

Project Title	Expanding FAIR solutions across EOSC
Project Acronym	FAIR-IMPACT
Grant Agreement No.	101057344
Start Date of Project	2022-06-01
Duration of Project	36 months
Project Website	fair-impact.eu

## D5.3 - Final recommendations on implementing and exposing FAIR assessment for data and code

### Guidelines and example implementations for existing and aspiring trustworthy digital repositories

---

Work Package	<b>WP5 - Metrics, certification, and guidelines</b>
Lead Author (Org)	<b>Wim Hugo (KNAW-DANS)</b>
Contributing Author(s) (Org)	<b>Robert Huber (UBremen), Robert Ulrich (KIT), Hervé L'Hours (UESSEX-UKDS), Oliver Parkes (UESSEX-UKDS), Joy Davidson (UEDIN/DCC), Parham Ramezani (LifeWatch)</b>
Due Date	<b>2025.05.31</b>
Date	<b>2025.05.27</b>
Version	<b>V1.0 - DRAFT not yet approved by the European Commission</b>
DOI	<b>10.5281/zenodo.15534285</b>

#### Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)

## Versioning and contribution history

Version	Date	Author	Notes
0.1	2025.02.28	Wim Hugo (KNAW-DANS), Maaïke Verburg (KNAW-DANS)	TOC and V0.1
0.2	2025.03.03	Wim Hugo (KNAW-DANS)	Inception, structuring
0.3	2025.04.08	Joy Davidson (UEDIN/DCC)	Support Action Description, Outcomes
0.4	2025.04.22	Wim Hugo (KNAW-DANS), Joy Davidson (UEDIN/DCC)	Landscape, Support Action
0.5	2025.04.23	Parham Ramezani (LifeWatch), Wim Hugo (KNAW-DANS)	Examples and Guidelines
0.6	2025.04.24	Wim Hugo (KNAW-DANS)	Guidelines, Appendices, Glossary
0.7	2025.04.25	Wim Hugo (KNAW-DANS), Robert Huber (UBREMEN), Hervé l'Hours (UESSEX-UKDS)	Conclusions and Recommendations Revised Guidelines, Appendices Support Action Outcome
0.8	2025.04.28	Hervé l'Hours (UESSEX-UKDS), Robert Ulrich (KIT)	Sub-editing & Comments
0.9	2025.04.30	Wim Hugo (KNAW-DANS)	Release for Internal Review
0.10	2025.05.15	Susanna-Assunta Sansone (Oxford University)	Internal Review Comments
0.11	2025.05.19	Mickey Lindlar (TIB)	Internal Review Comments
0.12	2025.05.22	Wim Hugo (KNAW-DANS), Robert Huber (UBREMEN) Robert Ulrich (KIT)	Corrections, Improvements
0.13	2025.05.28	Wim Hugo (KNAW-DANS)	Final corrections, revised executive summary.
1.0	2025.05.28	Maaïke Verburg (KNAW-DANS)	Version published on Zenodo

### Disclaimer

FAIR-IMPACT has received funding from the European Commission's Horizon Europe funding programme for research and innovation programme under the Grant Agreement no. 101057344. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.





# Table of Contents

<b>Executive Summary.....</b>	<b>7</b>
<b>1 Introduction.....</b>	<b>9</b>
1.1 Context.....	9
1.2 Landscape.....	10
1.2.1 Attributes and Levels of Granularity.....	10
1.2.2 Systemic Issues.....	10
1.2.3 Scope and Mechanism of Exposing Repository Attributes.....	11
1.3 Methodology.....	13
1.3.1 Case Studies Considered.....	14
1.3.2 Validation through FAIR-Impact Support Action.....	14
<b>2 Reference and Information Models.....</b>	<b>15</b>
2.1 DCAT Model.....	15
2.2 schema.org Model.....	17
2.3 Validation of Assertions.....	18
<b>3 Consolidated Guidance and Best Practices.....</b>	<b>19</b>
<b>4 Prototype Development.....</b>	<b>22</b>
<b>5 Conclusions and Recommendations.....</b>	<b>24</b>
<b>References.....</b>	<b>26</b>
<b>Appendices.....</b>	<b>29</b>
Appendix A: Prototype for DCAT Encodings of Repository Attributes.....	29
A.1 Exposing Repository Information with DCAT.....	29
A.2 Descriptive Metadata.....	29
A.3 Supported Standards.....	30
A.4 Policies and Principles.....	31
A.5 Certification and Quality Information.....	31
Appendix B: Prototype for schema.org Encoding of Repository Attributes.....	32
B.1 Exposing Repository Information with schema.org.....	32
B.2 Descriptive Metadata.....	32
B.3 Supported Standards.....	33
B.4 Policies and Principles.....	35
B.5 Certification and Quality Information.....	35
Appendix C: Prototype FAIRiCAT/ Signposting Implementation.....	36
Appendix D: Repository Attribute Prototype.....	39
Appendix E: Repository Certification Evidence.....	40
Appendix F: Lessons learned through the Support Action.....	42
Appendix G: Detailed Recommendations for Future Work.....	43
<b>Appendix H: Checklists for Implementation of Guidelines.....</b>	<b>45</b>
H.1 Standards for Exposing Repository Attributes.....	45
H.2 Encoding Repository Attributes.....	47

H.3 Exposing Repository Attributes.....	47
H.4 Implementation Mechanisms for Exposing Repository Attributes.....	48
H.5 Examples and Prototypes.....	49
H.6 Encoding Specific Attributes: FAIR and Trust Certification.....	49
H.7 Identifiers for Repositories.....	50

## List of Figures

Figure 1 - Levels of Granularity for Attribute Sets and Interactions with Validation Authorities.....	10
Figure 2 - Methodology.....	14
Figure 3 - Overview of the standard DCAT model.....	16
Figure 4 - Overview of the standard schema.org model.....	17
Figure 5 - Scope of Prototype Development.....	23
Figure A.1 - Example of dcat:service usage.....	30
Figure A.2 - Metadata standards and identifier types.....	31
Figure A.3 - Example of policy exposure.....	31
Figure A.4 - Example of certificate exposure.....	32
Figure B.1 - Example of exposure of available APIs.....	34
Figure B.2 - Example of standards exposure.....	34
Figure B.3 - Example of exposure of policy.....	35
Figure B.4 - Example of exposure of certification information. Please note that instead of the generic amt.coretrustseal.org URL, a valid reference to a certificate archived at the CoreTrustSeal certification repository should be used. At present, this can be tested via a Prototype.....	36
Figure C.1 - Typed Link Elements in HTML <head>.....	37
Figure C.2 - Example of a JSON-LD file containing schema.org encoded attributes.....	37
Figure C.3 - Example of a LinkSet file containing repository affordances.....	38
Figure C.4 - Example of Embedded JSON-LD based on DCAT Encoding.....	38
Figure D.1 - Dummy Repository Home Page.....	39
Figure D.2 - Example Result from the Repository Checker Application.....	40
Figure E.1 - Implementation of FAIRiCAT for a Prototype CoreTrustSeal Server.....	41
Figure E.2 - Search and Harvesting Affordances for the Prototype CoreTrustSeal Repository.....	41

## List of Tables

Table 1: A Sample of the Repository Attribute Definition Landscape.....	11
Table 2: Sample of Repository Information Guidelines Landscape.....	13
Table A.1 - Mapping of selected DRAWG attributes to recommended DCAT properties.....	29
Table A.2 - Mapping of dct:conformsTo to other properties for different DRAWG attributes.....	31
Table A.3 - Mapping of DRAWG attribute to properties.....	32
Table B.1 - Mapping of DRAWG attributes to schema.org properties.....	33
Table B.2 - Mapping of DRAWG attributes to schema.org properties.....	34

Table B.3 - Mapping of schema:publishingPrinciples to several DRAWG attributes.....	35
Table B.4 - Mapping of DRAWG certification attributes to schema.org properties.....	36
Table G.1 - Recommendations for Future Work.....	43
Table H.1 - Standards for Exposing Repository Attributes.....	46
Table H.2 - Best Practices: Encoding Repository Attributes.....	47
Table H.3 - Best Practices: Exposing Repository Attributes.....	47
Table H.4 - Best Practices: Implementation Mechanisms.....	48
Table H.5 - Best Practices: Implementation Mechanisms.....	49
Table H.6 - FAIR and Trust Certification.....	49
Table H.7 - Identifiers for Repositories.....	50

## ACRONYMS

Acronym	Description
WP	Work Package
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
TRSP	Technical Repository Service Provider
WG	Working Group
CAT	Compliance Assessment Toolkit
LTP	Long-Term Preservation
TF	Task Force
RDF	Research Description Framework
DQV	Data Quality Vocabulary
DDI	Data Documentation Initiative
PID	Persistent Identifier
DOI	Digital Object Identifier

## TERMINOLOGY

Term	Description
Digital Object	Any object that broadly includes research outputs and supporting materials - datasets, articles, reports, code, semantic artefacts, and similar materials. In this report, the focus is on datasets.
Catalogue	We base this definition loosely on the DCAT class, which describes a catalogue as a collection of datasets or data services. A catalogue can contain one or more datasets, data series, or services. Catalogues are used synonymously with ‘collections’.
Repository	A repository is not formally defined in the DCAT vocabulary, but is generally accepted to provide for one or more catalogues to be curated, maintained and made available for and on behalf of end users.

Term	Description
Registry	A registry provides an inventory of repositories for the benefit of an end user community.
Attribute	An attribute describes a characteristic, feature, or action associated with a digital object, a catalogue, or a repository.
Assessment	An assessment determines the alignment between attributes and some community expectation, such as FAIR.
Validation, Certification	Validation verifies or confirms the integrity of an assessment. Validation can take place in more than one way, including mechanisms for self-validation, certification, or community review.

## Executive Summary

---

Task 5.4 in FAIR-IMPACT had the objective of developing guidelines in respect of the sharing of repository attributes for a number of purposes, including sharing information on the characteristics of a repository, assisting end users with selection of repositories, or validation of such repository attributes. A special focus of this general set of guidelines were exposure of FAIR attributes of trustworthy repositories.

Claims in respect of trustworthiness of a repository and its certification as such, and of FAIR compliance in respect of the objects available from a repository (aggregated to reflect the status of FAIR in the repository) are used as examples throughout the report. These are important examples of repository attributes that may require validation, but certainly not the only ones. One can validate the claims in a number of ways, but this is not in scope for the report: we are concerned with the simplest and most expedient way in which such an attribute can be exposed to machines.

The report covers the following topics:

- **Context and Landscape:** an overview of the motivation for exposure and standardisation of repository attributes, a sample of efforts to inventorise and standardise repository attributes and their definitions, and the approach followed in this task of FAIR-IMPACT WP5 to develop guidelines and supporting prototypes.
- **Reference and information Models:** a section providing details on the available specifications for encoding repository attributes, and how these might be extended in future to link to the characteristics, activities and workflows commonly associated with or encountered in repositories.
- **Consolidated Guidance:** A summary of the guidelines, with an implementation checklist of informal recommendations for repositories presented in Appendix H.
- **Prototype Development:** Development of prototypes as demonstrators for implementation of the guidelines and the recommended specifications are covered in detail in this report. Five prototypes were developed to illustrate guideline implementation and serve as blueprints for repositories:
  - A DCAT-based implementation of repository attribute encodings (Appendix A)
  - A schema.org-based implementation to illustrate an alternative to DCAT (Appendix B)
  - A FAIRiCAT/ Signposting example is used to illustrate how machines, knowing only the URL for a repository, can be directed towards detailed attribute information for the repository and the objects it hosts (Appendix C).
  - A ‘Dummy Repository’ was developed to illustrate the behaviour of a repository that implements either DCAT and/ or [schema.org](https://schema.org) examples, together with FAIRiCAT (Appendix D)

- Appendix E describes a prototype modification to the CoreTrustSeal certification repository to expose the status of certification of repositories certified by them.

These prototypes were also used to obtain community feedback via a FAIR-IMPACT support action on relative ease of implementation.

- **Support Action:** the guidelines and prototypes reported here were validated with community input via a support action arranged by FAIR-IMPACT. The report provides a summary of the support action, participants, and the feedback received from participants (the latter presented in Appendix F).
- **Conclusions and Recommendations** are presented in the final section of the report, supported by detail on recommendations for future work in Appendix G.

While this is not the primary focus of the report, a subtitle was included in naming this deliverable to clearly mark and communicate the extra content and recommendations provided to existing and aspiring trustworthy digital repositories (TDRs) with regards to implementing and exposing trustworthiness status.



# 1 Introduction

---

## 1.1 Context

The objective of Task 5.4 (T5.4) in Work Package 5 in the FAIR-IMPACT project is the development of guidelines and mechanisms improving connections between repository registries and discovery portals (including EOSC Portal), repository trustworthiness assessments (e.g. CoreTrustSeal) and FAIR digital object assessments. The aim is to enable seamless discovery of repository attributes (including their trustworthiness) and dataset and code attributes (including their FAIRness) [1].

The report does not deal with exposure of trustworthiness or FAIR attributes only, but these were used as important examples, and an attempt was made to answer the following specific questions:

- How does a repository expose FAIR data assessment outcomes?
- How does a repository expose its FAIR-enabling qualities?
- How does a repository expose trustworthiness status?

In evaluating the answers to these questions, one should consider, inter alia, the mechanism for assertion of claims and information - typically by differentiating between self-declared, validated, and/ or verified information [16]. In addition, the FAIR and trustworthiness related attributes of a repository forms part of a wider need to share repository attributes and capabilities. The prototypes used to illustrate exposure of attributes allow for any attribute to be included.

The work reported in this document is presented as guidance for repositories on implementation of mechanisms to expose machine-actionable repository attributes and capabilities. This is supported by prototype implementations to illustrate the guidance.

Previous work reported by this task [16] identified (updated in this report) some principles for the guidance and prototype(s) to be developed:

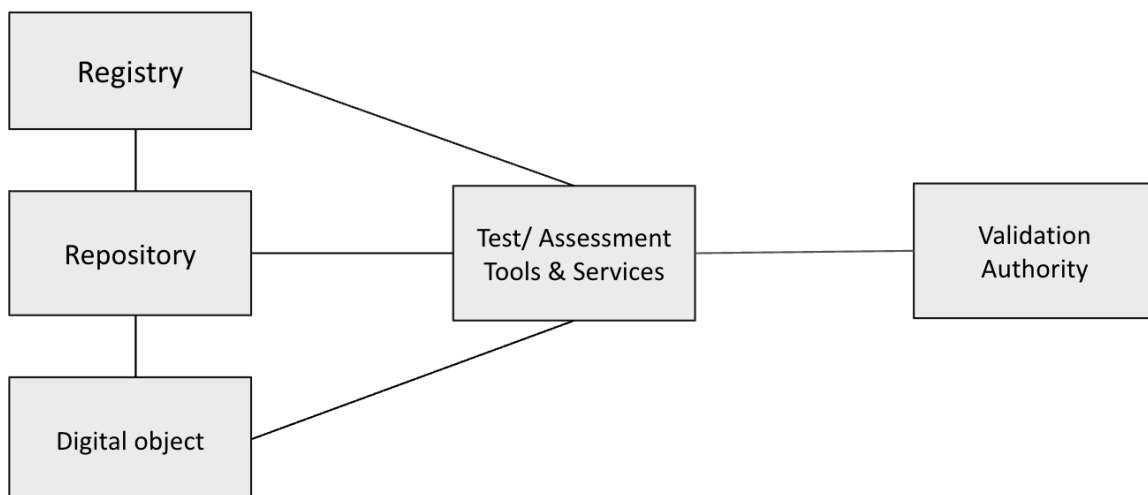
- The guidelines and accompanying prototype should focus on how to expose information and are neutral on which information should be exposed
- The model should be flexible enough to fit all relevant assertions of information
- The guidelines should be relevant to a broad range of repository and object assessment processes
- The final guidelines and prototype should be meaningful and usable by both human actors and machine agents

## 1.2 Landscape

The current research data environment is populated with digital objects, each with a range of characteristics, including those that, if exposed transparently and understandably, can inform perceptions and decisions related to FAIRness and trustworthiness [16]. These objects are generally hosted and possibly preserved in repositories, which may or may not organise the objects into collections of similar type, scientific domain, sensitivity, or other distinguishing feature. A repository may be curated at different levels of care, and is generally associated with an organisation, which may host multiple repositories. Repositories may also be co-hosted or co-curated by multiple organisations, or are truly federated. Repositories and organisations are often described in one or more registries. These arrangements result in **interrelated sets of digital objects and their attributes at varying granularity**.

### 1.2.1 Attributes and Levels of Granularity

It should be clear that attributes can be exposed for any of the levels of detail shown in **Figure 1** (amongst others). Moreover, most attributes can be aggregated to a higher level if required<sup>1</sup>. Assessment and subsequent validation can be performed for any level of detail (but usually for repositories or objects). Assessments and validations can be repeated over time and change as a result.



**Figure 1 - Levels of Granularity for Attribute Sets and Interactions with Validation Authorities**

(adapted and extended from [16], [22])

We will focus on the **exposure of repository attributes** in the remainder of this document.

### 1.2.2 Systemic Issues

Previous work undertaken by this task [16] identified several systemic issues related to the way in which repositories identify, characterise, and describe themselves. These include:

<sup>1</sup> Example: FAIR assessments for individual objects aggregated for a collection, repository, or organisation.

- It is **not simple to identify repositories** and their collections unambiguously, leading to issues with referencing, and there is no consistency in how repositories characterise themselves [16].
- The transparent disclosure of **self-declared assertions** about trust-related characteristics should be possible and supported in addition to exposing formal certification information [16].

In terms of FAIR assessment, there are also a number of systemic issues to consider:

- Many different FAIR assessment tools<sup>2</sup> are currently available for use, with a wide diversity of aims, audiences, purposes, target objects, execution types, and interpretations of FAIR [16], [24], [32]. Current efforts that address this diversity in part are under way in the FAIRCORE4EOSC [17] and in the OStrails [23] projects, providing harmonising information models and specifications on which applications can be developed.

### 1.2.3 Scope and Mechanism of Exposing Repository Attributes

The scope of repository attributes that are applicable depends on the context, and there have been several contributions to the definition of these attributes for the purposes of a specific application or context. Current initiatives that impact the definition of repository attributes are also under way. Table 1 provides a sample of the repository attribute definition landscape.

**Table 1: A Sample of the Repository Attribute Definition Landscape**

#	Context (1)	Description	Reference
1.1	DRAWG	The Digital Repository Attributes WG in RDA developed and published a survey of the most important repository attributes and how they might be encoded and offered to end users.	[11]
1.2	CoreTrustSeal	CoreTrustSeal defines a set of criteria and supporting guidelines for certification of repositories, indirectly thereby defining a set of repository attributes that are of interest.	[12]
1.3	nestor	Similarly, nestor defines criteria and guidelines for certification of repositories (largely focused on German institutions), and as a result indirectly defines repository attributes.	[13]
1.4	ISO 16363	ISO 16363:2025 offers the most comprehensive range of certification criteria, and thereby also defines a wide range of repository attributes indirectly.	[14]
1.5	TRSP WG	The Technical Repository Service Provider Working Group in RDA has embarked on a wide-ranging landscape analysis of repository and service attributes. These have not been	[15]

<sup>2</sup> <https://fairassist.org> lists more than 40.

		formally documented but work in progress is publicly available, and already covers several of the sources listed in the table (amongst others).	
1.6	FAIR Data Maturity Model	The indicators that are used in the FAIR data maturity model are derived from the FAIR principles and aim to formulate measurable aspects of each principle that can be used by evaluation approaches. As such, they imply availability of object or repository attributes to support such evaluation.	[24]
1.7	FAIRCORE4EOSC CAT	The Compliance Assessment Toolkit (CAT), developed by FAIRCORE4EOSC and released operationally at the end of March, 2025, can be configured for assessment and persistence of a variety of assessment profiles, including FAIR and TRUST-related assessments and appraisals. The CAT provides for persistence of assessment outcomes, and a mechanism for retrieval of such information via an API.	[17]
1.8	EOSC EDEN and FIDELIS	The EOSC EDEN and FIDELIS projects are closely aligned, and collaboratively work on the identification, description, and exposure/ assessment/ verification of repository attributes. These projects are focused on trustworthy repository-related perspectives, and cover topics for long-term preservation (EOSC EDEN) and trustworthy repository networking and federation (FIDELIS), for which there is clearly an overlap.	[30]
1.9	Existing Registries	Several existing registries define repository attributes in one way or another, and these are discussed in more detail in Section 6.4.	[2], [3], [5], [7]
1.10	EOSC LTP TF	The EOSC Long-Term Preservation Task Force provides an overview, inter alia, of preservation-related repository attributes pertinent to trustworthy repositories.	[22]
1.11	EOSC FAIR Metrics Task Force	The EOSC FAIR Metrics Task Force reported on the issues created by, and possible solutions to the diversity of tools and divergent outcomes of FAIR assessments performed by these tools. It does not in itself define a	[32]

The scope of repository attributes is by nature diverse, and multiple definitions can at times be found for the same concept. Work is clearly needed in future<sup>3</sup> to disambiguate and harmonise the inventory of repository attributes in use by communities, and to develop specifications for exposing these attributes to end users.

Several resources are already available in respect of guidelines for exposure of repository information and the objects it hosts , Table 2 provides a sample.

<sup>3</sup> Efforts to disambiguate and define repository attributes are under way in the EOSC EDEN and FIDELIS projects, as well as in the TRSP WG in RDA.

**Table 2: Sample of Repository Information Guidelines Landscape**

#	Context (1)	Description	Reference
2.1	FAIR Data Maturity Model	The indicators that are used in the FAIR data maturity model are derived from the FAIR principles and aim to formulate measurable aspects of each principle that can be used by evaluation approaches. Guidelines are provided in respect of implementation of measures.	[24]
2.2	FAIRCORE4EOSC CAT	The Compliance Assessment Toolkit (CAT), developed by FAIRCORE4EOSC and released operationally at the end of March, 2025, offers guidelines on the information model for assessment, and as a result provides standardised API templates for criteria-based repository attributes.	[17]
2.3	FAIR-IMPACT	FAIR-IMPACT has published a number of prior resources that are in scope for inclusion into the guidelines in this report.	[16], [18], [19]
2.4	FAIRsFAIR	One of the deliverables in the FAIRsFAIR project dealt explicitly with the features of FAIR repositories.	[20]
2.5	OSTRAILS	The OSTRAILS project is defining API specifications for invocation and exchange of FAIR-related test results and supporting guidance. The specifications are object-level and not repository-level, but nevertheless may be of use in standardising the aggregation of FAIR assessments to reflect repository status. This work is linked to repository-level assessments envisaged for CAT (see 2.2 and 1.7).	[23]

### 1.3 Methodology

The methodology followed by T5.4 is summarised in Figure 2. The task was approached in three phases.

- **Phase 1:** Examine current practices in respect of assessment and certification, as well as practices in respect of repository attribute encoding and exposure. Development of guidelines, recommendations, and best practices.

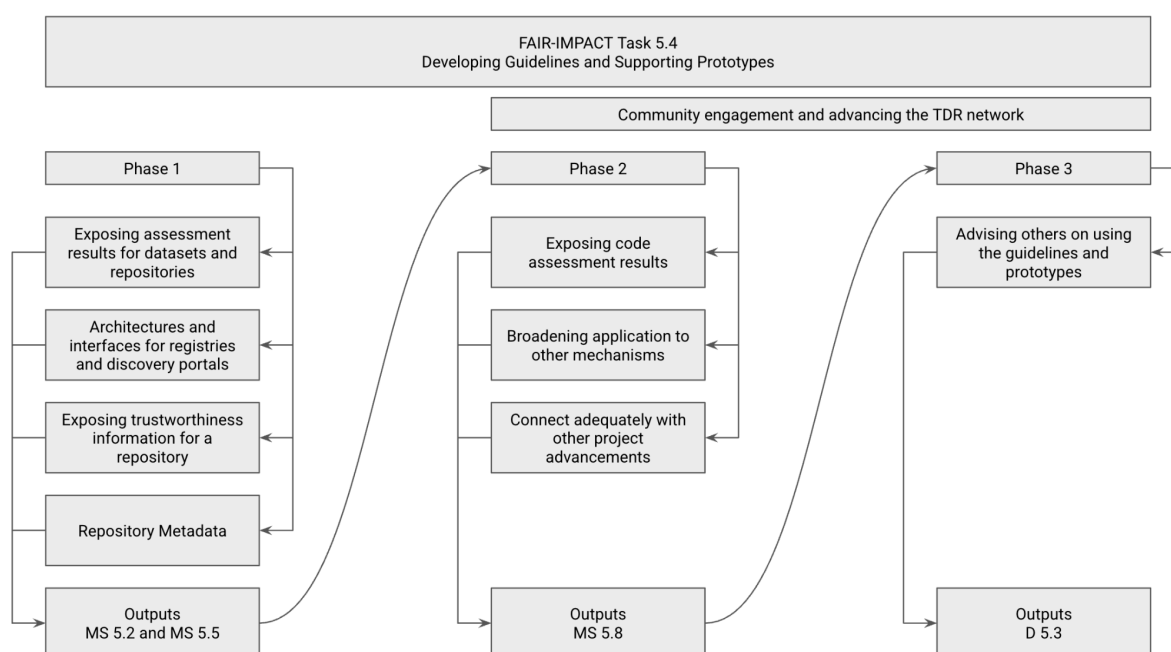
**Outputs:** milestone reports *M5.2 Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes* [16] and *M5.5 Initial repository registry support for discovery of repositories, policies, and interfaces* [18].

- **Phase 2:** Broadening the scope of the guidelines and testing them in part through prototypes, other FAIR-IMPACT outputs, and other mechanisms of assessment. Update and extend the guidance and develop supporting prototypes.

**Output:** milestone report *M5.8 Pilot for Exposing Repository Trustworthiness Status and FAIR Data and Code Assessment Outcomes in Generic and Disciplinary Registries and Portals* [29].

- **Phase 3:** participate in a Support Action to test and validate the guidelines and prototypes, and work on a final iteration of the guidelines and prototypes based on feedback. Gather lessons learned and recommendations for further work

**Output:** Deliverable D5.3 (this report)



**Figure 2 - Methodology**

### 1.3.1 Case Studies Considered

Early work in task T5.4 focused on detailed analysis of validation processes for trustworthy repositories (based on CoreTrustSeal as an example), and FAIR assessment (based on the F-UJI tool). The work resulted in a preliminary set of guidelines and principles to be applied in the development of prototypes [19].

### 1.3.2 Validation through FAIR-Impact Support Action

FAIR-IMPACT enabled a series of Support Actions, and task T5.4 was successful in applying for such support ('Testing the trustworthy and FAIR-enabling repositories prototype'). The support action is described in detail in Appendix F, and in the FAIR-Impact website.<sup>4</sup> Feedback on implementation of the guidelines, using the prototypes in Appendices A, B, and

<sup>4</sup> <https://fair-impact.eu/testing-tdr-and-fair-enabling-prototype>

C as examples, was provided by 8 participants across a variety of repositories and institutions:

- Cristiana Bettella, University of Padua
- Dieuwertje Bloemen, KU Leuven
- Juan Corrales, Consorcio Madroño
- Ioana Maria Cortea, National Institute for Research and Development in Optoelectronics - INOE 2000
- Monique Denissen, Austrian NeuroCloud
- Beth Knazook, Digital Repository of Ireland (observer)
- K  vin Salesse, IsoArch
- Socrates Varakliotis, UCL Advanced Research Computing

The feedback received has now been implemented in the guidelines and prototypes where applicable.

## 2 Reference and Information Models

The basic architecture for the encoding and exposure of repository attributes was confirmed in Milestone M5.8 [27], and the decision has been made to use RDF<sup>7</sup>, which allows for the presentation of information in ways that are understandable to both human actors and machine agents. To enable exploration at repository and digital object level, the use of [schema.org](http://schema.org)<sup>5</sup> or [DCAT](https://www.w3.org/TR/vocab-dcat-3/)<sup>6</sup> has been recommended [33]. DCAT is an established RDF vocabulary to describe digital catalogs. It is a mature standard and recommended by the World Wide Web Consortium (W3C). Schema.org is a standard initiated in 2011 by search engine operators, including Bing, Google and Yahoo. It is widely adopted in the World Wide Web. To expose quality information such as certificates of FAIR assessment results in RDF the use of the Data Quality Vocabulary (DQV) [11] was proposed which also is part of the W3C data on the web family of standards. This is the reference framework for the architecture of the prototype and the recommendations described below to expose repository information.

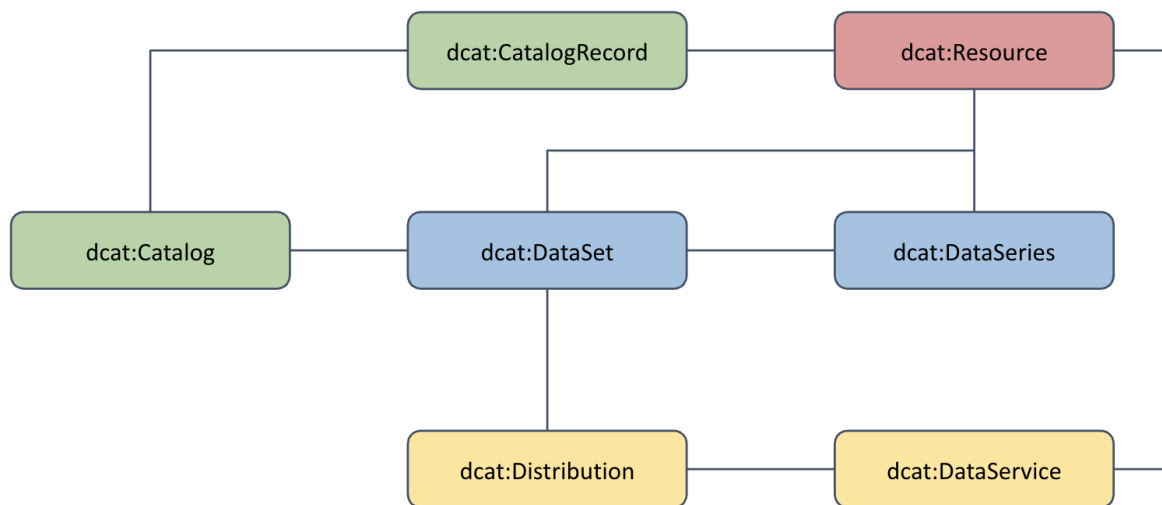
Section 4 discusses architecture in more detail.

### 2.1 DCAT Model

DCAT provides an applicable information model for description of (data) catalogues. A model based on DCAT (Figure 3) was used as a point of departure for identification and description of the actors and the relations between them. [18].

<sup>5</sup> <http://schema.org>

<sup>6</sup> <https://www.w3.org/TR/vocab-dcat-3/>



**Figure 3 - Overview of the standard DCAT model**

The classes (entities) identified in Figure 3 are discussed below.

**dcat:Catalog:** This class represents a collection of datasets or data services. A catalogue can contain one or more datasets, data series, or services (resources - dCat:Resource) and provides metadata about them, such as title, description, keywords, and access information.

**dcat:CatalogRecord:** This class is optional and describes a specific resource (dcat:Resource) within a catalogue. It typically is used to capture metadata such as its creation date, publisher, and other administrative details.

**dcat:Resource:** This is a generic class representing any entity that can be described in a data catalogue. It serves as a superclass for more specific types such as *Dataset*, *DataSeries*, *Distribution*, and *DataService*.

**dcat:Dataset:** This class represents a collection of data, often organised and presented in a structured format. A dataset typically includes metadata describing the data, such as its title, description, keywords, temporal and spatial coverage, licensing information, and access methods.

**dcat:DataSeries:** This class represents a collection of separate datasets that can be grouped or belong together.

**dcat:Distribution:** This class describes a specific way in which a dataset or data service is available, such as a file format, access method (e.g., download, API), or endpoint.

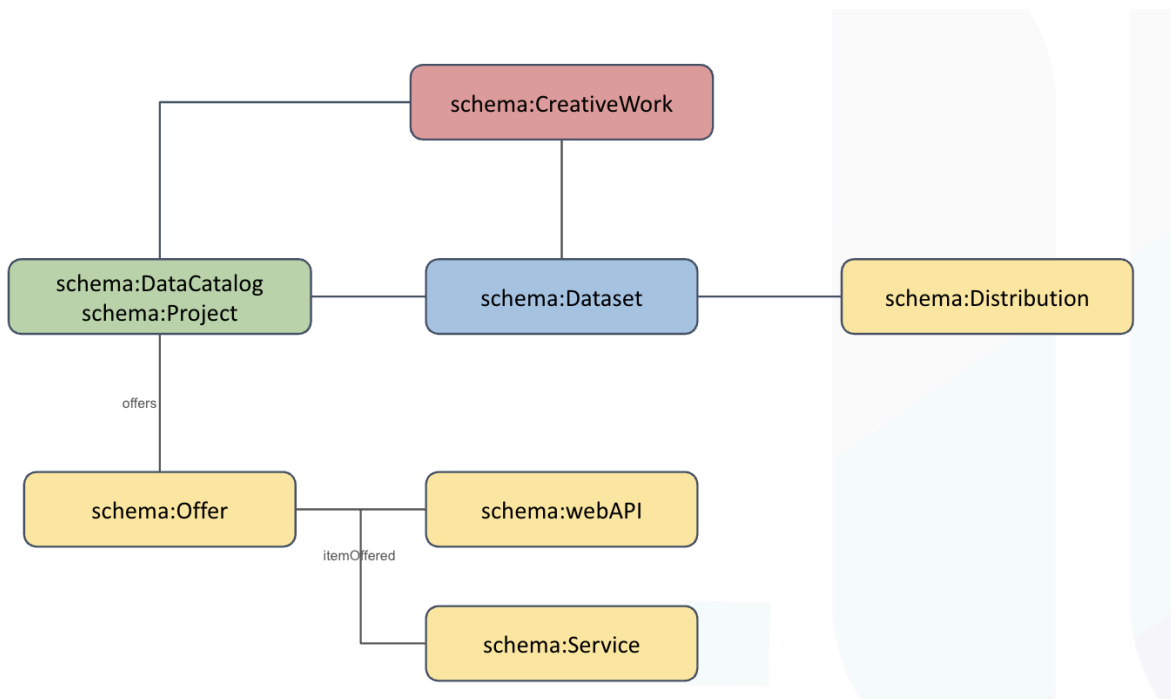
**dcat:DataService:** This class represents a service or API that provides access to data (a resource) via a distribution. It enables repositories to expose descriptions of their technical interfaces. The properties *dcat:endpointURL*, *dcat:endpointDescription*, *dcterms:conformsTo*, *dcat:servesDataset* provide automatic discovery of the provided



interfaces to access the datasets.

## 2.2 schema.org Model

schema.org offers an alternative to DCAT for the encoding of repository attributes. While DCAT and schema.org both have advantages and disadvantages (largely due to ease of encoding and availability of suitable vocabulary elements), the decision to develop prototypes and supporting guidance for both was informed by the widespread adoption of these standards in practice.



**Figure 4 - Overview of the standard schema.org model**

**schema:CreativeWork** - The most generic kind of creative work, including books, movies, photographs, software programs, etc. For the purposes of this report, it includes datasets, dataserries, and services.

**schema:Dataset** - A body of structured information describing some topic(s) of interest.

**schema:DataCatalog** - A collection of datasets. In this report, it corresponds to both a 'Repository' and a collection of datasets.

**schema: Project** - An enterprise (potentially individual but typically collaborative), planned to achieve a particular aim. Catalogues and Repositories can also be modelled as 'Projects' and some properties of a project apply to Repositories.

**schema:Offer** - An offer to transfer some rights to an item or to provide a service. In this context, the services and terms of use offered by a Repository.

**schema:Distribution** - A downloadable form of [a] dataset, at a specific location, in a specific format. This property can be repeated if different variations are available. There is no expectation that different downloadable distributions must contain exactly equivalent information (see also DCAT on this point). Different distributions might include or exclude different subsets of the entire dataset, for example.

**schema:WebAPI** - An application programming interface accessible over Web/Internet technologies, which may be used to access the offer(s) made by the Repository.

**schema:Service** - A service provided by an organisation (Repository), which may be used to materialise the offer(s) made by the Repository.

## 2.3 Validation of Assertions

Assertions made by repositories in respect of their characteristics and functions are recorded in metadata. The task developed an information model for description of the links between the metadata exposed by a repository and the process of assessment and validation undertaken by third parties, including validation authorities [19].

An external **validation authority** may assess the available evidence to deliver an evaluation such as the potential trustworthiness of a service by which an end user may answer the question *“Does this service meet my minimum criteria to trust what it provides?”*.

A metadata or data service has **characteristics** (e.g. an address, contact details, a preservation policy, etc.) which can be exposed as metadata and often exists on a website for the service. Again these can be described and exposed as metadata. However, can a user of the service trust the assertions made in the metadata? Are the contact details correct, is there an up-to-date preservation policy, or are the FAIR metrics scores asserted still valid? This kind of information, which is essential for facilitating trust in a service, is currently often lacking or difficult to discern.

The trust in the assertions made may be provided by a third party, a **validation authority**, who provides a **test** and **evidence** for the **asserted value** which may be valid for a fixed period to a **time limit**. For example, CoreTrustSeal<sup>7</sup> [12] could provide transparent evidence that the data service provider has a valid preservation policy as part of its certification of the data service, which is valid for a fixed time. If the service provider exposes FAIR assessment results for its digital objects, it is important to know which tool was used to create the score. Moreover, knowledge of how the score was created and which metrics were used may be essential to some users, for example when comparing research software tools and resources. Therefore, a FAIR assessment tool such as F-UJI [31] could be used by a **validation authority** on the scores it generates and could provide evidence in a transparent fashion with additional information on metrics used, software versioning, and domain-specificity of the metrics.

<sup>7</sup> <https://www.coretrustseal.org/>

This model has synergies and overlaps with models developed by FAIRCORE4EOSC [17] and OSTrails [23] (See Table 2). These models extend the notion of a validation authority and the associated tests by introducing concepts such as principles and criteria, metrics, benchmarks, tests, assessment rubrics, and guidance, and relationships between them. Future work could involve the alignment and consolidation of these models to produce a truly generic model for compliance assessment.

### 3 Consolidated Guidance and Best Practices

This is the second iteration of the guidance provided previously in FAIR IMPACT Milestone 5.2<sup>8</sup> [16]. The items of guidance below have been identified as essential for making information about repositories and digital objects transparent.

#### 1. Multiple outcomes relating to trustworthiness depend on the transparency of relevant information

---

The goals of transparency about repositories and digital objects to improve trustworthiness include clarity on repository activities and functions and the levels of curation and preservation in place. Transparent information is made more machine actionable through alignment with criteria and standards. Such information, shared by repositories as an authoritative source, supports validation and assessment by a variety of methods and tools. Together, these outcomes support clarity, assessment, assistance and the continuous improvement of repositories and the digital objects they hold.

*Appendix A.5: DCAT example, Appendix B.5: [schema.org](https://schema.org) example, Appendix D: repository implementation example, Appendix E: certification authority implementation example.*

#### 2. The functions and characteristics of repositories and objects should be shared

---

Precise descriptions of entities, including organisations, repositories, catalogues, services and objects, are essential to mutual understanding and interoperation. This includes the activities and functions undertaken by repositories and associated data and metadata services. Specific, detailed and accurate characteristics of digital objects are also required.

*Appendix A.2: DCAT Examples, Appendix B.2: [schema.org](https://schema.org) examples, Appendix D: repository implementation example*

#### 3. The levels of retention, curation and preservation provided should be clear

---

In addition to information about retention periods, repositories should be clear about the different levels of curation and preservation they provide across their digital object collections. At the digital object level it should be clear what levels of retention, curation and preservation are in place, and how and when these might change. Supporting information should include appraisal and selection criteria, re-appraisal schedules, preservation plans

---

<sup>8</sup> M5.2 - Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes (1.0). Zenodo. <https://doi.org/10.5281/zenodo.10058634>

etc.

*Appendix A.4: DCAT Example, Appendix B.5 [schema.org](https://schema.org) example, Appendix D: repository implementation example*

#### 4. Information to support transparency and trustworthiness should be expressed in line with defined standards and criteria

---

Transparent information exposed should take account of, and map to, existing standards and criteria. Defining and documenting how the information structure and content relates to existing efforts will minimise divergence and maximise interoperability.

In addition to descriptive information about objects at a generic level (e.g. Dublin Core or DataCite) provision should be made for more specialist characteristics such as those implied by the type of digital object or the target discipline (e.g. DDI for the social sciences). Repository activities and functions should be aligned at the functional level (e.g. CoreTrustSeal). This can be supported by prose information artefacts where necessary (e.g. policies and procedures), and ideally by more granular characteristics expressed as machine-readable metadata (e.g. DRAWG-compliant formats such as RDF based DCAT).

*Appendix H: Checklists and informal recommendations on implementation of references for repositories, supported licences, metadata schema, etc.*

#### 5. The level of support for humans and machines should be specified

---

The design and specification should take account of potential harvesters and consumers of information about repositories (e.g. re3data) or objects (e.g. F-UJI), including discovery services. The outcomes of validations and assessments of repositories and objects (e.g. FAIR assessments) should in turn be provided as standards-compliant (e.g. DQV) structured metadata. Alongside effective readable documentation in clear prose to support human actors this will enable machine agents to select and process the information.

Persistent identification, resolution and associated metadata are essential foundations. Organisations, such as DataCite or ORCID with established organisational structures, proven business model, and track records, enable sustainability and availability far into the future. They provide unique references and deduplication assistance, and in general improve findability and reuse within and between other services beyond a specific technology stack.

Defining how different assertions should be presented requires a balance between flexibility and providing enough guidance to ensure uniformity. For example, the following assertion types could be considered:

- Free-text assertions: statements in response to descriptive criteria about the required information.
- Controlled assertions: selections from ontologies, controlled vocabularies or semantic artifacts.

- Identification: Expose and reference PIDs with (meta)data (e.g. expose the DOI with the DCAT dataset description)
- Evidence artefacts: links to (authoritative) resources containing the asserted information (e.g. link to CoreTrustSeal to prove certification).

Communities, including information consumers, should work together to specify the desired scope of free-text assertions, the ontologies, controlled vocabularies and semantic artifacts used for the controlled assertions, or acceptable links to use for evidence artefacts.

*Appendix H: Checklists and informal recommendations on implementation of references for repositories, supported licences, metadata schema, etc.*

## 6. Information shared should be associated with an authoritative source

---

Registration and aggregation of information, alongside essential services like identification through the assignment of PIDs depend on a range of actors. Despite the clear benefits, much of the relevant information that needs to be available in a transparent way is available through multiple third party service providers. There are inevitable challenges in aligning accurate and timely information by navigating multiple services.

Ideally the source for information about repositories and the objects they hold should be the relevant organisation itself. This places the responsibility for maintaining information as accurate and up to date with the most authoritative source, permitting multiple other organisations to consume it for a variety of uses.

*Appendix H: Informal recommendations on referencing attributes with the origin, expressed as a PID unambiguously identifying the repository.*

## 7. Any potential methods for validating information should be documented

---

A subset of the information shared may be validated by third parties for accuracy or compared with standard criteria to assess conformance. Clear specification and documentation of such actions ensure uniformity in the interactions with human agents and/or machine agents. The following scenarios for validation may be considered:

- Acceptance of assertion: assertion is accepted without further validation.
- Direct machine-actionable validation: given that the assertion is machine-testable, the information presented is automatically validated in an established process.
- Machine-actionable validation through a third party: given a third party can be pointed to as the authority on the information asserted, they are called on to validate the information automatically through an established process.
- Validation through human action: given the assertion is not machine-testable, the consumer must take manual steps to validate the information presented.

- Validation through a mixture of human and machine action: given that the assertion is machine-testable, the choice may be made to also validate the information manually to ensure the content or quality of the supplied information (i.e. the machine tests the information is available, the human validates the content).

*Appendix H.6 provides informal recommendations on these aspects.*

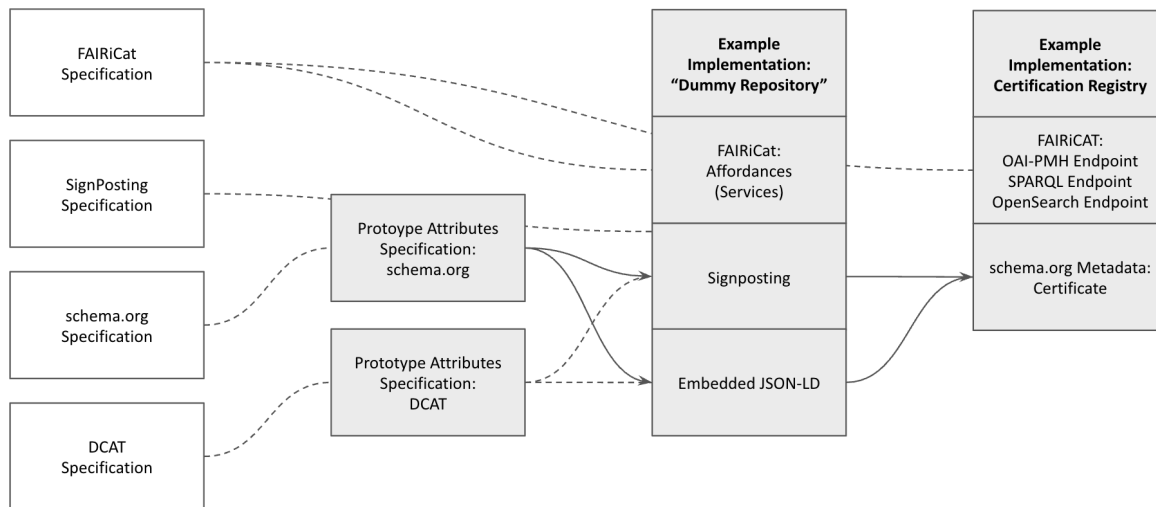
#### 8. Methods and tools for validation and assessment should provide support and guidance

Multiple methodologies, including associated automated, partially automated and self-assessment tools exist to support the validation of assertions or the assessment of compliance with criteria by repositories and objects. These include multiple FAIR evaluation tools whose implementations can vary. Ideally it would be possible to use multiple evaluation tools while being confident that they apply a common approach to validation. Tools should perform consistently against a range of benchmarking information sources, such as datasets and their metadata records. In addition to machine-readable outputs, tools should invest in human-friendly documentation to support assessment results, including the provision of support and guidance to increase the understanding of the validations and assessments and how future results can be improved.

The aim of the guidelines is to increase the transparency of information about repositories and other (meta)data services and the digital objects they hold, or link to, by exposing it in ways that enable humans and machines to harvest, exchange, and use the information. This provides an essential foundation for interoperability, demonstrating trustworthiness and allowing the (meta)data service to improve its presence in the wider environment and the European Open Science Cloud (EOSC) [19].

## **4 Prototype Development**

The task developed a portfolio of prototype implementations to support the guidelines and to facilitate testing and validation by the community. Figure 5 shows the scope of and relationship between these prototypes.



**Figure 5 - Scope of Prototype Development**

Four distinct prototype elements were developed:

- A prototype specification was developed for encoding repository attributes identified in the DRAWG [11] using the schema.org specification ([Appendix B](#)).
- Likewise, a prototype specification was developed for encoding repository attributes using the DCAT specification ([Appendix A](#)).
- An example implementation was developed to illustrate how repositories might implement the guidelines and prototype specifications ("Dummy Repository", described in [Appendix D](#)). The example has three distinct elements:
  - It shows an example of embedding repository attributes in the head section of an HTML page, based on the schema.org prototype specification.
  - It illustrates how the FAIRiCat specification may be used to point to a machine-actionable service inventory for the repository ('affordances'). This could be used to identify, for example, an OAI-PMH harvesting endpoint for the repository unambiguously ([Appendix C](#)).
  - Using the same specification, it indicates how Signposting may be used to point to a file containing the repository attributes, encoded in the schema.org prototype specification ([Appendix C](#)).
- Both the embedded JSON-LD and Signposting examples may also use the DCAT prototype specification (not implemented in the example repository).
- The JSON-LD and Signposting implementations both show examples of how to point to a certificate metadata record in the registry of a validation authority (in this case, a

prototype implementation based on Dataverse, showing CoreTrustSeal certificates, see [Appendix E](#)).

These prototypes are intended as examples for repositories to test the guidelines and specifications.

## 5 Conclusions and Recommendations

The following conclusions can be offered based on the work and experiences gained in T5.4:

- It is feasible to develop and publish specifications for exposure of repository attributes, based on a set of well-established and emerging standards that will aid in the machine-actionability, currency<sup>9</sup>, and validation of such attributes. The end users include the research ecosystem in general, and specifically registries of repositories, assessment and validation services and authorities, and depositors (individuals or organizations) that are looking for an appropriate repository for their outputs.
- The importance and utility of machine-readable metadata was confirmed through the support action, and illustrated that small, practical steps can significantly enhance metadata transparency.
- Detailed description of and clarification of the encoding options and best practices will be required for many attributes, for example unambiguously referencing the supported licences, PIDs, or metadata schema applicable to a particular repository or its collections. In some cases, it will be necessary to identify and recommend specific vocabularies or registries/ URIs to use for unambiguous references.
- The process whereby attributes at a given level of granularity are either aggregated upwards (for example computing a FAIR assessment for a repository from individual object scores), or inherited downwards (for example applying a repository-level attribute to all objects in a repository or a collection) needs to be further investigated, possibly with the addition of test, metric and benchmark definitions.
- The role of data stewards in maintaining and curating repository attributes was identified in support action feedback, as well as the fact that while implementation of the guidelines and specifications is not necessarily costly or complex, it requires the involvement of multiple skills in an organisation and a presumption of some knowledge of the applicable standards.
- Decisions on what metadata to include and how to structure the distribution of attributes in more complex cases (multiple collections, repositories, etc.) are not currently clear enough to prospective implementers. Additional formal documentation will assist in this decision-making process.
- A number of specific recommendations for future work and improvements were

<sup>9</sup> Currency is improved because attributes can be 'pulled' from source by machines. Updates are made in one location (locally) and harvested by registries, validation authorities, etc.



identified in addition to the implicit future improvements listed above. These are all described in more detail in Appendix G.

## References

- [1] FAIR-IMPACT (2025), WP5 - Metrics, certification and guidelines, <https://fair-impact.eu/wp5-metrics-certification-and-guidelines>
- [2] COAR (2025). A Global Repository Network, <https://coar-repositories.org/members/>
- [3] Strecker, D., Axtmann, A., Bertelmann, R., Cousijn, H., Elger, K., Ferguson, L. M., Fichtmüller, D., Jones, C., Lindenmann, I., Neidiger, C. Nguyen, T. B., Pal, J. K., Pampel, H., Petras, V., Schnepf, E., Semrau, A., Ulrich, R., Upmeier, A., Vierkant, P., Wang, H., Weickert, G., Weisweiler, N. L., Williams, S. C., Witt, M & Wright, S. J. (2023). Metadata Schema for the Description of Research Data Repositories : version 4.0. <https://doi.org/10.48440/re3.014>
- [4] re3data (2025), re3data API Documentation, <https://www.re3data.org/api/doc/>
- [5] Milo Thurston, Allyson Lister, Ramon Granell, Dominique Batista, Peter McQuilton, Philippe Rocca-Serra, Susanna-Assunta Sansone. (2022). FAIRsharing Application: Schema (1.0). Zenodo. <https://doi.org/10.5281/zenodo.6875036>
- [6] FAIRsharing, 2025: FAIRSharing API Documentation, [https://fairsharing.org/API\\_doc](https://fairsharing.org/API_doc)
- [7] EOSC (2025), Resource Hub, <https://open-science-cloud.ec.europa.eu/resources/datasources>
- [8] EOSC (2025), Metadata Entry for PANGAEA, <https://open-science-cloud.ec.europa.eu/resources/datasources/21.11166%2FThQswj>
- [9] Herbert Van de Sompel, Robert Huber, Patrick Hochstenbach, Michael L. Nelson, Martin Klein, Andrea Bollini, Enno Meijers, Petr Knoch, Paul Walk, Wim Hugo, Jim Meyers (2024), FAIRiCat: Supporting Discovery of a Repository's Interoperability Affordances, <https://signposting.org/FAIRiCat/>
- [10] van de Sompel, H. and Wilde, E. (2022), RFC 9264 Linkset: Media Types and a Link Relation Type for Link Sets <https://www.rfc-editor.org/rfc/rfc9264.html>
- [11] Witt, M., Cannon, M., Lister, A., Segundo, W., Shearer, K., Yamaji, K., & Research Data Alliance Data Repository Attributes Working Group. (2024). RDA Common Descriptive Attributes of Research Data Repositories (1.0). Zenodo. <https://doi.org/10.15497/RDA00103>
- [12] CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. <https://doi.org/10.5281/zenodo.7051012>
- [13] "nestor criteria : Catalogue of Criteria for Trusted Digital Repositories, Version 2, 2009, Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek, urn:nbn:de:0008-2010030806 <http://nbn-resolving.de/urn:nbn:de:0008-2010030806>, <https://d-nb.info/1189191830/34>
- [14] International Organisation for Standardization (2025). Audit and certification of trustworthy digital repositories. <https://www.iso.org/standard/87472.html>
- [15] TRSP WH (2025), Work In Progress, [https://drive.google.com/drive/folders/193ZpTeZRR6-VPAtSj-Gm8\\_ROuhuWUsgP?usp=drive\\_link](https://drive.google.com/drive/folders/193ZpTeZRR6-VPAtSj-Gm8_ROuhuWUsgP?usp=drive_link)
- [16] 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 (1.0). Zenodo. <https://doi.org/10.5281/zenodo.10058634>
- [17] FAIRCORE4EOSC (2025), The Compliance Assessment Toolkit, <https://cat.argo.grnet.gr/>
- [18] Ulrich, R., Verburg, M., Priddy, M., Huber, R., L'Hours, H., Neidiger, C., Meijas, G., & Davidson, J. (2024). M5.5 - Initial repository registry support for discovery of repositories, policies, and interfaces (1.0). Zenodo. <https://doi.org/10.5281/zenodo.10847707>
- [19] Maaïke Verburg, Hervé L'Hours, Robert Huber, Robert Ulrich, Mike Priddy, Joy Davidson, & Alejandra Gonzalez-Beltran. (2023). Introduction to the guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes (0.5). Zenodo. <https://doi.org/10.5281/zenodo.8224360>
- [20] 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 (1.0 DRAFT). FAIRsFAIR. <https://doi.org/10.5281/zenodo.3631528>
- [21] Exchanges and clarifications shared with participants during the FAIR-IMPACT support action.
- [22] Andreu, T., Anglada, L., Antos, D., Bähr, T., Brzeźniak, M., Burgi, P.-Y., Cavet, C., Celjak, D., Crépé-Renaudin, S., De Loof, C., Dillo, I., Dubois, O., Fernandes, R., Forsström, P.-L., Ganis, G., Gibney, E., Holl, A., L'Hours, H., Lamers, D., ... Wyns, R. (2023). EOSC Preservation: Overview Discussion Paper. Zenodo. <https://doi.org/10.5281/zenodo.7516259>
- [23] OSTrails (2025), OSTrails FAIR Assessment Output Specification, [https://github.com/OSTrails/FAIR\\_assessment\\_output\\_specification](https://github.com/OSTrails/FAIR_assessment_output_specification)
- [24] FAIR Data Maturity Model Working Group. (2020). FAIR Data Maturity Model. Specification and Guidelines (1.0). Zenodo. <https://doi.org/10.15497/rda00050>
- [25] FAIR-IMPACT T5.4 (2024), 3rd support call: Testing the trustworthy and FAIR-enabling repositories prototype, [https://docs.google.com/presentation/d/1IS\\_inpXp5IsGaGDYigv7SeYOYcCNLkfMLU0RuVmIclw/e/dit?usp=sharing](https://docs.google.com/presentation/d/1IS_inpXp5IsGaGDYigv7SeYOYcCNLkfMLU0RuVmIclw/e/dit?usp=sharing)
- [26] Huber, R. (2025), DCAT Demonstrator Source Code, [https://github.com/FAIRsFAIR/5\\_4\\_prototype/blob/main/guidelines/DCAT.md](https://github.com/FAIRsFAIR/5_4_prototype/blob/main/guidelines/DCAT.md)
- [27] Hugo, W. (2024). Case Study Analysis: FAIR-AWARE (0.4 Draft). Zenodo. <https://doi.org/10.5281/zenodo.13486326>
- [28] Huber, R. (2025), schema.org Demonstrator Source Code, [https://github.com/FAIRsFAIR/5\\_4\\_prototype/blob/main/guidelines/SCHEMAORG.md](https://github.com/FAIRsFAIR/5_4_prototype/blob/main/guidelines/SCHEMAORG.md)
- [29] Huber, R., Ulrich, R., L'Hours, H., Hugo, W., Neidiger, C., Parkes, O., Chue Hong, N., Meijas, G., & Dillo, I. (2024). M5.8 - Pilot for Exposing Repository Trustworthiness Status and FAIR Data and Code Assessment Outcomes in Generic and Disciplinary Registries and Portals (1.0). Zenodo. <https://doi.org/10.5281/zenodo.13861127>
- [29] Herbert Van de Sompel, Martin Klein, Shawn Jones, Michael L. Nelson, Simeon Warner, Anusuriya Devaraju, Robert Huber, Wilko Steinhoff, Vyacheslav Tykhonov, Luc Boruta, Enno Meijers, Stian Soiland-Reyes, Mark Wilkinson (2023), FAIR Signposting Profile,

<https://signposting.org/FAIR/>

- [30] EDEN-FIDELIS (2025), Repositories and Archives, <https://eden-fidelis.eu/repositories-archives>
- [31] Anusuriya Devaraju, & Robert Huber. (2020). F-UJI - An Automated FAIR Data Assessment Tool (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.4063720>
- [32] Wilkinson, M. D., Sansone, S.-A., Grootveld Marjan, Nordling, J., Dennis, R., & Hecker, D. (2022). FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons. Zenodo. <https://doi.org/10.5281/zenodo.7463421>
- [33] Gregory, A., Bell, D., Brickley, D., Buttigieg, P. L., Cox, S., Edwards, M., Doug, F., Gonzalez Morales, L. G., Heus, P., Hodson, S., Kanjala, C., Le Franc, Y., Maxwell, L., Molloy, L., Richard, S., Rizzolo, F., Winstanley, P., Wyborn, L., & Burton, A. (2024). WorldFAIR (D2.3) Cross-Domain Interoperability Framework (CDIF) (Report Synthesising Recommendations for Disciplines and Cross-Disciplinary Research Areas) (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.11236871>

## Appendices

### Appendix A: Prototype for DCAT Encodings of Repository Attributes

This section is taken from material maintained in GitHub [26].

#### A.1 Exposing Repository Information with DCAT

We follow the definition of data repository given by DCAT-AP<sup>10</sup> “a catalogue or repository that hosts the Datasets or Data Services being described.” and therefore model a repository as an instance of dcat:Catalog. To indicate the operational and organisational nature of a data repository we recommend to additionally use foaf:Project and/or foaf:Organisation for each instance of a data repository.

#### A.2 Descriptive Metadata

To provide basic descriptive metadata corresponding to DRAWG properties the use of DCAT properties as well as Dublin Core terms<sup>11</sup> is recommended. In addition, friend of a friend (foaf) properties<sup>12</sup> can be used in case catalogue instances are modelled as foaf:Project or foaf:Organization.

Most of these properties will be provided as Literals. However, we follow the recommendation of DCAT-AP to use instances of foaf:Agent (here foaf:Organization) for dct:publisher values. In addition, vcard:Organization<sup>13</sup> should be used as a type for dct:publisher. This allows it to include address information such as country name, which is required to provide DRAWG’s ‘Country’ property info.

**Table A.1 - Mapping of selected DRAWG attributes to recommended DCAT properties.**

DRAWG	DCAT
Repository Name	dct:title (foaf:name)
URL	dct:identifier (foaf:homepage)
Description	dct:description
Language	dct:language
Research Area	dcat:theme or dct:subject
Organisation	dct:publisher (a foaf:Organization, vcard:Organization)
Country	dct:publisher (a vcard:Organization) => vcard:country-name

<sup>10</sup><https://semiceu.github.io/DCAT-AP/releases/3.0.0/#Catalogue>

<sup>11</sup><https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

<sup>12</sup><http://xmlns.com/foaf/spec/>

<sup>13</sup><https://www.w3.org/TR/vcard-rdf/>

DRAWG	DCAT
Dataset Use License	dct:license
Terms of Access	dct:accessRights
Contact	dcat:contactPoint

### A.3 Supported Standards

To expose information about standards offered to support machine interoperability, we recommend using `dcat:service` which allows us to provide a list of instances of `dcat:DataService`. There, the properties `dcat:endpointURL` and `dct:conformsTo` shall be used to provide information about the service endpoint URL as well as the service type which should be the web link to the documentation of the standard the web API follows (see the Appendix and the living document<sup>14</sup>).

```

dcat:service: [
  - {
    @type: "dcat:DataService",
    dcat:endpointURL: "https://dummyrepository.org/.well-known/api-catalog",
    dct:conformsTo: "https://signposting.org/FAIRiCat/"
  },

```

Figure A.1 - Example of `dcat:service` usage

Supported standards which do not represent actionable services can be expressed using the `dct:conformsTo` property which shall point to an instance of `dct:Standard`. The property `rdfs:seeAlso` may be used to additionally indicate which standard type is listed. Here we recommend to use the FAIR Implementation Profile (FIP) vocabulary.<sup>15</sup>

To unambiguously identify a metadata standard we recommend to use a linked term from for example the DCC metadata list<sup>16</sup> or from FAIRsharing.<sup>17</sup> In case a XML standard is used the namespace identifier can also be used.

To indicate supported persistent identifier (PID) types, we recommend using the home URI of a PID system (e.g. DOI,<sup>18</sup> Handle,<sup>19</sup> etc.), which uniquely identifies a PID system.

<sup>14</sup><https://docs.google.com/spreadsheets/d/1mfVK5kEqTCi67R-6bmDi4-qfvi2pPHdLFnL-hqvZqEE/edit?gid=0#gid=0>

<sup>15</sup><https://w3id.org/fair/fip/>

<sup>16</sup><http://www.dcc.ac.uk/resources/metadata-standards>

<sup>17</sup><https://fairsharing.org>

<sup>18</sup><https://doi.org>

<sup>19</sup><https://handle.net>

```
dct:conformsTo: [
+ { ... },
+ { ... },
- {
  @type: "dct:Standard",
  @id: "http://www.dcc.ac.uk/resources/metadata-standards/dcat-data-catalog-vocabulary",
  rdfs:seeAlso: "https://w3id.org/fair/fip/latest/Metadata-schema"
},
- {
  @type: "dct:Standard",
  @id: "https://w3id.org",
  rdfs:seeAlso: "https://w3id.org/fair/fip/latest/Identifier-service"
}
],
```

**Figure A.2 - Metadata standards and identifier types**

## A.4 Policies and Principles

Similar to the way we recommend to indicate standards, the `dct:conformsTo` property should be used to link an instance of `dct:Policy` or subclasses of `dct:Policy`. Here we recommend to use `premis:PreservationPolicy`<sup>20</sup> to indicate the preservation and/or curation policy of a data repository.

**Table A.2 - Mapping of `dct:conformsTo` to other properties for different DRAWG attributes.**

DRAWG	DCAT
Curation	<code>dct:conformsTo =&gt; dct:Policy</code> or <code>premis:PreservationPolicy</code>
Terms of Deposit	<code>dct:conformsTo =&gt; dct:Policy</code>
Preservation	<code>dct:conformsTo =&gt; premis:PreservationPolicy</code>

```
- dct:conformsTo: [
- {
  @type: "dct:Policy",
  @id: "https://dummyrepository.org/policies/termsofdeposit.html"
},
- {
  @type: "premis:PreservationPolicy",
  @id: "https://dummyrepository.org/policies/preservationpolicy.html",
  rdfs:seeAlso: "https://w3id.org/fair/fip/latest/Metadata-preservation-policy"
},
]
```

**Figure A.3 - Example of policy exposure**

## A.5 Certification and Quality Information

The Data Quality Vocabulary (DQV) is part of the W3C data on the web best practices family of standards and recommended to be used within DCAT to indicate quality assessments and

<sup>20</sup>[https://id.loc.gov/ontologies/premis-3-0-0.html#c\\_PreservationPolicy](https://id.loc.gov/ontologies/premis-3-0-0.html#c_PreservationPolicy)

certificates. Therefore, we recommend to use the `dqv:hasQualityAnnotation` to indicate an instance of `dqv:QualityCertificate`, which links e.g. a `CoreTrustSeal` certificate of a data repository. The issuer of such a certificate can be indicated using the `dc:creator` property. The certificate itself has to be contained via a `oa:hasBody`<sup>21</sup> property since a `dqv:QualityCertificate` is a subclass of `oa:Annotation`. There, we recommend to use the DOI of a `CoreTrustSeal` certificate which is not shown in the example below (Figure A.5).

**Table A.3 - Mapping of DRAWG attribute to properties.**

DRAWG	DCAT
Certification	<code>dqv:hasQualityAnnotation =&gt; dqv:QualityCertificate</code>

```

- dqv:hasQualityAnnotation: {
  @type: "dqv:QualityCertificate",
  dct:creator: "CoreTrustSeal",
  oa:hasTarget: "https://dummyrepository.org",
  oa:hasBody: "https://amt.coretrustseal.org/certificates",
  oa:motivatedBy: "dqv:qualityAssessment"
},
  
```

**Figure A.4 - Example of certificate exposure**

## Appendix B: Prototype for schema.org Encoding of Repository Attributes

This section is taken from material maintained in GitHub [28].

### B.1 Exposing Repository Information with schema.org

Unlike ESIP's model,<sup>22</sup> which defines a data repository as an instance of `schema:Organization`, `schema:ResearchProject`, and `schema:Service`, we model a data repository (or at least the operational part of a data repository) simply as an instance of `schema:Project` and `schema:DataCatalog`. We chose `schema:Project` instead of `schema:ResearchProject` because a data repository usually does not perform research as the main purpose. Since every data repository should also have a catalogue of its datasets, we propose to use `schema:DataCatalog` in addition to `schema:Project`. If a data repository is an independent organisational or legal entity, `schema:Organization` can optionally be used as an additional type or in replacement of `schema:Project` to model a data repository. These schema.org types allow us to describe all essential DRAWG properties which we propose to map to schema.org as follows.

### B.2 Descriptive Metadata

<sup>21</sup><https://www.w3.org/community/openannotation/>

<sup>22</sup><https://github.com/ESIPFed/science-on-schema.org/blob/main/guides/DataRepository.md>



Since schema:DataCatalog is a subtype of schema:CreativeWork some descriptive properties are available which easily can be mapped to DRAWG.

**Table B.1 - Mapping of DRAWG attributes to schema.org properties.**

DRAWG	schema.org
Repository Name	schema:name
URL	schema:url
Description	schema:description
Language	schema:inLanguage
Research Area	schema:keywords
Organization	schema:publisher
Country	schema:publisher >= schema:address => schema:addressCountry
Dataset Use License	schema:license
Terms of Access	schema:conditionsOfAccess
Contact	schema:contactPoint

For the schema.org keywords property, it is possible to use the type schema:DefinedTerm which allows to use ontology terms unambiguously specifying research areas. Since the DRAWG property 'Country' is defined as *"The country in which the repository operates"*, we map this DRAWG property to the country information (schema:addressCountry) contained in a schema:publisher's schema:address property.

To indicate the contact information for a given repository we recommend to use the schema:contactPoint property which is part of schema:Project and allows to include detailed schema:ContactPoint properties such as phone, email, fax etc.

### B.3 Supported Standards

To expose information of available APIs supporting machine interoperability of a data repository, we propose to use the schema:offers property, which may list several instances of schema:Offer, which then links to instances of schema:WebAPI (a subclass of schema:Service) via their schema:itemOffered property. There, we follow the example of FAIRiCAT<sup>23</sup> and use the schema:documentation to describe the type of service. This should be the web link to the documentation of the standard the web API follows. The schema:url can be used to describe the endpoint URI of the API.

<sup>23</sup><https://signposting.org/FAIRiCat>

```
{
  @type: "schema:Offer",
  - schema:itemOffered: {
    @type: "schema:WebAPI",
    schema:url: "https://dummyrepository.org/services/static_oai.xml",
    schema:documentation: "https://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm"
  }
},
```

**Figure B.1 - Example of exposure of available APIs**

Similarly, other standards supported by a data repository can be described using schema.org as an schema:Offer, which then should be a schema:Service instead of a schema:WebAPI. We propose to use the property schema:serviceType to unambiguously indicate which DRAWG service category (persistent identifier or metadata standard) is described. We recommend using the FAIR vocabulary terms Identifier Service<sup>24</sup> and Metadata Schema<sup>25</sup> respectively to do so.

```
- {
  @type: "schema:Offer",
  - schema:itemOffered: {
    @type: "schema:Service",
    - schema:serviceOutput: {
      schema:identifier: "https://www.w3id.org"
    },
    schema:serviceType: "https://w3id.org/fair/fip/latest/Identifier-service"
  }
},
```

**Figure B.2 - Example of standards exposure**

We recommend using the schema:serviceOutput to indicate the PID or metadata standard a data repository supports.

Similarly as mentioned in the previous section, to indicate supported persistent identifier (PID) types, we recommend using the home URI of a PID system (e.g. DOI,<sup>26</sup> Handle,<sup>27</sup> etc.), which uniquely identifies a PID system.

For metadata standards, we recommend unique identifiers such as a FAIRsharing identifier (DOI), a DCC identifier, or the unique namespace or schema URI of e.g. XML metadata standards.

**Table B.2 - Mapping of DRAWG attributes to schema.org properties.**

<sup>24</sup><https://w3id.org/fair/fip/latest/Identifier-service>

<sup>25</sup><https://w3id.org/fair/fip/latest/Metadata-schema>

<sup>26</sup><https://doi.org>

<sup>27</sup><https://handle.net>

DRAWG	schema.org
Machine Interoperability	Schema:offers => schema:Offer => itemOffered => schema:WebAPI
Persistent Identifiers	Schema:offers => schema:Offer => itemOffered => schema:Service
Metadata	Schema:offers => schema:Offer => itemOffered => schema:Service

### B.4 Policies and Principles

Unfortunately, schema.org does not offer a generic way to describe policies such as dc:Policy via dc:conformsTo. However, schema:DataCatalog inherited the schema:publishingPrinciples from schema:CreativeWork which may serve to point to *“a document describing the editorial principles”*, which therefore is well suited to link to DRAWG terms of deposit etc.. We include the DRAWG ‘Curation’ property here because it may be explained in a dedicated data curation policy document.

Table B.3 - Mapping of schema:publishingPrinciples to several DRAWG attributes.

DRAWG	schema.org
Curation	schema:publishingPrinciples
Terms of Deposit	schema:publishingPrinciples
Preservation	schema:publishingPrinciples

To clarify which policy actually is described, we recommend to use the schema:additionalType property and here to use the values premis:PreservationPolicy to indicate the preservation policy.

```
- schema:publishingPrinciples: [
  - {
    @type: "schema:CreativeWork",
    schema:url: "https://dummyrepository.org/policies/termsofdeposit.html",
    schema:additionalType: "dct:Policy"
  },
  - {
    @type: "schema:CreativeWork",
    schema:url: "https://dummyrepository.org/policies/preservationpolicy.html",
    schema:additionalType: "premis:PreservationPolicy"
  }
],
```

Figure B.3 - Example of exposure of policy

### B.5 Certification and Quality Information

As mentioned above, we chose schema:Project to model a data repository which allows us to include additional DRAWG properties. From these, we recommend to use the schema:hasCertification property to indicate certification details such as a given CoreTrustSeal certification. This property requires the use of a schema:Certification instance which allows linking to a certification document via the schema:url property which should be the DOI of a CoreTrustSeal certificate. It further allows to include useful information about certificates such as audit date, validity and issuer.

```

schema:hasCertification: {
  @type: "schema:Certification",
  schema:url: "https://amt.coretrustseal.org/certificates",
  schema:certificationStatus: "schema:CertificationActive",
  - schema:issuedBy: {
    @type: "schema:Organization",
    schema:name: "CoreTrustSeal",
    schema:url: "https://www.coretrustseal.org"
  },
  schema:auditDate: "2024-09-09",
  schema:expires: "2024-12-31"
},

```

**Figure B.4 - Example of exposure of certification information<sup>28</sup>.** Please note that instead of the generic amt.coretrustseal.org URL, a valid reference to a certificate archived at the [CoreTrustSeal certification repository](#) should be used. At present, this can be tested via a [Prototype](#).

**Table B.4 - Mapping of DRAWG certification attributes to schema.org properties.**

DRAWG	schema.org
Certification	schema:hasCertification
Contact	schema:contactPoint

In addition, the use of schema:Project this would allow information about funding sources and other useful things which are relevant for the scientific community.

## Appendix C: Prototype FAIRiCAT/ Signposting Implementation

<sup>28</sup> Please note that instead of the prototype URL shown here, a valid reference to a certificate archived at the CoreTrustSeal certification repository (<https://dataverse.nl/dataverse/coretrustseal>) should be used in production implementations. The recommendations in this prototype must still be implemented by CoreTrustSeal.

For machines to find repository-related information reliably, standardisation is required, and the Signposting/ FAIRiCat specifications [9], [29] offer simple options for supporting this requirement. FAIRiCat is an emerging specification for encoding machine-actionable information about services and features offered by a repository ('affordances') in a standardised way.

The specifications rely on inclusion of either typed links in the HTTP Link headers and/or HTML <link> elements of the repository landing page. Alternatively, a pointer to a LinkSet, describing one or more of the typed links can also be implemented.

The prototype implementation in a 'dummy' repository, described in Appendix D, illustrates the second approach, by including link elements in the HTML header of the landing page. Figure C.1 shows this implementation, with two link elements of interest:

1. A link of type "describedby", referencing a JSON-LD file containing repository attributes, and
2. A link of type 'api-catalog', referencing a LinkSet describing repository affordances (services).

```
<title>A Dummy Data Catalog</title>
<link rel="describedby" type="application/ld+json" href="https://dummyrepository.org/metadata/catalog_metadata_schemaorg.json"/>
<link rel="api-catalog" type="application/linkset+json" href="https://dummyrepository.org/.well-known/api-catalog"/>
```

**Figure C.1 - Typed Link Elements in HTML <head>**

The content of the JSON-LD file ([schema.org example](https://dummyrepository.org/metadata/catalog_metadata_schemaorg.json)<sup>29</sup>) is shown in part in Figure C.2.

```
{
  "@context": {
    "dc": "http://purl.org/dc/terms/",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "schema": "http://schema.org/",
    "vcard": "http://www.w3.org/2006/vcard/ns#",
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "premis": "http://www.loc.gov/premis/rdf/v3/"
  },
  "@type": [
    "schema:DataCatalog",
    "schema:Project"
  ],
  "@id": "https://dummyrepository.org",
  "schema:url": "https://dummyrepository.org",
  "schema:name": "Dummy Data Repository",
  "schema:description": "Dummy Data Repository Description",
  "schema:contactPoint": "https://dummyrepository.org/contact",
  "schema:keywords": [
    "generic"
  ]
}
```

**Figure C.2 - Example of a JSON-LD file containing schema.org encoded attributes**

<sup>29</sup> [https://dummyrepository.org/metadata/catalog\\_metadata\\_schemaorg.json](https://dummyrepository.org/metadata/catalog_metadata_schemaorg.json)

The services offered by the repository can be found by navigating to the [LinkSet](#),<sup>30</sup> as illustrated partially in Figure C3. In this file, a number of services, including OAI-PMH and RSS endpoints are described.

```
{
  "linkset": [
    {
      "anchor": "https://dummyrepository.org/services/static_oai.xml",
      "service-doc": [
        {
          "href": "https://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm",
          "type": "text/html",
          "title": "Specification for an OAI Static Repository and an OAI Static Repository Gateway"
        }
      ],
      "service-meta": [
        {
          "href": "https://dummyrepository.org/services/static_oai.xml",
          "type": "application/xml"
        }
      ]
    },
    {
      "anchor": "https://dummyrepository.org/services/atom.xml",
      "service-doc": [
        {
          "href": "https://www.rssboard.org/rss-specification",
          "type": "text/html",
          "title": "RSS 2.0 Specification (Current)"
        }
      ]
    }
  ],
  {
    "anchor": "https://dummyrepository.org/sitemap.xml"
  }
}
```

**Figure C.3 - Example of a LinkSet file containing repository affordances**

The implementation also contains an example of embedded JSON-LD, containing DCAT-encoded repository attributes as an alternative to FAIRiCAT/ Signposting actionability (Figure C.4). The FAIRiCAT implementation, though, is more robust should a LinkSet be referenced from more than one landing page, since the information only has to be maintained in one location - the LinkSet/ JSON-LD files.

```
<script type="application/ld+json">
{
  "@context": {
    "dcat": "http://www.w3.org/ns/dcat#",
    "dct": "http://purl.org/dc/terms/",
    "foaf": "http://xmlns.com/foaf/0.1/",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "vcard": "http://www.w3.org/2006/vcard/ns#",
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "dq": "http://www.w3.org/ns/dq#",
    "oa": "http://www.w3.org/ns/oa#",
    "premis": "http://www.loc.gov/premis/rdf/v3/"
  },
  "@type": [
    "dcat:Catalog",
    "foaf:Project"
  ],
  "id": "https://dummyrepository.org"
}
```

**Figure C.4 - Example of Embedded JSON-LD based on DCAT Encoding**

Creating a simple utility application - a user interface and the ability to consume, edit, and export both FAIRiCat Linkset documents and DCAT/ schema.org documents - will ensure that Linksets and DCAT/ schema.org documents are interchangeable, quality assured, and

<sup>30</sup> <https://dummyrepository.org/well-known/api-catalog>

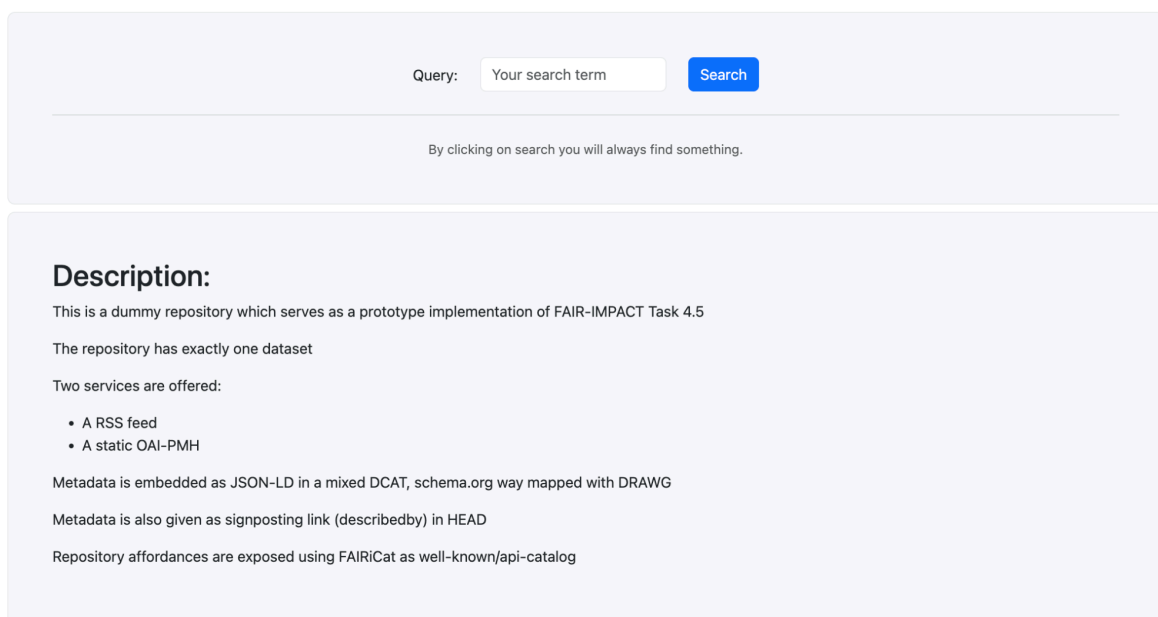
available for either implementation option. This could be considered for future development.

## Appendix D: Repository Attribute Prototype

A [prototype repository](https://dummyrepository.org/)<sup>31</sup> was developed to illustrate the implementation of the recommendations described in Appendices A, B, and C. The repository (Figure D.1) can be used to test these concepts without disturbing the operational systems of our use case partners. This prototype repository consists of a simple web based repository homepage as well as an integrated catalogue which contains a single test dataset. The repository provides DCAT as well as schema.org catalogue metadata compliant to the DRAWG specifications explained in Appendices A and B. The prototype repository webpage is publicly available and the source code of the prototype repository is available in Github.<sup>32</sup> It showcases implementations of the following:

- A mixture of DCAT and schema.org encoded attributes in a JSON-LD file.
- A dataset record with metadata encoded based on Signposting [29].
- A FAIRiCAT implementation [9] that exposes the services offered by the repository (RSS, OAI-PMH) as well as a pointer to the JSON-LD file.

## Dummy Data Repository



The screenshot shows the homepage of the Dummy Data Repository. At the top, there is a search bar with the placeholder text "Query: Your search term" and a blue "Search" button. Below the search bar, a message states: "By clicking on search you will always find something." The main content area is titled "Description:" and contains the following text:

This is a dummy repository which serves as a prototype implementation of FAIR-IMPACT Task 4.5

The repository has exactly one dataset

Two services are offered:

- A RSS feed
- A static OAI-PMH

Metadata is embedded as JSON-LD in a mixed DCAT, schema.org way mapped with DRAWG

Metadata is also given as signposting link (describedby) in HEAD

Repository affordances are exposed using FAIRiCat as well-known/api-catalog

**Figure D.1 - Dummy Repository Home Page**

The same application also provides a simple checker to test the extent to which any repository (including the dummy repository) implements the scope of DRAWG attributes using the guidelines in Appendices A and B, and whether these are automatically discoverable via FAIRiCAT and Signposting.

<sup>31</sup> <https://dummyrepository.org/>

<sup>32</sup> [https://github.com/FAIRsFAIR/5\\_4\\_prototype](https://github.com/FAIRsFAIR/5_4_prototype)



Figure D.2 shows the result for the dummy repository, indicating the scope of implementation and of attributes that could be successfully retrieved based on the guideline specifications.

<b>Repository Name</b> "Dummy Data Repository"	✓
<b>URL</b> "https://dummyrepository.org"	✓
<b>Description</b> "Dummy Data Repository Description"	✓
<b>Country</b> ["Germany"]	✓
<b>Language</b> "eng"	✓
<b>Organization</b> ["Dummy Publisher"]	✓
<b>Contact</b> "https://dummyrepository.org/contact"	✓
<b>Research Area</b> ["generic"]	✓
<b>Persistent Identifier</b> "https://w3id.org"	✓
<b>Machine Interoperability</b> "https://dummyrepository.org/sitemap.xml"	✓
<b>Metadata Standards</b> "http://www.dcc.ac.uk/resources/metadata-standards/dcat-data-catalog-vocabulary"	✓
<b>Curation / Preservation Policy</b> "https://dummyrepository.org/policies/preservationpolicy.html"	✓
<b>Terms of Deposit</b> "https://dummyrepository.org/policies/termsofdeposit.html"	✓
<b>Terms of Access</b> "open access"	✓
<b>License</b> "https://creativecommons.org/licenses/by/4.0/"	✓
<b>Certification</b> { "url": "https://amt.coretrustseal.org/certificates", "issuer": "CoreTrustSeal" }	✓

**Figure D.2 - Example Result from the Repository Checker Application**

## Appendix E: Repository Certification Evidence

In Appendices A and B, two options are described for the exposure of the typical attributes defined by DRAWG [11], including options for encoding certification status and evidence.

Doing so requires two complementary implementations to be available: a repository should expose certification status and evidence in a standardised manner (as exemplified in Appendix D), and certification (validation) authorities should enable machine navigation to the requisite evidence (this Appendix).

To illustrate this aspect by way of prototyping, we have implemented a prototype certification repository based on the [CoreTrustSeal certification repository](#),<sup>33</sup> hosted by DANS on behalf of CoreTrustSeal. The [prototype repository](#)<sup>34</sup> implements two of the guidelines:

1. It implements FAIRiCAT as a mechanism for directing machines to an OAI-PMH endpoint, where a list of all certified repositories can be obtained, and in turn, the list can be used to navigate to individual repository certification metadata records. This information can be found in the repository HEAD (Figure E.1).
2. These metadata records are of type schema:Certification, asserting that they are certificates (see example [here](#)<sup>35</sup>).

```
slavat@es-test:~/projects/dev/6.5/dataverse$ curl -I localhost:8080
HTTP/1.1 200 OK
Server: Payara Server 6.2024.6 #badassfish
X-Powered-By: Servlet/6.0 JSP/3.1 (Payara Server 6.2024.6 #badassfish Java/Eclipse Adoptium/17)
Set-Cookie: JSESSIONID=043da0abdf1d2e3a81c9806998e2; Path=/
Link: <http://localhost:8080/api/fairicat/api-info.json>;rel="api-catalog";type="application/json";profile="https://signposting.org/FAIRiCat/", <https://schema.org/AboutPage>;rel="type", <https://schema.org/Dataset>;rel="type", <https://schema.org/Certification>;rel="type", <https://schema.org/Certification>;rel="type"
Content-Security-Policy: frame-ancestors 'none'
Transfer-Encoding: chunked
Content-Type: text/html; charset=UTF-8
X-Frame-Options: SAMEORIGIN
```

**Figure E.1 - Implementation of FAIRiCAT for a Prototype CoreTrustSeal Server**

The LinkSet document referenced in the HEAD response from the server indicates how the OAI-PMH endpoint can be accessed. Other discovery protocols (e.g. OpenSearch, SPARQL) are also available.

```
{
  {
    "anchor": "http://dataverse-fairicat.dansdemo.nl/oai",
    "service-doc": [
      {
        "href": "https://www.openarchives.org/OAI/openarchivesprotocol.html",
        "type": "text/html",
        "title": "Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0"
      }
    ],
    "service-meta": [
      {
        "href": "http://dataverse-fairicat.dansdemo.nl/oai?verb=Identify",
        "type": "application/xml"
      }
    ]
  },
  {
    "anchor": "https://www.openarch.nl/search.php",
    "service-desc": [
      {
        "href": "https://www.openarchieven.nl/opensearch/nl.xml",
        "type": "application/xml"
      }
    ],
    "service-doc": [
      {
        "href": "https://github.com/dewitt/opensearch/blob/master/opensearch-1-1-draft-6.md",
        "type": "text/markdown",
        "title": "OpenSearch 1.1 Draft 6"
      }
    ]
  },
  {
    "anchor": "https://triplestore.netwerkdigitaalrfoed.nl/sparql",
    "service-doc": [
      {
        "href": "https://www.w3.org/TR/sparql11-query/",
        "type": "text/html",
        "title": "SPARQL 1.1 Query Language"
      }
    ]
  }
]
```

**Figure E.2 - Search and Harvesting Affordances for the Prototype CoreTrustSeal Repository**

<sup>33</sup> <https://dataverse.nl/dataverse/coretrustseal>

<sup>34</sup> <https://dataverse-fairicat.dansdemo.nl/>

<sup>35</sup>

<https://dataverse-fairicat.dansdemo.nl/api/datasets/export?exporter=schema.org&persistentId=doi%3A10.5072/FK2/7XSGCT>

## Appendix F: Lessons learned through the Support Action

A support action was defined to allow FAIR-IMPACT to work with data repositories who are interested in increasing their findability and interoperability by exposing machine actionable metadata about their repository and its contents. Participants with access to the repository hosting infrastructure and with sufficient technical expertise were invited to apply. In the support action, the participants learned about the updated Guidelines for repositories and registries on exposing repository trustworthiness status and FAIR data assessments outcomes<sup>36</sup> and tested a prototype developed by FAIR-IMPACT to check for exposed metadata. We received twelve applications to join the support action and eight teams were selected to receive support.<sup>37</sup>

### What worked well:

- Taking part in the support action highlighted the importance of machine-readable metadata and illustrated that **small, practical steps can significantly enhance metadata transparency**
- The support action provided a **useful starting point for understanding and implementing FAIR** and the Research Data Alliance (RDA) Data Repository Attributes Working Group recommendations. The prototype also provided good entry level information to learn more about **existing web standards** and how they can be used to provide information about data repositories. This information can help to ensure that better metadata is provided at both the repository and at the dataset level.
- The support action provided participants with a **deeper understanding of metadata standards** and highlighted the crucial role they play in enhancing repository interoperability
- The participants **valued the concrete examples** that were shared
- **Peer exchange and advice** from support providers was useful to suggest new vocabularies and potential solutions
- Participation helped some to identify **longer-term improvements** that could be made in the repository workflow
- Participation highlighted the **need for dedicated data stewards** within organisation/service to ensure long-term sustainability of FAIR practices

### Challenges encountered:

- Knowledge of the repository service and the broader repository landscape is key as are having familiarity with the schema.org, DCAT formats and JSON files. The **required knowledge is often not found in one individual**. Many participants had to

<sup>36</sup> <https://doi.org/10.5281/zenodo.10058634>

<sup>37</sup> For more on the support action and the list of participants, see <https://fair-impact.eu/testing-tdr-and-fair-enabling-prototype>

join up with different roles across their service to find the right knowledge while for others this required additional time and self-directed learning.

- Additional **time and effort was needed to reorganise and/or create repository website content** to include links to the relevant policies and documents
- Implementing FAIR principles, particularly metadata standards, requires both time and specialized skills. **Collaboration with IT teams is essential** to ensure technical feasibility.
- Deciding exactly **what metadata to expose and how best to represent it** was a challenge. This is particularly true when the repository supports several different options. While the Support Action provided clear code examples and easy-to-use testing tools, additional formal documentation would also have helped in this decision-making process.
- Several encountered **difficulties when documenting urls**. In particular, providing a clear documentation url for an OAI endpoint was not easy.
- Several participants had problems with writing the JSON-LD file as for some fields, there was a **lack of clarity about what values were allowed or recommended**.
- The checker highlighted that problems existed but in some cases it was **not clear what the specific issue was**.
- In some cases, the repository provided **metadata in multiple standards**, but the checker could only recognize the first standard
- The checker needs should differentiate between inherent identifiers supplied by the repository and compatible identifiers that can be provided by depositors.
- Not all relevant policies or documentation is available in a standardised or machine readable format. Using **free text as unstructured information is sometimes the only option**. While not ideal, it was deemed better than providing no information at all.
- As long as no machine-readable policies are available, a **vocabulary** that better identifies the policy type (as subclasses of dc:Policy) of linked policy documents would be advantageous. However, this has yet to be developed.
- As repositories evolve, the guidelines for repository metadata will also evolve, making it **challenging to stay up to date**.

## ***Appendix G: Detailed Recommendations for Future Work***

This section contains a consolidation of recommendations for future work and/ or application using the guidelines developed in the task.

***Table G.1 - Recommendations for Future Work***

#	Aspect	Description	Reference
RF-1	Reporting repository FAIR outcomes - inheritance of FAIR attributes from objects or collections	<p>“An object’s FAIR score is partially dependent on the many FAIR-enabling activities undertaken by a repository. Because of this, FAIR assessment scores across many objects held in the same repository are often similar or identical. ...”</p> <p>Continue work in forums such as the EOSC Task Forces or RDA to define such attributes in detail. The EOSC EDEN and FIDELIS projects may also contribute to the work</p>	[16], [32]
RF-2	Changes in FAIRness of an object over time	<p>“ ... Since the use of FAIR assessment tools is intended to provide a starting point from which improvements can be made, it is desirable that multiple assessments can be done for one object over time, and that this history of scores and their changes can be exposed for the object ...”</p> <p>Continue work in forums such as the EOSC Task Forces or RDA to define such attributes in detail. The EOSC EDEN and FIDELIS projects may also contribute to the work</p>	[16], [32]
RF-3	Generating FAIR Implementation Profiles	<p>“ A FAIR Implementation Profile (FIP) is a report made by an organisation or wider community of practice detailing specific choices and implementations made related to the FAIR principles. ... If a repository were to expose information about their FAIR-enabling practices using the guidelines and model, there is a potential to harvest this information and generate a FIP from it ...”</p> <p>Continue work in forums such as the EOSC Task Forces or RDA to define such attributes in detail. The EOSC EDEN and FIDELIS projects may also contribute to the work</p>	[16], [32]
RF-4	Including Repository Information into DMPs	<p>“... There is potential to incorporate guidance into existing questions about data storage, curation, preservation, and access that are included in most data management planning (DMP) tools. By incorporating guidance into the DMP tools directly, users will be better supported to make informed choices about data deposits from the earliest stage of their research... ”</p>	[16]
RF-5	Harvesting of repository information requirements to support assessment and certification	<p>“The exposure of information following the proposed guidelines and model could be used to aid the completion of applications for assessment by a certification body (validation authority). ...”</p> <p>The EOSC EDEN and FIDELIS projects may contribute to this aspect.</p>	[16]
RF-6	Increased ease of executing landscape analyses	<p>“ ... it also supports ... the ability to zoom out to get a view of the landscape as a whole... ”</p> <p>The EOSC EDEN and FIDELIS projects may also contribute to this aspect.</p>	[16]
RF-7	Aggregation of FAIR	FAIR assessments apply at the level of an individual	This

	Assessments	digital object, and there is no clear guidance on aggregation of these results to reflect a measure of FAIR at collection, repository, or organisation level. Work is needed to develop community consensus on the tests, metrics and benchmarks required to implement such aggregation.	document
RF-8	Extended DCAT Model	Future work may include the further development of an extended DCAT model that describes the landscape of research data management in a standardised way.	[18], [27]
RF-9	Harmonised Compliance Assessment Model	Future work could involve the alignment and consolidation of models developed here with those published by FAIRCORE4EOSC and OSTrails to produce a truly generic model for compliance assessment.	[17], [19], [23]
RF-10	LinkSet and JSON-LD Editor Support	Creating a simple utility application - a user interface and the ability to consume, edit, and export both FAIRiCat Linkset documents and DCAT/ schema.org JSON-LD documents - will ensure that Linksets and DCAT/ schema.org documents are interchangeable, quality assured, and available for either implementation option. This could be considered for future development. Such an editor can incorporate the functionality currently offered in the Checker prototype.	[29]
RF-11	Structuring Complex Attribute Sets	Decisions on what metadata to include and how to structure the distribution of attributes in more complex cases (multiple collections, repositories, etc.) are not currently clear enough to prospective implementers. Additional formal documentation will assist in this decision-making process.	This document
RF-12	Harmonise and improve vocabulary usage	Vocabularies and ontologies that are used to make repository attributes machine-readable should be standardised or harmonised via suitable mappings. This applies for example to vocabularies for classifying fields of science or policy documents, which are not yet available to a sufficient extent.	This document
RF-13	Harmonisation of Registry Practices	Engagement with these registries with a view to harmonisation of encoding practices, and publication of schemata for their repository attribute data in cases where this is lacking is a consideration for future work.	This document

## Appendix H: Checklists for Implementation of Guidelines

For guidelines presented in the report, the prototypes described in the Appendices represent options for its implementation. The tables below present checklists of typical considerations should the example of the prototypes be followed.

### H.1 Standards for Exposing Repository Attributes




**Table H.1 - Standards for Exposing Repository Attributes**

#	Purpose	Description	References
H.1.1	Secure a Repository PID	Repositories SHOULD have a PID, and this PID unambiguously identifies the attributes associated with the repository.	[20]
H.1.2	Make use of Repository Registries	A repository SHOULD register itself in an appropriate registry for repositories of a similar nature, and the identifier provided by the registry MAY be used as a PID for the repository - see detailed recommendation in 3.6.	[20]
H.1.3	Use standards to encode attributes Attribute Encoding: DCAT	DCAT MAY be used as a standard for encoding repository attributes.	[16], [18], [25], [26]
H.1.4	Attribute Encoding: schema.org	schema.org MAY be used as a standard for encoding repository attributes.	[16], [18], [25]
H.1.5	FAIRiCAT	FAIRiCAT MAY be used as a specification for directing machines to locations where repository attributes are found, or to direct machines to affordances offered by the repository.	[9], [25]
H.1.6	LinkSet	The LinkSet specification MAY be used in conjunction with FAIRiCAT to encode and describe repository affordances.	[10], [25]
H.1.7	Linkset Reference to Repository Attributes	The LinkSet, if used, MAY include a reference to a JSON resource (web-accessible file or service) that exposes repository attributes using the DCAT or schema.org specification.	[11], [25]
H.1.8	Metadata and Harvesters	Appropriate standards SHOULD be used to expose [object] metadata, FAIR assessment results, and catalogue information towards harvesters and discovery services.	[16] (Guideline 6)
H.1.9	DQV (Data Quality Vocabulary)	FAIR assessments (and potentially other assessments) SHOULD be encoded using the DQV specification as a basis. (1)	[16] (Guideline 6)
H.1.10	Signposting	Repositories MAY implement Signposting to assist with FAIR evaluation.	[29]

Explanatory notes:

1. To standardise outputs of FAIR metrics and associated assessment results the use of Data Quality Vocabulary (DQV)<sup>38</sup> is recommended since this may be used to embed

<sup>38</sup> <https://www.w3.org/TR/vocab-dqv>

FAIR assessment results within the metadata of assessed data sets via DCAT as recommended by the W3C ‘Data on the Web Best Practices’.<sup>39</sup>

2. The Signposting FAIR profile can be implemented for individual digital objects (e.g. datasets) to make their FAIR evaluation machine actionable.

## H.2 Encoding Repository Attributes

**Table H.2 - Best Practices: Encoding Repository Attributes**

#	Best Practice	Description	References
H.2.1	Vocabularies and Registries	Registries and repositories MUST use published and publicly available vocabularies and registries/ URIs for the encoding of attribute values for all attributes that are not descriptive text or evidence (1).	[16] See also 3.3.6
H.2.2	Structured Attributes in a Public Schema	All registry information about repositories SHOULD be offered as structured data conformant with a publicly available schema	

Notes:

1. Examples include lists of supported licences, supported file types, supported metadata schemata, supported disciplines, repository typology, levels of curation, and similar.

## H.3 Exposing Repository Attributes

**Table H.3 - Best Practices: Exposing Repository Attributes**

#	Best Practice	Description (Explanatory Notes)	References
H.3.1	Transparency and Precision	Repositories MUST provide precise descriptions of the particular resources in scope for the repository. (1)	[16] (Guideline 1)
H.3.2	Repository Attribute Exposure	Repositories SHOULD offer attributes by applying one of the recommended mechanisms (see 3.4) and applying applicable standards (see 3.1) (2)	[16] (Guideline 2)
H.3.3	Provide Sources for Attributes and a Single Source of Truth	Information SHOULD be exposed by, and/or provide references to, an originating source.(3)	[16] (Guideline 3)
H.3.4	Clarity of Assertion	Attributes MAY be grouped into one of the following classes: <ul style="list-style-type: none"> <li>• Free-text assertions: statements in response to descriptive criteria about the required information.</li> <li>• Controlled assertions: selections from ontologies</li> </ul>	[16] (Guideline 4)

<sup>39</sup> <https://www.w3.org/TR/vocab-dcat-3/#quality-information>



		or controlled vocabularies. <ul style="list-style-type: none"> <li>● Identification: Expose and reference PIDs or URIs.</li> <li>● Evidence artefacts: links to resources containing the asserted information.</li> </ul> (4)(5)(6)	
--	--	---	--

Explanatory notes and examples:

1. Exposing information on organisations, services, and objects, as well as their functions and characteristics, implies the precise descriptions of the particular resources. For example, a dataset should be recognisable as a dataset by the information consumer and a repository should identify itself as a repository and data catalogue. The denotation should be as detailed as possible, yet referencing superclasses (e.g. 'Resource' for a dataset, or 'DataService' for a catalogue, depending on the standard used).
2. Transparent information exposed should take account of, and map to, existing standards and criteria. While perfect alignment may not be possible, defining and documenting how the information structure and content relates to existing efforts will minimise divergence and maximise interoperability.
3. Much of the relevant information that needs to be available in a transparent way to inform trust is currently available at multiple third party service providers. This guideline therefore specifies that an information provider should be placed in control of its own information and permit multiple other organisations to consume it for a variety of uses.
4. Information consumers or communities could work to specify the scope of content of free-text assertions, the ontologies and controlled vocabularies used for the controlled assertions, or acceptable links to use for evidence artefacts.
5. Persistent identification, resolution and associated metadata are essential foundations for this guideline.
6. To facilitate meaningful validation actions, supporting information and documents should be linked and exposed. As an example, when a repository claims certification, it should provide a link to the certificate at the certification authorities' site.

#### H.4 Implementation Mechanisms for Exposing Repository Attributes

**Table H.4 - Best Practices: Implementation Mechanisms**

#	Best Practice	Description	References
H.4.1	FAIRiCAT	Repositories MAY enable machine navigation to a description of affordances (e.g. services) using the FAIRiCAT specification. The recommended FAIRiCAT implementation mechanism is discussed in detail in Appendix C.	<a href="#">Appendix C</a>

H.4.2	Local LinkSet	Repositories MAY enable machine navigation to a description of affordances (e.g. services) using the LinkSet standard. The recommended local LinkSet implementation is specified in Appendix C.	<a href="#">Appendix C</a>
H.4.3	Signposting	Repositories MAY enable machine navigation to repository attributes encoded in DCAT or schema.org using the Signposting specification.	<a href="#">Appendix D</a>
H.3.4	Embedded JSON-LD	Repositories MAY enable machine navigation to repository attributes encoded in DCAT or schema.org using embedded JSON-LD.	<a href="#">Appendix D</a>
H.3.5	Local Repository Attributes	Repositories SHOULD maintain their repository attributes by applying one of two recommended standards for such encoding: DCAT or schema.org	<a href="#">Appendix A</a> <a href="#">Appendix B</a>
H.3.6	Validation Authority - Repository Certification Status	Validation Authorities SHOULD offer repository certification status by applying the recommended mechanisms.	<a href="#">Appendix E</a>

## H.5 Examples and Prototypes

**Table H.5 - Best Practices: Implementation Mechanisms**

#	Prototype	Description	References
H.5.1	FAIRiCAT	The recommended FAIRiCAT implementation mechanism is discussed in detail in Appendix C.	<a href="#">Appendix C</a>
H.5.2	Local LinkSet	The recommended local LinkSet implementation is specified in Appendix C.	<a href="#">Appendix C</a>
H.5.3	Example DCAT Encoding	An example DCAT encoding for repository attributes is shown in Appendix A	<a href="#">Appendix A</a>
H.5.4	Example schema.org encoding	An example schema.org encoding for repository attributes is shown in Appendix B	<a href="#">Appendix B</a>
H.5.5	Example Certification Evidence	A prototype showing exposure of certification evidence via machine-actionable links	<a href="#">Appendix E</a>

## H.6 Encoding Specific Attributes: FAIR and Trust Certification

**Table H.6 - FAIR and Trust Certification**

#	Best Practice	Description	References
H.6.1	Exposing FAIR	FAIR Assessment SHOULD be aggregated for a	[16]

	Assessment Status	repository from assessments of objects in the repository.	(Guideline 7)
H.6.2	Exposing Certification Status	Repositories that are formally certified as trustworthy SHOULD expose a link to evidence of such certification.	[16] <a href="#">Appendix E</a>
H.6.3	Validation Type	Guideline 7 defines a range of possibilities for the type of validation associated with an attribute, these SHOULD be used to quality the attribute.	(Guideline 7)

Explanatory notes:

1. Since there are several FAIR evaluation tools, each of which evaluates the various FAIR implementation options somewhat differently, multiple evaluation tools should be used which should be calibrated against a selection of FAIR benchmarking standard datasets, such as the currently prepared set of benchmarks for FAIR signposting.<sup>40</sup>

## H.7 Identifiers for Repositories

**Table H.7 - Identifiers for Repositories**

#	Best Practice	Description	References
H.7.1	re3data Identifier	The re3data identifier and associated DOI are referencing the repository description in re3data, not the repository itself. It MAY be used as a reference to a repository.	[21]
H.7.2	FAIRSharing Identifier	The FAIRSharing identifier also refers to the entry in FAIRSharing, and not to the repository itself. It MAY be used as a reference to a repository.	[21]
H.7.3	RRID Identifier	The RRID (Research Resource ID) refers directly to the repository. It SHOULD be used as a reference to a repository.	[21]
H.7.4	Organisations	Some organisations are equivalent to a repository, but many organisations operate more than one repository, and in a few cases a repository is operated by multiple organisations. A RoR SHOULD be used to reference the organisation(s) involved,	[21]

<sup>40</sup> <https://s11.no/2022/a2a-fair-metrics/>