University of Oxford

ORA FAIR Assessment (Datasets) – Conceptual Requirements

doi:

Version number	1.0
Status	Ready
Author(s)	Allyson Lister, Tom Mitchell, Thomas Wrobel
Contributor(s)	Jason Partridge

Revision History

VERSION	STATUS	DATE	DESCRIPTION	AUTHOR(S)
1.0	Ready		First draft	Allyson Lister, Tom Mitchell, Thomas Wrobel

Author List

ORGANISATION	NAME AND ORCID	CONTACT INFORMATION	ROLE / ATTRIBUTION - HTTPS://CREDIT.NISO.ORG/
University of Oxford	Allyson Lister, https://orcid.org/0000-0002-7702-4495	allyson.lister@oerc.ox.ac.uk	FAIRsharing Content and Community Lead Writing – original draft, Writing – review & editing
University of Oxford	Tom Mitchell, 0009-0007-5451-0486	tom.mitchell@bodleian.ox.ac.uk	Open Access Service Manager Writing – original draft, Writing – review & editing
University of Oxford	Tom Wrobel, 0000-0003-3302-161X	thomas.wrobel@bodleian.ox.ac.uk	Open Access Services Development Lead Writing – original draft,

			Writing – review & editing

Contributor List

Organisation	Name and ORCID	Contact Information	Role / attribution - https://credit.niso.org/
University of Oxford	Jason Partridge https://orcid.org/0000-0001-9819-9991	jason.partridge@bodleian.ox.ac.uk	Writing – review & editing

Abbreviations and Symbols

General Abbreviations

Abbreviation	Meaning
GUPRI	Globally unique, persistent and resolvable
GUID	Globally unique and resolvable. Note that often people use this (mistakenly) interchangeably with GUPRI.

Responsibilities for implementation: The symbols in this table provide information on whether the principle is satisfied at an infrastructure level or dependent upon record-level metadata (submitter-dependent). Some principles may involve both levels.

SYMBOL	MEANING
	**Repository-implemented ("pass by design"):** The principle is fulfilled through repository infrastructure, configuration, or policy. Compliance does not depend on individual submitter behaviour.
	**Submitter-dependent compliance pathway:** The principle requires action or completeness at the record level. The outcome of the evaluation depends on the dataset owner or submitter.

Alignments with existing generic metrics: The symbols in this table provide information on whether an established generic metric is sufficient, requires adaptation, or is insufficient.

SYMBOL	MEANING
	**Generic metric is sufficient:** An existing generic metric fully supports evaluation within this benchmark.
	**Generic metric requires specialisation:** Requires a specialised test or adaptation to meet community-specific requirements.
	**Specialised definition provided:** No *suitable* generic metric exists; a community-specific definition or test is supplied.
	**Not applicable:** The principle does not apply within the scope of this community benchmark.
	Work in progress: Definition or implementation for this principle is not yet complete.

FAIR Definition: Summary

This summary outlines how ORA supports the FAIR Principles in practice. Many FAIR features are provided automatically by the repository, while others rely on information supplied at submission. Understanding this distinction can help researchers make small but meaningful choices that improve the visibility, usability and impact of their work.

The **key for each of the symbols** is available in the "Abbreviation and Symbols" table at the top of this document.

This **FAIR Benchmark for Institutional Repository Datasets** is a structured, practical benchmark designed to help institutional repositories and their researchers assess and improve the FAIRness of dataset metadata by providing consistent, transparent evaluation criteria aligned with community standards and best practices.

Findable

- 💾 💡 **F1 – Globally unique, persistent and resolvable identifiers**: Every ORA record is assigned a globally unique, persistent and resolvable identifier so it can be reliably found and cited over time.
- 💾 👤 💡 **F2 – Metadata that supports discovery**: Depending on the completeness of the submitted record, ORA records will include one or more machine-readable metadata fields that enable discovery through search tools and indexing services. A variety of fields are tested.
- 💾 ✅ 💡 **F3 – Clear links between records and data**: Each ORA record is designed to include its own metadata identifier as well as clearly link to the data or files it describes. If external datasets are not referenced, then this metric fails.
- 💾 ❎ 💡 **F4 – Searchable**: ORA records are searchable within ORA and are also indexed by major web search engines, helping others discover the work.

Accessible

- 💾 💡 **A1.1 – Open, widely supported protocols**: Metadata is delivered using standard web technologies (http or https) that are openly specified and broadly supported.

- 💾 ✅ **A1.2 – Clear access conditions**: ORA metadata is openly accessible, with no authorisation/authentication requirements. (This check only evaluates the metadata and not the data.)
- 💾 ✅ **A2 – Database persistence policy**: Even if data files are withdrawn or restricted, the ORA record describing them remains accessible so that the existence and context of the work are preserved.

Interoperable

- 💾 ✅ 💡 **I1 – Standard representation format** (semantic): ORA metadata is provided in a standard, linked data format that can be processed and exchanged by computer systems.
- 💾 👤 💡 **I2 – Use of FAIR terminologies** (semantic): ORA supports the use of recognised subject vocabularies (FAST and ISO 639-2). Including at least one term from these vocabularies to helps improve interoperability and discovery. Note that until FAST URLs are provided (and not just the subject labels) the metrics for FAST will fail.
- 👤 💡 **I3 – Connections to related research**: ORA records can include structured links to outputs that are related but that do not directly support the dataset itself (both links to other resources and links to other versions of the same resource). Including these links strengthens connections within the research ecosystem.

Reusable

- 💾 👤 ✅ **R1.1 – Clear licence**: Every ORA record is covered by a licence explaining how the data and metadata may be used. Dataset submitters may choose a specific licence, or the repository's default licence will apply.
- 👤 💡 **R1.2 – Provenance information**: ORA allows dataset submitters to include information about creators, funders, institutional affiliation, related publications and origin of content. The more complete this information, the easier it is for others to understand and trust the work.
- 💾 👤 💡 **R1.3 – Alignment with recognised standards**: ORA supports widely used metadata structures and subject vocabularies appropriate for a general repository. Including recognised subject terms and language codes as well as using a structured metadata format helps align records with broader community best practices.

Introduction and Task Management

This document describes a FAIR benchmark and its associated components, tailored to the community needs of the University of Oxford's institutional repository – Oxford University Research Archive (ORA). This document is a human-readable, narrative description of this community's requirements for FAIR assessment of archival digital research objects across all subject areas, without the inclusion of technical details. Initially, the requirements will focus on the metadata describing datasets, with plans to expand to other types of research outputs in the future. It follows the Assess-IF and is used to create the associated technical components. Datasets in ORA are classified into two types: those archived within ORA, and those stored in external repositories but referenced in ORA. Metadata for both types will be evaluated, either upon deposit as new records or retrospectively for records created prior to the introduction of the FAIR evaluation workflow.

The following arguments must be provided to this benchmark:

1. **URI (mandatory)**
   In the MVP implementation of this benchmark, any URI may be passed (the record landing page URL, its DOI URI or its ARK URI). When resolved programmatically via content negotiation, the machine actionable metadata in JSON-LD for this record must be returned. Future versions of the benchmark may allow other formats. Note that the passed URI will **not** be used in the assessment process; when identifiers must be assessed as part of FAIR, the metrics will look within the metadata and not at the passed URI.
   -- OR --
2. **The metadata to evaluate (future work)**
   Although not expressly included as a requirement for the MVP, in future we will incorporate functionality that would allow us to pass the metadata directly rather than a URI.

This document was created using a template found at https://doi.org/10.5281/zenodo.18244844. This template can be copied and used within any community to begin preparations for a FAIR benchmark specific to your needs.

FAIR Assessment Conceptual Components - Roadmap		
Task	Status	Notes
1: Benchmark – narrative requirements	**Completed**	
2: All metrics - narrative requirements	**Completed**	
3: Related standards, databases and collections - registration with FAIRsharing	**Completed**	No additional records required.
4: Completed first draft of document	**Completed**	
5: Benchmark - registration with FAIRsharing	**Completed**	https://fairsharing.org/7598
6: All metrics - registration with FAIRsharing	**Completed**	See links within this document.

Overview

The Assessment Interoperability Framework (Assess-IF) being used by this community aims to create a uniform approach for defining and running tests (and specifically FAIR tests in the context of this document) by:

1. **Giving the community the power to choose what FAIR means to them:** Until now, using a FAIR tool such as those listed at https://fairassist.org/tools meant either selecting from among the benchmarks for FAIR that a given tool provides (which may be unclear or hard to understand exactly what's being tested) or a formal collaboration with tool(s) to ensure they provide a custom FAIR assessment solution for your needs. The Assess-IF allows you to explicitly state how your community defines FAIR, and exactly how it should be tested, measured, weighted and scored.
2. **Encouraging transparency in assessment behaviours:** To date, FAIR assessment has been characterised by disparate behaviours and scores arising from independently authored FAIR Assessment tools, which themselves vary in the level of transparency they exhibit for their tests and benchmarks. The Assess-IF methodology encourages all tools and assessment platforms to register their assessment components (e.g. benchmarks, metrics) in a searchable manner, with rich metadata explaining what is being tested, why, and how. The Assess-IF allows you not only to define FAIR according to your community requirements but also provides a way of making those definitions clear, visible, reusable and extendable. Benchmarks that cannot be discovered via a search, and do not describe themselves sufficiently, do not themselves meet the minimum expectations for FAIRness and make it harder to understand if a particular tool interprets FAIR the way your community requires it to be interpreted.
3. **Defining a consistent report format for assessments**: Because the reports are designed to be understood by any tool, the user can have a wider range of independent tools to select from. Being able to move these reports from one tool to another also allows the user to contextualise their tests, and to explore results from different perspectives.
4. **Enabling Automatic FAIR Testing via APIs**: by creating shared APIs among the different domains (FAIR, SKG, DMP, maDMP) it becomes possible to "embed" assessments into existing tooling. For example, a DMP authoring tool will be able to "call out" to a FAIR Assessment tool to obtain real-time information about the FAIRness of a digital object while the DMP is being written or updated.

Summary of component types

The Assess-IF has seven main “components”, split according to three main categories: **conceptual** components define FAIR for a community in human-readable terms, **software** components provide the actual tests and interpretations of those tests, and **data** components store the results of running an assessment. This template helps you to fully specify the conceptual components for your community definition of FAIR and gives you jumping off points to create the software components and ultimately run assessments to produce data in the form of test results. As such, we have provided a short description of each component so that you can understand the wider framework at a very high level.

The **conceptual components** are:

1. **Dimensions/principles**: Dimensions (for this template, our dimensions are the FAIR Principles) are designed to be subject- and implementation-agnostic criteria or high-level goals that may be refined for particular communities as described in this template. More information on the FAIR principles can be found in the FAIR Principles FAIRsharing record (https://doi.org/10.25504/FAIRsharing.WWI10U), which provides a wealth of information about them as well as linking to individual records describing all sub-principles.
2. **Benchmarks**: Community-specific groupings of metrics. They provide a narrative describing how a community defines FAIR. Communities can choose the granularity of their benchmarks with regards to subject area and digital object type. Specifically, this means that benchmarks may be agnostic of object type or subject area, or may be scoped as tightly as required by the benchmark authors.
3. **Metrics**: Narrative description that a Test must wholly implement. Each metric should implement exactly one dimension (e.g. one sub-principle from the FAIR Principles). Metrics may be domain-agnostic or not.

A detailed understanding of the remaining components (below) is not required for the purposes of this template, but a short description of each is provided for completeness.

The **software components** are:

1. **Tests**: The instantiation of metrics, executed to assess a digital object in accordance with the metric linked to the test.

2. **Scoring Algorithms**: The instantiation of benchmarks, executed to create community-specific value judgements on the outcomes of tests. They specify the exact set of tests which are to be run by the assessment service together with weightings for each of these tests. Algorithms contextualise the sum of all test results for a given benchmark into a final quantitative assessment result.

The **data components** are:

1. **Test Results**: The output of running a test over a digital object. A test result also contains provenance metadata about the process followed to create it.
2. **Test Result Sets**: A set of test results, together with their respective metadata.
3. **Benchmark Scores**: Obtained after executing a scoring algorithm over a set of test results. The benchmark score includes a value, a log and a link to the test results used to obtain the score.

FAIR Benchmark - Institutional Repository Datasets

Benchmark Name and URL	**FAIR Benchmark - Institutional Repository Datasets (FB - Inst Repo Datasets)** *https://doi.org/10.25504/FAIRsharing.e829b4*
Description	The 'FAIR Benchmark – Institutional Repository Datasets' provides a structured framework for evaluating and improving the metadata of research datasets in institutional repositories. It operationalizes the FAIR principles into practical, community-endorsed assessments. The benchmark serves two primary audiences: 1) Repository Teams, who integrate these tests into institutional workflows to ensure systemic compliance; and 2) Researchers, who use the evaluation feedback to identify improvements and deepen their understanding of data FAIRness. By offering a transparent, cross-disciplinary approach, the benchmark distinguishes between repository-level infrastructure and record-specific properties. The benchmark is designed to be applied to institutional repository records once they have been made publicly available.
Applicability	*What digital object(s) will be covered?* **Datasets** *What subject area(s) are covered by this benchmark?* **Subject agnostic**
Related Resources	This benchmark relates to the following high-level standards, databases and/or policies: • FAIRsharing: https://doi.org/10.25504/FAIRsharing.2abjs5 • ORA: https://doi.org/10.25504/FAIRsharing.rkwr6y
Guidance	More information on this benchmark may be found in its initial homepage, though a more formal project page may be created in future.

FAIR Principles F1: (Meta)data are assigned globally unique and persistent identifiers

Each data object and its metadata must be assigned an identifier that is globally unique to prevent ambiguity and persistent to ensure long-term reference. This identifier must remain stable over time (persist) and support unambiguous citation, linking, and machine processing. An additional property supported by the GO FAIR Foundation is that the identifier is also 'resolvable' by machines.

For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.a2cea7.

This principle is very commonly split into its distinct aspects of the identifiers being checked for 1) persistence, 2) resolvability and 3) global uniqueness. This benchmark utilises the three **generic metrics** for these three aspects; these metrics are summarised below.

Alignment with generic metrics	**The generic metrics are sufficient for our needs.** Institutional repository records will each be assigned an identifier, which will have various qualities of being globally unique, persistent and resolvable. Most repositories will mint more than one type of identifier (e.g. UUIDs and DOIs). This can be because institutional repositories often function both as an institutional repository and as a data catalogue. As an institutional repository, they may mint DOIs (or use other appropriate identifier schemas) for data files directly archived within it. As a catalogue, they may record DOIs that reference datasets held in external repositories, and those DOIs resolve to external sources rather than to the repository itself. This metric will evaluate the metadata identifier passed to the FAIR tool to determine if the identifier schema used is globally unique, persistent and resolvable. ORA datasets will all have UUIDs and ARKs, and some will have DOIs. Therefore, all ORA dataset records are expected to pass.
Generic metric names and summaries	**FAIR Metric – F1 – Metadata - Metadata contains identifier that is guaranteed globally unique: https://fairsharing.org/8205** *This metric evaluates whether the metadata retrieved from the provided URI satisfies the EOSC-PID aligned FAIRsharing definition of global uniqueness.* **FAIR Metric – F1 – Metadata - Metadata contains identifier that is guaranteed persistent https://fairsharing.org/8204** *This metric evaluates whether the metadata retrieved from the provided URI satisfies the EOSC-PID aligned FAIRsharing definition of persistence.* **FAIR Metric – F1 – Metadata - Metadata contains identifier that is guaranteed resolvable https://fairsharing.org/8203** *This metric evaluates whether the metadata retrieved from the provided URI satisfies the EOSC-PID aligned FAIRsharing definition of resolvability.*

FAIR-F2: data are described with rich metadata

Data must be accompanied by sufficiently detailed metadata to enable discovery. Richness implies structured, domain-relevant descriptive attributes that support machine discovery. Discovery (rather than understanding or reuse) is key for this principle, as it is part of the Findability principle; other principles (e.g. R1.3) deal more with metadata richness for reuse and understanding.

⏫ For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.e05e98.

💡 There is no single way to define discovery metadata. What is one community's discovery metadata is another's reusability metadata; what are one community's discovery metadata fields are irrelevant to another. ORA has defined three **specialised** metrics for this principle.

FAIR Metric – F2 - Metadata – discovery-oriented metadata fields

Metric descriptions	**FAIR Metric – F2 - Metadata – discovery-oriented metadata fields (FM:F2:M:DiscoveryFields).** https://doi.org/10.25504/FAIRsharing.82c497 **What is being measured?** This metric evaluates whether a metadata record includes a core set of mandatory descriptive elements that are essential for basic discovery. Specifically, it checks the resolved metadata for the presence of the following four fields: title, contributor names, summary/abstract/description, and publication date (defined as the date the record was first made publicly available). To pass, all of these fields must be present and populated within a structured, common format such as schema.org JSON-LD, DataCite XML, or Dublin Core XML. If any of these fields are empty, the evaluation is expected to fail. **Why should we measure it?** Principle F2 is specifically concerned with the "richness" of metadata in the context of findability, rather than domain-specific completeness or reusability. While the exact indexing algorithms of major search engines are proprietary, these specific metadata elements are recognised as key drivers for discovery (based on feedback from the research community) and are often ignored by metrics associated with other FAIR principles. Measuring these fields as a mandatory group ensures that a record contains the fundamental descriptive features necessary for a search engine or repository interface to successfully index and present the resource to a user. **FAIR Metric – F2 - Metadata – tagging to aid discovery (FM:F2:M:DiscoveryTags).** https://fairsharing.org/8021 **What is being measured?**

	This dimension evaluates the metadata for the presence of at least one keyword or tag of any kind (free-text or controlled). With regards to Findability (as opposed to e.g., Interoperability), the presence of any type of tag (irrespective of whether that tag is part of a controlled vocabulary) is the key feature for this metric. The metric expects the identifier to point to structured metadata (e.g., schema.org, DataCite, or DC) and verifies that the keyword/subject property within that metadata is not empty. The metric passes if at least one tag is identified and fails if the keyword attribute is missing or null. **Why should we measure it?** Keywords and tags are essential for enabling filtered discovery and faceted search. While a title or abstract provides narrative context, tags provide ways for search engines and repository interfaces to categorize the data. In a generalist institutional repository, where datasets span many disciplines, these tags often serve as the primary mechanism for users to narrow down search results to relevant topics. Measuring this ensures that researchers are providing the basic "labels" required for their data to be categorized. **FAIR Metric – F2 - Metadata – has publisher information (FM:F2:M:DiscoveryPublisher).** https://fairsharing.org/8022 **What is being measured?** This metric evaluates whether the metadata includes explicit information regarding the organisation responsible for publishing the metadata record. It looks for a structured "publisher" field within the record. In the context of an institutional repository, this is typically the institution itself, or an external repository (like Zenodo) if the record is registering an object hosted elsewhere. The metric will fail if this value is not present. **Why should we measure it?** Publisher information is a discovery attribute that allows users and machines to filter resources by their "home" or source of authority, and to group resources based on their point of origin.

FAIR-F3: metadata clearly and explicitly include the identifier of the data it describes

Metadata must explicitly reference the identifier of the data object to which they relate. This bidirectional clarity ensures reliable linking between metadata records and the data they describe, enabling accurate indexing and retrieval. Additionally, metadata must explicitly reference its own identifier.

⏫ For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.820324.

☑ This benchmark utilises the **generic metric** for evaluating the metadata identifier, and a **specialised** metric for evaluating the data identifier(s).

Alignment with generic metrics	ORA satisfies this principle through two complementary mechanisms. Future work may focus on resolving the listed identifiers to ensure they are correct. 1. **Inclusion of the ORA record identifier within metadata**: The UUID, ARK - and DOI where present - is explicitly referenced within the metadata record. This ensures that the metadata contains the persistent identifier of the ORA record it describes, thereby satisfying the core requirement of F3. 2. **Inclusion of data/file identifiers within metadata**: Where ORA stores binary files associated with an ORA record, the schema.org JSON-LD metadata includes a file URI that identifies the stored digital object. In addition, ARKs assigned to data identifiers are included within the metadata where applicable. Where no ORA identifier exists, there should be at least one external data identifier (e.g. a DOI linking to a Zenodo object). Because ORA metadata consistently embeds the identifiers of the digital objects it archives, the only ORA records that might fail are those where the submitter did not provide an external identifier for the research object.

Metric names and summaries	**FAIR Metric – F3 – Metadata - includes its own identifier** https://doi.org/10.25504/FAIRsharing.3df457 *This metric assesses whether the metadata explicitly contains its own identifier in a structured, machine-actionable format. This ensures that the metadata itself is machine-actionable and can be retrieved, accessed, and managed as an independent entity in an indexing system.* **FAIR Metric – F3 – Metadata - uses common formats to reference data identifier(s)** https://doi.org/10.25504/FAIRsharing.0a2061 **What is being measured?** This metric evaluates whether the metadata record (structured as one of a specified list of common formats) explicitly includes the identifier of the research object it describes. **Why should we measure it?** F3 requires that metadata include the identifier of the research object it describes. To align with FAIR, if a metadata record is separated from its associated research object (which is common in harvesting and indexing), the connection between them must be explicitly preserved. This metric is concerned with the explicit linkage aspect of FAIR; F3 ensures that a machine-actionable relationship exists, allowing harvesters to unambiguously navigate from a descriptive metadata record back to the research object itself. Because many research objects cannot be arbitrarily extended to include references to their metadata, in many cases the only means to discover the metadata related to a research object will be to search based on the identifier of the research object itself. This metric evaluates whether the identifier provided as input resolves to metadata that contains a research object identifier. Further, that research object identifier must be distinguished from the numerous other fields and values that will be present in the metadata via one of the following formats: schema.org JSON-LD format, DataCite XML, DC XML.

FAIR-F4: (meta)data are registered or indexed in a searchable resource

Data and metadata must be registered in repositories or services that enable search and discovery. Indexing in structured, searchable resources ensures that identifiers and metadata can be located and queried programmatically.

⏫ For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.0c0d21.

☑ The metric relating to this principle is **generic**.

FAIR Metric – F4 – Metadata - indexed in a searchable resource

☑ While the existing **generic** metric is sufficient for our needs from a narrative perspective, it is likely that in addition to its associated generic test a new **specialised test** will be required to meet the requirements of ORA records that have been newly deposited (and therefore not necessarily indexed yet by Google).

Alignment with generic metric	Principle F4 requires that metadata be registered in a searchable resource to ensure discovery via query rather than just direct identifier. While the generic metric for F4 is narratively sound, a specialised implementation is required for the ORA context. Because external search engines (e.g., Google) involve indexing latency, a record's "Findability" should not be penalised during the window between deposit and external crawling. For this benchmark, a record satisfies F4 if it is discoverable via ORA's internal search interface. Since ORA is routinely indexed by major web search engines, internal discoverability serves as a reliable proxy for future global discoverability. Accordingly, an ORA record will be considered to satisfy F4 if it is discoverable through the ORA search interface, regardless of its immediate indexing status in external search engines. The generic test can still be run, but a specialised test should also be implemented.
Generic metric name and summary	**FAIR Metric – F4 – Metadata - indexed in a searchable resource** https://doi.org/10.25504/FAIRsharing.fe8b9b *F4 checks whether metadata is actually discoverable through web-based search engines, not just published online. This metric tests whether the metadata can be found via automated searches using the resource's identifier or descriptive content (such as title or keywords), confirming that it is indexed and accessible within the research ecosystem.* *An additional test will be implemented that will specifically query ORA for discoverability of the metadata record.*

FAIR-A1: (meta)data are retrievable by their identifier using a standardized communications protocol

Identifiers must resolve via established communication protocols that enable retrieval of data or metadata. Standardisation ensures interoperability and predictable access mechanisms across systems.

⏫ For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.7014eb.

☑ The metrics relating to this principle are **generic** and are implemented via sub-principles (see the next two sections).

This principle is split into two distinct sub-principles, which require that the protocol being used to retrieve the (meta)data 1) is openly specified, thus making them open, free and universally implementable (A1.1, https://doi.org/10.25504/FAIRsharing.7612c1), and 2) clearly states the procedures for access, including authentication/authorisation where required (A1.2, https://doi.org/10.25504/FAIRsharing.8e0027). As such, the **generic metrics** in use for A1 are also split between these sub-principles.

This benchmark utilises a **specialised metric** for A1.1 and the **generic metric** for A1.2.

FAIR-A1.1: the protocol is open, free, and universally implementable

The retrieval protocol must be openly documented and implementable without proprietary barriers. Openness ensures equitable and scalable access across technical environments.

⏫ For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.7612c1.

💡 This is the first of two sub-principles of A1; this benchmark utilises a **specialised metric**.

ORA record identifiers (ARKs) resolve using standard HTTP(S) protocols. These protocols are openly specified web standards and are free to implement. When an ORA record identifier is resolved, metadata can be retrieved either as a human-readable web page or programmatically via content negotiation (e.g. JSON-LD). Both modes of access rely exclusively on standard web protocols. Although access to certain data files may be mediated or subject to restrictions, this benchmark concerns metadata retrieval only. Metadata associated with ORA records is accessible via standard HTTP(S) resolution without reliance on proprietary protocols. Accordingly, ORA record identifiers are retrieved using open, free and universally implementable protocols, and ORA records are therefore expected to pass the A1.1 evaluation.

Metric description	**FAIR Metric – A1.1 – Metadata - HTTP(S) retrieval protocol** https://doi.org/10.25504/FAIRsharing.79aee0 **What is being measured?** This metric evaluates whether the protocol used to retrieve metadata referenced by the provided identifier is either HTTP or HTTPS and therefore openly specified, free to implement, and universally implementable. **Why should we measure it?** Principle A1.1 requires that metadata be retrievable by an open, free, and universally implementable protocol. The intent is to ensure that access to metadata is not gated by proprietary technologies, non-standardized interfaces, or restrictive licensing that prevents automated integration and universal access. Protocols such as HTTP(S) are the gold standard for this requirement because they are openly specified web standards, free for any developer to implement, and universally supported across virtually all modern software environments and programming languages. Protocols that are closed source or that have royalties associated with them could prevent users from being able to obtain the resource. By mandating these standards, we ensure that metadata remains accessible to the widest possible range of humans, harvesters and research discovery services. This metric evaluates whether the metadata access protocol satisfies the requirements for open, free, and universal implementation. The evaluation checks for the use of the HTTP or HTTPS protocols.

FAIR-A1.2: the protocol allows for an authentication and authorization procedure, where necessary

Access mechanisms must support controlled access when required by ethical, legal, or contractual constraints. Authentication and authorisation should not prevent metadata from remaining discoverable.

⏫ For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.8e0027.

☑ This is the second of two sub-principles of A1; this benchmark utilises this **generic metric**.

Alignment with generic metric	This generic metric **is sufficient** for our needs. A1.2 requires that the protocol used to retrieve (meta)data clearly describe any necessary authentication and authorisation procedures. The purpose of this requirement is to ensure that access conditions are transparent and machine-interpretable where restrictions apply. For ORA records, no authentication or authorisation is required to access metadata. When an ORA record identifier is resolved, metadata is openly retrievable via standard HTTP(S) protocols, without login or access mediation. ORA metadata remains openly accessible regardless of any restrictions that may apply to the underlying data files. Because ORA metadata can be retrieved without authentication and without additional access procedures, ORA records are expected to pass the A1.2 evaluation.
Generic metric name and summary	**FAIR Metric – A1.2 – Metadata - retrieval protocol supports auth https://doi.org/10.25504/FAIRsharing.0d94b0** *This metric evaluates whether the metadata retrieval protocol supports authentication and authorization mechanisms for controlled access to restricted content. If a FAIRsharing database record is available, the metric first checks whether the database declares open access; if access is restricted, it then verifies that the protocol supports appropriate authentication procedures to ensure transparent and defined access conditions.*

FAIR-A2: metadata are accessible, even when the data are no longer available

Metadata must remain accessible even if the underlying data are restricted, withdrawn, or no longer available. Persistent metadata ensures that records of existence, provenance, and context are preserved over time.

⏫ For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.0c0d21.

☑ The metric relating to this principle is **generic**; this benchmark utilises this **generic metric**.

Alignment with generic metric	This generic metric **is sufficient** for our needs. A2 requires that metadata remain accessible even if the digital object itself is no longer available. The purpose of this principle is to ensure the long-term discoverability and interpretability of records, independent of the continued availability of the underlying data files. ORA's approach to long-term preservation and metadata persistence is defined in the Bodleian Libraries preservation policy, as published on the Bodleian website and curated within ORA's FAIRsharing record (https://doi.org/10.25504/FAIRsharing.rkwr6y). The presence of a formally stated preservation policy satisfies the requirements of A2 for the purposes of this benchmark. While the policy is currently under review, and ORA has achieved CORE Trust Seal certification (2029-02-09 - Oxford University Research Archive (ORA) - CoreTrustSeal Requirements 2023-2025 - Core... - which provides more detailed documentation of preservation practices), these developments do not affect the pass/fail determination under this metric. The evaluation is based on the existence of a publicly declared preservation policy. Accordingly, all ORA records are expected to pass.
Generic metric name and summary	**FAIR Metric – A2 – Metadata - Database persistence policy https://doi.org/10.25504/FAIRsharing.d1a07b** *This metric evaluates whether the hosting database or repository has declared a formal data preservation policy listed within its FAIRsharing record, providing evidence of commitment to long-term metadata availability even when the research object itself is no longer accessible. By checking database-level preservation policies rather than record-level declarations, the metric addresses the common challenge that persistence information is rarely included in individual metadata records.*

FAIR-I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

Data and metadata must be expressed using formal knowledge representation languages that are machine-readable and widely adopted. Use of shared syntactic and semantic frameworks enables cross-platform integration.

⏫ For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.ec5648.

☑ The metrics relating to this principle are **generic**. Each community should choose either the *syntactic* (https://doi.org/10.25504/FAIRsharing.686c44, with a looser interpretation of FAIR I1) or *semantic* (https://doi.org/10.25504/FAIRsharing.3260c8, with a more precise definition of FAIR I1) version of the metric.

ORA provides content in JSON-LD, and as such the **semantic version** is applicable for this benchmark. We also implement the semantic database-level metric (https://doi.org/10.25504/FAIRsharing.9114a7) to provide additional evaluation in case the generic metric cannot correctly evaluate the record metadata.

Alignment with metrics	This generic metric **is sufficient** for our needs, and is supplemented by an additional metric. I1 requires that (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation. The purpose of this principle is to ensure that metadata can be interpreted consistently by machines using standard, non-proprietary representation formats. ORA exposes dataset metadata in schema.org JSON-LD when identifiers are resolved via content negotiation. JSON-LD is a formally defined W3C standard that provides a structured, validate-able, machine-readable representation of metadata. It is openly specified, widely adopted and supported across software environments. Schema.org JSON-LD represents metadata as linked data, using IRIs to identify properties and, where applicable, controlled vocabulary terms. This ensures that metadata elements are not merely structured but are also semantically grounded in defined schemas and vocabularies. The use of schema.org JSON-LD ensures that metadata is expressed in a formal and broadly applicable knowledge representation language. This satisfies the requirement that metadata be encoded using a shared, machine-interpretable standard rather than an ad hoc or proprietary format. Because ORA metadata is consistently exposed using schema.org JSON-LD, a formal and widely supported representation language, ORA records are expected to pass the I1 evaluation.
Metric names and summaries	**FAIR Metric – I1 – Metadata - Uses Knowledge representation language (semantic)** https://doi.org/10.25504/FAIRsharing.3260c8 *This metric evaluates whether the metadata retrieved upon resolution of the provided identifier is expressed in a semantic (grounded/linked data) format, following FAIRsharing's guidance on knowledge representation languages. By checking that metadata uses formally defined, machine-readable semantic structures, the metric ensures that metadata can support automated reasoning and interoperability across systems.* **FAIR Metric – I1 – Metadata - Database-level knowledge representation languages (semantic)** https://doi.org/10.25504/FAIRsharing.9114a7 *This metric evaluates whether the hosting database declares support for metadata formats that use semantic (grounded/linked data) knowledge representation languages, as recorded in its FAIRsharing registry entry. By checking database-level format declarations rather than examining individual metadata records, the metric provides an overview of the repository's capacity to deliver semantically structured metadata across all its content.*

FAIR-I2: (meta)data use vocabularies that follow FAIR principles

Controlled vocabularies and terminologies used to describe data must themselves be FAIR, meaning they are uniquely identified, accessible, and well-governed. This ensures semantic consistency and long-term interpretability.

For more information on the principle that this metric must implement, including documentation and justifications its implementation, please see https://doi.org/10.25504/FAIRsharing.96d4af.

There is a choice of two **generic** metrics that implement this principle: the *syntactic* (https://doi.org/10.25504/FAIRsharing.0273a2) and *semantic* (https://doi.org/10.25504/FAIRsharing.9b4569) versions.

ORA must be able to evaluate for specific terminologies. Therefore we have created our own **specialised** metrics.

FAIR Metric – I2 – Metadata – FAST subject heading is linked data

Metric description	**https://doi.org/10.25504/FAIRsharing.24acf4** This metric evaluates whether the metadata contains at least one FAST (Faceted Application of Subject Terminology) subject vocabulary term that resolves to content expressed in a linked data format. By specifically checking for FAST terms with formal semantic structure, the metric ensures that subject descriptors support automated reasoning and semantic interoperability, providing a community-specific implementation of the broader I2 principle for repositories using FAST vocabularies.
Related resources	FAIR for this principle relates to the following standards, databases and/or policies: • *https://doi.org/10.25504/FAIRsharing.p5df9c (FAST)*

FAIR Metric – I2 – Metadata – ISO language is linked data

Metric description	**https://doi.org/10.25504/FAIRsharing.18539f** This metric evaluates whether the metadata contains at least one ISO 639-2 / 639:2023 language code that resolves to content expressed in a linked data format. By specifically checking for ISO 639-2 / 639:2023 codes with formal semantic structure, the metric ensures that language descriptors support automated reasoning and semantic interoperability, providing a community-specific implementation of the broader I2 principle for repositories using ISO 639-2 / 639:2023.

Related resources	FAIR for this principle relates to the following standards, databases and/or policies: • https://doi.org/10.25504/FAIRsharing.dc041b *(ISO 639:2023)*

FAIR-I3: (meta)data include qualified references to other (meta)data

Data and metadata must include structured, typed links to related resources. Qualified references clarify the nature of relationships, enabling machines to understand context, provenance, and dependency.

For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.ae22b8.

Although a **generic** metric relating to this principle exists, it is insufficient for our needs, and we have created two **specialised** metrics for this principle, each testing a particular aspect of the outward links that are relevant to this benchmark.

FAIR Metric – I3 – Metadata – Qualified References to Related Research Objects

Metric description	**https://doi.org/10.25504/FAIRsharing.6cb5e5** I3 requires that (meta)data include qualified references to other (meta)data. The purpose of this principle is to ensure that digital objects are connected to the wider research ecosystem through explicit, semantically defined relationships. The existing generic I3 metric evaluates identifier references that have different domains from the root domain of the input identifier. However, this approach is insufficient for many resources (e.g. institutional repositories), where meaningful relationships may occur within the same domain and must therefore be identified through explicitly defined relationship types rather than domain comparison alone. The FAIR Metric – I3 – Metadata – Qualified References to Related Research Objects (FM:I3:RRO) metric evaluates whether the metadata retrieved upon identifier resolution contains at least one qualified, semantically defined link to another research object that provides contextual information but is not part of its direct supporting information. Implementations of this metric should test for: • **A qualified relationship to a related research object**, such as a publication or other research output, expressed using a defined relationship type(s) (e.g., "related to"), excluding relationships that indicate direct support (e.g., "supported by this record"). • **The relationship must represent contextual linkage** (providing relevant information on the context of the repository record) rather than representing direct supporting information required for understanding of the record. A repository record will pass this metric if at least one such qualified, contextual reference to another research object is present in the metadata.

FAIR Metric – I3 – Metadata – Qualified References to Versioned Records

Metric description	https://doi.org/10.25504/FAIRsharing.e10b26 I3 requires that (meta)data include qualified references to other (meta)data. The purpose of this principle is to ensure that digital objects are connected to the wider research ecosystem through explicit, semantically defined relationships. The existing generic I3 metric evaluates identifier references that have different domains from the root domain of the input identifier. However, this approach is insufficient for many resources (e.g. institutional repositories), where meaningful relationships may occur within the same domain and must therefore be identified through explicitly defined relationship types rather than domain comparison alone. This metric evaluates whether the metadata retrieved upon identifier resolution contains at least one qualified, semantically defined link to another version of the repository record. Implementations of this metric should test for: • **A qualified relationship indicating another version of the repository record** (e.g. previous version, newer version, or alternative DOI version), expressed using defined relationship type(s). • **The relationship must explicitly describe the version connection and must not represent other types of relationships** such as supporting files associated with the record. A repository record will pass this metric if at least one explicitly defined version relationship is present in the metadata.

FAIR-R1: (meta)data are richly described with a plurality of accurate and relevant attributes

Data and metadata must contain detailed, accurate, and context-specific attributes that enable interpretation and reuse. Description must go beyond minimal metadata to include elements that support validation and reproducibility.

⏫ For more information on the top-level principle R1, including documentation and justifications its implementation, please see https://doi.org/10.25504/FAIRsharing.821487.

☑ Metric R1.1 is **generic**.

💡 Metrics R1.2 and R1.3 are **specialised**. This means that we expect communities to create their own metrics for R1.2 and 1.3 in ways appropriate to their needs.

This principle is split into three distinct sub-principles, which require that the (meta)data 1) is released with a clear usage licence (R1.1, https://doi.org/10.25504/FAIRsharing.7612c1), 2) are linked to detailed provenance (R1.2, https://doi.org/10.25504/FAIRsharing.8e0027), and 3) meet relevant domain-specific community standards (R1.3, https://doi.org/10.25504/FAIRsharing.8e0027). There are two **generic metrics** available for R1.1: https://fairsharing.org/7843 for the database-level licence metric, and https://fairsharing.org/7844 for the metadata-level licence metric. While it is preferrable to use the metadata-level licence metric, in cases where the licence information is not provided within the resolved metadata, the database-level metric may provide some extra information.

This benchmark utilises both **generic metrics** for R1.1. The R1.2 and R1.3 card sets describe this benchmark's **specialised** metrics for those sub-principles.

FAIR-R1.1: (meta)data are released with a clear and accessible data usage license

A clear, machine-readable licence must specify the conditions under which data and metadata may be reused. Explicit licensing removes ambiguity and enables lawful and transparent reuse.

⏫ For more information on this principle, including documentation and justifications for the implementation of this principle, please see https://doi.org/10.25504/FAIRsharing.aff99f.

☑ There are two **generic metrics** available for R1.1: https://fairsharing.org/7843 for the database-level licence metric, and https://fairsharing.org/7844 for the metadata-level licence metric. While it is preferrable to use the metadata-level licence metric, in cases where the licence information is not provided within the resolved metadata, the database-level metric may provide some extra information.

This benchmark utilises the **generic metrics** for R1.1.

FAIR Maturity Indicator - Metadata contains link to license

Alignment with generic metrics	This generic metric **is sufficient** for our needs, **although it is unclear if a specialised test may need to be implemented for full alignment with ORA requirements**. R1.1 requires that (meta)data be released with a clear and accessible data usage licence. The purpose of this principle is to ensure that reuse conditions are explicitly stated and legally unambiguous. Within ORA, all records are subject to licence terms, and therefore R1.1 is satisfied by design at the repository level. Two licence contexts apply: 1. **Metadata licence:** ORA metadata is licensed under CC0 and details of this licencing is available via the ORA API webpage (https://ora.ox.ac.uk/api). This ensures that metadata is openly reusable. Although certain elements such as abstracts or summary documentation may remain subject to copyright and therefore removed from the CC0 version, the metadata itself is explicitly licensed CC0. As a result, the metadata component of this metric will always pass. 2. **Data licence:** Each ORA record is associated with a data licence. Where a depositor specifies a licence, it is recorded within the metadata. Where no licence is specified by the depositor, the default ORA deposit licence (https://ora.ox.ac.uk/terms_of_use) applies. The default deposit licence permits data download but does not grant reuse rights. Although this default licence may not be explicitly listed within the metadata, it is publicly declared and applies to all deposited data. For the purposes of this benchmark, and with regards to the data licence, the presence of either an explicit licence in the metadata or the application of the default ORA deposit licence satisfies the requirements of R1.1. Accordingly, ORA records are expected to pass the R1.1 evaluation both in the context of metadata and data licencing.
Generic metric names and summaries	**FAIR Metric – R1.1 – Metadata - discoverable data license in metadata** https://doi.org/10.25504/FAIRsharing.714d4e *This metric evaluates whether the metadata retrieved upon resolution of the provided identifier contains an explicit license declaration that clearly specifies usage conditions for the research object. By verifying that license information is present at the individual record level, the metric ensures that reuse conditions are transparently stated and legally unambiguous for each specific digital object.* **FAIR Metric - R1.1 - Metadata - Database-level licenses** https://doi.org/10.25504/FAIRsharing.908aa1 *This metric evaluates whether the hosting database or repository has declared at least one license in its FAIRsharing record, providing database-level information about usage conditions that may apply to content within the repository. While this approach does not specify which license applies to individual research objects, it provides useful contextual information about data reuse conditions, particularly when license information is not included in individual metadata records.*

FAIR-R1.2: (meta)data are associated with detailed provenance

Data and metadata must include structured provenance information describing origin, processing history, and responsibility. Provenance enables trust, reproducibility, and informed reuse decisions.

These metrics are **specialised**. (Remember that the metric for R1.1 was **generic**, and the metric for next sub-principle – R1.3 - is also **specialised**.)

For more information on the principle that this metric must implement, including documentation and justifications its implementation, please see https://doi.org/10.25504/FAIRsharing.3e9860.

Five Specialised R1.2 Metrics

Metric descriptions	R1.2 requires that (meta)data be associated with detailed provenance. The purpose of this principle is to ensure that the origins, contributors, funding context and related research outputs of a digital object are clearly described, enabling informed reuse and proper attribution. Within ORA, provenance is represented through qualified references to external identifiers and related research objects. For the purposes of this benchmark, a qualified reference is defined as a reference in which the metadata explicitly describes the type of relationship between the ORA record and the referenced entity. Evaluation under this metric assesses the presence of the following provenance-related cross references. Each provenance dimension is evaluated independently and contributes cumulatively to the overall R1.2 score. For each dimension, the presence of at least one qualified reference is sufficient to pass that dimension. Each reference type will be evaluated independently. 1. **ORCID identifiers for people** **https://doi.org/10.25504/FAIRsharing.342aaa:** Individuals associated with an ORA record may be identified using ORCIDs. For datasets assigned a DOI, at least one creator (dc:creator) must be present. The presence of ORCIDs linked to individuals, together with defined roles, constitutes qualified provenance references. 2. **Funder identification** (ROR ID): Funding organisations may be identified using ROR identifiers. Where a funder role is assigned, this creates a qualified reference describing the financial provenance of the record. Two metrics for this content: 1) https://doi.org/10.25504/FAIRsharing.607ad4: Funding information is defined in metadata and 2) https://fairsharing.org/8185: metric that checks for a ROR ID (the previously listed metric does not check for RORs, just absence/presence of a funder). 3. **Institutional identification** https://doi.org/10.25504/FAIRsharing.99349b: Author affiliation is defined in metadata: Non-UOXF institutions associated with contributors may or may not be identified using ROR identifiers. Role assignments create qualified relationships linking individuals and their affiliated organisations. Currently, this is only tracked for the University of Oxford, but we may expand this to external institutions in the future. 4. **Linked publications** **https://doi.org/10.25504/FAIRsharing.653df6:** Where an ORA record includes a linked publication in the "related items" field with the relationship type "supported by this record," this constitutes a qualified reference connecting the dataset to its scholarly context. 5. **Original content source** **https://doi.org/10.25504/FAIRsharing.4264f7:** Where applicable, metadata may describe the original source of the record (e.g., OpenAlex ingestion, direct deposit, or deposit via internal Oxford systems). This provides additional provenance information regarding how the record entered ORA. For this benchmark, R1.2 compliance will be assessed by evaluating the presence of these qualified provenance references. For each of these reference types, where this information is present, at least one source declaration is sufficient to pass this metric. The overall level of FAIRness under R1.2 increases as additional provenance dimensions are satisfied. Possible additional features after MVP:

	• We may decide to weight different outward links in different ways. • The MVP metric will return a pass if at least one of the cross references is present. Would we ever be interested in providing a higher 'score' if eg all authors had ORCIDs? This evaluates completeness of coverage rather than just presence. • Optionally apply CREDIT taxonomy • Would we want to evaluate the GUPRI qualities of the external cross reference identifiers? Note: While https://doi.org/10.25504/FAIRsharing.0a3749 (dataset cited in publication via reverse search in OpenAlex) exists, this is unsuitable for an institutional repository. Instead, we wish to check that the ORA record has an appropriate publication.
Related resources	• *ROR registry (https://doi.org/10.25504/FAIRsharing.1jKfji)* • *Dublin Core (https://doi.org/10.25504/FAIRsharing.3nx7t)* • *ORCID registry (https://doi.org/10.25504/FAIRsharing.nx58jg)*

FAIR-R1.3: (meta)data meet domain-relevant community standards

Data and metadata must conform to recognised standards and community norms appropriate to their domain. Alignment with disciplinary standards ensures interoperability, interpretability, and comparability within and across research communities.

💡 These metrics are **specialised**. (Remember that the metric for R1.1 was **generic**, and the metric for the previous sub-principle – R1.2 - is also **specialised**.)

⏫ For more information on the principle that this metric must implement, including documentation and justifications its implementation, please see https://doi.org/10.25504/FAIRsharing.87d197.

R1.3 requires that (meta)data meet domain-relevant community standards. The purpose of this principle is to ensure that digital objects are described using standards that are recognised and adopted within the relevant research community. As an institutional repository, ORA is subject agnostic. Unlike discipline-specific repositories, ORA cannot enforce domain-specific community standards across all submissions. Instead, alignment with R1.3 is evaluated through the application of three metrics testing for cross-disciplinary or subject-agnostic standards that appropriate to institutional repositories.

FAIR Metric – R1.3 – Metadata – ORA institutional repository relevant standards

Metric description	**FAIR Metric – R1.3 - Metadata – use of FAST subject vocabulary (FM:R1.3:M:UseFAST).** https://doi.org/10.25504/FAIRsharing.6dc3a5 **What is being measured?** This dimension evaluates whether the digital object referenced by the identifier includes at least one resolvable FAST (Faceted Application of Subject Terminology) subject heading. The test inspects the subject or keyword fields of the metadata record to verify the presence of a persistent, resolvable FAST URI that successfully links to the recognized community authority. **Why should we measure it?** FAIR Principle R1.3 requires metadata to follow domain-relevant community standards. While applicable to a variety of communities, the use of FAST is particularly useful for institutional repositories and other generalist contexts, as it provides a standardized, cross-disciplinary schema. Ensuring these identifiers are resolvable makes the subject description machine-actionable and provides alignment with a recognized global standard.
Related resources	• *FAST: https://doi.org/10.25504/FAIRsharing.p5df9c*

Metric description	**FAIR Metric – R1.3 - Metadata – use of ISO 639 language (FM:R1.3:M:UseISO639).** https://doi.org/10.25504/FAIRsharing.eb0cef **What is being measured?** This dimension evaluates whether the metadata for the digital object includes at least one resolvable language attribute defined by the ISO 639 standard. The assessment validates the value held in the language field against the official ISO 639 registry, searching for a valid, standardized code (specifically 639-2 / 639:2023) or a URI pointing to the ISO registry. **Why should we measure it?** Linguistic interoperability is essential for the international discovery and reuse of research data. For institutional repositories and other generalist contexts, the use of ISO 639 codes ensures that datasets are correctly indexed by global aggregators and assistive technologies. Measuring this alignment confirms that the repository is following a recognized community standard for linguistic description, satisfying the requirements of R1.3.
Related resources	• *ISO 639:2023:* https://doi.org/10.25504/FAIRsharing.dc041b

Metric description	**FAIR Metric – R1.3 - Metadata – use of recognised and structured generic metadata format (FM:R1.3:M:UseStuctGenericFormat).** https://doi.org/10.25504/FAIRsharing.389fad **What is being measured?** This dimension evaluates whether the dataset metadata is structured according to predefined, widely adopted metadata schemas appropriate to institutional repositories and other generalist contexts, such as schema.org (JSON-LD) or DataCite. The test checks the record's landing page for embedded or linked structured data that can be successfully parsed against the declared community schema. **Why should we measure it?** Beyond individual fields, the overall structure of a metadata record must meet community expectations to ensure machine-readability. Implementing structured formats allows institutional repositories and other generalist contexts to provide record metadata that can be seamlessly consumed by major search engines and international research graphs. Measuring the use of such formats ensures that the record is not merely a human-readable web page, but a structured data object that complies with modern, domain-relevant standards for data exchange.
Related resources	• *schema.org:* https://doi.org/10.25504/FAIRsharing.hzdzq8 • *JSON-LD:* https://doi.org/10.25504/FAIRsharing.8f9bbb