How to find a trustworthy repository?

Milica Ševkušić, EIFL

Outline

- Practical advice
- The underlying principles and standards (TRAC, TRUST, FAIR)
- Certification Schemes
- Resources
- Training tips

OpenAIRE

Science Europe's [Practical Guide to the International Alignment of Research Data Management](#) (2021)

1. Provision of Persistent and Unique Identifiers (PID)
   a. Allow data discovery and identification
   b. Enable searching, citing, and retrieval of data
   c. Provide support for data versioning

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

3. 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

4. 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)

Practical advice (OpenAIRE)

"The general steps for finding a data repository are:

1. use a disciplinary repository if there is one;

2. alternatively, use the institutional repository, if you have one where the data will also be available for the long term;

3. use the catch-all repository Zenodo, maintained by CERN;

4. or search in a global registry - re3data or FAIRsharing - for a fitting repository (they provide several filtering options)."

"However, if you focus on **repositories that are certified as being trustworthy**, you simplify your selection process. So, if you don't have a disciplinary repository, and use the re3data portal for your search, we recommend that you filter on "Certificate" and look for the red icon (which identifies the repository as having a certificate). Searching in FAIRsharing you can look to the section "Certifications and community badges" to find if the repository is certified."

https://www.openaire.eu/find-trustworthy-data-repository

OpenAIRE

Suggest a repository

re3data.org Registration Policy

To be registered in re3data.org a research data repository must

- be run by a legal entity, such as a sustainable institution (e.g. library, university)
- clarify access conditions to the data and repository as well as the terms of use
- **have focus on research data**

A research data repository is a subtype of a sustainable information infrastructure which provides long-term storage and access to research data that is the basis for a scholarly publication. Research data means information objects generated by scholarly projects for example through experiments, measurements, surveys or interviews.

A research data repository listed in re3data.org is either:

- a data provider if it offers research data and its metadata (ideally exposing metadata via interfaces),

and/or

- a service provider (e.g. a portal) if it harvests the metadata of research data from data providers as a basis for building value-added services.

For the re3data.org research data repository registration workflow see our FAQ.

If you find a research data repository missing from our list, we are open to suggestions. Please use the following form:

Filter

Subjects ⊞
Content Types ⊞
Countries ⊞
AID systems ⊞
API ⊞
Certificates ⊞
Data access ⊞
Data access restrictions ⊞
Database access ⊞
Database access restrictions ⊞
Database licenses ⊞
Data licenses ⊞
Data upload ⊞
Data upload restrictions ⊞
Enhanced publication ⊞
Institution responsibility type ⊞
Institution type ⊞
Keywords ⊞
Metadata standards ⊞
PID systems ⊞
Provider types ⊞
Quality management ⊞
Repository languages ⊞
Software ⊞
Syndications ⊞
Repository types ⊞
Versioning ⊞

Required information

Repository name	
Repository name language	English
Repository url	
Description	

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Search... 🔍 Search

The underlying principles and standards

- Reference Model for an Open Archival Information System (OAIS) (2002)
- TRAC checklist (2007)
- TRUST principles (2020)
- Certification schemes

Purpose: to enable FAIR!

Reference Model for an Open Archival Information System (OAIS)

- Conceptual model for a digital archive dedicated to **preserving** and **maintaining** access to digital information in the long term

- Produced under the leadership of the Consultative Committee for Space Data Systems (CCSDS); major input from archives and libraries

- Approved as ISO Standard 14721 in 2002: https://www.iso.org/standard/57284.html
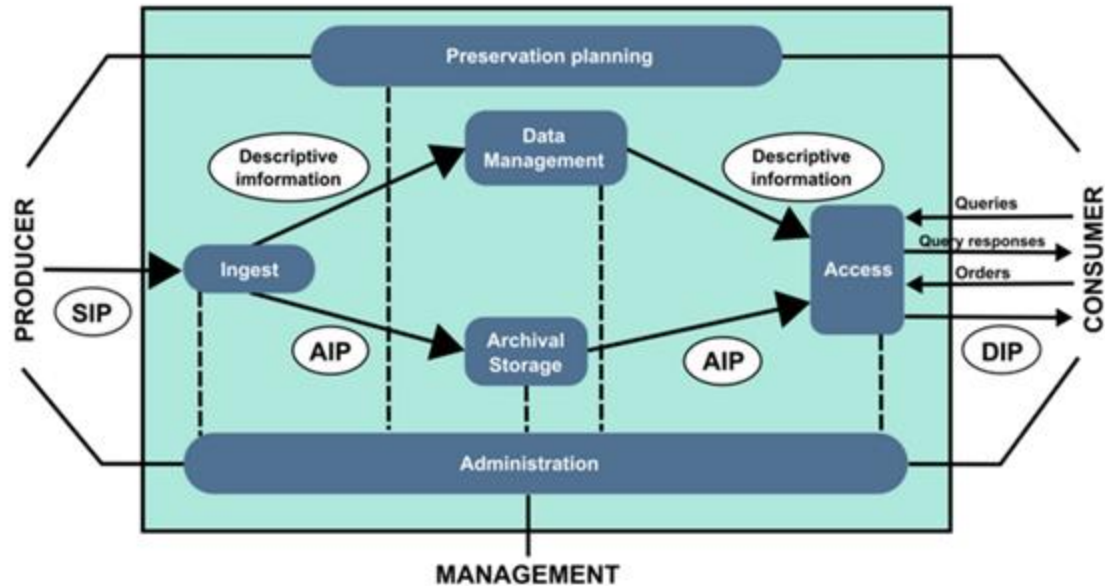
OAIS Reference Model, source: Wikimedia Commons, CC BY-SA 4.0 International

OpenAIRE

Trusted Repository Audit and Certification (TRAC)

- TRAC criteria measure the ability of a repository to preserve digital content in a way that serves the repository's designated community.
- Based on the OAIS reference model / ISO 14721:2012 standard
- Developed through the Consultative Committee for Space Data Systems (CCSDS) and a task force under the auspices of OCLC's Research Libraries Group (RLG) and the National Archives and Records Administration (NARA)
- Revised by the Center for Research Libraries and the Research Libraries Group (RLG) (2005-2006)
- The final version of the TRAC checklist was published in 2007.

OAIS + focus on governance

OpenAIRE

TRUST principles

guiding principles to demonstrate the trustworthiness of digital repositories

Principle	Guidance for repositories
Transparency	To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.
Responsibility	To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.
User Focus	To ensure that the data management norms and expectations of target user communities are met.
Sustainability	To sustain services and preserve data holdings for the long-term.
Technology	To provide infrastructure and capabilities to support secure, persistent, and reliable services.

Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* **7**, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

OpenAIRE

Certification schemes

- Integrated frameworks for auditing and certifying (digital repositories)
- The status of a "trusted digital repository"
  - Mission to provide trustworthy long-term access to curated digital resources
  - Monitoring, planning and maintenance
  - Awareness of (and mechanisms to deal with) threats and risks within systems
  - Regular checks and/or certification

OpenAIRE

European Framework for Audit and Certification of Digital Repositories

Formal

Extended

Core

ISO
http://www.iso16363.org/

nestor Seal 2018
https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/nestor_Siegel/siegel.html

CORE TRUST SEAL
https://www.coretrustseal.org

Core Trust Seal

- Core level certification based on the Data Seal of Approval (DSA) and World Data System of the International Science Council (WDS) Core Trustworthy Data Repositories Requirements (two catalogues merged under the umbrella of the Research Data Alliance)

- CoreTrustSeal - international, community based, non-governmental, and non-profit organization
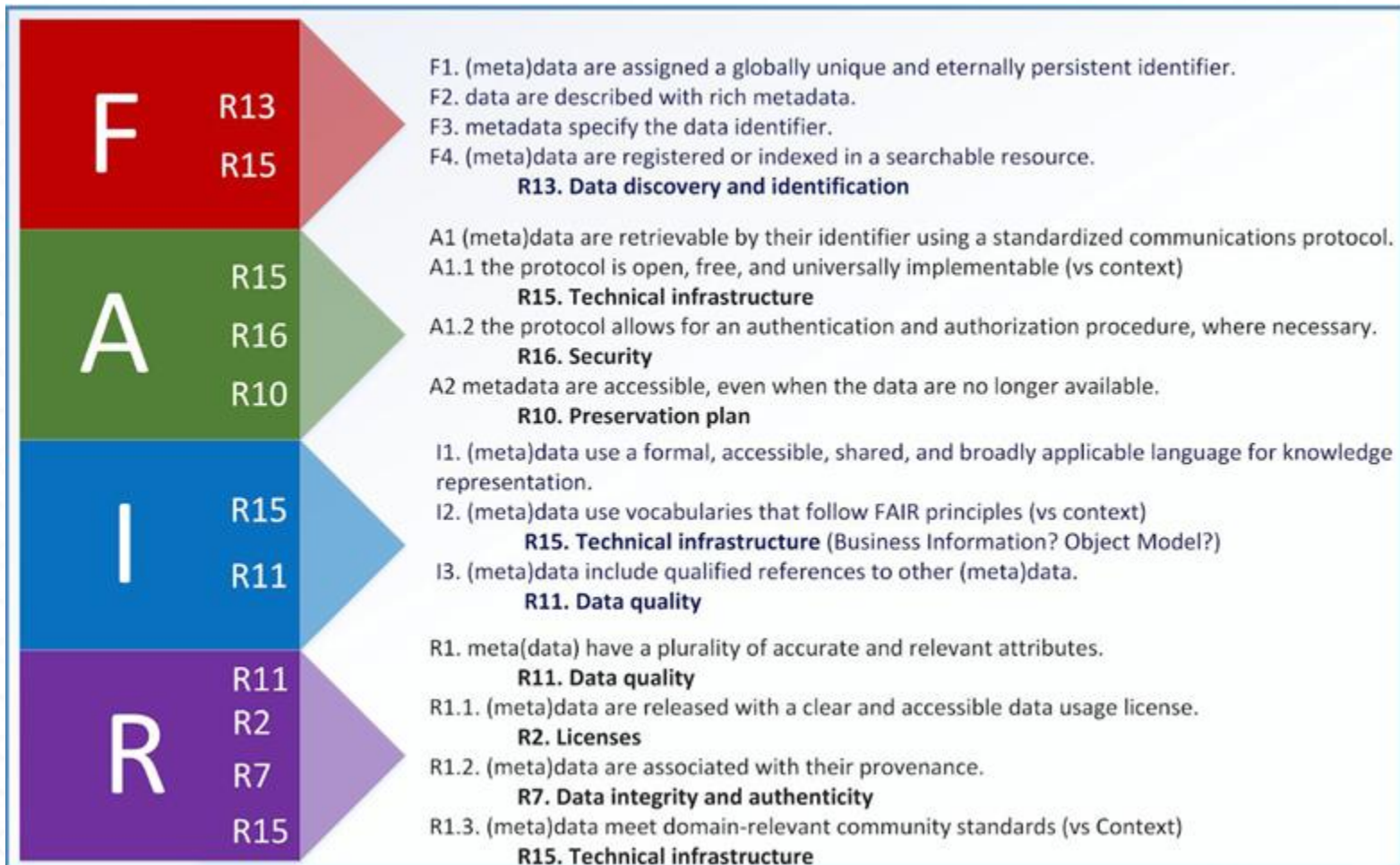
- Requirements: https://doi.org/10.5281/zenodo.7051012

CTS requirements

CoreTrustSeal+FAIRenabling Capability Maturity (CapMat)
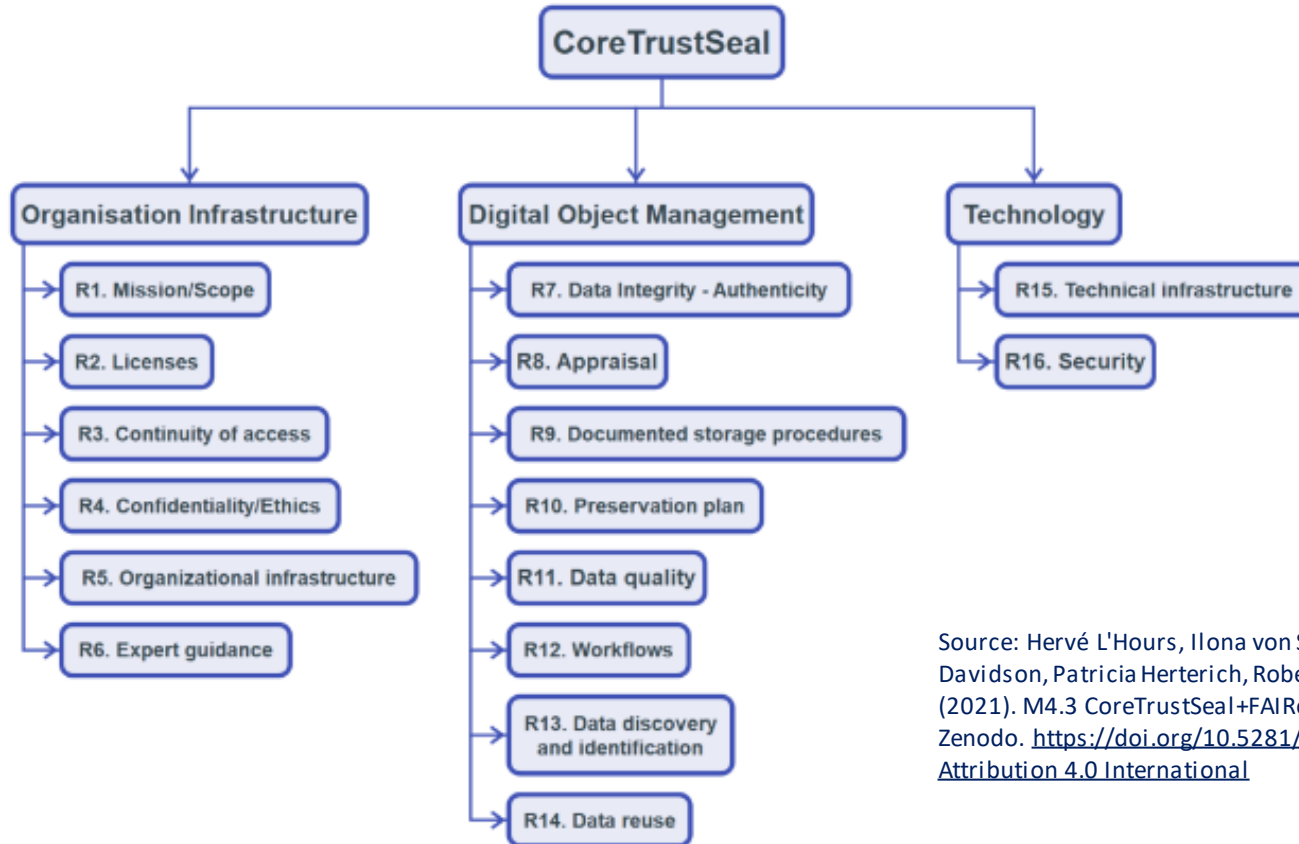
Hervé L'Hours, Maaike Verburg, Jerry de Vries, Linas Cepinskas, Ilona von Stein, Robert Huber, Joy Davidson, Patricia Herterich, & Benjamin Mathers. (2022). Report on a maturity model towards FAIR data in FAIR repositories (D4.6) (V2.0). Zenodo. https://doi.org/10.5281/zenodo.6699520


F — R13, R15
- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. metadata specify the data identifier.
- F4. (meta)data are registered or indexed in a searchable resource.
  - **R13. Data discovery and identification**

A — R15, R16, R10
- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
- A1.1 the protocol is open, free, and universally implementable (vs context)
  - **R15. Technical infrastructure**
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
  - **R16. Security**
- A2 metadata are accessible, even when the data are no longer available.
  - **R10. Preservation plan**

I — R15, R11
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles (vs context)
  - **R15. Technical infrastructure** (Business Information? Object Model?)
- I3. (meta)data include qualified references to other (meta)data.
  - **R11. Data quality**

R — R11, R2, R7, R15
- R1. meta(data) have a plurality of accurate and relevant attributes.
  - **R11. Data quality**
- R1.1. (meta)data are released with a clear and accessible data usage license.
  - **R2. Licenses**
- R1.2. (meta)data are associated with their provenance.
  - **R7. Data integrity and authenticity**
- R1.3. (meta)data meet domain-relevant community standards (vs Context)
  - **R15. Technical infrastructure**

NESTOR Seal 2018

- Extended self-assessment process for digital archives developed and offered by nestor on the basis of the DIN 31644 standard "Criteria for trustworthy digital archives"
- Granted to core certification repositories which also perform a structured, externally reviewed and publicly available self-audit based on DIN 31644/nestorSeal
- International collaboration
- Catalogue of criteria: 34 criteria written by the German nestor group and adopted in Germany as DIN 31644
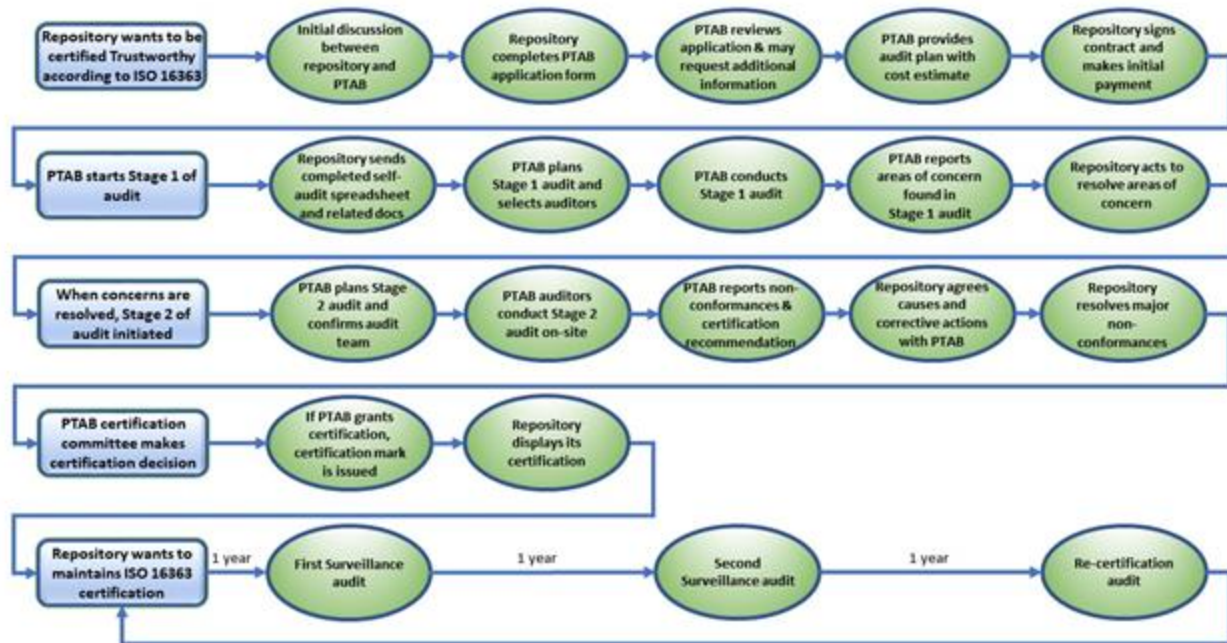
Resources:
- Info
- Explanatory notes on the nestor Seal for Trustworthy Digital Archives (Version 2, in German)
- Model contract (in German)
- Assessment form for obtaining the nestor Seal for Trustworthy Digital Archives
- Explanatory notes on the assessment form

https://files.dnb.de/nestor/materialien/nestor_mat_08-eng.pdf
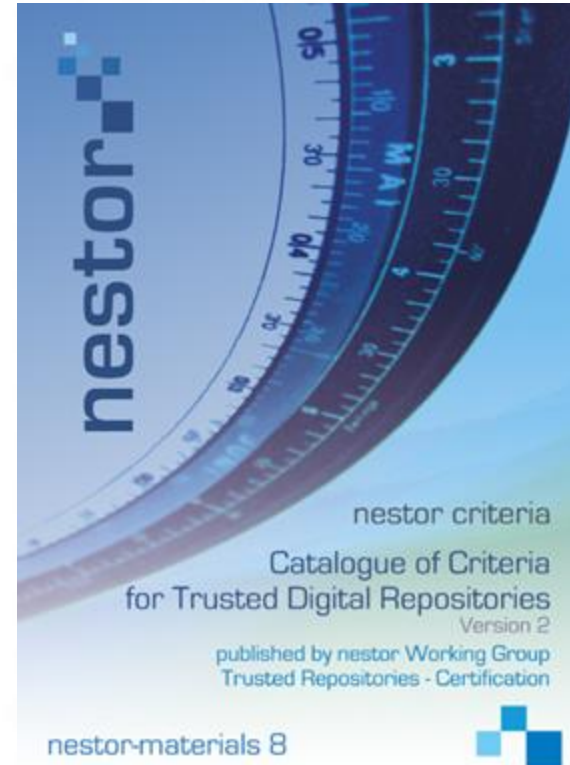
ISO 16363: formal certification

- Based on Open Archival Information System (OAIS) and Trusted Repository Audit and Certification (TRAC)
- More than 100 indicators
- Granted to repositories which in addition to casic certification obtain **full external audit** and certification based on ISO 16363

PTAB = PTAB - Primary Trustworthy Digital Repository Authorisation Body

OpenAIRE

Resources

- Philipp Conzett, Ingrid Dillo, Francoise Genova, Natalie Harrower, Vasso Kalaitzi, Mari Kleemola, Amela Kurta, Pedro Principe, Olivier Rouchon, Hannes Thiemann, & Maaike Verburg. (2022). Towards a European network of FAIR-enabling Trustworthy Digital Repositories (TDRs) - A Working Paper (v2.0). Zenodo. https://doi.org/10.5281/zenodo.7034315

- European Commission, Directorate-General for Research and Innovation, Jones, S., Aronsen, J., Beyan, O.et al., Recommendations on certifying services required to enable FAIR within EOSC – , Genova, F.(editor), Publications Office, 2021, https://data.europa.eu/doi/10.2777/127253

- https://www.sshopencloud.eu/certification-faqs

- **Materials from the re3data COREF / CoreTrustSeal Workshop on Quality Management at Research Data Repositories**

Training tips

Researchers

- Explain the practical aspects of TDRs
- Focus on FAIR
- Show how to find trusted repositories
- Show how to find certified repositories

Repository owners / managers

- Untangle the complexity
- Focus on FAIR
- Explain the benefits of certification
- Explain requirements
- Show where to find supporting materials, use cases and advice

THANKS

Web
www.openaire.eu

Email
milica.sevkusic@eifl.net

Twitter
@openaire_eu
@lessormore4