and data use agreements).”
- “Terms of Use: Information provided to end users of a data set that outline the requirements or conditions for use (e.g., a Creative Commons License).”

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF15 Rights was not found to be addressed in the gathered Agri-Food resources except, marginally, in **Top et al., 2022:** AgroDataCube demonstrates repository transparency in rights management by publicly documenting licenses and access conditions on its web page. “The web page of AgroDataCube also provides information about the used license and the access token needed for data retrieval (‘R’).”

Biomedical Sciences

The TTRAM Activity/Function AF15 Rights is addressed in the following Biomedical Sciences resources:

Sansone & Rocca-Serra, 2016: Ownership of open content standards can be problematic, and the legal framework to encourage maintenance, contribution, and development is very immature.

Karsch-Mizrachi et al., 2025: The article explains that INSDC is founded on the principle of open access, with all members committing to provide free and unrestricted access to nucleotide sequence data. However, it also acknowledges that certain types of data, such as controlled-access human data, are managed separately by each member and are not shared across the collaboration. Additionally, the article discusses concerns around equitable benefit-sharing and the development of the CARE principles for Indigenous Data Governance, showing awareness of the ethical dimensions of data rights.

Wilkinson et al., 2016: Rights are addressed by explicitly stating that scholarly digital objects, including their metadata, must be released with a clear and accessible data usage license. This is a core component of the Reusability (R) principle.

Funded by the European Union

Climate Science

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC manages permissions, prohibitions, and obligations governing how digital objects are accessed and used by individuals and organizations primarily through licensing agreements. Data providers must decide under which license to publish their data.

Linguistics

CLARIN B and E centres are required "to make clear statements about their policy of offering data and services and their treatment of IPR issues" (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023, p. 3).

For CLARIN B centres, the following requirement applies (Wittenburg, Van Uytvanck, Zastrow, Straňák, et al., 2023):

- 3.a Data offering & IPR. Requirement: Each centre needs to make clear statements about their policy of offering data and services and their treatment of IPR issues.

CLARIN centres are recommended to offer their resources and services according to formal legal terms and conditions, which include (CLARIN Legal and Ethical Issues Committee, n.d.):

1) The **general terms of service** regulating access to CLARIN resources;
2) The **end-user licenses** further specifying the access conditions for each resource;
3) The **deposition agreements** that are signed by resource providers when depositing a resource to a CLARIN centre.

These are formulated in a number of legal documents that have been made available by the Finnish CLARIN consortium (see links below).

Terms of Service

- The general conditions under which end-users may access the resources distributed by CLARIN are stated in the **CLARIN Terms of Service**.
- The **model document** CLARIN-TOS-v1.0 includes the legal terms and conditions holding between a *CLARIN centre* and its end-users with regard to resource access.

End-User Licenses

- All **resources distributed by CLARIN are accompanied with licenses**, which may include more specific conditions than those stated in the CLARIN Terms of Service. In fact, the same language resource can be distributed with more than one license, depending on the end user's role or intended use.
- In order to facilitate the management of language resources, CLARIN uses a classification system (see below), according to which licenses are divided into three main categories:
    - CLARIN PUB ('**public use**')

Funded by the European Union

  - CLARIN ACA ('**academic use**')
  - CLARIN RES ('**restricted use**').
- In addition, CLARIN uses a set of labels (aka '*laundry tags*') that correspond to conditions of use that are frequently associated with the distribution of language resources.

Deposition License Agreements

- When a resource provider (aka '*depositor*') wants to include a new resource into the CLARIN domain, a specific **deposition license agreement** must be made between the depositor and the CLARIN centre that will host the resource.
- In the deposition agreement, the depositor also specifies the license(s) with which the resource will be distributed to end-users. Depending on the category under which the selected end-user license falls (see previous section), a minimal set of access conditions have to apply. These conditions are stated in the following three model agreements; additional or more specific conditions can also be agreed upon between the centre and the resource depositors. The minimal model agreements below can be used as checklists.
  - CLARIN-DELA-PUB-v1.0 (for public use resources)
  - CLARIN-DELA-ACA-v1.0 (for academic use resources)
  - CLARIN-DELA-RES-v1.0 (for restricted use resources).

Social Sciences

The CESSDA Data Access Policy (Bolton, 2022) requires that an agreement between the SP and the Data Owner covering data access arrangements must be in place for each data collection within each SP's holdings (Principle 3).

The CESSDA Data Archiving Guide addresses the A|F Rights in several of its chapters. First, it states that archiving processes in European archives are governed by both national laws and international regulations, and that the main legal concerns include copyright law and the sharing of personal data (i.e. GDPR). These frameworks impose obligations on how digital objects are managed within repositories. (CESSDA Training Team, 2025a, Chapter 1.9 What are relevant legislations concerning data archiving?).

Next, the guide discusses data access. Data access and data usage permissions are managed through access conditions, and the guide outlines three levels of access, depending on the nature of the data and the agreements made. The levels are: (1) Open access to data, (2) Restricted data access and (3) Controlled data access. (CESSDA Training Team, 2025a, 2.4 Data Access Policies). Finally, the guide addresses data dissemination principles. Data dissemination principles reflect the archive's position on safeguarding and sharing data. These principles can also refer to explicit policies and legal framework, such as GDPR compliance. (CESSDA Training Team, 2025a, 2.5 Data Dissemination Policies and Principles).

3.3.4. RESOURCES (AF16)

Funded by the European Union

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Resources is described as follows: “The organisational hierarchies and processes by which the organisational entity's human and financial resources are obtained and managed. This includes a knowledge of and an effective deployment of human, financial and energy assets.” and suggests for transparent information “[b]udget and finance strategies, plans, reports and projections.” (L’Hours et al., 2025c, p. 20)

The A|F Resources is dealt with in **CoreTrustSeal** requirement R05 Governance & Resources, which requires repositories to have “adequate funding and sufficient numbers of staff managed through a clear system of governance to effectively carry out the mission” (CoreTrustSeal Standards and Certification Board, 2022, p. 16). For the Resources part of this requirement, CoreTrustSeal asks for applicants to provide evidence for the following items (CoreTrustSeal Standards and Certification Board, 2022, p. 16):

- Timescales for provision and renewal of funding for operational costs and recruitment; it is understood that permanent, ongoing funding cannot be perfectly quantified or guaranteed.
- Evidence that the repository is, or is hosted by, a recognized institution (supporting long-term stability and sustainability) appropriate to its Designated Community.
- Demonstrate that the repository can meet its obligations, including sufficient funding, staff resources, IT resources, and a budget for external engagement when necessary.

The **nestor Seal** requirement C8 Funding corresponds to this A|F (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** the high-level governance of the OAIS is carried out by Management, which in the model’s logic is not part of the OAIS, but an element of the OAIS environment. “Management is often the primary source of funding for an OAIS and therefore should agree the budget for the Archive’s activities and may provide guidelines for resource utilization (personnel, equipment, facilities)” (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 2/11). The day-to-day utilization of these resources within the OAIS is managed within the Administration functional entity: “The Establish Standards and Policies function [...] receives approvals and budget information and policies such as the OAIS charter, scope, resource utilization guidelines, and pricing policies from Management” (Consultative Committee for Space Data Systems (CCSDS), 2024, p. 4/13).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** policy (White House Office of Science and Technology Policy, 2022) addresses the financial and management aspects of resources primarily through the requirement for Long-term Organizational Sustainability, which mandates that the repository have a plan for long-term management of data built on a stable technical infrastructure and funding plans.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that

Funded by the European Union

are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term (European Commission, 2024).

The **Repository Features to Help Researchers: An invitation to a dialogue** (Cannon et al., 2021) includes the following questions on funding: Who funds the operation (organisations) of the repository and the type of funding (e.g. grants, donations, memberships). Does the repository provide information on its sustainability plans?

The **Current State and Future Directions for Open Repositories in Europe** (Shearer et al., 2023) states that increased staffing at these repositories could help to address many of the challenges being experienced and ensure there is widespread adoption of good practices and next generation repository functionalities. Shared infrastructure models, which have already been adopted in a few countries, are another approach that offer economies of scale and could relieve some of the burden from individual institutions. Several respondents provided more information about their sustainability challenges, grouped into several categories:

- Time and resource requirements to properly curate metadata and content
- Replacement of repositories with CRIS systems, which do not fully support the needs for managing a variety of content types
- Complexities of regular software upgrades
- High cost of employing outside companies to support software upgrades and ongoing maintenance of the system
- Lack of expected functionalities of the repository platforms
- Understaffing

In a survey, respondents perceptions of repository sustainability where 97% of respondents (351) felt their repository was either “very” or “somewhat” sustainable, with only 13 respondents (3%) indicating that it was “not sustainable”. The 3% of respondents that felt their repository was unsustainable came from different countries and repository types, so no geographic generalisations could be inferred.

The **REPOSITORIES: Key Infrastructure For Maintaining European Research Excellence** (Shearer et al., 2025) states regarding maintenance with sustainable funding and adequately staffed. Treat repositories as critical research infrastructure and ensure dedicated institutional budget lines and sustainable funding models.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that resource availability (staffing, funding) was a major factor in certification readiness. Smaller repositories often struggled with the resource demands of certification. Diversity of organisations combined with the fact that certification, particularly for the first time, demands time from the repositories points to the difficulty of finding certification solutions that are suitable for all organisations.

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) states that repositories must demonstrate effective resource

Funded by the European Union

management by maintaining a long-term plan for managing and funding the repository, while also designating at least one staff member with the explicit responsibility of managing the services to support sustainability.

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF16 Resources was not found to be addressed in the gathered Agri-Food resources.

Biomedical Sciences

The TTRAM Activity/Function AF16 Resources is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: The article partially covers the "Resources" function, focusing on governance and financial sustainability. However, it does not delve into operational resource management or energy use, which may be outside the scope of ELIXIR evaluation framework.

Field et al., 2011: Funding bodies for GSC include NERC International Opportunities Fund and National Science Foundation (NSF) grants.

Yilmaz et al., 2011: The standards themselves are centrally maintained in a relational database system at the Max Planck Institute for Marine Microbiology Bremen on behalf of the GSC, utilizing established processes like a public issue tracking system and annual releases to manage and deploy these valuable knowledge assets to the scientific community.

Rehm et al., 2021: GA4GH is an expert-driven structure involving over 1000 individuals from more than 90 countries. It has twenty four Driver Projects committing at least two full-time equivalents (FTEs) to standards development across its eight Work Streams. Financial resources are secured from a wide array of international funding bodies and institutions.

Climate Science

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that financial resources come through external or third-party funding and associated grants from data projects during the metadata submission phase in MetaXA.

Linguistics

CLARIN B and E centres are required "to make explicit statements to the CLARIN boards about its technological and funding support state and its perspectives in these respects" (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023, p. 3).

For CLARIN B centres, the following requirement applies (Wittenburg, Van Uytvanck, Zastrow, Straňák, et al., 2023):

Funded by the European Union

- 2.d Continuity of access and funding support. Requirement: Each centre needs to make explicit statements about perspectives of continuity of access and funding support to continue activities as an active CLARIN centre.

Social Sciences

The CESSDA Resource Directory (CESSDA, 2025) contains links to resources on the topic of "Organisation: Staffing, Management and Finances" that are relevant to the topic of Resources (https://www.cessda.eu/Resource-Directory?tree=17,20). The CESSDA Resource Directory (RD) lists resources with the intention to "help to build sustainable and mature data archives and support the development of new services and features within existing data archives. Information on relevant documents, training materials, tools and support services are collected, selected and reviewed, making the RD a curated inventory of existing resources" (CESSDA, 2025).

3.3.5. PEOPLE & EXPERTISE (AF17)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) People & Expertise is described as follows: "The management of human resources to fulfil the mission. Includes ensuring that sufficient skills are available, internally or externally (see AF20 "External Engagement")." and suggests for transparent information "Skills Statement, Skills Development Plan, Staff Roles & Skills List" (L'Hours et al., 2025c, p. 21).

CoreTrustSeal address the A|F People & Expertise in their requirement R06 Expertise & Guidance, which requires that "[t]he repository adopts mechanisms to secure ongoing expertise, guidance and feedback-either in-house, or external" and for which evidence for the following items needs to be provided (CoreTrustSeal Standards and Certification Board, 2022, pp. 16–17):

- That guidance and expertise reflects the scientific scope of the repository, if relevant.
- The repository aligns internal recruitment and external engagement with the services it offers.
- The repository ensures that its staff have access to ongoing training and professional development.
- The range and depth of expertise of both the organisation and its staff, including any relevant affiliations (e.g. national or international bodies), is appropriate to the mission.
- In-house advisers, or external advisory committees that include technical, curation, data science, data security, and disciplinary experts.
- How the repository communicates with experts for advice.

The **nestor Seal** criterion "C9 Personnel" requires that "[s]ufficient numbers of appropriately qualified staff are available. Updated job descriptions exist which set out the required qualifications of the digital archive personnel and contain an organisational chart and/or a staff development plan based on the tasks and objectives of the digital archive" add citation.

Funded by the European Union

The **Desirable Characteristics of Data Repositories for Federally Funded Research** policy (White House Office of Science and Technology Policy, 2022) states that the repository provides or facilitates expert curation and quality assurance to improve the accuracy and integrity of datasets and metadata.

The **Recommendations Consultation. EOSC-A Long Term Data Preservation Task Force** (Andreu et al., 2023) formulates that all of the preservation-specific and supporting research data management roles across the data lifecycle require sustained training based on a rich knowledge base of preservation information. Clear responsibilities must be in place for developing standards and guidance, for communication and for training.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term (European Commission, 2024).

The **Update of the Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements** states that the lack of a public policy for preservation, curation and security of the contents is the most frequent reason, followed by not adhering to a specific metadata standard (Lazzeri, 2024).

The **Current State and Future Directions for Open Repositories in Europe** (Shearer et al., 2023) states that as the needs of the user community expand and evolve with open science becoming mainstream, there will be an increasing strain on repository staff. Low staffing levels are due to the fact that repositories have not been a high priority service for universities and that they are also competing with the commercial sector for skilled technical staff.

The **REPOSITORIES: Key Infrastructure For Maintaining European Research Excellence** (Shearer et al., 2025) states regarding trained professionals ensure repository managers and librarians receive ongoing training in metadata curation, research data management, and emerging AI applications. Establish professional development programmes and participate in international capacity-building initiatives.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) recommend that repositories express information about Skills Statements, Skills Development Plans, and Staff Roles & Skills Lists to demonstrate the availability of sufficient expertise. Furthermore, these guidelines note that specialist repositories, in particular, require specific skills in data formats, research methodologies, metadata, and ontologies related to their disciplines or content types to execute their functions effectively. specific skills in data formats, research methodologies, metadata, and ontologies related to their disciplines or content types to execute their functions effectively.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that repository certification is a key part of building a trusted FAIR data ecosystem. It needs to be applied in a way that acknowledges the differences in

Funded by the European Union

goals, practices and maturity of data repositories and other service providers. Interoperability, standards, automation and technology are all parts of the solution, but reusability of data and long term preservation of understandability is ultimately dependent on domain and disciplinary expertise. The report highlights the importance of staff expertise in data curation, preservation, and certification. Peer support and training helped build capacity in less mature repositories. Ensuring the sustainable management of trust is not solely dependent on assessment or certification, as trust goes beyond the technical aspects of repositories and also involves people.

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF17 People & Expertise is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData highlights the importance of identifying and crediting repository personnel and ensuring teams include the expertise necessary to maintain and support databases. See AF01, Identification & Contact.

Sen et al., 2020: WheatIS points to the need for skilled data curators and technical staff within each contributing repository to maintain consistent standards and interoperability; as well as the need for a good range of different expertises to keep the network running. “It needed technical expertise to build and maintain a strong computational infrastructure and create data formats to make data sets readable; scientific expertise to understand different types of wheat data sets (including genetic, genomic, phenotypic, and metabolic); outreach capability to help build relationships to add new nodes with new data sets; and leaders who not only motivate and manage personnel, but also work with the Wheat Initiative and the broader wheat community to promote and support WheatIS. The need for dedicated and competent personnel with complimentary and overlapping expertise was crucial. For WheatIS, or for any scientific community for that matter, the critical question is the type of the expertise needed and how much time the experts can devote to a fledgling community.”

Biomedical Sciences

The TTRAM Activity/Function AF17 People & Expertise is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: Effective management of human resources is essential for fulfilling the mission of an ELIXIR Core Data Resource. This includes ensuring that the resource has access to the necessary expertise and skills. Resources support ELIXIR CDRs’ operations, maintain scientific quality, and respond to evolving community needs. This also involves strategic engagement with user communities, helping to ensure that the resource remains relevant and scientifically robust.

Sansone & Rocca-Serra, 2016: Ideally, standards should be implemented by dedicated experts in tools, services, and infrastructure to make them "invisible" to regular users

Funded by the European Union

Field et al., 2011: GSC develops standards through community-based activities. It achieves this by forming working groups and actively encouraging the wider scientific community to join and contribute their expertise.

Yilmaz et al., 2011: MIxS is maintained and authored by representatives from genome sequencing centers, maintainers of major resources, and principal investigators of both large- and small-scale sequencing projects.

Climate Science

The Activity/Function (A|F) People & Expertise is not addressed by any of the reviewed resources from Climate Science.

Linguistics

The Activity/Function (A|F) People & Expertise is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The Activity/Function (A|F) People & Expertise is not addressed by any of the reviewed resources from Social Sciences.

3.3.6. THIRD PARTY DEPENDENCIES (AF18)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Third Party Dependencies is described as follows: “Identifying and managing third-party organisations, hosts, partners, and other actors that are essential to operations, including the delivery of data or metadata services.” and suggests for transparent information “[l]ists of partners, Partnership Statement, Partner Management Plan, Contracts/suppliers mapped to the functions and activities they provide for the organisation (e.g. a storage provider).”

The A|F Third Party Dependencies is explicitly dealt with in item “(6) Cooperation and outsourcing to third parties, partners and host organisations” in the R0. Background Information & Context section of **CoreTrustSeal**. In cases where the applicant does not have have direct control of a repository function and/or supporting evidence, CoreTrustSeal requires repositories to list the relevant host organisation, partner or other third party, “[d]escribe the function or service they provide, the nature of the relationship or agreement (contractual, Service Level Agreement, Memorandum of Understanding, etc.) and whether they have any relevant certifications” .

The **OAIS Reference Model** discusses this topic in chapter 6 “Archive Interoperability”. It distinguishes independent, cooperating, and federated archives as well as different “styles of resource sharing” ranging from “all in-house” to “distributed” (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 6).

Funded by the European Union

The **Recommendations Consultation. EOSC-A Long Term Data Preservation Task Force** (Andreu et al., 2023) formulates that roles and responsibilities including for complex partnerships, third party relationships and outsourcing should be understood and transparent. Technical repository service providers' (storage providers, ARCHIVER22 etc) portfolio of service offerings should be clear and comparable for client end-users. Where responsibility is distributed, accountability should remain clear, including accountability for (meta)data loss or destruction.

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), respondents were asked who provides their technical infrastructure. The responses to this question were more varied, with two out of the three non-repository service providers and four of the repositories indicating that their technical infrastructure was provided by 'the organisation/the repository itself'. The remaining non-repository service provider and one repository indicated that their technical infrastructure was provided by 'a third party (outsourced)'. Of the remaining six repositories, three selected that their technical infrastructure was provided by 'the host organisation' and the other three repositories indicated that they 'shared responsibility' with regards to their technical infrastructure.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) note that repositories depend on a wider partnership of metadata and data services, such as storage providers, to high-performance computing, and multiple registries, all of which play a critical role in research data infrastructure. Consequently, the guidelines specify that Security must extend to the boundary between the repository and external entities, including these third-party dependencies.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states in chapter 4.2 TDR Partnership Models and Outsourcing that repositories must document third-party relationships and ensure quality and security in outsourced services. Complex partnerships models and outsourcing pose challenges for certification. While outsourcing is not an obstacle to certification as such, it requires repositories outsourcing their functions to explicitly state these functions and related documentation.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) rely on Third Party Dependencies for the delivery of data citation services, emphasizing that repositories harvest citations from and submit data to essential external actors like DataCite and Crossref. Additionally, the initiative relies on various external sources and third-party aggregators—including Dimensions, Europe PMC, NASA ADS, and the Chan Zuckerberg Initiative—to collect citation data and contribute to centralized resources like the Data Citation Corpus.

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF18 Third Party Dependencies was not found to be addressed in the gathered Agri-Food resources.

Funded by the European Union

Biomedical Sciences

The TTRAM Activity/Function AF18 Third Party Dependencies is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: While ELIXIR CDRs coexist with a diverse array of databases, they often serve as **foundational infrastructure**, with other resources depending on them to deliver key data or metadata services. This **reach-through effect** highlights the strategic importance of ELIXIR CDRs and their role in enabling the functionality and reliability of the wider network of life science data resources.

Field et al., 2011: GSC working groups operate via various public data portals and software suites, such as GOLD and MG-Rast, and by working closely with related communities and initiatives like the MIBBI initiative and BioSharing.

Yilmaz et al., 2011: Third-party dependencies include the International Nucleotide Sequence Database Collaboration (INSDC), comprising GenBank, the European Nucleotide Archive (ENA), the DNA Data Bank of Japan (DDBJ), and the Sequence Read Archive (SRA), for submitting, validating, and hosting compliant data. Development and ongoing maintenance involve contributions from various research consortia and institutes such as the Human Microbiome Project, Terragenome Consortium, ICoMM, and the Max Planck Institute for Marine Microbiology, as well as developers of supporting tools like MetaBar, QIIME, and ISA.

Barrett et al., 2012: The BioProject database was established to facilitate the organization and classification of project data submitted to the NCBI, EBI, and DDBJ databases. It captures descriptive information about research projects that result in high-volume submissions to archival databases, ties together related data across multiple archives, and serves as a central portal to inform users of data availability.

Sensitive human data, such as clinical samples, are not stored in BioSample. Instead, they continue to be deposited in NCBI's dbGaP database. The dbGaP and BioSample databases collaborate to present sanitized versions of these data in BioSample—versions where sensitive attributes are omitted. This approach allows users to locate relevant data in BioSample and then apply to dbGaP for access to the full descriptions as necessary.

To submit data, users must authenticate via the NCBI Submission Portal, which supports several login options, including various National Institutes of Health accounts and a general NCBI PDA (Primary Data Archives) account. Once submitted, BioProject and BioSample records can be accessed through Entrez links from other NCBI databases, enabling seamless navigation across related data resources.

Climate Science

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) explains that the organizational infrastructure relies on the Earth System Grid Federation (ESGF), which utilizes distributed data centers (like PCMDI, BADC, and WDCC) to host data and function as gateways for data publication and download. Furthermore, the development of the

Funded by the European Union

core metadata standards exploited collaborative work with the US Earth System Curator project, and the central metadata harvesting tool (the CMIP5 Questionnaire) was deployed and run by the British Atmospheric Data Centre (BADC).

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that DataCite for Digital Object Identifier (DOI) is a third party dependency for publication.

Linguistics

The Activity/Function (A|F) Third Party Dependencies is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The Activity/Function (A|F) Third Party Dependencies is not addressed by any of the reviewed resources from Social Sciences.

3.3.7. CONTINUITY OF SERVICE (AF19)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Continuity of Service is described as follows: “Ensuring business continuity by maintaining normal operations and preparing for disruptions. This includes non-technical processes (for technical, see AF29 “Technical Infrastructure”) related to the organisation, digital object management (see AF09 "Workflows"), and security (see AF30 "Security"). It also covers recovery plans for events like service interruptions and disasters. Continuity includes succession planning for the end of a service.” and suggests the following items for transparent information (L’Hours et al., 2025c, pp. 21–22):

- Business Continuity Plan
- Disaster Recovery Plan
- Succession Plan

The A|F Continuity of Service has its **CoreTrustSeal** equivalent in the R03 requirement, with the identical name. CoreTrustSeal R03 requires the repository to have “a plan to ensure ongoing access to and preservation of its data and metadata”, covering “the stable management of repository services over time (business continuity) and the response when services have problems (disaster recovery)”, and “includes preparations for handover of digital objects and services to another repository (succession planning)”, applying to the “deposit, storage, preservation, and access services offered by the repository to depositors and users” (CoreTrustSeal Standards and Certification Board, 2022, p. 13). Applicants are asked to provide evidence addressing the following items (CoreTrustSeal Standards and Certification Board, 2022, pp. 13–14):

- The functions and services offered by the repository to depositors and users.
- The approach to rapid changes of circumstance and long-term planning.

Funded by the European Union

- The options for relocation or transition of the activity to another repository. For example, the case of cessation of funding due to an unexpected withdrawal of funding, or a shift of host institution interests.
- The repository approach to managing policies, procedures and other business information over time.

The **nestor Seal** addresses aspects of this A|F in "C12 Crisis / successorship management" (nestor Certification Working Group, 2025), which requires that the digital archive is in possession of a plan which ensures continuation of the preservation tasks even beyond the archive's own existence. The digital archive should have made contingency plans. In such a case the preservation work must be continued in a different organisational framework, thereby ensuring that the set tasks can be carried out in full. Where this is not possible, any deficiencies should be documented. The digital archive should take precautions to ensure that the transition process can be defined, planned and implemented in good time.

The **OAIS Reference Model** refers to continuity and success planning in the context of the mandatory responsibilities. As part of the responsibility of "follow[ing] established preservation policies and procedures", it states: "The Archive should have a formal Succession Plan, contingency plans, and/or escrow arrangements in place in case the Archive ceases to operate or the governing or funding institution substantially changes its scope. The formal Succession Plan should include planning for how the AIPs will be findable once they have been moved to a successor repository" (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 3, p. 6).

The **EOSC Federation Handbook** addresses the A|F Continuity of Service in its section 5.2.2.1 FAIR Data repositories, requiring data repositories to "have a long-term support sustainability plan for at least the next five to ten years, with the goal to be available for even longer for certain datasets (e.g. unique high quality datasets which are of high value and unique i.e. cannot be reproduced)" (EOSC Association, 2025, p. 32).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** policy (White House Office of Science and Technology Policy, 2022) states that the repository has a plan for long-term management of data, including maintaining integrity, authenticity, and availability of datasets; has contingency plans to ensure data are available and maintained during and after unforeseen events.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. Trusted repositories have specific provisions in place and offer explicit information online about their policies, which define their services (e.g. acquisition, access, security of content, long-term sustainability of service including funding, etc) (European Commission, 2024).

Funded by the European Union

The **D1.3 Recommendations for a FAIR EOSC - White Paper of the FAIR-IMPACT Synchronisation Force** (Grootveld, 2025) states that the sustainability of such initiatives, networks of TDRs, is problematic and there is a need to enable more active cooperation between local, European and global initiatives.

The exclusion criteria from the **Research Data College Working Group** rule out subject-specific repositories that provide “no guarantee about the sustainability of the infrastructure” (Lamotte et al., 2024, p. 10).

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) approaches the continuity of service concept by establishing the Essential indicator A2-01M, which requires that metadata is guaranteed to remain available even after the data is no longer available, thereby addressing the long-term continuity of resource information.

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), nine out of 14 respondents had a business continuity plan as well as a preservation policy, procedure or plan in place.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.4, Trustworthy Digital Archive Capabilities, that formal policies and strategies to ensure data backups, retention policies, long-term financial planning, and exit strategies are insufficient for most disciplines and must be established. Backups are an important task for many repositories, and concrete technical implementation plans (with options for transfer to third parties if necessary) that support data package transfer across all disciplines must be rolled out. Few infrastructures operate under sustainable funding models, where some repositories may receive funding from one or two close partner institutions, but at the same time lack a long-term strategy for operational upkeep and maintenance that can be in conflict with their guaranteed retention period. Formal exit strategies with concrete plans, partners, and workflows must be established in order to prevent data loss and additionally have grave consequences for researchers.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that trust is not contingent on certificates alone but is earned through interaction with and accountability to the designated community. This is why cooperation through peer networks can be an important route to growth and sustainability. Such networks could be starting points for shared responsibility for sustainable data and enable ensuring continuity of repositories. However, the current models of trust support that largely rely on project-based funding are not favourable to ensuring sustainability, which is why more enduring solutions are required. " Formal agreements with host institutions are often lacking, which can hinder certification.

Within AF19 Continuity of Service, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

Funded by the European Union

- "Succession Planning: Planning for contingency, and/or escrow arrangements, in the case that the repository (or other entity responsible) ceases to operate or the institution substantially changes its scope."
- "Cease Data Curation: Plan for any contingencies that will ultimately terminate access to the data. For example, providing tombstones or metadata records for data that have been deselected and removed from stewardship."

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF19 Continuity of Service is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData underscores the need for repositories to ensure long-term sustainability and prevent service interruption through structured maintenance and continuity planning. "Long-term sustainability and maintenance of databases is critical to ensure continued access to data and tools; mechanisms must be put in place to guarantee that curated data and metadata remain available to the community."

Marrano et al., 2025: Repositories are encouraged to implement succession planning and establish formal mechanisms for continuity through trusted institutional or national archives. "Repositories should plan for succession by ensuring that data and metadata can be transferred to trusted archives or institutional repositories when services evolve or close."

Šestak & Copot, 2023: The authors associate continuity with repository resilience, emphasizing long-term institutional commitment and redundant hosting arrangements. "A sustainable agri-data ecosystem requires long-term institutional commitment, redundancy in data hosting, and management plans that can withstand policy or funding disruptions."

Ali & Dahlhaus, 2022: The paper points to the necessity of integrating repositories into resilient data infrastructures capable of maintaining access and services through coordinated management and technical partnerships. "Reliable access to hydrological and agricultural data requires consistent infrastructure and continuity mechanisms across institutions to ensure the ongoing availability of shared data services."

Biomedical Sciences

The TTRAM Activity/Function AF19 Continuity of Service is addressed in the following Biomedical Sciences resources:

Sansone & Rocca-Serra, 2016: Business models are mentioned as a need to tackle sustainability.

Funded by the European Union

Yilmaz et al., 2011: The standard's core data is securely and stably maintained in a relational database at the Max Planck Institute on behalf of the Genomic Standards Consortium, ensuring versioning and future programmatic access to guarantee its sustained utility and accessibility.

Climate Science

To ensure Continuity of Service for the core scientific metadata, **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) states that the organizational infrastructure acknowledged that Controlled Vocabulary (CV) must evolve, improve, and be reused in future projects, thus requiring a plan for its long-term preservation. This continuity and succession planning involves the intention to establish an international governance committee under IS-ENES2 to manage the future evolution, preservation, and ongoing maintenance of the CV and its associated information pipeline.

Linguistics

For CLARIN B centres, the following requirement applies (Wittenburg, Van Uytvanck, Zastrow, Straňák, et al., 2023):

- 2.d Continuity of access and funding support. Requirement: Each centre needs to make explicit statements about perspectives of continuity of access and funding support to continue activities as an active CLARIN centre.

Social Sciences

Best practice: “The Data Preservation Alliance for the Social Sciences (Data-PASS) is a voluntary partnership of organizations created to archive, catalog, and preserve data used for social science research” (Data-PASS, 2025, p. 1). The Data-PASS Memorandum of Understanding states that one partner’s materials can be transferred to one of the other partners in case they can no longer retain their collections and have no other successor (see Data-PASS, 2025, Section ‘9. Transfer Protocols’).

3.3.8. EXTERNAL ENGAGEMENT (AF20)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) External Engagement is described as follows: “Organisational interaction with external parties, including individuals, partner organisations, funders, and other relevant bodies.” and suggests for transparent information “[c]ommunity Skills List, Community Engagement Plan, engagement with policy makers and funders, surveys of depositor and user needs” (L’Hours et al., 2025c, p. 22).

CoreTrustSeal partially covers the A|F External Engagement through its requirement R06 Expertise & Guidance, which requires repositories to “adopt[...] mechanisms to secure ongoing expertise, guidance and feedback-either in-house, or external” (CoreTrustSeal Standards and Certification Board, 2022, pp. 16–17). While the FIDELIS TTRAM A|F covers external engagement at large, the engagement part of the CoreTrustSeal R06 requirement focuses on expertise and guidance, as

Funded by the European Union

expressed in the following items that applicants are asked to address (CoreTrustSeal Standards and Certification Board, 2022, p. 17):

- The repository aligns internal recruitment and external engagement with the services it offers.
- The range and depth of expertise of both the organisation and its staff, including any relevant affiliations (e.g. national or international bodies), is appropriate to the mission.
- In-house advisers, or external advisory committees that include technical, curation, data science, data security, and disciplinary experts. How the repository communicates with experts for advice.

Other aspects of the CoreTrustSeal requirement R06 Expertise & Guidance relate to the A|F People & Expertise (AF17).)

The **Desirable Characteristics of Data Repositories for Federally Funded Research** policy (White House Office of Science and Technology Policy, 2022) emphasizes that repositories must facilitate Broad and Measured Reuse by monitoring the user community and having mechanisms to measure attribution, citation, and reuse of data, which requires ongoing engagement with external users and stakeholders. Additionally, the policy's overall purpose is to provide a consistent set of characteristics for repositories to inform external data repository developers and managers and improve coordination across agencies to enhance compliance and open science infrastructure.

The **Recommendations for Services in a FAIR data ecosystem** presents four recommendations that stand out as being assigned at least medium priority by all, and top priority by two different groups. This is one: Foster global collaboration on FAIR implementation challenges and emerging solutions through organisations such as the Research Data Alliance (Bangert et al., 2019).

The **REPOSITORIES: Key Infrastructure For Maintaining European Research Excellence** (Shearer et al., 2025) states regarding connection to national and international repository networks that actively participating in repository networks and global initiatives enhance visibility, sharing of best practices, and align with evolving open science policies.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) emphasize that repositories depend on a wider partnership of metadata and data services, such as storage providers and multiple registries, which play a critical role in research data infrastructure. Consequently, the guidelines recommend that repositories must transparently share information related to their trustworthiness and services to foster mutual trust between human actors (e.g., researchers or funders) and machine agents, which is a critical precursor to trusted relationships with these external parties.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) recommend repositories submit data citations to DataCite and harvest citations from third-party aggregators such as Crossref, Dimensions, Europe PMC, and NASA ADS repositories submit data citations to DataCite and harvest citations from third-party aggregators such as Crossref, Dimensions, Europe PMC, and NASA ADS, while also aligning their recommendations with external community work like the FORCE11 Data

Funded by the European Union

Citation Principles. Furthermore, GREI encourages ongoing engagement by inviting repositories and interested parties to contribute data citations to the Data Citation Corpus and provide feedback during its development.

The **Core Preservation Process CPP-018 Community Watch** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "The TDA monitors its Designated Community in order to identify its evolving needs and knowledge." (EOSC EDEN T1.2 et al., 2025).

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF20 External Engagement is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData explicitly recommends that repositories establish active communication with their user and depositor communities to strengthen engagement and accountability. "A strong collaborative community of database developers, curators, and users is critical to the success and sustainability of biological databases."

Drakos et al., 2015: agINFRA directs repositories to maintain continuous collaboration with partner institutions, funders, and data providers to ensure services meet the needs of external stakeholders. "Ensure stakeholder needs are met with regards to data management and sharing; involve as many agricultural data sources as possible to provide maximum value."

Sen et al., 2020: WheatIS demonstrates repository-level engagement by coordinating global partners and researchers to maintain interoperable data services and standards. "The Wheat Initiative tasked the WheatIS EWG's to provide the international wheat research community with easy access to wheat genetics, phenotype with environmental information, genomic data and bioinformatics tools, and to support and promote the diverse wheat databases internationally."

Marrano et al., 2025: The authors underline that repositories must develop community engagement plans and communicate with funders to secure sustainable partnerships. "Engagement with funders, community stakeholders, and policy makers is essential for repositories to align objectives and ensure the long-term sustainability of services."

Šestak & Copot, 2023: External engagement is linked to repository resilience and policy impact, emphasizing that communication with policymakers and institutions strengthens sustainability. "A sustainable agri-data ecosystem depends on coordinated action among repositories, research institutions, and policymakers to ensure shared understanding and effective use of open data resources."

Ali & Dahlhaus, 2022: The paper identifies engagement with research networks and environmental agencies as necessary for repositories to maintain relevant and interoperable data services. "The

Funded by the European Union

reliability of agricultural data infrastructures depends on strong engagement with research networks, water authorities, and other agencies that generate or maintain data."

Biomedical Sciences

The TTRAM Activity/Function AF20 External Engagement is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: The article includes community-related indicators, such as size and diversity of the user base, engagement with scientific communities, and responsiveness to user needs. ELIXIR CDRs are often part of international consortia or collaborative networks, and the article highlights the importance of these relationships in sustaining and evolving the resource. The article discusses the role of host institutions and funding bodies, which are external entities essential to the resource's sustainability. The presence of an independent, international Scientific Advisory Board is another form of external engagement, providing strategic input from outside experts.

Field et al., 2011: GSC acts as an international body for external engagement with major data repositories like the International Nucleotide Sequence Database Collaboration (INSDC), as well as collaborating extensively with numerous partner organizations and the wider scientific community through community-led surveys and open calls for contributions to develop, maintain, and ensure the widespread adoption of its MIxS standard.

Barrett et al., 2012: The article highlights NCBI's role within the International Nucleotide Sequence Database Collaboration (INSDC), working alongside DDBJ (Japan) and EBI (Europe). This reflects sustained engagement with global partner organizations. NCBI collaborates with ATCC and Coriell to create Reference BioSample records for widely used biological materials, facilitating standardized reuse across the research community.
The article references the Genomics Standards Consortium, whose MiXS checklists are integrated into BioSample workflows, demonstrating alignment with community-driven metadata standards.

Karsch-Mizrachi et al., 2025: The article describes how INSDC interacts with a broad range of external parties, including current and prospective member organizations, standards bodies, and the wider scientific community. To reflect this commitment, INSDC Executive Committee will establish an ad hoc International Stakeholders Committee (ISC) which will include a diverse group of external experts who are expected to promote the principles and activities of INSDC and provide input on different stakeholders' needs.

Rehm et al., 2021: GA4GH collaborates with external standards development organizations like Health Level Seven (HL7), ISO, and Open Biological and Biomedical Ontology Foundry (OBO). GA4GH also engages with diverse data-hosting environments, to ensure global interoperability and uptake of its secure data sharing solutions.

Climate Science

Funded by the European Union

The METAFOR project: preserving data through metadata standards for climate models and simulations (Callaghan et al., 2010) advocates Community consultation and Intergovernmental Panel on Climate Change (IPCC) involvement.

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) describes the engagement in extensive External Engagement by conducting a wide consultation process with more than 35 climate modeling experts from 13 research centers representing 6 countries to define and establish scientific consensus on the Controlled Vocabulary (CV) content and granularity. This effort, which was supported by the EU 7th Framework Programme and involved collaboration with the US Earth System Curator project, included planning to establish an international governance committee under IS-ENES2 to manage the CV's future evolution and preservation.

The **NetCDF Climate and Forecast (CF) Metadata Conventions** (Eaton et al., 2024) state that the CF framework promotes the processing and sharing of climate and forecast data across diverse sources and applications, and is designed to be backward compatible with other conventions like COARDS while also explicitly incorporating standards such as UGRID to enhance interoperability.

Linguistics

The Activity/Function (A|F) External Engagement is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The CESSDA Resource Directory (CESSDA, 2025) contains links to resources on the topic of “User Support and Communication” that are relevant to the topic of External Engagement (https://www.cessda.eu/Resource-Directory?tree=11). The CESSDA Resource Directory (RD) lists resources with the intention to “help to build sustainable and mature data archives and support the development of new services and features within existing data archives. Information on relevant documents, training materials, tools and support services are collected, selected and reviewed, making the RD a curated inventory of existing resources” (CESSDA, 2025).

3.3.9. RELEASE & PUBLISHING (AF21)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Release & Publishing is described as follows: “The processes by which the organisation creates, manages, approves, releases, updates and withdraws information artefacts (not deposited digital objects) such as publications, policies, guidelines, and other content related to the repository activities and functions. It includes publishing information that provides evidence for external assessments (see AF24 "Criteria, Assessment, Improvement") and supports external engagement (see AF20 "External Engagement").” and suggests for transparent information “[r]ecords management policy, records retention schedule, publication approval process” (L’Hours et al., 2025c, p. 22).

Funded by the European Union

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) generally support the release and publishing activity by recommending that information exposed by repositories should include both internal information artefacts (like policies and procedures) and external information artefacts (like guidance and criteria), which are typically released through publishing processes. Moreover, the guidelines emphasize that repositories should expose information related to trustworthiness and certification status, which necessitates the organizational release of evidence for external assessment.

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF21 Release & Publishing is addressed in the following Agri-Food resources:

Caracciolo et al., 2020: The Agrisemantics group directs repositories to publish semantic artefacts, metadata guidelines, and controlled vocabularies to promote transparency and reuse.

Marrano et al., 2025: Marrano et al. highlight the importance of repositories releasing and maintaining formal records and documentation to evidence compliance and improvement processes. “Repositories are expected to publish and regularly update their internal policies, governance documents, and operational guidelines to provide transparency and facilitate external assessment.”

Šestak & Copot, 2023: Sestak et al. connect release and publishing to open-science communication, advocating for repositories to publicly share reports, assessments, and guidance to ensure credibility and traceability. “Repositories contribute to a sustainable open-data environment by publishing documentation and assessment results that support transparency and accountability across the agri-data ecosystem.”

Ali & Dahlhaus, 2022: The authors refer to the role of repositories in releasing documentation and technical specifications to support data integration within hydrological and agricultural systems. “Documentation and publication of technical standards and access mechanisms are essential for maintaining transparency and reproducibility across environmental and agricultural data infrastructures.”

Biomedical Sciences

The TTRAM Activity/Function AF21 Release & Publishing is addressed in the following Biomedical Sciences resources:

Field et al., 2011: GSC release and publish in the Standards in Genomic Sciences journal. This journal functions as a formal voice for the GSC, supporting the publication of standardized genome, metagenome, and pan-genome reports, and other standards-supportive publications such as Standard Operating Procedures (SOPs)

Funded by the European Union

Karsch-Mizrachi et al., 2025: The article discusses the publication of formal documents such as the Founders Arrangement, the Membership Arrangement, and the Membership Acceptance and Performance Guidelines, which are made available through the INSDC website. These documents provide transparency into the collaboration's governance, policies, and operational expectations, supporting external engagement and enabling external assessment of the repository's practices. The article also mentions updates to the INSDC website, including new sections that describe the mission, governance model, and plans for global participation, further demonstrating the repository's commitment to releasing information artefacts that support its strategic and operational goals.

Rehm et al., 2021: GA4GH releases its products through a multi-stage approval process involving its Work Streams, specialized committees, and the Steering Committee, with development work often conducted publicly on platforms like GitHub.

Climate Science

The Activity/Function (A|F) Release & Publishing is not addressed by any of the reviewed resources from Climate Science.

Linguistics

The Activity/Function (A|F) Release & Publishing is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The Activity/Function (A|F) Release & Publishing is not addressed by any of the reviewed resources from Social Sciences.

3.3.10. INTEROPERABILITY (AF22)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Interoperability is described as follows: "The protocols and processes by which people, processes, technologies and digital objects effectively interact with others, both within and across organisational entities." and suggests for transparent information "Interoperability policy, plan, procedures. Links to and documentation of any machine accessible routes to data and/or metadata, e.g. APIs, OAI-PMH, aligned with Technical Infrastructure (see AF29 "Technical Infrastructure"). Crosswalks between relevant metadata standards and ontologies." (L'Hours et al., 2025c, p. 23)

The **OAIS Reference Model** discusses issues around interoperability in Chapter 6 "Archive Interoperability" (Consultative Committee for Space Data Systems (CCSDS), 2024).

PAIS provides "a standard method for formally defining the digital information objects to be transferred by an information Producer to an Archive and for effectively packaging these objects in

Funded by the European Union

the form of Submission Information Packages (SIPs)" and an XML schema for this purpose (Consultative Committee of Space Data Systems (CCSDS), 2014, Chapter 1, p. 1).

PAIMAS defines the steps and phases by which producers and archives interact. It has the purpose "to identify, define and provide structure to the relationships and interactions between an information Producer and an Archive. This Recommendation defines the methodology for the structure of actions that are required from the initial time of contact between the Producer and the Archive until the objects of information are received and validated by the Archive" (Consultative Committee for Space Data Systems (CCSDS), 2004, Chapter 1, p. 2) . 1-2

The **REPOSITORIES: Key Infrastructure For Maintaining European Research Excellence** (Shearer et al., 2025) states regarding interoperability with scholarly communication and research systems enable seamless data exchange with institutional research information systems (CRIS), funder databases, and scholarly publishing platforms.

The **D1.3 Recommendations for a FAIR EOSC - White Paper of the FAIR-IMPACT Synchronisation Force** (Grootveld, 2025) emphasizes that enhancing these legal and organizational aspects requires scaling up knowledge via a central support program offering training and advice, particularly for Research Infrastructures that often face challenges due to fragmented setups. Furthermore, to advance legal and organizational interoperability, continued collaboration across Horizon Europe projects and the active advocacy of Creative Commons licenses as the default option are key recommendations

The **O'FAIRe makes you an offer: Metadata-based Automatic FAIRness Assessment for Ontologies and Semantic Resources** (Amdouni et al., 2022) assessed by the O'FAIRe tool, the average normalized score for Interoperability (I) was notably low at 42, highlighting that the implementation of standardized protocols and processes for interaction is challenging.

Interoperability is a strong focus of the **FAIRsFAIR project D2.3** report (Behnke et al., 2020), covered by the mention of permanent identifiers as the manifestation of a data policy, the support to standard formats, the intensive use of metadata at the level of the repository itself but also for the digital objects, the granularity being under the responsibility of the repository.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) recommend that transparent information exposed by repositories should take account of, and map to, existing standards and criteria like DCAT and schema.org, and emphasize information exposed by repositories should take account of, and map to, existing standards and criteria like DCAT and schema.org, and emphasize that using DCAT (Data Catalog Vocabulary) is key for fostering interoperability between data catalogues on the web and maximizing machine-actionability for harvesters.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) directly address Interoperability (I), defining several indicators as Important that relate to these organizational processes, such as requiring metadata (RDA-I1-01M) and data (RDA-I1-01D) to use knowledge representation expressed in standardized formats and for metadata (RDA-I2-01M)

Funded by the European Union

to use FAIR-compliant vocabularies. Crucially, while the Interoperable area has 12 indicators, the FDMM notes that Essential indicators are completely absent from this FAIR area, which means a high level of FAIRness could be achieved even in the absence of capacity in these interoperability aspects.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) mandate the standardized storage of citation relationships using DataCite metadata fields (e.g., relatedIdentifier and relationType). Furthermore, the protocols require repositories to actively interact with other entities by submitting collected data citations to DataCite and harvesting information from external sources such as Crossref, Dimensions, and Europe PMC.

In the **D4.5 Report on Completed FAIR Data Standard Adoption and Certifications of Data Repositories in the Region** (Alaterä et al., 2022), it is recommended that data be structured using a formal language suitable for knowledge representation, along with a community-defined ontology or data template. Additionally, it is considered best practice for metadata to use terms that resolve to linked FAIR data, and for the metadata identifier itself to resolve to and utilize FAIR vocabularies, while also linking to third-party resources. Overall, datasets should be potentially representable as Linked Data and include references to other relevant metadata.

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF22 Interoperability is addressed in the following Agri-Food resources:

Drakos et al., 2015: agINFRA defines interoperability as a core mission for agricultural repositories and provides explicit guidance for implementation, through the "interconnect [ion of] agricultural data repositories through extended metadata [and the] advanced implementation and adoption of European standards and specifications."

Sen et al., 2020: WheatIS demonstrates operational interoperability through federated metadata harvesting and standardized data formats.

Caracciolo et al., 2020: The Agrisemantics group insists on semantic interoperability via shared vocabularies and metadata crosswalks.

Harper et al., 2018: AgBioData identifies interoperability as one of its fundamental repository recommendations, promoting adoption of shared formats and APIs.

Marrano et al., 2025: Marrano et al. stress that repositories must document and publish interoperability procedures and crosswalks between metadata schemas. "Repositories should maintain transparent documentation of interoperability protocols and crosswalks among metadata standards and ontologies to ensure consistency across federated infrastructures."

Šestak & Copot, 2023: The authors frame interoperability as a sustainability enabler, recommending coordinated data-exchange standards across Agri-food infrastructures. "A sustainable agri-data

Funded by the European Union

ecosystem requires interoperable repositories built on open protocols and harmonized metadata frameworks to enable seamless information flow between research domains."

Ali & Dahlhaus, 2022: Interoperability is described as a key repository responsibility for connecting agricultural and hydrological data services. "Interoperability between repositories is essential for the integration of environmental and agricultural datasets; standardized metadata and API access ensure discoverability and reuse."

Biomedical Sciences

The TTRAM Activity/Function AF22 Interoperability is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: ELIXIR Core Data Resources (CDRs) are evaluated on their use of community-agreed standards for data formats, metadata schemas, and ontologies. The CDRs are expected to provide programmatic access (e.g., APIs), enabling interaction with other tools, platforms, and services. The article discusses how ELIXIR CDRs often serve as foundational resources for other databases and services. This reach-through dependency highlights the importance of interoperability across organisational entities. Lastly, Interoperability is one of the four FAIR principles, which are central to the evaluation framework. CDRs must ensure that both data and metadata are machine-readable and semantically linked, enabling effective interaction.

Lin et al., 2024: In the article, the authors emphasise the importance for project repositories to apply extra efforts to ensure data adherence to field-specific standards for increased interoperability and reusability. Besides, FAIR Principles are considered as one of the desirable characteristics for all repositories to serve as guidelines to enhance data's discoverability, interoperability, and reusability.

Sansone & Rocca-Serra, 2016: Describes a need to overcome the current fragmented and overlapping efforts due to a lack of central authority and coordination across different organizational types and domains in life science.

Yilmaz et al., 2011: MIxS is compliant with many submission tools e.g., GenBank, EBI-ENA, SRA tools, MetaBar, QIIME, ISA, and designed to enable the seamless interaction and integration of sequence data and contextual metadata across different organizational entities and platforms.

Barrett et al., 2012: BioProject and BioSample records are reciprocally linked to experimental data stored in multiple archival databases, including GenBank, SRA, GEO, and dbGaP, enabling seamless navigation and integration across NCBI resources. These databases are part of the International Nucleotide Sequence Database Collaboration (INSDC), which includes DDBJ (Japan) and EBI (Europe), with data exchanged regularly among partners—demonstrating robust interoperability across organizational boundaries. The BioSample database supports semantic interoperability through the use of controlled vocabularies and standardized checklists, such as the MiXS standards developed by the Genomics Standards Consortium. Additionally, collaboration with external providers like ATCC and Coriell to create standardized sample records further enhances interoperability by enabling consistent referencing and integration of datasets across institutions and platforms.

Funded by the European Union

Karsch-Mizrachi et al., 2025: The article describes how INSDC operates as a coordinated collaboration among three major international organizations—NCBI, DDBJ, and EMBL-EBI—who exchange data daily to ensure each site maintains a complete and synchronized dataset. The article also discusses the alignment of INSDC standards with external bodies such as the Genomic Standards Consortium, PHA4GE, and GA4GH, which facilitates interoperability across different data systems and communities. Furthermore, the development of minimal standards and metadata checklists supports consistent interaction between data submitters, repository systems, and users, enabling effective integration and reuse of data across organizational boundaries.

Wilkinson et al., 2016: The article addresses interoperability for scholarly digital objects (data, algorithms, tools, and workflows) and states they must utilize formal, shared, and broadly applicable languages and vocabularies, along with qualified references to other (meta)data, to enable effective interaction across diverse systems and entities.

Rehm et al., 2021: The activity and function is not explicitly covered in the article. However, technical standards and policy frameworks, such as APIs and data formats are developed by GA4GH. These advocate seamless and responsible interaction among people, processes, technologies, and digital objects across diverse institutional and national boundaries, and facilitate federated data analysis and exchange.

Climate Science

The METAFOR project: preserving data through metadata standards for climate models and simulations (Callaghan et al., 2010) states CIM is designed for interoperability.

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) describes the established interoperability protocols and processes by adopting the CF convention for data formatting and developing a metadata information pipeline to bridge different systems. This pipeline automatically converts the human-readable Controlled Vocabulary (CV) into machine-readable formats like XML and OWL ontologies, enabling tools like the CMIP5 Questionnaire to interact effectively with distributed organizational entities such as the Earth System Grid Federation (ESGF) gateways via broadcast protocols like "atom feeds".

Linguistics

The Activity/Function (A|F) Interoperability is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

It is recommended that repositories include metadata files or tags using a generic schema, as this improves interoperability and increases the likelihood that non-specialised infrastructures can reuse key citation elements and other metadata. (Bornatici et al., 2025, p. 8).

3.3.11. LEGAL & ETHICAL (AF23)

Funded by the European Union

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Legal & Ethical is described as follows: “The legal and ethical obligations that the organisation must be aware of and adhere to.” and suggests for transparent information “[l]egal and/or ethical compliance statements, policies and procedures, ethical guidance lists, applicable legislation lists. (L’Hours et al., 2025c, p. 23)”

The A|F Legal & Ethical has its counterpart in the **CoreTrustSeal** requirement R04 with the same name, requiring repositories to “ensure[...] to the extent possible that data and metadata are created, curated, preserved, accessed and used in compliance with legal and ethical norms”, evidenced by addressing the following items (CoreTrustSeal Standards and Certification Board, 2022, p. 15):

- How the repository identifies and manages relevant legal and ethical standards that impact operations.
- Compliance with specific legal and/or ethical discipline or domain standards.
- Information requested from depositors to confirm that data collection or creation was carried out in accordance with legal and ethical criteria in the relevant geographical location or discipline (e.g. Ethical Review Committee/Institutional Review Board or Data Protection legislation).
- Any data or metadata with disclosure risk e.g. depositor/user information, personal, cultural, or environmental information

Repositories that “hold data or metadata with disclosure risk”, CoreTrustSeal requires applicants to address and provide evidence for the following items (CoreTrustSeal Standards and Certification Board, 2022, p. 15):

- Special procedures applied to manage disclosure risk
- Conditions of distribution, access protection and use
- Processes to review disclosure risk and to take the necessary steps to either anonymize files or to provide access in a secure way
- Staff training in the management of digital objects with disclosure risk.
- Guidance provided on the responsible deposit, download, and use of disclosive or potentially disclosive data and metadata.

The **nestor Seal** addresses this A|F in criterion C7 Legal conformity requiring that the archive “monitors and documents conformity with relevant regulations concerning the ingest, archiving and use of digital objects. These include: data protection, protection of the rights of affected parties, confidentiality regulations, copyright and usage rights, internal and external compliance” (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** the “Establish Standards and Policies” function is part of the Administrative functional entity. The function “is responsible for establishing and maintaining the Archive system standards and policies” and receives input from various other OAIS functional entities (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4, p. 13).

Funded by the European Union

The **EOSC Federation Handbook** briefly addresses the A|F Legal & Ethical, stating that "[i]n the case of sensitive data, [FAIR Data repositories] should be as open as possible but as closed as necessary" (EOSC Association, 2025, p. 32).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** (White House Office of Science and Technology Policy, 2022) state that the repository has documented capabilities for ensuring that administrative, technical, and physical safeguards are employed to comply with applicable confidentiality, risk management, and continuous monitoring requirements for sensitive data. Retention Policy. It further states that the repository provides documentation on policies for data retention.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. It further provides broad, equitable and ideally open access to content free at the point of use, as appropriate, and respects applicable legal and ethical limitations. They meet generally accepted international and national criteria for security to prevent unauthorized access and release of content and have different levels of security, depending on the sensitivity of the data being deposited, to maintain privacy and confidentiality (European Commission, 2024).

The **Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets** (Peng et al., 2022) concerning legal and ethical obligations emphasize that metadata must include essential information about the license under which the data can be reused (RDA-R1.1-01M), as data reuse is otherwise impossible. Achieving a high level of reusability also depends on the essential compliance of both data and metadata with domain-relevant community standards (RDA-R1.3-01M, RDA-R1.3-01D).

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), ten out of 14 respondents provided legal and ethical information.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.1, Contextual Data Quality, that awareness about GDPR concerns surrounding sensitive data, as well as consideration of the CARE principles must increase, especially for disciplines focused in the Humanities and Social Sciences. However, universal tools to ensure GDPR-compliance and/or CARE-adherence are recommended to be rolled out across all disciplines. Some disciplines state that sensitive data is not handled and therefore legal guidelines such as GDPR requirements are not applicable. Similarly, the impact of some research on Indigenous persons societies, environments and spaces may not be immediately recognized, and therefore the CARE requirements are not considered. However, growing novel cross-disciplinary research topics (e.g. medical geology, integrations of computational and medical sciences), as well as the increasing recognition that much scientific research is being conducted within Indigenous societies, environments and spaces, requires widespread adoption of these policies.

Funded by the European Union

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) emphasize that data citations are a critical mechanism to provide credit for the data producer, thereby fulfilling the ethical obligation of recognizing the individual(s) or organization(s) that collected and shared the data used in citing work. Furthermore, promoting data citation increases the rigor and reproducibility of research, aligning with the broader ethical goal of ensuring research integrity and transparent documentation of data sources.

Within AF23 Legal & Ethical, the **Data Curation Network** defines the following Curation Activity (Johnston et al., 2016): "Risk Management: The process of reviewing data for known risks such as confidentiality issues inherent to human subjects data, sensitive information (e.g., sexual histories, credit card information) or data regulated by law (e.g. HIPAA, FERPA) and taking actions to reject or facilitate remediation (e.g., de-identification services) when necessary."

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF23 Legal & Ethical is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData stresses that repositories must define ethical and legal policies to protect data integrity and users' rights while promoting openness. "AgBioData recommends that data repositories adopt clear policies for data use, attribution, and sharing, including licensing terms and ethical considerations for sensitive data."

Marrano et al., 2025: Marrano et al. explicitly recommend that repositories publish legal compliance and ethical-use statements to support accountability and transparency. "Repositories should provide openly accessible statements describing applicable legal obligations, ethical guidelines, and data-use conditions to ensure transparent governance."

Šestak & Copot, 2023: The authors relate legal and ethical awareness to sustainability, arguing that transparent governance includes ethical responsibility for open-data practices. "A sustainable agri-data ecosystem depends on clear articulation of ethical principles and compliance with national and international legal frameworks for data management."

Ali & Dahlhaus, 2022: Legal and ethical compliance is addressed in the context of repository collaboration across jurisdictions, emphasizing licensing and responsible data sharing. "Reliable access to hydrological and agricultural data requires consistent legal frameworks, appropriate licensing, and ethical safeguards for data use."

Biomedical Sciences

The TTRAM Activity/Function AF23 Legal & Ethical is addressed in the following Biomedical Sciences resources:

Funded by the European Union

Durinx et al., 2017: As part of their legal and governance infrastructure, ELIXIR Core Data Resources are expected to maintain transparency and compliance with international standards through publicly available policies. This includes a privacy policy that clearly outlines how personal data and cookies are handled, ensuring users are informed about data protection and security practices. Additionally, resources should have an ethics policy in place that aligns with relevant international standards and best practices, reinforcing their commitment to responsible data stewardship and ethical conduct in service delivery.

Lin et al., 2024: According to the article, repositories holding human data must comply with additional characteristics:

- Fidelity to Consent using documented procedures to ensure access and use of data are consistent with any restrictions imposed by participant consent.
- Restricted Use Compliant enforcing restrictions and ensuring access and use of data are consistent with participant consents.
- Privacy protecting human subjects' data from inappropriate access.
- Plan for Breach handling security breaches and unauthorized access to data.
- Download Controls controlling over data downloads including audits of access.
- Violations addressing the violations of terms of use by user or mismanagement by the repository.
- Request Review establishing and maintaining a transparent process.

Sansone & Rocca-Serra, 2016: The legal framework supporting the ownership, maintenance and development of open content standards is very rudimentary. A new framework is needed to address technical and legal restrictions, particularly related to proprietary vocabularies. Interoperability standards should ensure the availability of information, which includes important aspects such as data permissions, data protection, patient consent, anonymization and encryption.

Rehm et al., 2021: GA4GH addresses legal and ethical obligations via its Regulatory & Ethics Work Stream (REWS), which develops policy frameworks and guidance, including model consent clauses and machine-readable consent standards like the Data Use Ontology (DUO).

Climate Science

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCCs legal and ethical obligations include facilitating proper data citation through DataCite DOI publication, advising on licensing issues like Creative Commons licenses, and enforcing data archiving regulations via submission agreements that also cover authorship and intellectual property.

The **NetCDF Climate and Forecast (CF) Metadata Conventions** (Eaton et al., 2024) state that the CF conventions is dedicated to the public domain, adhering to the Creative Commons Zero v1.0 Universal Deed.

Linguistics

Funded by the European Union

For CLARIN B centres, the following requirement applies (Wittenburg, Van Uytvanck, Zastrow, Straňák, et al., 2023):

- 3.b Privacy statement (unchanged). Requirement: The centre has to implement the GÉANT Data Protection Code of Conduct (DP-CoC) for each of its federated Service Providers.

Social Sciences

Data archives must assess legal and ethical issues before accepting data for distribution for secondary use. If there's a risk that the data includes copyrighted content or sensitive personal information without proper consent, the archive should discuss these concerns with the researcher and may need to reject the dataset. (CESSDA Training Team, 2025a, Chapter 3.2.3 Negotiate data sharing Legal and ethical aspects). If the data archive does not accept personal data, the reviewer should:

- check data for direct personal identifiers,
- ask the researchers if data are completely anonymised or if there is a 'code key',
- assess risks of disclosure created by the detailed background of data becoming available

(CESSDA Training Team, 2025a, Chapter 3.4.2 Control of the submitted material).

Data archiving must comply with national and international regulations. Curators should pay special attention to intellectual property rights and personal data, ensuring that both the data and related documentation are properly anonymised. (CESSDA Training Team, 2025a, Chapter 4.4.1 Risks for integrity (checks for compliance with General Data Protection Regulation-GDPR))

3.3.12. CRITERIA, ASSESSMENT, IMPROVEMENT (AF24)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Criteria, Assessment, Improvement is described as follows: "The processes used to assess compliance with standards, policies, and procedures, along with identifying actions for continuous and targeted improvement." and suggests for transparent information "Change Management procedures, Self-Assessments, Certification" (L'Hours et al., 2025c, p. 23).

The **EOSC Federation Handbook**, puts forward the following recommendations for FAIR Data repositories in the EOSC Federation when it comes to the A|F Criteria, Assessment, Improvement (EOSC Association, 2025, p. 32):

- Measure the FAIRness of their data through FAIR metrics
- Demonstrate compliance to the FAIR principles by implementing at least the Findable and Accessible guidelines
- Have a (documented) plan for the continuous improvement of the FAIR compliance of the data in the repository in order to approach full compliance with the FAIR principles in their domain e.g. via a dedicated committee or community effort tasked with improving the FAIRness of the data by reviewing metadata standards, with regular review processes.

Funded by the European Union

EOSC-federated data repositories are also strongly encouraged to be certified by a community-endorsed certification scheme, such as CoreTrustSeal, Nestor Seal, ISO 16363 certification (EOSC Association, 2025, p. 33).

The **Current State and Future Directions for Open Repositories in Europe** (Shearer et al., 2023) states that most repositories do not make use of existing certification frameworks, either because the assessment process is too resource intensive or existing requirements are deemed as unattainable. An alternative, perhaps lighter-weight self- assessment framework may be more widely applicable for the majority of repositories, perhaps based on the "COAR Community Framework for Good Practices in Repositories" 9 (which has already been adapted into a self- assessment tool in the Japanese context). Certification patterns are aligned at the national level - that is, there are a few countries where certification rates are much higher - therefore certification may be most effectively propagated via national agencies or communities.

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.4, Trustworthy Digital Archive Capabilities, that awareness of repository certification must be increased within the target researcher communities.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) propose that a repository should execute FAIR assessments on its entire holdings to gain insights into overall scores and identify areas that need improvement. Furthermore, exposing metadata related to certification requirements (like those of CoreTrustSeal) allows Validation Authorities to harvest this information, which simplifies the assessment process for trustworthiness and reduces administrative overhead.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that CoreTrustSeal provides a structured framework for assessment. Repositories used self-assessment to identify gaps and improve practices. Although some participants opted for self-assessment without intention to apply for formal certification due to resources available or not being in scope for CoreTrustSeal, the process enabled them to improve their documentation, policies and procedures. Sustainable management of trust goes beyond assessment, evaluation and certification while these are valuable in demonstrating trustworthiness to users and stakeholders.

The **FAIR Data Maturity Model. Specification and Guidelines** (FAIR Data Maturity Model Working Group, 2020) directly support assessment and improvement by providing a common set of indicators and priorities designed to normalize FAIRness assessment across methodologies. Furthermore, the FDMM proposes evaluation methods, such as "Measuring progress," specifically intended to help data providers and publishers conduct self-assessments and determine where practices can be improved to achieve a higher level of FAIRness.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) outline a set of best practices and specific metadata standards for data citation that repositories should implement, thus providing criteria for compliance. The initiative itself represents an action for improvement within

Funded by the European Union

the scholarly ecosystem by focusing on the development of the Data Citation Corpus, a centralized resource intended to compile and make data citation information readily available and openly accessible to the community.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** report (Kleemola et al., 2022) states that CoreTrustSeal certification is used to assess and improve repository practices. Repositories benefited from self-assessment and peer support to enhance documentation, metadata quality, and compliance with standards.

Within AF24 Criteria, Assessment, Improvement, the **Data Curation Network** defines the following Curation Activity (Johnston et al., 2016): "Repository Certification: The technical and administrative capacities of the repository undergo review through a transparent and well-documented process by a trusted third-party accreditation body (e.g., TRAC, or Data Seal of Approval)."

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF24 Criteria, Assessment, Improvement is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData calls for repositories to evaluate and refine their data management practices through community standards and feedback cycles. "AgBioData recommends that databases assess and update their practices regularly to ensure data quality, interoperability, and long-term usability."

Caracciolo et al., 2020: The Agrisemantics group encourages repositories to iteratively assess semantic quality and adopt improvement mechanisms aligned with FAIR. "We recommend systematic monitoring and evaluation of semantic artefacts and their use in Agri-food data repositories to ensure ongoing FAIR compliance."

Marrano et al., 2025: Marrano et al. emphasize repositories' responsibility to document and publish assessment results and certification outcomes to demonstrate compliance and progress. "Repositories are expected to undertake regular self-assessment against international standards, document improvement actions, and publish results to support transparency."

Šestak & Copot, 2023: The authors link continuous improvement with sustainability, recommending that repositories monitor performance indicators and evolve with user and policy demands. "A sustainable agri-data ecosystem relies on regular evaluation of repository performance, ensuring adaptation to policy changes, user needs, and technological advances."

Ali & Dahlhaus, 2022: The paper points to assessment processes as necessary for repository credibility, particularly in multi-institutional data infrastructures. "Continuous assessment of data quality and repository practices is essential to maintain trust and usability across integrated environmental and agricultural data systems."

Funded by the European Union

Biomedical Sciences

The TTRAM Activity/Function AF24 Criteria, Assessment, Improvement is addressed in the following Biomedical Sciences resources:

Sansone & Rocca-Serra, 2016: Tracking the lifecycle of content standards is considered challenging, but initiatives like BioSharing use four specific indicators (Ready, Under Development, Uncertain, Obsolete) to assess a standard's readiness for implementation or use.

Field et al., 2011: GSC fulfills its mission by fostering widespread adoption and collaboration across the scientific community.

Yilmaz et al., 2011: MIxS standard is defined through GSC for core contextual data requirements and the use of specific ontologies for consistent reporting. Compliance is automatically assessed and validated by major databases like GenBank, EBI-ENA, and SRA via provided tools and templates.

Karsch-Mizrachi et al., 2025: The article explains that INSDC has developed a Maturity Model that uses a series of objective criteria to increase levels of maturity in all of the categories in each of the five areas:

- Governance and Institutional Context
- Technical Infrastructure
- Data Operations
- Communications and Engagement
- Quality

Rehm et al., 2021: GA4GH processes are piloted by its twenty four Driver Projects providing feedback. Interoperability testing is carried out through initiatives like FASP that identify areas for specification updates.

Climate Science

The **Development and exploitation of a controlled vocabulary in support of climate modelling** (Moine et al., 2014) describes established rigorous processes for assessing compliance by incorporating a mindmap validator and Schematron-based validation within the metadata information pipeline, ensuring that the collected climate model descriptions adhered to predefined encoding rules, CIM syntax, and parameter coherency. For continuous improvement and targeted enhancement of the metadata standards, the project planned to establish an international governance committee to manage the evolution and preservation of the controlled vocabulary, recognizing the necessity of reinvesting lessons learned from the complexity of the initial harvesting procedure into future projects.

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC's organizational infrastructures multi-step submission process includes Technical Quality Assurance by data management to verify metadata consistency and data integrity against defined criteria like CF Conventions.

Funded by the European Union

Linguistics

CLARIN B and E centres are required "to [...] participate in a quality assessment procedure as proposed by the CoreTrustSeal or the nestor Seal" (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023, p. 3). CLARIN B centres "cannot be certified [...] until the CoreTrustSeal assessment has been successfully concluded (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023).

Social Sciences

The CESSDA Resource Directory (CESSDA, 2025) contains links to resources on the topic of "Organisation: Certification" that are relevant to the topic of Criteria, Assessment and Improvement (https://www.cessda.eu/Resource-Directory?tree=17,21). The CESSDA Resource Directory (RD) lists resources with the intention to "help to build sustainable and mature data archives and support the development of new services and features within existing data archives. Information on relevant documents, training materials, tools and support services are collected, selected and reviewed, making the RD a curated inventory of existing resources" (CESSDA, 2025).

3.3.13. ANALYSIS & IMPACT (AF25)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Analysis & Impact is described as follows: "The analysis of internal data and external information to verify and demonstrate that the organisation is effectively fulfilling its mission (see AF02 "Mission & Scope")." and suggests for transparent information "[b]usiness information analyses, identification and validation of external impact. Dashboards or interfaces for metrics and usage information." (L'Hours et al., 2025c, p. 24)

In the **OAIS Reference Model** functions relevant to this A|F are part of the Administration functional entity, which issues report requests to other OAIS functional entities and receives reports from these in return (Consultative Committee for Space Data Systems (CCSDS), 2024, Chapter 4.2.3.6).

The measurement of repository impact and digital object is also mentioned in the **FAIRsFAIR project D2.3** report (Behnke et al., 2020), with the suggestion to use a publication tracker for associated datasets and allow citation of reuse of partial data or single elements of datasets.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) center on the mandate to implement data metrics that enable reporting on the reach and impact of NIH-funded research data. Data citations are identified as a crucial component of these measures, signaling dataset usage and providing valuable evidence for research evaluation frameworks.

Within AF25 Analysis & Impact, the **Data Curation Network** defines the following Curation Activity (Johnston et al., 2016): "Use Analytics: Monitor and record how often data are viewed, requested, and/or downloaded. Track and report reuse metrics, such as data citations and impact measures for the data over time."

Funded by the European Union

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF25 Analysis & Impact is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData urges repositories to demonstrate their impact through documented community uptake and usability of their databases. “Metrics on data use, user engagement, and citation should be collected to demonstrate the value and impact of databases within the broader agricultural genomics community.”

Drakos et al., 2015: agINFRA calls for monitoring repository outcomes to verify how interoperability and shared services contribute to community impact.

Caracciolo et al., 2020: The Agrisemantics group links impact assessment with FAIR uptake, recommending that repositories measure how semantic resources improve findability and reuse. “We recommend systematic evaluation of the impact of semantic artefacts and FAIR implementation to ensure repositories contribute effectively to Agri-food data integration.”

Marrano et al., 2025: Marrano et al. emphasize that repositories should analyze internal performance data and publish dashboards to demonstrate mission fulfillment. “Repositories are encouraged to monitor and publicly report indicators of usage, data quality, and service efficiency through dashboards or open reporting tools.”

Ali & Dahlhaus, 2022: Repositories are positioned as measurable components in hydrological and agricultural data infrastructures, where impact is reflected in accessibility and integration. “The performance of agricultural and hydrological data repositories can be evaluated through their interoperability, accessibility, and contribution to evidence-based environmental management.”

Biomedical Sciences

The TTRAM Activity/Function AF25 Analysis & Impact is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: Understanding the impact of ELIXIR Core Data Resources involves examining their unique contributions to the scientific ecosystem. A useful approach is the counterfactual perspective—considering what would happen if the resource had never existed or were to disappear without replacement. Many ELIXIR CDRs are globally unique, and their absence would significantly disrupt dependent resources and scientific workflows. These resources also play a vital role in accelerating science, by setting standards, promoting data and software reuse, enhancing research efficiency, and enabling the extension of technical solutions across disciplines. To communicate their value effectively, resources may use translational figures—familiar metrics or examples that resonate with stakeholders and help illustrate the resource’s core function and broader relevance.

Funded by the European Union

Lin et al., 2024: The article outlines a list of commonly collected repository metrics. According to the authors, these metrics are essential for monitoring the scientific impact of data usage, thereby supporting the continued operation and success of a repository. Moreover, they offer systematic parameters for evaluating the costs and benefits—essentially the return on investment—for various stakeholders, including managers, research institutions, funding agencies, and research communities. While data metrics are a key component of repository metrics, the two serve distinct purposes. Repository metrics are aggregate indicators that reflect the overall access, usage, and impact of the repository's services across all hosted data. They provide a holistic view of the repository's value and influence. In contrast, data metrics focus on individual datasets, offering granular insights into their reuse, value, and alignment with FAIR Principles over time.

Yilmaz et al., 2011: MIxS is based on analysis of community needs and existing data gaps identified through surveys. The widespread adoption of these standards by major data providers and the INSDC serves as a demonstration of their success.

Karsch-Mizrachi et al., 2025: The article describes how INSDB is fulfilling its mission by documenting its growth in data volume, its alignment with international standards, and its efforts to expand membership and representation.

Rehm et al., 2021: GA4GH's Federated Analysis System Project tests implementations against real-world scenarios, with learnings feeding back to refine specifications and ensure interoperability and solutions. Community engagement through the Genomics in Health Implementation Forum (GHIF) helps verify that standards meet actual data sharing needs.

Climate Science

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states that WDCC ensures it effectively fulfills its mission of long-term data accessibility and re-usability through Technical Quality Assurance (TQA).

Linguistics

The Activity/Function (A|F) Analysis & Impact is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The CESSDA Resource Directory (CESSDA, 2025) contains links to resources on the topic of "Organisation: Monitoring" that are relevant to the topic of Analysis & Impact (https://www.cessda.eu/Resource-Directory?tree=17,18). The CESSDA Resource Directory (RD) lists resources with the intention to "help to build sustainable and mature data archives and support the development of new services and features within existing data archives. Information on relevant documents, training materials, tools and support services are collected, selected and reviewed, making the RD a curated inventory of existing resources" (CESSDA, 2025).

3.3.14. TRAINING (AF26)

Funded by the European Union

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Training is described as follows: “Leveraging internal expertise to train data and metadata producers, owners, depositors, and users. Training can cover the entire digital object management lifecycle, from the conception of research to the reuse of data and metadata. It also includes training for internal staff, as well as for peer organisations, partners, and third parties.” and suggests for transparent information “[l]ink to primary location where the organisation provides training materials and/or details of training programmes. (L’Hours et al., 2025c, p. 24)”

The A|F Training is partially covered by the **CoreTrustSeal** requirement R06 Expertise & Guidance (“The repository adopts mechanisms to secure ongoing expertise, guidance and feedback - either in-house, or external.”), which among other things asks the repository to provide evidence that it “ensures that its staff have access to ongoing training and professional development” (CoreTrustSeal Standards and Certification Board, 2022, pp. 16–17).

The **EOSC Federation Handbook** addresses the A|F Training in its chapter on research training (5.2.6), which stresses that “[t]raining material for research services and other services is essential for scientists to use them properly” and lists the following inclusion criteria for training resources to be registered in the EOSC Federation (EOSC Association, 2025, p. 36):

- Specify the learning outcomes, resource type (e.g. recorded lesson, textbook, activity plan, etc.), content resource type (e.g. video, slides, audio, etc.), and estimated duration (e.g. estimated work hours).
- Be in at least one of the European languages except from metadata information, which shall be available in English.
- Incorporate information about the expected level of training and expertise to be achieved (beginner, intermediate, advanced, all) and required qualifications to access the training resource.

The EOSC Federation Handbook also encourages providers to use the Quality Assurance Certification Framework produced by Skills4EOSC.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) recommend that guidance should be incorporated directly into Data Management Planning (DMP) tools. This guidance helps users make informed choices about data deposit concerning access, storage, curation, and preservation from the earliest stage of their research.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) explains that training was provided through webinars, documentation, and one-on-one support. Repositories gained skills in self-assessment and certification preparation.

Funded by the European Union

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF26 Training is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData directly calls for repositories to provide training and documentation to curators and users. “Training users and curators to adopt metadata standards and best practices is essential to ensure data quality and consistency across AgBioData repositories.”

Drakos et al., 2015: agINFRA promotes training as part of repository collaboration, ensuring users and partner institutions can effectively use shared services. “Provide tools, training and support for collaboration between European institutions in data-intensive research, validating the approach by enabling users to interact with each other and the data.”

Caracciolo et al., 2020: The Agrisemantics group links training with the adoption of semantic technologies, encouraging repositories to educate staff and users in metadata and ontology management. “Promote training in semantics and metadata to ensure repositories can effectively implement and maintain FAIR-compliant systems.”

Marrano et al., 2025: Marrano et al. stress that repositories should document and publish training programmes and materials for transparency and replication. “Repositories should provide openly accessible training materials and descriptions of their training programmes for curators, depositors, and users.”

Šestak & Copot, 2023: The authors connect training to repository sustainability, suggesting that education strengthens capacity for open data stewardship. “Sustainable agri-data infrastructures depend on continuous training and knowledge sharing among repository staff and the broader research community.”

Ali & Dahlhaus, 2022: The paper highlights training in data management and interoperability as key to maintaining repository quality and coordination across environmental and agricultural data systems. “Training and documentation are necessary to ensure consistent data management practices and interoperability across distributed repositories.”

Biomedical Sciences

The TTRAM Activity/Function AF26 Training is addressed in the following Biomedical Sciences resources:

Durinx et al., 2017: Under the Quality of Service indicators, the article asks whether the resource undertakes training, alongside helpdesk support and user feedback mechanisms. The article emphasizes engagement with the scientific community, which may include training as part of outreach and capacity building. However, there is no reference to lifecycle-based training (e.g., from

Funded by the European Union

data creation to reuse), nor mention to internal staff training, or training for third-party partners or metadata producers.

Sansone & Rocca-Serra, 2016: Need to foster collaboration beyond the pharmaceutical and biotech industries, including others key stakeholders such as publishers, librarians. Need for education, documentation, hackathons, training and courses materials (and events) targeting both producers and consumers of standards, and set to create a new career path.

Climate Science

The Activity/Function (A|F) Training is not addressed by any of the reviewed resources from Climate Science.

Linguistics

The Activity/Function (A|F) Training is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

CESSDA Training Resources provides a wide range of training material focused on key aspects of research data, including data discovery, data management, data analysis, and data preservation. The platform offers content in various formats such as webinars, presentations, slides, and videos. (CESSDA Training Team, 2025b).

The final CESSDA recommendation (8.) for data repositories regarding data citation encourages raising awareness and enhancing user guidance about data citation for data users and data producers. This can be done through, for example, general information, dataset documentation, teaching materials, and seminars. (Bornatici et al., 2025, pp. 8–9).

3.3.15. RESEARCH & DEVELOPMENT (R&D) (AF27)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Research & Development (R&D) is described as follows: “Projects and other activities outside current operations, routine maintenance and upgrades. This can include new approaches to data, metadata, business processes, technology, security, and research infrastructure.” and suggests for transparent information “[d]ocumentation of project, specification, development and delivery processes. Links to information about projects the organisation is involved with. How products move from ‘in development’ to ‘in production’.” (L’Hours et al., 2025c, p. 25)

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) focus on establishing standardized data citation practices and implementing data metrics to report on the reach and impact of NIH-funded research data across the ecosystem. This includes supporting the development of the Data Citation Corpus, which is a collaborative project by DataCite to create a centralized

Funded by the European Union

resource that compiles data citations from various sources, addressing existing challenges in consistency and accessibility.

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF27 Research & Development (R&D) is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData positions repositories as innovation platforms, encouraging active participation in projects that develop new standards and tools for agricultural genomics. “AgBioData encourages collaboration and innovation among database developers to pilot new technologies and standards that improve interoperability and data reuse.”

Caracciolo et al., 2020: The Agrisemantics group calls on repositories to engage in R&D that advances semantic technologies, ontologies, and FAIR data alignment. “We recommend repositories participate in semantic research and development to test and deploy new metadata standards and ontologies for the Agri-food domain.”

Marrano et al., 2025: Marrano et al. underline the importance of documenting R&D processes and ensuring that prototype services are transitioned into production systematically. “Repositories should make R&D activities transparent, describing the development process, testing stages, and how prototypes move to production environments.”

Šestak & Copot, 2023: The authors link R&D to sustainability, recommending that repositories continuously innovate to adapt to new technologies and policy contexts. “A sustainable agri-data ecosystem relies on ongoing research and development in repository technology, ensuring adaptation to emerging scientific and policy challenges.”

Ali & Dahlhaus, 2022: R&D is highlighted as necessary for improving repository technologies and integration capabilities across environmental and agricultural systems. “Research and development are essential to enhance interoperability, develop new data services, and maintain repository relevance within multidisciplinary infrastructures.”

Biomedical Sciences

The TTRAM Activity/Function AF27 Research & Development (R&D) is addressed in the following Biomedical Sciences resources:

Sansone & Rocca-Serra, 2016: New funding frameworks need to be created to provide catalytic support for activities necessary to: research new or apply existing methods to develop, extend, refine and harmonize interoperability standards, and also related tools and educational material.

Funded by the European Union

Field et al., 2011: GSC actively leads the creation and evolution of new data and metadata standards, such as the extension Minimum Information About a Marker Gene Sequence (MIMARKS). R&D includes establishing new business processes and technology by collaborating with major public databases.

Climate Science

The **Development and exploitation of a controlled vocabulary in support of climate modelling** effort (Moine et al., 2014) established a sophisticated information pipeline and tool chain to automatically convert the newly developed controlled vocabulary (CV) from human-readable mindmaps into machine-readable formats like XML and OWL ontologies, ensuring data preservation, reuse, and extensibility.

Linguistics

The Activity/Function (A|F) Research & Development (R&D) is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The Activity/Function (A|F) Research & Development (R&D) is not addressed by any of the reviewed resources from Social Sciences.

3.4.Technology

Technology addresses the repository's technical infrastructure, including storage, integrity measures, and IT service management, to support reliable and scalable digital object handling and interoperability.

3.4.1. STORAGE & INTEGRITY (AF28)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Storage & Integrity is described as follows: "The methods used to store and replicate digital objects' data and metadata, including the validation of replicated copies and the restoration of data from backups in case of errors." and suggests for transparent information "[s]torage Documentation, integrity measures used, approach to integrity checking" (L'Hours et al., 2025c, p. 26).

The A|F Storage & Integrity has its equivalent in the **CoreTrustSeal** requirement R14 (CoreTrustSeal Standards and Certification Board, 2022, pp. 24–25), which requires repositories to apply "documented processes to ensure data and metadata storage and integrity", which means that repositories "[i]n addition to maintaining 'archival' copies of digital objects [...] need to store data and metadata from the point of deposit, for curation and preservation, and for access by users", and that "[f]or each storage location, measures should be in place to ensure that unintentional or unauthorised changes can be detected and correct versions of data and metadata recovered". To

Funded by the European Union

assess this requirement, CoreTrustSeal applicants are asked to address and provide evidence for the following items:

- Processes and documents to ensure that the repository staff have a clear understanding of all storage locations and how they are managed.
- The repository's strategy for multiple copies.
- The risk management techniques used to inform the strategy.
- Procedures for handling and monitoring deterioration of storage media.
- Procedures to ensure that data and metadata are only deleted as part of an approved and documented process.
- Any checks (i.e. fixity checks) used to verify that a digital object has not been altered or corrupted from deposit to use.

The **nestor Seal** addresses this A|F through "C15 Integrity: Functions of the archival storage", requiring that "archival storage provides functions necessary for checking and maintaining the integrity of the representations by the administration of the digital archive. The functions include recording of the archival information packages onto storage media, long-term storage, restoration of the archival information packages and all changes to the packages" (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** this A|F is covered in particular by the Archival Storage functional entity (chapter 4.2.3.4) as well as the Common Services described in chapter 4.2.3.2 (see "Security Services" in particular).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** (White House Office of Science and Technology Policy, 2022) state that the repository has a plan for long-term management of data, including maintaining integrity, authenticity, and availability of datasets; has contingency plans to ensure data are available and maintained during and after unforeseen events.

The **EU Annotated Grant Agreement** states repositories displays specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term (European Commission, 2024).

The **REPOSITORIES: Key Infrastructure For Maintaining European Research Excellence** (Shearer et al., 2025) states that resilient & secure against cyber threats and technological changes can be mitigated through ensuring regular backups, infrastructure monitoring, and compliance with best practices in long-term data integrity and protection.

In **Repositories and Beyond: Analysis of Survey for SSHOC Organisations** (Ala-Lahti et al., 2022), four out of the 11 repositories did not use checksums. It was not possible to ascertain the exact stage at which the checksums were applied, since some work is needed to define the terminology used to describe data integrity measures.

Funded by the European Union

The **D3.1—Report on Discipline Requirements and Needs** (Andreassen et al., 2025) states in section 4.3.4, Trustworthy Digital Archive Capabilities, that formal policies and strategies to ensure data backups, retention policies, long-term financial planning, and exit strategies are insufficient for most disciplines and must be established. Backups are an important task for many repositories, and concrete technical implementation plans (with options for transfer to third parties if necessary) that support data package transfer across all disciplines must be rolled out. Few infrastructures operate under sustainable funding models, where some repositories may receive funding from one or two close partner institutions, but at the same time lack a long-term strategy for operational upkeep and maintenance that can be in conflict with their guaranteed retention period. Formal exit strategies with concrete plans, partners, and workflows must be established in order to prevent data loss and additionally have grave consequences for researchers.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) primarily through the context of external dependencies and planning. Specifically, the guidelines emphasize that repositories depend on a wider partnership of services, including storage providers, which are critical components of research data infrastructure. Furthermore, to support proper management of this activity, guidance related to storage (alongside access, curation, and preservation) should be incorporated into Data Management Planning (DMP) tools to help users make informed deposit choices from the earliest stage of their research.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) explains that outsourced services such as consolidated storage (bit-level integrity) can be offered in the same way across disciplines.

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) states that to uphold Storage and Integrity, repositories must perform integrity checks on resources regularly to detect unauthorized changes or accidental damage and must keep at least one copy of the contents stored in a separate location from the original repository. The repository should also record the checksum when a resource is submitted or modified and utilize a business continuity plan that outlines the response procedures for handling cyber-attacks or natural disasters, supporting the restoration aspect of integrity.

The **Core Preservation Process CPP-001 Checksum Generation and Recording** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA [...] records checksums for every File.” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-002 Checksum Validation** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA validates checksums against those stored in the Information Package at Ingest or Access.” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-003 Integrity Checking** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA supports periodic integrity checking, reporting any damaged or missing Files.” (EOSC EDEN T1.2 et al., 2025).

Funded by the European Union

The **Core Preservation Process CPP-004 Data Corruption Management** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA replaces damaged Files from replicated copies and reports on actions taken.” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-011 Replication** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA automatically manages the replication of Information Packages to multiple storage locations (potentially in different geographical locations).” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-027 File Repair** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “As a way of handling validation errors or rendering errors, the TDA corrects structural issues identified in its Files.” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-030 Refreshment** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA replaces the Information Packages on a storage medium, copying them to a new medium and discarding the old one at the end of a storage medium's life cycle.” (EOSC EDEN T1.2 et al., 2025).

Within AF28 Storage & Integrity, the **Data Curation Network** defines the following Curation Activities (Johnston et al., 2016):

- “File validation: A computational process to ensure that the intended data transfer to a repository was perfect and complete using means such as generating and validating file checksums (e.g., test if a digital file has changed at the bit level) and format validation to ensure that file types match their extensions.”
- “File Audit: Periodic review of the digital integrity of the data files and taking action when needed to protect data from digital erosion (e.g., bitrot) and/or hardware failure.”
- “Secure Storage: Data files are properly stored in a well-configured (in terms of hardware and software) storage environment that is routinely backed-up and physically protected. Perform routine fixity checks (to detect degradation or loss) and provide recovery services as needed.”

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF28 Storage & Integrity is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData stresses the importance of maintaining integrity through reliable infrastructure and documentation. “Databases should adopt data validation and version control practices to ensure that stored information remains consistent, reproducible, and trustworthy.”

Sen et al., 2020: WheatIS highlights federated data storage as essential to maintaining the integrity and accessibility of distributed repositories.

Funded by the European Union

Marrano et al., 2025: Marrano et al. recommend documenting repository storage architectures, backup procedures, and integrity verification processes. “Repositories should make storage documentation publicly available, describing backup frequency, replication strategies, and checksum validation for data integrity.”

Šestak & Copot, 2023: The authors link integrity with sustainability, stressing that redundant storage and verification underpin long-term repository resilience. “Reliable storage and periodic integrity verification are fundamental to the resilience of agricultural data infrastructures, preventing data loss and maintaining community trust.”

Ali & Dahlhaus, 2022: The paper emphasizes the need for integrated storage and integrity strategies. “Reliable access to hydrological and agricultural data requires robust storage systems, replication mechanisms, and integrity validation across institutions.”

Biomedical Sciences

The TTRAM Activity/Function AF28 Storage & Integrity is addressed in the following Biomedical Sciences resources:

Lin et al., 2024: Repositories should ensure Security and Integrity of data/metadata using documented and appropriate measures.

Sansone & Rocca-Serra, 2016: Optimal interoperability for digital objects requires standardized identifiers and descriptions (metadata) implemented by services capable of storing and managing these objects, including versioning. A key conclusion is the significant lack of adequate tools and services to manage and validate compliance with content standards.

Yilmaz et al., 2011: The MIxS standard itself, which defines the reporting guidelines, is maintained in a relational database system at the Max Planck Institute for Marine Microbiology Bremen, providing a secure and stable mechanism for updating and versioning the standard's checklists.

Karsch-Mizrachi et al., 2025: The article does not provide extensive technical detail. However, it explains that INSDC members exchange data daily to ensure that each site maintains a complete and synchronized copy of the repository.

Climate Science

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states WDCC fixes data entities to remain unchanged after publication. Data integrity is maintained through Technical Quality Assurance checks which verify data and metadata accessibility, correct file sizes, and consistency between variable descriptions and data, with a system that supports versioning and new editions for updates rather than direct file modification.

The **NetCDF Climate and Forecast (CF) Metadata Conventions** (Eaton et al., 2024) state that the CF convention is designed to promote the processing and sharing of files created with the NetCDF Application Programmer Interface, enabling software tools with powerful extraction, regridding, and

Funded by the European Union

display capabilities. The conventions prioritize making datasets self-describing, easily parsable by programs, and maintaining backward compatibility with previous versions and other standards like COARDS.

Linguistics

The Activity/Function (A|F) Storage & Integrity is not addressed by any of the reviewed resources from Linguistics.

Social Sciences

The Activity/Function (A|F) Storage & Integrity is not addressed by any of the reviewed resources from Social Sciences.

3.4.2. TECHNICAL INFRASTRUCTURE (AF29)

Domain-agnostic resources

The **FIDELIS TTRAM** defines the repository Activity/Function (A|F) Technical Infrastructure is described as follows: "The comprehensive provision of hardware, software, and IT service management to support the organisation's entity. Monitoring the user community technical needs and broader technical landscape to update and maintain levels of technical service (see AF08 "ReUse")." and suggests for transparent information "[t]echnical Infrastructure Statement, Technical Infrastructure Plan, IT Service Management methodologies, 'Service Catalogues' (cf: FitSM), repository-related software and source code. (L'Hours et al., 2025c, p. 26)"

The A|F Technical Infrastructure is covered by the **CoreTrustSeal** requirement R15 (CoreTrustSeal Standards and Certification Board, 2022, pp. 25–26), which requires the repository to be "managed on well-supported operating systems and other core infrastructural software and hardware appropriate to the services it provides to its Designated Community", to be documented with evidence for the following items:

- The repository software used for deposit, curation, preservation and access management. Whether it is community supported, open source, or locally developed.
- Any IT service management approach followed and the functions this approach specifies (e.g. systems documentation, software inventories, code repositories, infrastructure development planning).
- Any international, community or other technical infrastructure standards in place and how compliance is monitored.
- The version control systems used for repository generated software.
- Measures taken to ensure that availability, bandwidth, and connectivity are sufficient to meet the needs of the Designated Community.
- Processes in place to monitor and manage the need for technical change, including in response to the changing needs of Preservation (R10), and Reuse (R13) by the Designated Community.

Funded by the European Union

The **nestor Seal** addresses this A|F in “C33 IT infrastructure” (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** this A|F is covered by the “Common Services” (chapter 4.2.3.2) which include Operating System services, Network services and Security.

The **Desirable Characteristics of Data Repositories for Federally Funded Research** (White House Office of Science and Technology Policy, 2022) state that the repository has a plan for long-term management of data, building on a stable technical infrastructure and funding plans.

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term (European Commission, 2024).

The **Current State and Future Directions for Open Repositories in Europe** (Shearer et al., 2023) states that many repositories have difficulty keeping up with the newest version of their software platform, which can have an effect on the service provision, as requirements change and user expectations evolve. In addition, there is an inherent tension in the repository ecosystem where - on the one hand - there is a need to ensure widespread interoperability and maintain ease of upgrades by not introducing special functionalities - and on the other hand - being responsive to the needs of various local and national communities that request certain tailored services (for example, local languages). Maintaining this balance can present a challenge for repositories, as they seek to provide a high quality service to their local communities while maintaining a modern repository platform.

In the **FAIRsFAIR project D2.3** report (Behnke et al., 2020), it is recommended to connect computing infrastructures and data repositories to avoid commuting data.

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) emphasize that some FAIR assessment metrics, particularly those associated with technical implementations (e.g., file formats and metadata schemas), will evolve over time. Therefore, digital objects not undergoing active preservation to address changes to technology risk a deteriorating FAIR score over time. objects not undergoing active preservation to address changes to technology risk a deteriorating FAIR score over time.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that infrastructure is a key component of certification. Repositories must demonstrate reliable systems for ingest, storage, access, and preservation.

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) states that Technical Infrastructure should be built on well-supported, open source software and provide a responsive, mobile-friendly user interface. Furthermore, the repository must support standardized technical protocols such as OAI-PMH for

Funded by the European Union

harvesting metadata and SWORD for mediated submission, while ensuring resources are stored in machine-readable, non-proprietary formats to facilitate reuse.

The **GREI Data citation best practices for repositories** (Puebla et al., 2024) dictate the required technical infrastructure by mandating the consistent use of DataCite metadata fields to store citation relationships. This necessary technical capacity must support the submission of collected data citations to DataCite and enable the harvesting of citation data from various external sources and aggregators, such as Crossref and Dimensions.

The **Core Preservation Process CPP-006 AIP Batch Export** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “As part of an exit strategy, the TDA batch exports Information packages and all associated Metadata in a manageable format/structure for Ingest into another TDA.” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-008 File Format Identification** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA identifies file formats to the appropriate level of precision, based on an existing registry (IANA MIME types, PRONOM, etc.).” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-009 Metadata Extraction** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA extracts characteristics (such as size, image dimensions, video codec, audio run time, creating application).” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-010 Format Validation** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA validates Files against File format specifications.” (EOSC EDEN T1.2 et al., 2025).

The **Core Preservation Process CPP-016 Metadata Ingest and Management** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): “The TDA ingests and manages all required Metadata including Metadata appropriate for specific content types (e.g. geospatial, audio visual).” (EOSC EDEN T1.2 et al., 2025).

Within AF29 Technical Infrastructure, the **Data Curation Network** defines the following Curation Activity (Johnston et al., 2016): “Technology Monitoring and Refresh: Formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements of the digital infrastructure and hardware storing the data.”

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF29 Technical Infrastructure is addressed in the following Agri-Food resources:

Funded by the European Union

Harper et al., 2018: AgBioData emphasizes that repositories need scalable, sustainable technical infrastructure to support growth and FAIR data management. “Databases should adopt modern, scalable technical architectures that allow long-term maintenance and community access, ensuring continued usability of data and tools.”

Caracciolo et al., 2020: The Agrisemantics group recommends that repositories adopt technical infrastructures capable of supporting semantic interoperability and FAIR-aligned tools. “Repositories should implement and maintain infrastructures that support machine-actionable metadata, persistent identifiers, and semantic alignment with community standards.”

Marrano et al., 2025: Marrano et al. stress transparent documentation of repositories’ technical infrastructure, including hardware, software, and service management frameworks. “Repositories are encouraged to publish their technical infrastructure plans, outlining software environments, IT service management, and updates to ensure long-term reliability.”

Ali & Dahlhaus, 2022: The authors highlight that repositories supporting agricultural and environmental data need integrated, cloud-based infrastructure for stability and scalability.

Biomedical Sciences

The TTRAM Activity/Function AF29 Technical Infrastructure is addressed in the following Biomedical Sciences resource:

Karsch-Mizrachi et al., 2025: The article does not provide detailed specifications of hardware or IT service management. However, the described infrastructure and coordination among members demonstrate a robust technical foundation that supports the repository’s operations and responsiveness to evolving scientific and technological demands.

Climate Science

The **WDCC User Guide for Data Publications** (Long Term Archive (LTA) group, 2024) states WDCC technical infrastructure includes hosting on systems like Levante and supporting open-source data formats like NetCDF, GRIB, and Zarr, along with the MetaXA graphical user interface for metadata submission. This is complemented by an IT service management, offering user guidance and automated technical quality assurance.

Linguistics

CLARIN B and E centres are required “to have a proper and clearly specified repository system” (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023, p. 3).

Social Sciences

The CESSDA Resource Directory (CESSDA, 2025) contains links to resources on the topic of “Technical Infrastructure” that are relevant to this A|F (https://www.cessda.eu/Resource-Directory?tree=28). The CESSDA Resource Directory (RD) lists resources with the intention to “help to build sustainable

Funded by the European Union

and mature data archives and support the development of new services and features within existing data archives. Information on relevant documents, training materials, tools and support services are collected, selected and reviewed, making the RD a curated inventory of existing resources" (CESSDA, 2025).

3.5.Security

Security focuses on safeguarding digital assets and systems through comprehensive security policies, risk management, and compliance, ensuring trust in repository operations.

3.5.1. SECURITY (AF30)

Domain-agnostic resources

The **FIDELIS TTRAM**: defines the repository Activity/Function (A|F) Security is described as follows: "Ensuring security across the organisation's infrastructure, digital object management, and systems. Security also extends to the boundary between the repository and external entities, including users, depositors and dependencies on third parties." and suggests for transparent information "Information Security Statement, Information Security Plan, related certifications (e.g. ISO27001)" (L'Hours et al., 2025c, p. 27).

The A|F Security has its equivalent in the **CoreTrustSeal** requirement R14 (CoreTrustSeal Standards and Certification Board, 2022, pp. 26–27), which requires repositories to "protect[...] the facility and its data, metadata, products, services, and users", and asks applicants to address and provide evidence for the following items:

- The levels of security required for different data and metadata and environments, and how these are supported.
- The IT security system, employees with roles related to security (e.g. security officers), and any risk analysis approach in use.
- Measures in place to protect the facility. How the premises where digital objects are held are secured.
- Any security-specific standards the repository references or complies with.
- Any authentication and authorization procedures employed to securely manage access to systems in use.

The **nestor Seal** addresses aspects of this A|F in "C34 Security", which requires the organisation and the infrastructure protect the digital archive and its archived information objects and representations (nestor Certification Working Group, 2025).

In the **OAIS Reference Model** this A|F is covered by the "Common Services" (chapter 4.2.3.2) which include Operating System services, Network services and Security. In addition, Physical Access Control is a function in the Administration functional entity (chapter 4.2.3.6).

The **Desirable Characteristics of Data Repositories for Federally Funded Research** (White House Office of Science and Technology Policy, 2022) state that the repository has documented measures in

Funded by the European Union

place to meet well established cybersecurity criteria for preventing unauthorized access to, modification of, or release of data, with levels of security that are appropriate to the sensitivity of data (e.g., the NIST Cybersecurity Framework: https://www.nist.gov/cyberframework).

The **EU Annotated Grant Agreement** states repositories display specific characteristics of organisational, technical and procedural quality, such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. Trusted repositories have specific provisions in place and offer explicit information online about their policies, which define their services (e.g. acquisition, access, security of content, long-term sustainability of service including funding, etc). They meet generally accepted international and national criteria for security to prevent unauthorized access and release of content and have different levels of security, depending on the sensitivity of the data being deposited, to maintain privacy and confidentiality (European Commission, 2024).

The **Update of the Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements** states the lack of a public policy for preservation, curation and security of the contents is the most frequent reason, followed by not adhering to a specific metadata standard (Lazzeri, 2024).

The **REPOSITORIES: Key Infrastructure For Maintaining European Research Excellence** (Shearer et al., 2025) states that resilient & secure against cyber threats and technological changes can be mitigated through ensuring regular backups, infrastructure monitoring, and compliance with best practices in long-term data integrity and protection.

For depositing personal data that cannot be anonymised, the exclusion criteria from the **Research Data College Working Group** rule out subject-specific repositories that are “located outside the European Union [...], with the exception of Switzerland, Great Britain, Japan and Argentina, which are deemed to be GDPR- compliant” (Lamotte et al., 2024, p. 11).

The **M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes** (Verburg et al., 2023) specify that repositories rely on a wider partnership of data and metadata services, such as storage providers and registries, and therefore the security provision must explicitly extend to the boundary between the repository and these external entities. Transparent information related to security, such as an Information Security Statement, an Information Security Plan, and related certifications like ISO27001, should be exposed by repositories.

The **D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories** (Kleemola et al., 2022) states that information security is another challenge when outsourcing, particularly for repositories that store sensitive digital objects and/or significant amounts of personal data (e.g., data concerning producers, repositories, researchers using the datasets etc.). A breach of information security (even when the repository is not responsible) can result in the repository sustaining significant reputational damage and possible legal consequences.

Funded by the European Union

The **COAR Community Framework for Good Practices in Repositories, Version 2** (Confederation of Open Access Repositories, 2022) states that ensuring security in repositories requires the application of security practices to prevent unauthorized manipulation of resources and the regular performance of integrity checks to detect unauthorized changes or accidental damage to digital objects. Security across the organization's infrastructure is also supported by maintaining a business continuity plan that details the response and procedures for handling cyber-attacks or natural disasters.

The **Core Preservation Process CPP-007 Virus Scanning** sets the following good-practice baseline expectation for Trustworthy Digital Archives (TDA): "Information packages are virus checked, with appropriate facilities for quarantine." (EOSC EDEN T1.2 et al., 2025).

Domain-specific resources

Agri-food

The TTRAM Activity/Function AF30 Security is addressed in the following Agri-Food resources:

Harper et al., 2018: AgBioData underlines the need for security policies that balance openness with data protection. "Repositories must define access control mechanisms and data security policies to ensure that sensitive information is protected while maintaining data availability for legitimate users."

Caracciolo et al., 2020: The Agrisemantics group links repository security to responsible stewardship of semantic and metadata resources. "Ensuring reliable and secure access to semantic resources is essential for maintaining their integrity and long-term usability within Agri-food repositories."

Marrano et al., 2025: Marrano et al. recommend formal documentation of information security procedures and external certification. "Repositories should maintain and publish an Information Security Statement, describing access control, risk management, and, where applicable, ISO 27001 or equivalent certification."

Šestak & Copot, 2023: The authors connect security with repository sustainability and resilience. "Secure data infrastructures are foundational to the resilience of agri-data ecosystems, protecting data integrity and ensuring continuity of service."

Ali & Dahlhaus, 2022: Security is discussed in the context of shared environmental–agricultural infrastructures. "Secure access and controlled sharing mechanisms are essential for maintaining trust between repositories and data providers in integrated agricultural systems."

Biomedical Sciences

The TTRAM Activity/Function AF30 Security is addressed in the following Biomedical Sciences resources:

Funded by the European Union

Lin et al., 2024, emphasize the critical importance of ensuring the security and integrity of data and metadata in repositories. They advocate for the use of documented and appropriate measures, considering these essential characteristics of trustworthy repositories. For repositories that manage human data, the authors highlight the need for a comprehensive breach response plan to address potential security incidents and unauthorized access. Furthermore, the article promotes the TRUST Principles, particularly the role of technology in delivering secure, persistent, and reliable services. Repositories are encouraged to adopt relevant and effective technologies and practices to mitigate security threats. Notably, technology infrastructure and security are identified as one of the three core assessment areas in evaluating repository trustworthiness and in certification standards.

Climate Science

The Activity/Function (A|F) Security is not addressed by any of the reviewed resources from Climate Science.

Linguistics

CLARIN B and E centres are required "to join the national identity federation where available and join the CLARIN service provider federation to support single identity and single sign-on operation based on SAML2.0 and trust declarations. In case all resources at a centre are open, setting up a Service Pro

For **CLARIN** B and E centres, the following requirements apply when it comes to the AF Security (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023, p. 3):

- Centres need to adhere to the security guidelines, i.e. the servers need to have accepted certificates.
- Centres need to join the national identity federation where available and join the CLARIN service provider federation to support single identity and single sign-on operation based on SAML2.0 and trust declarations. In case all resources at a centre are open, setting up a Service Provider is optional.

For CLARIN B centres, the following requirement applies (Wittenburg, Van Uytvanck, Zastrow, Straňák, et al., 2023):

- 4. Server Certificates. Requirement: Centres need to adhere to the security guidelines, i.e. the servers need to have accepted certificates.
- 5. Federated Identity Management. [Note: if a centre only provides and will provide fully open resources, this requirement is not applicable]. Requirement: Centres need to join the national identity federation where available and join the CLARIN service provider federation to support single identity and single sign-on operation based on SAML2.0 and trust declarations.
- 9. Attribute Checker (optional). Requirement: Centres can opt to configure the Shibboleth SP Attribute Checker which assists in case of failed logins.
- 10. Attribute Aggregator (optional). Requirement: Centres can opt to configure the Attribute Aggregator provided by Lindat which provides insights into failed login attempts.

Funded by the European Union

Social Sciences

The Activity/Function (A|F) Security is not addressed by any of the reviewed resources from Social Sciences.

3.6. Other capabilities and characteristics

Some frameworks include other capabilities and characteristics than the ones in the FIDELIS TTRAM. These are briefly addressed in this section.

The **Data Curation Network** includes peer-review among its curation activities and describes it as follows (Johnston et al., 2016): "The review of a data set by an expert with similar credentials and subject knowledge as the data creator for the purposes of validating the soundness and trustworthiness of the file contents."

Within the field of **Linguistics**, CLARIN B centres are required to have to be recognized by a CLARIN ERIC member or observer country or have to be established as a third party (Wittenburg, Van Uytvanck, Zastrow, & Offersgaard, 2023).

The exclusion criteria from the **Research Data College Working Group** rule out subject-specific repositories that practice an excessive pricing policy, i.e., "repositories where every low-volume data deposit automatically incurs a fee [...]", whereas "repositories that may require a financial contribution in return for depositing significant volumes of data (more than 50 GB) have not been excluded" (Lamotte et al., 2024, p. 11).

4. Analysis and Discussion

This chapter discusses key aspects of the reviewed and mapped repository capabilities and characteristics presented in Chapter 3. In Section 4.1, we briefly describe the characteristics of the reviewed resources. Section 4.2 summarises identified commonalities across the reviewed resources, whereas Section 4.3 presents specificities that stand out across resources.

4.1. Resource Characteristics

During the work on this report, a total of 85 resources were reviewed and added to the mapping spreadsheet. Of these, 75 were included in the present report.

4.1.1. Network Beacon Communities

The resource mapping conducted in Task 5.2 of the FIDELIS project reviewed a total of 75 resources, which were categorized by their relevance to the six designated network beacon communities. These communities represent key scientific domains within the FIDELIS network and serve as focal points for aligning repository practices with the EOSC framework. The distribution of resources across the network beacon communities is as follows:

- Agri-food: 10 resources
- Biomedical Sciences: 4 resources

Funded by the European Union

- Climate Science: 4 resources
- Linguistics: 7 resources
- Social Sciences: 8 resources
- Generic (domain-agnostic): 33 resources
- Cross-domain resources (relevant to multiple communities):
    - Biomedical Sciences, Climate Science, Linguistics, Social Sciences, Physical Sciences, Other: 1 resource
    - Biomedical Sciences, Generic: 1 resource
    - Linguistics, Social Sciences, Other: 2 resources
    - Other discipline, Biomedical Sciences: 4 resources
    - Social Sciences, Generic: 1 resource

Notably, while the Physical Sciences community was not directly represented due to capacity limitations, it was indirectly covered through cross-domain resources. The Generic category, comprising 33 resources, reflects the strong presence of domain-agnostic standards, frameworks, and best practices applicable across all repository types.

This distribution highlights the broad engagement of the FIDELIS project with both domain-specific and cross-cutting repository practices, ensuring that the resulting recommendations and frameworks (e.g., TTRAM) are grounded in a diverse and representative evidence base.

4.1.2. Resource Types

The reviewed resources can be categorized into the following **types**, often with overlapping classifications:

- **Standards**: Formalized technical or procedural specifications (e.g., MIxS, CF Metadata Conventions, OAIS-RM)
- **Recommendations**: Community or project-based guidance (e.g., FAIRsFAIR D2.3, COAR Community Framework)
- **Best Practices**: Descriptive or prescriptive practices from domain-specific initiatives (e.g., AgBioData, WheatIS)
- **Modes of Federation**: Descriptions of federated repository models (e.g., CLARIN Centre Registry, INSDC)
- **Solutions**: Technical or organizational tools and frameworks (e.g., TTRAM, AgroDataCube, GA4GH APIs)
- **Certification Bodies and Frameworks**: CoreTrustSeal, nestor
- **Guidelines and Handbooks**: EOSC Federation Handbook, FAIR Data Maturity Model, EU Annotated Grant Agreement
- **Surveys and Landscape Analyses**: SSHOC, EOSC-Nordic, FAIRsFAIR, OpenAIRE/COAR studies

These resource types were mapped against the **30 TTRAM Activities and Functions (AF01–AF30)**, providing a structured overview of how each resource related to repository capabilities and characteristics in the following five main areas:

Funded by the European Union

- Context
- Digital Object Management
- Organisational Infrastructure
- Technology
- Security

The remainder of Section 4 provides a comprehensive analysis and discussion of the mapping results, organised according to the five main TTRAM areas.

4.2.Context

The Context area of the FIDELIS Transparent Trustworthy Repository Attributes Matrix (TTRAM) encompasses foundational elements that define a repository's identity, mission, and scope. This includes how repositories present themselves (AF01: Identification & Contact) and articulate their purpose and responsibilities (AF02: Mission & Scope). The analysis of 75 resources across domain-agnostic and domain-specific contexts reveals both shared practices and community-specific nuances.

4.2.1. Identification & Contact (AF01)

Commonalities: Across all Network Beacon Communities (NBCs), a strong emphasis is placed on transparency in repository identification. Domain-agnostic standards such as **CoreTrustSeal** (R0) and the **FAIRsFAIR D2.3** report recommend the inclusion of persistent identifiers (e.g., DOIs, ROR IDs), clear contact information, and registration in trusted registries like **re3data**. These practices are foundational for discoverability and trust.

Specificities by NBC:

- **Agri-food** repositories, as highlighted in **Harper et al. (2018)** and **Drakos et al. (2015)**, stress the importance of naming responsible individuals and providing direct contact methods. Registration in agricultural aggregators like **CIARD RING** and **AgriVIVO** is also emphasized.
- **Biomedical Sciences** resources such as **Durinx et al. (2017)** and **Lin et al. (2024)** underscore the need for detailed repository metadata, including organizational affiliations and support contacts, especially for ELIXIR Core Data Resources.
- **Linguistics** repositories, particularly **CLARIN** centres, are required to visibly reference their affiliation with CLARIN and register in the **Centre Registry**, which feeds into **re3data**.
- **Social Sciences** repositories, as per **CESSDA** guidelines, are encouraged to register in multiple registries (e.g., **OpenDOAR**, **FAIRsharing**) and provide persistent identifiers.

Resource Types: Standards and best practices dominate this AF, with **CoreTrustSeal**, **FAIRsFAIR**, and **OAIS-RM** providing comprehensive guidance. Domain-specific recommendations often take the form of best practices and community guidelines.

4.2.2. Mission & Scope (AF02)

Funded by the European Union

Commonalities: Most repositories, regardless of domain, articulate a mission aligned with long-term preservation, open access, and FAIR principles. The **CoreTrustSeal R01**, **nestor Seal**, and **OAIS Reference Model** provide structured expectations for defining mission, designated communities, and scope of services. The **EU Annotated Grant Agreement** and **FAIRsFAIR** further emphasize the importance of transparency in repository responsibilities and alignment with community needs.

Specificities by NBC:

- **Agri-food** repositories, such as those discussed by **Drakos et al. (2015)** and **Harper et al. (2018)**, often tailor their missions to specific crops, data types, or stakeholder groups. The emphasis is on aligning repository goals with agricultural research and FAIR data sharing.
- **Biomedical Sciences** resources, notably **Durinx et al. (2017)** and **Lin et al. (2024)**, distinguish between deposition databases and knowledge bases, with missions focused on supporting clinical and genomic research. Repositories are expected to define their scientific scope, user communities, and data types (e.g., nucleotide sequences, clinical data).
- **Linguistics** repositories, particularly **CLARIN** centres, must define their role within the infrastructure and specify the services and data they offer to the research community.
- **Social Sciences** repositories, as outlined in the **CESSDA Data Management Expert Guide**, focus on integrating data into the research lifecycle, supporting teaching, learning, and policy-making.

Resource Types: This AF is supported by a wide range of resource types:

- **Standards**: OAIS-RM, CoreTrustSeal, nestor Seal
- **Best Practices**: AgBioData, WheatIS, CLARIN Requirements
- **Recommendations**: EU Annotated Grant Agreement, FAIRsFAIR, Research Data College Working Group

Discussion: The Context area reveals a high degree of alignment across domains in terms of transparency and mission articulation. However, the **specificities** of each NBC – such as the need for domain-specific identifiers, community engagement strategies, and tailored mission statements –highlight the importance of contextualizing repository practices. The **Generic** resources provide a robust foundation, but domain-specific adaptations are essential for meaningful alignment with community expectations and EOSC integration.

4.3.Digital Object Management

The **Digital Object Management** area of the TTRAM encompasses the full lifecycle of digital objects, from their conception and creation to deposit, curation, discovery, access, reuse, and preservation. This area is the most densely populated in the mapping, reflecting its centrality to repository operations. The analysis of 75 reviewed resources reveals a rich landscape of domain-agnostic standards and domain-specific practices, with varying levels of maturity and specialization across the Network Beacon Communities (NBCs).

Funded by the European Union

4.3.1. Commonalities Across Communities and Resource Types

Across all NBCs, several foundational practices are widely adopted:

- **Persistent Identifiers (PIDs)**: The use of DOIs, Handles, and other PIDs is nearly universal (AF06), supporting discovery, citation, and long-term access. Resources such as the **FAIRsFAIR D2.3** report and **CoreTrustSeal R06** emphasize the importance of PID policies and machine-actionable metadata.
- **Metadata Standards**: Repositories commonly adopt community-specific or general standards (e.g., **Dublin Core**, **DataCite**, **DDI**, **CMDI**, **MIxS**, **CF Metadata Conventions**) to ensure interoperability and reuse (AF05, AF06). The **FAIR Data Maturity Model** and **Global Community Guidelines** further reinforce the need for rich, structured, and machine-readable metadata.
- **Access and Reuse Policies**: Most repositories provide clear licensing and access conditions, often aligned with **FAIR**, **TRUST**, and **CARE** principles (AF07, AF08). The **EU Annotated Grant Agreement** and **Desirable Characteristics of Data Repositories for Federally Funded Research** highlight the need for transparent access and reuse frameworks.
- **Curation and Quality Assurance**: Initial and ongoing curation practices are widely implemented, with varying levels of sophistication (AF05). The **CoreTrustSeal R10**, **Data Curation Network**, and **COAR Community Framework** provide structured guidance on quality assurance, metadata enhancement, and compliance with standards.
- **Preservation Planning**: While less uniformly implemented, many repositories articulate preservation strategies, often referencing **OAIS**, **CoreTrustSeal R09**, and the **LoRCaP** model (AF10). The **EOSC Long-Term Data Preservation Task Force** and **M5.2 Guidelines** stress the importance of transparency in preservation levels and actions.

4.3.2. Specificities by Network Beacon Community

Agri-food

Agri-food repositories emphasize early engagement with data producers (AF03), often providing guidance on data management plans, metadata standards, and semantic interoperability. Resources such as **Harper et al. (2018)** and **Drakos et al. (2015)** recommend the use of controlled vocabularies and registration in agricultural metadata aggregators like **CIARD RING** and **AgriVIVO**. The **AgroDataCube** exemplifies advanced API-based access and reuse mechanisms (AF07, AF08), while **Caracciolo et al. (2020)** and **Sen et al. (2020)** stress the importance of semantic standards and federated metadata harvesting (AF05, AF06).

Biomedical Sciences

Biomedical repositories demonstrate high maturity in digital object workflows. Standards such as **MIxS** (Yilmaz et al., 2011), **MIMARKS**, and infrastructures like **INSDC** (Karsch-Mizrachi et al., 2025), **GA4GH** (Rehm et al., 2021), and **BioProject/BioSample** (Barrett et al., 2012) provide detailed guidance on metadata, submission protocols, and reuse. These repositories often manage sensitive data, necessitating tiered access models and secure environments (AF07, AF08). Provenance and

Funded by the European Union

authenticity are rigorously maintained through accessioning systems, audit trails, and versioning (AF11).

Climate Science

Climate repositories, such as **WDCC** and **CMIP5**, focus on standardization and quality control. The **CF Metadata Conventions**, **NetCDF** formats, and **CIM** metadata model are widely adopted (AF05, AF06). The **WDCC User Guide** outlines detailed workflows for data submission, quality assurance, and long-term preservation (AF04, AF10). Controlled vocabularies and metadata validation tools (e.g., **Schematron**) are used to ensure compliance and discoverability (AF06). The **METAFOR project** and **Moine et al. (2014)** highlight the importance of metadata coherence and semantic consistency.

Linguistics

Linguistics repositories, particularly **CLARIN** centres, implement the **CMDI** metadata standard and maintain registries of recommended file formats (AF04, AF06). The **Tromsø Recommendations** and **CLARIN Data Citation Guidelines** provide detailed guidance on citation practices and metadata completeness (AF06, AF08). **CLARIN SIS** distinguishes between recommended, acceptable, and discouraged formats, guiding deposit and appraisal (AF04). However, fewer resources address preservation and provenance in depth.

Social Sciences

Social science repositories, such as those in the **CESSDA** network, emphasize data quality, documentation, and ethical reuse. The **CESSDA Data Management Expert Guide** and **Data Archiving Guide** provide comprehensive support for pre-ingest, deposit, and curation (AF03–AF05). Repositories implement structured access levels (open, restricted, controlled) and support reuse through rich metadata and persistent identifiers (AF07, AF08). The **CESSDA Metadata Model (CMM)** and **DDI** standard are central to metadata practices (AF06), while provenance and versioning are addressed through changelogs and structured file naming (AF11).

4.3.3. Resource Types

The resources reviewed in the mapping exercise span a diverse array of types, each contributing uniquely to the understanding and implementation of trustworthy digital repository practices. These types include standards, best practices, recommendations, solutions, and certification frameworks. Their distribution across the TTRAM activities and Network Beacon Communities reveals both foundational and community-specific contributions to digital object management.

Standards

Standards provide formalized, often internationally recognized, specifications that underpin repository operations:

- **OAIS Reference Model (CCSDS, 2024)**: A foundational framework referenced across multiple TTRAM activities, particularly for preservation (AF10), provenance (AF11), and workflows (AF09).

Funded by the European Union

- **MIxS and MIMARKS (Yilmaz et al., 2011; Field et al., 2011)**: Widely adopted in Biomedical Sciences for metadata completeness and contextual data capture (AF03, AF05).
- **CF Metadata Conventions and NetCDF (Eaton et al., 2024)**: Core to Climate Science repositories for ensuring metadata consistency and machine-readability (AF05, AF06).
- **CMDI (CLARIN)** and **DDI (CESSDA)**: Central to Linguistics and Social Sciences, respectively, for structured metadata and interoperability (AF06).
- **DataCite Metadata Schema** and **Dublin Core**: Used across domains for persistent identification and citation (AF06, AF08).

Best Practices

Best practices offer community-driven guidance on implementing repository functions effectively:

- **AgBioData (Harper et al., 2018)**: Provides detailed recommendations for Agri-food repositories on metadata standards, user support, and sustainability (AF01–AF08).
- **WheatIS (Sen et al., 2020)**: Demonstrates federated metadata harvesting and standardization across distributed Agri-food repositories (AF04–AF06).
- **CLARIN SIS (CLARIN Standards and Interoperability Committee, n.d.).**: Offers a structured approach to file format recommendations, guiding deposit and appraisal processes in Linguistics (AF04).
- **CESSDA Data Management Expert Guide (2022)** and **Data Archiving Guide (2025a)**: Comprehensive resources for Social Sciences, covering pre-ingest, deposit, curation, access, and reuse (AF03–AF12).

Recommendations

These resources synthesize expert consensus and policy guidance to support repository alignment with broader frameworks:

- **FAIRsFAIR D2.3 (Behnke et al., 2020)**: Recommends machine-readable metadata, PID assignment, and FAIR-aligned repository practices (AF03–AF08).
- **FAIR Data Maturity Model (RDA, 2020)**: Provides indicators for assessing FAIRness, particularly in metadata quality, reuse, and provenance (AF05–AF08).
- **EOSC Federation Handbook (EOSC Association, 2025)**: Offers high-level guidance on repository onboarding, FAIR compliance, and technical interoperability (AF06–AF10).
- **Global Community Guidelines (Peng et al., 2022)**: Emphasize quality documentation and domain-relevant standards for reuse and compliance (AF03–AF08).

Solutions

Funded by the European Union

Practical implementations and tools that exemplify repository capabilities:

- **AgroDataCube (Top et al., 2022)**: A model Agri-food repository offering API-based access, standardized formats, and licensing transparency (AF07, AF08).
- **GA4GH APIs (Rehm et al., 2021)**: Enable secure, federated access and reuse of genomic data, with support for data discovery and interoperability (AF06–AF08).
- **WDCC Workflows (Long Term Archive Group, 2024)**: Provide a detailed blueprint for Climate Science repositories on data submission, quality control, and preservation (AF04–AF10).

Certification Frameworks

These define criteria and processes for assessing and certifying repository trustworthiness:

- **CoreTrustSeal (2022)**: Covers a wide range of TTRAM activities, including deposit, appraisal, curation, access, reuse, and preservation (AF04–AF12).
- **nestor Seal (2025)**: Offers detailed criteria for preservation, authenticity, and legal compliance, particularly relevant to long-term digital stewardship (AF04–AF12).

4.3.4. Discussion

The Digital Object Management area reveals a high level of maturity and diversity. While foundational practices such as PIDs and metadata standards are widely adopted, domain-specific repositories often go further, implementing advanced workflows, semantic interoperability, and secure access models. The **Biomedical Sciences** and **Climate Science** communities stand out for their technical sophistication, while **Agri-food** and **Social Sciences** emphasize community engagement and ethical reuse. **Linguistics** repositories contribute significantly to citation and metadata practices, though preservation and provenance are less developed.

This area also illustrates the interplay between **generic** and **domain-specific** resources. Generic standards provide a baseline, but domain-specific solutions are essential for addressing unique data types, user needs, and ethical considerations. The mapping underscores the importance of supporting repositories in adopting both foundational and community-tailored practices to ensure alignment with EOSC and the broader goals of open and FAIR science.

4.4.Organisational Infrastructure

The **Organisational Infrastructure** area of the TTRAM encompasses the governance, policy frameworks, staffing, partnerships, and sustainability mechanisms that underpin the operations of trustworthy digital repositories. This area is critical for ensuring that repositories are not only technically sound but also institutionally robust, transparent, and aligned with the needs of their designated communities. The mapping reveals a strong emphasis on governance and policy frameworks across all Network Beacon Communities (NBCs), with notable differences in how repositories manage resources, expertise, and external relationships.

Funded by the European Union

4.4.1. Commonalities Across Communities and Resource Types

Across all Network Beacon Communities (NBCs), several organisational features are consistently emphasized:

- **Governance Structures (AF13)**: Most repositories operate under clear governance models, often involving host institutions, consortia, or national infrastructures. The **CoreTrustSeal R05** and **nestor Seal C10** require documentation of governance bodies, decision-making hierarchies, and institutional affiliations.
- **Policy and Standards Management (AF14)**: Repositories adopt internal policies for data acquisition, curation, access, and preservation. The **FAIRsFAIR D2.3** report and **FAIR Data Maturity Model** recommend aligning policies with FAIR principles and ensuring machine-actionable documentation.
- **Rights Management (AF15)**: Licensing and rights are managed through deposit agreements, standard licenses (e.g., Creative Commons), and access control mechanisms. The **CLARIN Legal and Ethical Issues Committee** provides model agreements, while **CESSDA** emphasizes GDPR compliance and structured access levels.
- **Staffing and Expertise (AF17)**: Repositories invest in skilled personnel and ongoing training. The **D8.3 Trustworthy Digital Repository status update** highlights the importance of domain-specific expertise for long-term preservation and trust.
- **Continuity of Service (AF19)**: Business continuity and disaster recovery planning are increasingly formalized. The **CoreTrustSeal R03**, **nestor Seal C12**, and **OAIS RM** stress the need for succession planning and risk mitigation.
- **External Engagement (AF20)**: Repositories engage with user communities, funders, and policy bodies. The **EOSC Federation Handbook** and **M5.2 Guidelines** recommend transparent communication and participation in federated infrastructures.

4.4.2. Specificities by Network Beacon Community

Agri-food

Agri-food repositories emphasize collaborative governance and sustainability. **Harper et al. (2018)** advocates for visible and accountable repository teams, while **Drakos et al. (2015)** promotes shared services and standards across institutions. **Caracciolo et al. (2020)** underscores the need for semantic stewardship and policy alignment with FAIR principles. **Marrano et al. (2025)** and **Šestak & Copot (2023)** stress the importance of succession planning and institutional resilience.

Biomedical Sciences

Biomedical repositories demonstrate advanced governance and policy frameworks. **Durinx et al. (2017)** describes ELIXIR Core Data Resources as having independent advisory boards, sustainable funding, and formal governance. **Karsch-Mizrachi et al. (2025)** outlines the INSDC's structured

Funded by the European Union

governance through Executive and Implementation Committees. Rights management is critical, with repositories implementing tiered access and secure environments for sensitive data.

Climate Science

Climate repositories operate within federated infrastructures like **ESGF**, with distributed governance and shared standards. The **METAFOR project** and **CF Metadata Conventions** illustrate community-driven policy development. The **WDCC User Guide** details institutional responsibilities and continuity mechanisms, including DOI assignment and long-term data stewardship.

Linguistics

CLARIN centres follow a rigorous certification process, with requirements for governance, IPR policies, and continuity planning. The **CLARIN Terms of Service**, **Deposition Agreements**, and **End-User Licenses** provide a structured framework for rights management. Centres are also required to make explicit statements about funding and access continuity (AF19).

Social Sciences

CESSDA repositories prioritize legal compliance, policy transparency, and community engagement. The **CESSDA Data Archiving Guide** outlines policies for data acquisition, preservation, and dissemination. The **CESSDA Resource Directory** offers curated tools for governance, staffing, and financial planning. **Bornatici et al. (2025)** and **Kleemola et al. (2025)** recommend extending PID coverage and standardizing metadata across repositories.

4.4.3. Resource Types

The organisational infrastructure of trustworthy digital repositories is supported by a diverse set of resource types. These include formal standards, best practices, recommendations, solutions, and certification frameworks. Each type contributes differently to shaping governance, policy development, staffing, rights management, and sustainability strategies across domains.

Standards and Frameworks

Standards provide structured, often internationally recognized, criteria for assessing and guiding repository operations:

- **CoreTrustSeal**: A widely adopted certification framework that addresses governance (R05), rights management (R02), continuity (R03), and expertise (R06), among others. It is referenced across all NBCs and serves as a benchmark for organisational maturity.
- **nestor Seal**: Offers detailed criteria for governance (C10), funding (C8), personnel (C9), and crisis/succession planning (C12), with particular relevance for repositories in Germany and beyond.
- **OAIS Reference Model**: Provides a conceptual framework for repository functions, including administrative and preservation planning responsibilities (AF13, AF14, AF19).

Funded by the European Union

Best Practices and Guides

Best practices offer practical, community-informed approaches to implementing organisational functions:

- **AgBioData (Harper et al., 2018)**: Recommends visible governance structures, responsive user support, and sustainability planning for Agri-food repositories.
- **CESSDA Data Archiving Guide (2025a)**: Covers policies for data acquisition, preservation, access, and dissemination, with guidance on legal compliance and user engagement.
- **CLARIN Centre Requirements**: Define expectations for governance, IPR handling, continuity of access, and external engagement for certified centres in the Linguistics community.
- **WheatIS (Sen et al., 2020)**: Demonstrates federated governance and shared metadata standards across international Agri-food repositories.

Recommendations

These resources synthesize expert consensus and policy guidance to support alignment with broader frameworks:

- **FAIRsFAIR D2.3 (Behnke et al., 2020)**: Emphasizes the need for machine-readable policies, PID coverage, and alignment with FAIR principles.
- **FAIR Data Maturity Model**: Provides indicators for assessing organisational readiness, including staffing, rights, and policy transparency.
- **EOSC Federation Handbook (EOSC Association, 2025)**: Recommends long-term sustainability planning, alignment with EOSC onboarding criteria, and transparent governance.
- **EU Annotated Grant Agreement (European Commission, 2024)**: Outlines expectations for repository quality, including services, access, security, and sustainability.
- **Research Data College Working Group (Lamotte et al., 2024)**: Offers exclusion criteria for subject-specific repositories, emphasizing openness, rights clarity, and preservation responsibilities.
- **M5.2 Guidelines for Repositories and Registries (Verburg et al., 2023)**: Advocate for exposing trustworthiness metadata, including governance, skills, and preservation levels.

Solutions and Modes of Federation

These represent operational models and infrastructures that embody organisational best practices:

- **INSDC (Karsch-Mizrachi et al., 2025)**: A globally coordinated federation of nucleotide sequence databases with formal governance, shared policies, and sustainability mechanisms.

Funded by the European Union

- **GA4GH (Rehm et al., 2021)**: Provides a structured alliance model with work streams, advisory boards, and driver projects that shape standards and policy frameworks in genomics.
- **CLARIN ERIC**: Offers a legal and organisational structure for Linguistics repositories, supporting certification, policy harmonization, and community engagement.
- **CESSDA ERIC**: Coordinates Social Sciences repositories across Europe, providing shared services, training, and policy development support.

4.4.4. Discussion

The Organisational Infrastructure area reflects the increasing institutional maturity of digital repositories. While domain-agnostic frameworks such as **CoreTrustSeal** and **OAIS** provide a solid foundation, domain-specific repositories have developed tailored governance models, policy frameworks, and engagement strategies.

The **Biomedical Sciences** and **Climate Science** communities exemplify federated governance and international collaboration. **Agri-food** and **Social Sciences** emphasize community engagement and sustainability, often in the context of project-based funding. **Linguistics** repositories benefit from the structured support of **CLARIN ERIC**, which provides templates, legal frameworks, and certification pathways.

A recurring challenge is ensuring long-term sustainability, particularly for repositories reliant on short-term funding. The mapping highlights the need for clear succession planning, transparent governance, and investment in human expertise. External engagement is not only a mechanism for trust-building but also a strategic tool for aligning repository services with evolving community needs and policy frameworks.

4.5.Technology

The **Technology** area of the FIDELIS Transparent Trustworthy Repository Attributes Matrix (TTRAM) encompasses two key Activities and Functions (AFs): **Storage & Integrity (AF28)** and **Technical Infrastructure (AF29)**. These functions are foundational to the operational reliability, scalability, and trustworthiness of digital repositories. The mapping of resources across domain-agnostic and domain-specific contexts reveals both common technical expectations and community-specific implementations.

4.5.1. Storage & Integrity (AF28)

Commonalities:
Across all reviewed resources, repositories are expected to implement robust strategies for ensuring the integrity and long-term preservation of digital objects. Domain-agnostic frameworks such as **CoreTrustSeal**, **OAIS Reference Model**, and the **nestor Seal** emphasize the importance of:

- Redundant storage systems and replication strategies
- Regular fixity checks and integrity verification

Funded by the European Union

- Documentation of restoration procedures and risk mitigation plans

These practices are essential for maintaining the authenticity and usability of digital objects over time. The **CoreTrustSeal LoRCaP** model further distinguishes between deposit compliance and active preservation, highlighting the need for repositories to take long-term responsibility for the renderability and understandability of data and metadata.

Specificities by NBC:

- **Biomedical Sciences** repositories (e.g., ELIXIR CDRs, INSDC) demonstrate strong commitments to long-term data stewardship, often supported by institutional mandates and international collaboration. INSDC, for example, ensures integrity through structured submission systems and validation rules.
- **Climate Science** repositories such as WDCC implement multi-stage quality control processes, including both technical and scientific assurance. Metadata standards like CF Metadata Conventions support self-describing datasets and long-term interpretability.
- **Agri-food** repositories (e.g., AgroDataCube, WheatIS) emphasize the use of open formats and web technologies (e.g., REST, GeoJSON) to ensure accessibility and integrity. Authors stress the importance of maintaining semantic artefacts and metadata consistency over time.

4.5.2. Technical Infrastructure (AF29)

Commonalities:
Technical infrastructure underpins the repository's ability to deliver services reliably and securely. Domain-agnostic resources recommend:

- Scalable and interoperable systems
- Support for machine-actionable metadata and persistent identifiers
- Use of standardized protocols (e.g., OAI-PMH, APIs) for metadata harvesting and federated search
- Documentation of infrastructure components and service continuity plans

Repositories are increasingly expected to align with FAIR and TRUST principles, ensuring that infrastructure supports discoverability, accessibility, and reuse.

Specificities by NBC:

- **Biomedical Sciences** infrastructures (e.g., GA4GH, ELIXIR) rely on APIs and standardized data models to support secure data sharing and federated access. Tools like Beacon and htsget exemplify technical solutions for genomic data interoperability.
- **Climate Science** infrastructures leverage distributed systems such as ESGF for data publication and access. Metadata harvesting tools (e.g., CMIP5 Questionnaire) and controlled vocabularies ensure consistency across nodes.

Funded by the European Union

- **Agri-food** repositories adopt interoperable infrastructures that support semantic alignment and machine-readable metadata. AgroDataCube, for instance, integrates catalogue entries with persistent identifiers and standardized access protocols.

4.5.3. Discussion

The Technology area reflects the increasing sophistication and standardization of repository infrastructure across disciplines. While domain-agnostic frameworks provide essential guidance, domain-specific implementations demonstrate tailored approaches to integrity assurance and technical service delivery.

Repositories in **Biomedical Sciences** and **Climate Science** exemplify federated infrastructures and international coordination, while **Agri-food** repositories highlight the importance of semantic interoperability and open technologies. Across all communities, the emphasis on machine-actionable metadata, persistent identifiers, and standardized protocols supports the FIDELIS goal of enabling federation and discoverability within the EOSC ecosystem.

A recurring challenge is ensuring that technical infrastructure remains adaptable to evolving community needs and technological landscapes. The mapping underscores the need for repositories to document their infrastructure, monitor risks, and plan for continuity and scalability. These practices are essential for building trust and supporting long-term preservation and reuse.

4.6.Security

The **Security** area of the FIDELIS Transparent Trustworthy Repository Attributes Matrix (TTRAM) is represented by a single Activity and Function (AF30), which addresses the repository's ability to protect digital objects and associated metadata from unauthorized access, loss, or corruption. This includes both technical and procedural safeguards, as well as alignment with legal and ethical frameworks such as GDPR and CARE principles.

4.6.1. Commonalities Across Communities and Resource Types

Security is a cross-cutting concern that affects all aspects of repository operations, from ingest and storage to access and reuse. Domain-agnostic frameworks such as **CoreTrustSeal**, **OAIS Reference Model**, and the **EU Annotated Grant Agreement** emphasize the need for:

- Access control mechanisms tailored to data sensitivity
- Documentation of security policies and procedures
- Compliance with legal and ethical standards (e.g., GDPR, CARE)
- Risk management and contingency planning

Repositories must demonstrate that they meet generally accepted international and national criteria for security, and that they implement differentiated levels of protection based on the nature of the data they manage.

4.6.2. Specificities by Network Beacon Community

Funded by the European Union

- **Social Sciences** repositories (e.g., CESSDA) emphasize anonymization, informed consent, and access control as key elements of secure data sharing. The CESSDA Data Management Expert Guide and Data Archiving Guide provide detailed guidance on protecting personal data and ensuring compliance with GDPR.
- **Biomedical Sciences** repositories increasingly address ethical obligations related to sensitive human data, including controlled access environments and authentication protocols. The GA4GH framework and INSDC member practices reflect growing attention to privacy and data governance.
- **Climate Science** repositories (e.g., WDCC) implement technical quality assurance and secure metadata workflows, though explicit security policies are less frequently documented.
- **Agri-food** repositories show emerging awareness of ethical and legal considerations, particularly in relation to Indigenous data and environmental monitoring. However, formal security frameworks are less consistently applied.

4.6.3. Resource Types

The resources mapped under the Security function reflect a combination of technical standards, ethical frameworks, and operational guidelines that support secure repository practices across disciplines.

Standards and Frameworks

- **CoreTrustSeal R03 and R02**: Address continuity of service and rights management, including security provisions and compliance monitoring.
- **OAIS Reference Model**: Defines Access Rights Information as part of Preservation Description Information, ensuring traceability and controlled access.
- **EU Annotated Grant Agreement**: Requires repositories to prevent unauthorized access and release of content, with explicit documentation of security policies.

Best Practices and Guidelines

- **CESSDA Training Materials**: Provide operational guidance on anonymization, informed consent, and access control, particularly for personal and sensitive data.
- **Global Community Guidelines**: Emphasize licensing, metadata completeness, and ethical reuse as prerequisites for secure data management.
- **D3.1 Report on Discipline Requirements and Needs**: Highlights the need for universal GDPR and CARE compliance tools across disciplines, especially for repositories handling sensitive or Indigenous data.

Solutions and Ethical Frameworks

Funded by the European Union

- **CARE Principles**: Increasingly relevant for repositories managing Indigenous data, emphasizing collective benefit, authority to control, responsibility, and ethics.
- **GREI Data Citation Best Practices**: Link ethical obligations to proper attribution, reproducibility, and transparency in data reuse.

4.6.4. Discussion

Security is both a technical and ethical imperative for trustworthy digital repositories. The mapping reveals that while domain-agnostic standards provide a solid foundation, implementation varies significantly across disciplines. Repositories in the **Social Sciences** and **Biomedical Sciences** demonstrate more mature approaches to security, particularly in handling sensitive data and aligning with legal frameworks. In contrast, **Agri-food** and **Climate Science** repositories show uneven adoption of formal security policies, though awareness is growing.

The integration of ethical principles such as CARE, alongside legal compliance with GDPR, is essential for ensuring that repositories not only protect data but also respect the rights and expectations of their designated communities. As cross-disciplinary research expands, repositories must adopt flexible and transparent security models that accommodate diverse data types and user needs.

FIDELIS can support this effort by promoting harmonized security practices, encouraging the adoption of community-specific standards, and facilitating the development of shared tools for compliance and risk management. Security must be embedded in the federation model to ensure trust, sustainability, and responsible data stewardship across the EOSC ecosystem.

Funded by the European Union

5. Summary and Main Implications for FIDELIS

This report presents a detailed mapping of repository capabilities and characteristics across the European digital repository landscape, structured according to the **FIDELIS Transparent Trustworthy Repository Attributes Matrix (TTRAM)**. The mapping exercise, based on desk research, provides a foundational landscape analysis that informs how FIDELIS can support repositories in becoming and remaining **Trustworthy Digital Repositories (TDRs)** aligned with the **European Open Science Cloud (EOSC)**.

5.1.1. Key Findings

The mapping exercise conducted in Task 5.2 has yielded several key findings that illuminate the current landscape of repository capabilities and characteristics.

1. **Comprehensive Coverage of Repository Functions**
   The mapping spans all 30 TTRAM Activities and Functions (AFs), grouped into five thematic areas: Context, Digital Object Management, Organisational Infrastructure, Technology, and Security. This structure enabled a granular analysis of repository practices, standards, and gaps.

2. **Domain-Agnostic and Domain-Specific Insights**
   Resources were reviewed across five Network Beacon Communities – Agri-food, Biomedical Sciences, Climate Science, Linguistics, and Social Sciences – revealing both shared challenges and discipline-specific needs. While some AFs (e.g., Identification & Contact, Discovery & Identification) are well-covered across domains, others (e.g., Training, R&D, Security) show uneven implementation and documentation.

3. **Alignment with Existing Standards and Certification Frameworks**
   The mapping integrates and cross-references key frameworks such as **CoreTrustSeal**, **nestor Seal**, **OAIS Reference Model**, and **FAIR principles**, providing a robust foundation for future certification and interoperability efforts.

4. **Emerging Gaps and Needs**
   Several areas require further attention:

   - **Semantic interoperability** and machine-actionable metadata remain inconsistent.
   - **Preservation planning** and **succession strategies** are often underdeveloped.
   - **Training and support** for repository staff and users is unevenly documented.
   - **Security policies**, especially in domains handling sensitive data, need clearer articulation.

5.1.2. Implications for FIDELIS

Building on the findings of this mapping exercise, several strategic implications emerge for how FIDELIS can support repositories in strengthening trustworthiness, interoperability, and alignment with EOSC.

Funded by the European Union

1. **Informing the Design of the FIDELIS Network**
   The mapping provides a baseline for identifying repository strengths and weaknesses, which can guide the development of targeted support mechanisms within the FIDELIS Network. This includes recommendations for onboarding, certification pathways, and interoperability frameworks.

2. **Enhancing the TTRAM Framework**
   Insights from the mapping suggest refinements to TTRAM, such as clearer definitions for underrepresented AFs, improved guidance for semantic artefact alignment, and integration of FAIR assessment metrics.

3. **Supporting Repository Self-Assessment and Improvement**
   The structured mapping and resource annotations can be repurposed into tools for repository self-assessment, benchmarking, and strategic planning. This aligns with FIDELIS's goal of fostering a culture of continuous improvement and transparency.

4. **Facilitating Cross-Domain Learning and Federation**
   By documenting domain-specific practices and challenges, the report enables cross-pollination of ideas and solutions among disciplines. This supports FIDELIS's mission to federate repositories across scientific domains and EOSC nodes.

5. **Strengthening Engagement with Policy and Standards Bodies**
   The mapping highlights the importance of coordinated engagement with external stakeholders, including standards organizations, funders, and policy makers. FIDELIS can act as a bridge between repositories and these bodies, promoting alignment and reducing duplication of effort.

Funded by the European Union

6. References

Akdeniz, E., & Moilanen, K. (2023). *CMM CESSDA Metadata Model*. https://doi.org/10.5281/ZENODO.7528240

Ala-Lahti, H., Mathers, B. J., L'Hours, H., Kleemola, M., & Alaterä, T. J. (2022). *Repositories and Beyond: Analysis of Survey for SSHOC Organisations*. https://doi.org/10.5281/zenodo.6325149

Alaterä, T., Kleemola, M., Ala-Lahti, H., & Jerlehag, B. (2022). *D4.5 Report on completed FAIR data standard adoption and certifications of data repositories in the region*. https://zenodo.org/records/7303538

Ali, B., & Dahlhaus, P. (2022). The Role of FAIR Data towards Sustainable Agricultural Performance: A Systematic Literature Review. *Agriculture*, *12*(2), 309. https://doi.org/10.3390/agriculture12020309

Amdouni, E., Bouazzouni, S., & Jonquet, C. (2022). O'FAIRe makes you an offer: Metadata-based automatic FAIRness assessment for ontologies and semantic resources. *International Journal of Metadata, Semantics and Ontologies*, *16*(1), 16–46. https://doi.org/10.1504/IJMSO.2022.131133

Andreassen, H. N., Berez-Kroeker, A. L., Collister, L., Conzett, P., Cox, C., De Smedt, K., McDonnell, B., & Group, R. D. A. L. D. I. (2019). *Tromsø recommendations for citation of research data in linguistics*. https://doi.org/10.15497/rda00040

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). *D3.1—Report on Discipline Requirements and Needs*. https://zenodo.org/records/15789261

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., Holl, A., Lamers, D., Lammert, A., L'Hours, H., … Wyns, R. (2023). *Recommendations Consultation. EOSC-A Long Term Data Preservation Task Force*. Zenodo. https://doi.org/10.5281/zenodo.10014698

Bangert, D., Hermans, E., van Horik, R., de Jong, M., Koers, H., & Mokrane, M. (2019). *Recommendations for Services in a FAIR data ecosystem*. Zenodo. https://doi.org/10.5281/zenodo.3585742

Barrett, T., Clark, K., Gevorgyan, R., Gorelenkov, V., Gribov, E., Karsch-Mizrachi, I., Kimelman, M., Pruitt, K. D., Resenchuk, S., Tatusova, T., Yaschenko, E., & Ostell, J. (2012). BioProject and BioSample databases at NCBI: Facilitating capture and organization of metadata. *Nucleic Acids Research*, *40*(D1), D57–D63. https://doi.org/10.1093/nar/gkr1163

Behnke, C., Bonino, L., Coen, G., Le Franc, Y., Parland-von Essen, J., Riungu-Kalliosaari, L., & Staiger, C. (2020). *D2.3 Set of FAIR data repositories features*. https://zenodo.org/records/5361952

Bolton, S. (2022). *CESSDA Data Access Policy* (Version 2). Zenodo. https://doi.org/10.5281/zenodo.6722000

Funded by the European Union

Bornatici, C., Jernung, A., Alaterä, T. J., Tveit Sandberg, L., Strand, K., Štebe, J., & Trtíková, I. (2025). *CESSDA Recommendations on Data Citation: Practical Recommendations for Key Stakeholders*. https://zenodo.org/records/15043854

Callaghan, S. A., Treshansky, A., Moine, M., Guilyardi, E., Alias, A., Balaji, V., Bojariu, R., Cofiño, A. S., Denvil, S., Elkington, M., Ford, R., Kolaninski, M., Lautenschlager, M., Lawrence, B., Steenman-Clark, L., & Valcke, S. (2010). The METAFOR project: Preserving data through metadata standards for climate models and simulations. *Proceedings of the 1st International Digital Preservation Interoperability Framework Symposium*, 1–8. https://doi.org/10.1145/2039263.2039269

Cannon, M., Graf, C., McNeice, K., Chan, W. M., Callaghan, S., Carnevale, I., Cranston, I., Edmunds, S. C., Everitt, N., Ganley, E., Hrynaszkiewicz, I., Khodiyar, V. K., Leary, A., Lemberger, T., MacCallum, C. J., Murray, H., Sharples, K., Soares E Silva, M., Wright, G., … (Moderator) Sansone, S.-A. (2021). *Repository Features to Help Researchers: An invitation to a dialogue*. Zenodo. https://doi.org/10.5281/zenodo.4683794

Caracciolo, C., Aubin, S., Jonquet, C., Amdouni, E., David, R., Garcia, L., Whitehead, B., Roussey, C., Stellato, A., & Villa, F. (2020). 39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition. *Data Science Journal*. https://doi.org/10.5334/dsj-2020-047

CESSDA. (2025). *CESSDA Resource Directory*. https://www.cessda.eu/Resource-Directory

CESSDA Training Team. (2022). *CESSDA Data Management Expert Guide*. https://dmeg.cessda.eu/

CESSDA Training Team. (2025a). *CESSDA Data Archiving Guide*. Version 4.0. https://dag.cessda.eu

CESSDA Training Team. (2025b). *CESSDA Training Resources*. https://www.cessda.eu/Training-Resources

CLARIN Legal and Ethical Issues Committee. (n.d.). *CLARIN Licensing Framework*. Retrieved 4 October 2025, from https://www.clarin.eu/content/clarin-licensing-framework

CLARIN Standards and Interoperability Committee. (n.d.). *CLARIN Standards Information System*. Retrieved 30 April 2025, from https://standards.clarin.eu/sis/

Confederation of Open Access Repositories. (2022). *COAR Community Framework for Good Practices in Repositories*, Version 2. Zenodo. https://doi.org/10.5281/zenodo.7108101

Consultative Committee for Space Data Systems (CCSDS). (2004). *Producer-Archive Interface Methodology Abstract Standard* (Magenta Book). https://ccsds.org/Pubs/651x0m1.pdf

Consultative Committee for Space Data Systems (CCSDS). (2024). *Reference Model for an Open Archival Information System (OAIS)*. CCSDS. https://ccsds.org/wp-content/uploads/gravity_forms/5-448e85c647331d9cbaf66c096458bdd5/2025/01//650x0m3.pdf

Consultative Committee of Space Data Systems (CCSDS). (2014). *Producer-Archive Interface Specification (PAIS). Recommended Standard*. https://ccsds.org/wp-content/uploads/gravity_forms/5-448e85c647331d9cbaf66c096458bdd5/2025/01//651x1b1.pdf

Funded by the European Union

Conzett, P., & De Smedt, K. (2022). Guidance for Citing Linguistic Data. In A. L. Berez-Kroeker, B. McDonnell, E. Koller, & L. B. Collister (Eds), *The Open Handbook of Linguistic Data Management* (p. 143–155). The MIT Press. https://doi.org/10.7551/mitpress/12200.003.0015

Conzett, P., Duvaud, S., Jouneau, T., Kallio, J., Klemetsen, T., Recker, J., Straňák, P., & Tujunen, S. (2025). *Digital repository capabilities and characteristics mapping spreadsheet* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.17471456

CoreTrustSeal Standards and Certification Board. (2022). *CoreTrustSeal Trustworthy Digital Repositories Requirements 2023-2025 Extended Guidance*. https://doi.org/10.5281/zenodo.7051096

CoreTrustSeal Standards and Certification Board. (2024). *Curation & Preservation Levels: CoreTrustSeal Position Paper*. Zenodo. https://doi.org/10.5281/ZENODO.11476980

Currie, A., & Kilbride, W. (2021). *FAIR Forever? Long Term Data Preservation Roles and Responsibilities*, Final Report. Zenodo. https://doi.org/10.5281/zenodo.4574234

Data-PASS. (2025). *Data-PASS Memorandum of Understanding*. https://docs.google.com/document/d/1ygSMnT9C47qKO_UzfXQkldG0d3RjfS_zDAmk3o72Vsk/edit?usp=sharing

Drakos, A., Protonotarios, V., & Manouselis, N. (2015). agINFRA: A research data hub for agriculture, food and the environment. *F1000Research*. https://doi.org/10.12688/f1000research.6349.2

Durinx, C., McEntyre, J., Appel, R., Apweiler, R., Barlow, M., Blomberg, N., Cook, C., Gasteiger, E., Kim, J.-H., Lopez, R., Redaschi, N., Stockinger, H., Teixeira, D., & Valencia, A. (2017). Identifying ELIXIR Core Data Resources. *F1000Research*, 5, 2422. https://doi.org/10.12688/f1000research.9656.2

Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., Pamment, A., Juckes, M., Raspaud, M., Blower, J., Horne, R., Whiteaker, T., Blodgett, D., Zender, C., Lee, D., … Bartholomew, S. L. (2024). *NetCDF Climate and Forecast (CF) Metadata Conventions*. https://zenodo.org/records/14275599

EOSC Association. (2025). *EOSC Federation Handbook*. Zenodo. https://doi.org/10.5281/zenodo.14999577

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). *M1.1 Report on Identification of Core Preservation Processes*. Zenodo. https://doi.org/10.5281/zenodo.16992452

European Commission. (2024). *EU Grants. AGA – Annotated Grant Agreement. EU Funding Programmes 2021-2027*. Version 1.0. European Commission. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf

FAIR Data Maturity Model Working Group. (2020). *FAIR Data Maturity Model. Specification and Guidelines*. https://doi.org/10.15497/rda00050

Field, D., Amaral-Zettler, L., Cochrane, G., Cole, J. R., Dawyndt, P., Garrity, G. M., Gilbert, J., Glöckner, F. O., Hirschman, L., Karsch-Mizrachi, I., Klenk, H.-P., Knight, R., Kottmann, R., Kyrpides, N., Meyer,

Funded by the European Union

F., Gil, I. S., Sansone, S.-A., Schriml, L. M., Sterk, P., … Wooley, J. (2011). The Genomic Standards Consortium. *PLOS Biology*, 9(6), e1001088. https://doi.org/10.1371/journal.pbio.1001088

Grootveld, M. (2025). *D1.3 Recommendations for a FAIR EOSC - White Paper of the FAIR-IMPACT Synchronisation Force*. https://doi.org/10.5281/zenodo.14979704

Harper, L., Campbell, J., Cannon, E. K. S., Jung, S., Poelchau, M., Walls, R., Andorf, C., Arnaud, E., Berardini, T. Z., Birkett, C., Cannon, S., Carson, J., Condon, B., Cooper, L., Dunn, N., Elsik, C. G., Farmer, A., Ficklin, S. P., Grant, D., … Main, D. (2018). AgBioData consortium recommendations for sustainable genomics and genetics databases for agriculture. *Database*, 2018, bay088. https://doi.org/10.1093/database/bay088

Jahn, N., Laakso, M., Lazzeri, E., & McQuilton, P. (2023). *Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements*. Zenodo. https://doi.org/10.5281/zenodo.7728016

Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., & Stewart, C. (2016). *Definitions of Data Curation Activities used by the Data Curation Network*. https://hdl.handle.net/11299/188638

Jouneau, T., Verburg, M., Horton, L., van Geest, G., Tujunen, S., Kallio, J., Paulsen, T., Duvaud, S., Recker, J., Holthe-Tveit, Å. J., Thorpe, D. E., Conzett, P., Kleemola, M., Kalaitzi, V., Huber, R., Jonquet, C., Aguilar, F., Forshaug, A. K., Alaterä, T. J., & Esteves, E. (2025). *FIDELIS landscape survey analysis*. https://doi.org/10.5281/zenodo.15744995

Karsch-Mizrachi, I., Arita, M., Burdett, T., Cochrane, G., Nakamura, Y., Pruitt, K. D., Schneider, V. A., & of the International Nucleotide Sequence Database Collaboration, on behalf. (2025). The international nucleotide sequence database collaboration (INSDC): Enhancing global participation. *Nucleic Acids Research*, 53(D1), D62–D66. https://doi.org/10.1093/nar/gkae1058

Kleemola, M., Ala-Lahti, H., Alaterà, T., L'Hours, H., Mathers, B. J., Broeder, D., Horik, R. van, Jeriehag, B., Degl'Innocenti, E., Sanesi, M., & Koski, N. (2022). *D8.3 Trustworthy Digital Repository status update and certification solutions for SSHOC repositories*. https://zenodo.org/records/6530203

Kleemola, M., Wong, M. W., Wolff-Boenisch, B., Verburg, M., L'Hours, H., Saji, A., & Jakobsen, M. (2025). *The battle for trust: Advancing the social science data ecosystem*. Zenodo. https://doi.org/10.5281/ZENODO.15125538

Lamotte, F. de, Stoll, V., Arènes, C., Herbet, M.-E., Debard, S., Genova, F., Hadrossek, C., Lerigoleur, E., Leroux, G., Ohanessian, G., Pierkot, C., & Stahl, M. (2024). *Selecting a trustworthy subject-specific repository for self-depositing data: Methodology and analysis of existing services* [Report]. Comité pour la Science Ouverte. https://doi.org/10.52949/81

Lazzeri, E. (2024). *Update of the Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements*. Zenodo. https://doi.org/10.5281/zenodo.13919643

L'Hours, H., Kleemola, M., & Recker, J. (2024). *CoreTrustSeal Levels of Curation and Preservation: Implied Repository and Object Metadata Characteristics*. https://doi.org/10.5281/ZENODO.12701324

Funded by the European Union

L'Hours, H., Kleemola, M., Parkes, O., Recker, J., Duvaud, S., van Horik, R., Alaterä, T. J., Liberante, F., Conzett, P., Kaartinen, H., Bäckman, S., & Esteves, E. (2025a). *FIDELIS TTRAMatrix v01.00 Introduction and Overview*. Zenodo. https://doi.org/10.5281/zenodo.17144159

L'Hours, H., Kleemola, M., Parkes, O., Recker, J., Duvaud, S., van Horik, R., Alaterä, T. J., Liberante, F., Conzett, P., Kaartinen, H., Bäckman, S., & Esteves, E. (2025b). *FIDELIS TTRAMatrix Design Statement*. Zenodo. https://doi.org/10.5281/zenodo.17144178

L'Hours, H., Kleemola, M., Parkes, O., Recker, J., Duvaud, S., van Horik, R., Alaterä, T. J., Liberante, F., Conzett, P., Kaartinen, H., Bäckman, S., & Esteves, E. (2025c). *FIDELIS TTRAMatrix v01.00 Guide*. Zenodo. https://doi.org/10.5281/zenodo.17144141

Lin, D., McAuliffe, M., Pruitt, K. D., Gururaj, A., Melchior, C., Schmitt, C., & Wright, S. N. (2024). Biomedical Data Repository Concepts and Management Principles. *Scientific Data*, 11(1), 622. https://doi.org/10.1038/s41597-024-03449-z

Lokers, R., Janssen, S., Bulens, J., Knapen, R., & Chelle, M. (2018). *e-ROSA D2.2—Impact Chains and Science Solutions*. https://doi.org/10.5281/zenodo.1313554

Long Term Archive (LTA) group. (2024, June 11). *WDCC User Guide for Data Publications* (Version 1.4). https://www.wdc-climate.de/docs/WDCC_User_Guide.pdf

Marrano, A., Cabugos, L., Hafner, A., Kapoor, B., McNamara, J., O'Donnell, M., Reiser, L., Tello-Ruiz, M. K., Zhang, H., & Staton, M. (2025). A teaching and training framework to promote findable, accessible, interoperable, and reusable data generation in agriculture. *Database*, 2025, baaf034. https://doi.org/10.1093/database/baaf034

Matthiesen, M., & Leonardič, J. (2025). *CLARIN Data Citation Guidelines version 1.0* (No. CE-2025-2608; Version 1.0). https://doi.org/10.34733/doc-189

Moine, M.-P., Valcke, S., Lawrence, B. N., Pascoe, C., Ford, R. W., Alias, A., Balaji, V., Bentley, P., Devine, G., Callaghan, S. A., & Guilyardi, E. (2014). Development and exploitation of a controlled vocabulary in support of climate modelling. *Geoscientific Model Development*, 7(2), 479–493. https://doi.org/10.5194/gmd-7-479-2014

nestor Certification Working Group. (2025). *Explanatory notes on the nestor Seal for Trustworthy Digital Archives*. Version 2.2 (No. 17; Nestor Materials). https://nbn-resolving.org/urn:nbn:de:0008-2501311511155.033954116333

OpenAIRE. (2022). *OpenAIRE Guidelines for Data Archives 2.0*. https://guidelines.readthedocs.io/en/latest/data/index.html

Peng, G., Lacagnina, C., Downs, R. R., Ganske, A., Ramapriyan, H. K., Ivánová, I., Wyborn, L., Jones, D., Bastin, L., Shie, C., & Moroni, D. F. (2022). Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets. *Data Science Journal*, 21(1). https://doi.org/10.5334/dsj-2022-008

Puebla, I., Neumann, S., Stathis, K., Gauthier, J., Hamelers, A., Nielsen, L. H., Belletti, L., Elliott, F., & Hahnel, M. (2024). *GREI Data citation best practices for repositories*. https://doi.org/10.5281/zenodo.10562429

Funded by the European Union

Rehm, H. L., Page, A. J. H., Smith, L., Adams, J. B., Alterovitz, G., Babb, L. J., Barkley, M. P., Baudis, M., Beauvais, M. J. S., Beck, T., Beckmann, J. S., Beltran, S., Bernick, D., Bernier, A., Bonfield, J. K., Boughtwood, T. F., Bourque, G., Bowers, S. R., Brookes, A. J., … Birney, E. (2021). GA4GH: International policies and standards for data sharing across genomic research and healthcare. *Cell Genomics*, 1(2), 100029. https://doi.org/10.1016/j.xgen.2021.100029

Sansone, S.-A., & Rocca-Serra, P. (2016). *Review: Interoperability standards*. https://doi.org/10.6084/M9.FIGSHARE.4055496.V1

Science Europe. (2021). *Practical Guide to the International Alignment of Research Data Management (Extended Version)*. https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf

Sen, T. Z., Caccamo, M., Edwards, D., & Quesneville, H. (2020). Building a successful international research community through data sharing: The case of the Wheat Information System (WheatIS). *F1000Research*. https://doi.org/10.12688/f1000research.23525.1

Shearer, K., Nakano Koga, S. M., Rodrigues, E., Manola, N., Pronk, martine, & Proudman, V. (2023). *Current State and Future Directions for Open Repositories in Europe*. Zenodo. https://doi.org/10.5281/zenodo.10255559

Shearer, K., Rodrigues, E., Manola, N., Malaguarnera, G., Pronk, M., & Proudman, V. (2025). *REPOSITORIES: Key Infrastructure For Maintaining European Research Excellence*. Zenodo. https://doi.org/10.5281/zenodo.15119849

Top, J., Janssen, S., Boogaard, H., Knapen, R., & Şimşek-Şenel, G. (2022). Cultivating FAIR principles for agri-food data. *Computers and Electronics in Agriculture*, 196, 106909. https://doi.org/10.1016/j.compag.2022.106909

Verburg, M., Ulrich, R., L'Hours, H., Huber, R., Priddy, M., Davidson, J., Gonzalez-Beltran, A., Meijas, G., & Neidiger, C. (2023). *M5.2—Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes*. https://zenodo.org/records/10058634

White House Office of Science and Technology Policy. (2022). *Desirable Characteristics of Data Repositories for Federally Funded Research (NSTC Reports)*. Executive Office of the President of the United States. https://doi.org/10.5479/10088/113528

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wittenburg, P., Van Uytvanck, D., Zastrow, T., & Offersgaard, L. (2023). *CLARIN Centre Types* (Version 1.0). http://hdl.handle.net/11372/DOC-77

Wittenburg, P., Van Uytvanck, D., Zastrow, T., Straňák, P., Broeder, D., Schiel, F., Boehlke, V., Reichel, U., & Offersgaard, L. (2023). CLARIN B Centre Checklist (Version 7.4.1). http://hdl.handle.net/11372/DOC-78

Funded by the European Union

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., … Glöckner, F. O. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology*, 29(5), 415–420. https://doi.org/10.1038/nbt.1823

Funded by the European Union