FAIR Metrics Evaluation Results

Supplementary Information for "A design framework and exemplar metrics for FAIRness"

In order to assess the usability of the core set of FAIR metrics, and to test the intended universal applicability of the FAIR metrics for the evaluation of FAIRness on diverse digital resources, a short questionnaire was prepared, prompting researchers to manually evaluate a digital resource of their choice. Participation was voluntary. The aim was to find a number of participants who, when taken together, would demonstrate application of the FAIR metrics on multiple different kinds of digital resource. Here we present 10 independent evaluations (performed by 12 researchers) covering:

(1) multiple data repositories (public and commercial);
(2) an index to a large data repository;
(3) dataset composed of nanopublications (early mover in FAIR data)
(4) dataset composed of software.

The data evaluated here span humanities, life science, and earth systems knowledge domains.

Table 1 lists the the digital resources evaluated, and gives brief description of the data types and participant metadata.

Below is the "**Questionnaire for (manual) evaluation of a digital resource**" that participants used to complete their FAIR metrics evaluations. There were 15 questions (requiring 22 responses) corresponding to each of the FAIR metrics. Participant responses are collated here in a single document. In cases where the digital resource did not comply with the FAIR metric, participants might have responded with "none" or left the question unanswered, which we indicate with NRP (No Response Provided).

Table 1: Description of Digital Resources Evaluated

Digital Resource Evaluated	Participant Evaluators
Dataverse Repository https://dataverse.harvard.edu https://doi.org/10.25504/FAIRsharing.t2e1ss Dataverse is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data available to others, and allows you to replicate others' work more easily. Researchers, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility. Dataset analyzed for FAIRness: a replication study of historical research in archeology	Name: Julian Gautier Title: Product Research Specialist Affiliation: Institute for Quantitative Social Science, Dataverse Network, Harvard University Email: juliangautier@g.harvard.edu Link: https://www.iq.harvard.edu/people/julian-gautier ORCID: 0000-0001-9229-5317 Name: Derek Murphy Title: Usability Researcher Affiliation: Institute for Quantitative Social Science, Harvard University Email: dlmurphy@g.harvard.edu

	Link: https://www.iq.harvard.edu/people/derek-murphy ORCID: 0000-0002-2850-0846
Dryad http://datadryad.org https://doi.org/10.25504/FAIRsharing.wkggtx A curated resource that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad provides a general-purpose home for a wide diversity of datatypes. Dataset analyzed for FAIRness: A dataset of 204 million gene-disease associations discovered by text mining scientific literature.	Name: Annika Jacobson Affiliation: Department of Human Genetics, Leiden University Medical Center, The Netherlands Email: A.Jacobsen@lumc.nl Link: Biosemantics at LUMC ORCID: 0000-0003-4818-2360
Nanopublication Dataset A dataset of 204 million gene-disease associations discovered by text mining scientific literature, exposed as machine-readable nanopublications. The nanopublication schema is described at http://nanopub.org/wordpress/	Name: Annika Jacobson Affiliation: Department of Human Genetics, Leiden University Medical Center, The Netherlands Email: A.Jacobsen@lumc.nl Link: Biosemantics at LUMC ORCID: 0000-0003-4818-2360 Name: Rajaram Kaliyaperumal Title: Scientific Programmer Affiliation: Department of Human Genetics, Leiden University Medical Center, The Netherlands Email: R.Kaliyaperumal@lumc.nl Link: Biosemantics at LUMC ORCID: 0000-0002-1215-167X
Zenodo Repository https://www.zenodo.org https://doi.org/10.25504/FAIRsharing.wy4egf The OpenAIRE project to be a catch-all repository for EC funded research. CERN, an OpenAIRE partner and pioneer in open source, open access and open data, provided this capability and Zenodo (launched in May 2013). Dataset analyzed for FAIRness: This is an example of applying the FAIR Metrics to a software Digital Resource. Xenon software is a flagship of the Netherlands eScience Center. Xenon is a unified programming interface to many remote computation and data storage facilities, hiding complicated protocols and tool specific details when using applications. https://www.esciencecenter.nl/technology/software/xenon	Name: Mateusz Kuzak Title: Scientific Community Manager Affiliation: Dutch Techcentre for Life Science Email: mateusz.kuzak@dtls.nl ORCID: 0000-0003-0087-6021 Name: Carlos Martinez-Orti Title: escience research software engineer Affiliation: Netherlands eScience Center Email: c.martinez@esciencecenter.nl ORCID: 0000-0001-5565-7577
Yale Institution for Social and Policy Studies (ISPS) The ISPS KnowledgeBase is the gateway to all ISPS	Name: Katherine Thornton Title: Postdoctoral Associate

data, projects, and publications. It is an integrated database which provides a one-stop-shop for ISPS-related research products. Dataset analyzed for FAIRness: Data is of scholars responding to a survey about their experiences trying to replicate published quantitative work.	Affiliation: Department of Computer Science, Yale University Email: katherine.thornton@yale.edu ORCID: 0000-0002-4499-0451
Figshare http://figshare.com/ https://doi.org/10.25504/FAIRsharing.drtwnh Figshare is a repository where users can make all of their research outputs available in a citable, shareable and discoverable manner. Dataset analyzed for FAIRness: Science Stories SPARQL Queries: A set of SPARQL queries used in the Science Stories web application.	Name: Katherine Thornton Title: Postdoctoral Associate Affiliation: Department of Computer Science, Yale University Email: katherine.thornton@yale.edu ORCID: 0000-0002-4499-0451 Name: Kenneth Seals-Nutt Afiliation: Department of Computer Science, Yale University Email: kenneth.seals-nutt@yale.edu ORCID: 0000-0002-5926-9245
Single Cell Portal (BETA) URL? The SCP is a visualization portal for single cell RNA-seq data. Now featuring 32 studies with 394,070 cells. Dataset analyzed for FAIRness:	Name: Eric Weitz Title: Senior Software Engineer Affiliation: Broad Institute, 415 Main Street, Cambridge, MA 02142 Email: eweitz@broadinstitute.org ORCID: Name: Timothy Tickle Title: - Affiliation: Broad Institute, 415 Main Street, Cambridge, MA 02142 Email: ttickle@broadinstitute.org ORCID: Name: Jonathan Bistline Title: - Affiliation: Broad Institute, 415 Main Street, Cambridge, MA 02142 Email: bistline@broadinstitute.org ORCID:
The Sea Data Net, Common Data Index (CDI) http://www.seadatanet.org/Data-Access/Common-Data-Index-CDI https://doi.org/10.25504/FAIRsharing.yzagph The Common Data Index (CDI) service gives users a highly detailed insight in the availability and geographical spreading of marine data sets that are managed by the SeaDataNet data centres. Moreover it provides a unique interface for requesting access, and if granted, for downloading data sets from the distributed data centres across Europe.	Name: Peter Thijsse, MSc. Title: Project Manager Affiliation: MARIS BV, Koningin Julianalaan 345 A, 2273 JJ, Voorburg Email: peter@maris.nl Link: https://www.maris.nl/ ORCID: 0000-0001-9214-3217

The CDI metadata directory describes the archived "raw but validated" monitoring and research datasets held at the datacenters in the SeaDataNet infrastructure. The CDI directory contains > 2 million of these index files, and are the basis for discovery, ordering and downloading the data. The CDI service gives users a highly detailed insight in the availability and geographical spreading of marine data sets, that are managed by the SeaDataNet data centres. Moreover it provides a unique interface for requesting access, and if granted, for downloading data sets from the distributed data centres across Europe.	
Wikidata http://wikidata.org/ https://doi.org/10.25504/FAIRsharing.6s749p Free knowledge database project hosted by Wikimedia and edited by volunteers. Wikidata is a free, open, collaborative, multilingual, secondary knowledge base, collecting structured data that can be read and edited by both humans and machines. Dataset analyzed for FAIRness:	Name: Andra Waagmeester Title: Data scientist Affiliation: Micelio www.micel.io Email: andra@micel.io Link: https://www.wikidata.org/wiki/Q19845625 ORCID: 0000-0001-9773-4008

Table 2. Summary of FAIR metrics self-scoring.

Green = passes FAIR Metric
Red = fails FAIR Metric
Yellow = problementatic (for example, incorrectly interpreted question)
Gray = Can not be evaluated

IRI = Respondent gives an IRI
none = Respondent answered "none"
NRP = No Response Provided

FM	Question	Dataverse	Dryad	Nano-pub	Zenodo	Yale ISPS	Figshare	Broad's SCP	SeaData Net's CDI	Wikidata
IRI Exists	1	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI
F1A	2	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI
F1B	3	IRI	IRI	IRI	NRP	none	IRI	IRI	IRI	IRI
F2A	4A	IRI	IRI	IRI	IRI	none	none	IRI	IRI	IRI
F2A	4B	IRI	none	IRI	IRI	"Multiple"	none	IRI	IRI	IRI
F3	5A	IRI	IRI	IRI	IRI	none	NRP	IRI	IRI	IRI
F3	5B	IRI	IRI	IRI	IRI	IRI	IRI	IRI	none	IRI
F4	6A	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI
F4	6B	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI
A1.1	7A	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI	IRI
A1.1	7B	true	true	true	true	true	true	true	true	true
A1.1	7C	true	true	true	true	true	true	true	true	true
A1.2	8A	false	false	false	false	false	false	false	true	false
A1.2	8B	N/A	N/A	N/A	N/A	NRP	NRP	NRP	link	N/A
A2	9	IRI	IRI	none	IRI	none	IRI	none	IRI	NRP
I1	10	IRI	IRI	IRI	IRI	none	none	NRP	IRI	IRI

I2	11	IRI	IRI	IRI	none	none	none	IRI	IRI	IRI
I3	12	NRP	IRI	IRI	none	none	none	NRP	NRP	IRI
R1.1	13	IRI	IRI	IRI	IRI	IRI	IRI	NRP	IRI	IRI
R1.2	14A	IRI	IRI	IRI	IRI	none	none		NRP	NRP
R1.2	14B		none		none	none	none			
R1.3	15	NRP			none	none	none	NRP		

Questionnaire for (manual) evaluation of a digital resource

Introduction

You will be asked to provide various information at each step. In most cases the information you provide will be a URL that will be resolved, and the returned data will be used to validate your adherence to the intent of the Principle being evaluated.

What resource should I be evaluating?

FAIR evaluations are quite "granular" - we evaluate the content returned from resolving some kind of Internationalized Resource Identifier (https://en.wikipedia.org/wiki/Internationalized_Resource_Identifier) - generally, a URL. So in most cases we are evaluating **only one thing** - the "thing" at that URL.   Moreover, what resource you evaluate will depend on what kind of FAIR data provider you are, and/or what specific type of entry you are interested in. For example, a data repository owner may choose to evaluate an example of the provenance metadata from a typical deposit in their repository (am I collecting metadata about my users' deposits, in a FAIR manner?), whereas a *de novo* data creator (e.g. a biologist) may wish to evaluate the FAIRness of the contextual metadata they have captured about their data, or they may wish to evaluate the FAIRness of the data itself. The important thing is to recognize **what portion of the data/metadata <u>you are responsible for,</u>** and evaluate that - clearly you cannot be responsible for things that are out of your control (though if you are a special-purpose repository, you could certainly provide guidance for your community members to help them create more FAIR data deposits!). For example, Zenodo is not responsible for the FAIRness of the *data* deposited by Zenodo users; they are, however, responsible for the collection and publication of the metadata related to that deposit (even if it is duplicated within the deposited data itself!). Similarly, data providers are responsible for both the data, and the provision of a wide variety of provenance and other contextual metadata; if these are represented in different files - possibly deposited in different locations - **they will want to evaluate these files independently**.

Finally, it is entirely up to the provider to decide which of their resources they wish to evaluate. The temptation will (obviously) be to select the one that you know to be "best" or "the most FAIR"! However, we would encourage providers to select both the "best" and the "worst" and evaluate these separately. FAIRness is NOT a competition, it is an aspiration, and the Metrics (and this evaluation) are not intended to judge you, they are intended to guide you. By evaluating the "worst" of your resources, you may use the Metrics as a guideline for improving the FAIRness of those data, thus improving your overall FAIRness.

With that introduction, please select the resource (metadata or data) and answer the following requests for information (where we ask for a IRI, this will often mean "a URL").

How to interpret your result
It is important to reiterate that FAIR is intended to be aspirational. As such, it is not possible to "pass" or "fail" this evaluation! Individual components of this evaluation questionnaire may reveal high levels of FAIR Principles-compliance, while other components may show low or no compliance. In the latter case, the purpose of this evaluation is merely to provide an opportunity to consider how to improve the FAIRness of that (or that type of) resource in the future. Note also that, while many of the answers are self-provided, and cannot be rigorously evaluated by an automated system (or even someone outside of your specific domain), it is likely that these evaluations will be made public, and thus the answers provided will be available for inspection by others in your community! This "peer review" acts as an additional layer of validation for the results of this FAIR evaluation.

Question 1: RESOURCE ID, Please provide the IRI of the resource to be evaluated:
   ● Dataverse http://dx.doi.org/10.7910/DVN/S2ZPXL
   ● Dryad https://doi.org/10.5061/dryad.gn219
   ● Nanopublication
     https://datadryad.org/bitstream/handle/10255/dryad.91060/gda-np.nq.gz?sequence=1 (This is
     the link behind the "View/Open" button for file *gad-np.nq.gz* 30GB)
   ● Zenodo http://doi.org/10.5281/zenodo.1188072
   ● Yale ISPS http://sbx.isps.yale.edu/research/data/d108
   ● Figshare https://figshare.com/articles/Science_Stories_SPARQL_Queries/5942305
   ● Broad's SCP https://portals.broadinstitute.org/single_cell (Users of Single Cell Portal with
     sufficient permissions to run workflows in our system get analysis.json files as a metadata file
     describing their submission. Currently, these files are accessible only to the study owner by
     default. However, study owners can export their analysis.json and upload it as a file to their study
     -- that is what I have done here. A specific instance of the analysis.json resource was created for
     this evaluation.
     https://portals.broadinstitute.org/single_cell/data/public/fair-evaluation?filename=analysis.json)
   ● SeaDataNet's CDI https://www.seadatanet.org/Metadata/CDI-Common-Data-Index
   ● Wikidata https://www.wikidata.org/wiki/Special:EntityData/Q2013.ttl
   ● Euretos

Question 2: FM F1A, Identifier Persistence:

Please provide the IRI for a registered identifier schema for your resource's IRI (e.g. DOI, HTTP):

- Dataverse https://www.doi.org/hb.html
- Dryad https://www.doi.org/hb.html
- Nanopublication https://www.ietf.org/rfc/rfc3986.txt
- Zenodo https://www.doi.org/hb.htm
- Yale ISPS https://www.w3.org/wiki/UriSchemes/http
- Figshare https://www.w3.org/wiki/UriSchemes/http
- Broad's SCP https://tools.ietf.org/html/rfc7230
- SeaDataNet's CDI http://seadatanet.maris2.nl/v_cdi_v3/browse_step.asp (but no identifier schema yet => http://cdi.seadatanet.org/record/id coming soon)
- Wikidata https://www.wikidata.org/wiki/Help:Namespaces
  There are:
    - Main namespace: Example: http://www.wikidata.org/wiki/Q142
    - Property: Example: http://www.wikidata.org/wiki/Property:P102
    - Wikidata: Example: http://www.wikidata.org/wiki/Wikidata:Glossary
    - User: Example: http://www.wikidata.org/wiki/User:Lydia_Pintscher_(WMDE)
    - Talk namespace: Example: http://www.wikidata.org/wiki/Q142 → http://www.wikidata.org/wiki/Talk:Q142
    - Special pages: Example: http://www.wikidata.org/wiki/Special:Search
- Euretos

Question 3: FM F1B, Identifier Persistence:

Please provide the IRI to the document describing the persistence policy for the identifier of this (meta)data (this may be a document from your service provider, e.g. Zenodo, UniProt, etc.):

- Dataverse https://dataverse.org/best-practices/harvard-dataverse-preservation-policy
- Dryad (from doi) https://www.doi.org/doi_handbook/6_Policies.html (from the Terms of Use) https://datadryad.org/pages/policies#preservation
- Nanopublication The nanopubs do not have a link to such a policy document. So I assume as an implicit default, these nanopubs linked to here in Dryad will use the Dryad persistence policy.
- Zenodo NRP
- Yale ISPS none
- Figshare https://figshare.com/blog/Ensuring_persistence_on_figshare/25
- Broad's SCP https://cloud.google.com/storage/docs/
- SeaDataNet's CDI https://www.seadatanet.org/content/download/1695/10119/file/SeaDataNet+Data+Policy.pdf?version=1 (Linked from this page https://www.seadatanet.org/Data-Access/Data-policy)
- Wikidata http://www.wikidata.org/entity/Pxxxx, http://www.wikidata.org/entity/Qxxxx, https://www.mediawiki.org/wiki/Wikibase/DataModel
- Euretos

Question 4: FM-F2, Metadata for discoverability:

(A) Please provide the IRI to a document that contains machine-readable metadata for the digital resource:

- Dataverse
  https://dataverse.harvard.edu/api/datasets/export?exporter=ddi&persistentId=doi%3A10.7910/DVN/S2ZPXL (note this is XML, not semantically enabled like RDF)
- Dryad https://datadryad.org//resource/doi:10.5061/dryad.gn219?show=full
- Nanopublication
  https://datadryad.org/bitstream/handle/10255/dryad.91060/gda-np.nq.gz?sequence=1
- Zenodo http://data.datacite.org/10.5281/zenodo.1188072 (sigle url pointing to different formats, requires content negotiation: curl -LH "Accept: application/vnd.datacite.datacite+xml" it would be nice if zenodo provided separate url for each type of the metadata)
- Yale ISPS none
- Figshare none
- Broad's SCP
  https://raw.githubusercontent.com/HumanCellAtlas/metadata-schema/4.6.1/json_schema/analysis.json
- SeaDataNet's CDI http://seadatanet.maris2.nl/v_cdi_v3/print_xml.asp?n_code=2626292
- Wikidata https://www.wikidata.org/entity/Q2013.ttl
- Euretos

(B) Please provide the IRI for the file format of this metadata:
- Dataverse https://www.w3.org/2001/XMLSchema-instance
- Dryad none
- Nanopublication http://www.nanopub.org/nschema
- Zenodo https://schema.datacite.org
- Yale ISPS Multiple file formats
- Figshare none
- Broad's SCP https://tools.ietf.org/html/rfc7159
- SeaDataNet's CDI http://vocab.nerc.ac.uk/collection/L24/current/
- Wikidata https://www.wikidata.org/wiki/Special:EntityData/Q114409.ttl
- Euretos

Question 5: FM-F3, Identifier in Metadata:
(A) Please provide the IRI of the metadata:
- Dataverse
  https://dataverse.harvard.edu/api/datasets/export?exporter=ddi&persistentId=doi%3A10.7910/DVN/S2ZPXL
- Dryad https://datadryad.org//resource/doi:10.5061/dryad.gn219?show=full
- Nanopublication
  https://datadryad.org/bitstream/handle/10255/dryad.91060/gda-np.nq.gz?sequence=1 (each nanopub will have its own Provenance and Publication Information metadata)
- Zenodo http://data.datacite.org/10.5281/zenodo.1188072
- Yale ISPS none
- Figshare NRP
- Broad's SCP
  https://portals.broadinstitute.org/single_cell/data/public/fair-evaluation?filename=analysis.json
  (user must be authorized)
- SeaDataNet's CDI http://seadatanet.maris2.nl/v_cdi_v3/print_xml.asp?n_code=2626292
- Wikidata https://www.wikidata.org/wiki/Special:EntityData/Q2013.ttl
- Euretos

(B) Please provide the IRI of the data described by that metadata: *(note that **one, or the other**, of these **MUST** be the same as the RESOURCE ID provided at the beginning of this questionnaire)*

- Dataverse http://dx.doi.org/10.7910/DVN/S2ZPXL
- Dryad https://datadryad.org/bitstream/handle/10255/dryad.91060/gda-np.nq.gz?sequence=1
- Nanopublication
  https://datadryad.org/bitstream/handle/10255/dryad.91060/gda-np.nq.gz?sequence=1
- Zenodo http://doi.org/10.5281/zenodo.1188072
- Yale ISPS http://sbx.isps.yale.edu/research/data/d108
- Figshare https://figshare.com/articles/Science_Stories_SPARQL_Queries/5942305
- Broad's SCP
  https://raw.githubusercontent.com/HumanCellAtlas/metadata-schema/4.6.1/json_schema/analysis.json
- SeaDataNet's CDI None (later a PID will become available for data replicated in cloud)
- Wikidata http://www.wikidata.org/
- Euretos

Question 6: FM-F4, Indexed in Searchable Resource:

Please provide the URL to a search engine, and the query that will be executed to discover your RESOURCE ID (found **in the first page** of the search):

(A) URL:

- Dataverse   https://www.google.com/
- Dryad https://www.google.nl/
- Nanopublication https://www.google.nl/
- Zenodo https://www.google.com/
- Yale ISPS www.google.com
- Figshare www.google.com
- Broad's SCP   https://portals.broadinstitute.org/single_cell
- SeaDataNet's CDI   http://seadatanet.maris2.nl/v_cdi_v3/search.asp
- Wikidata http://query.wikidata.org
- Euretos

(B) Search query/terms:

- Dataverse   Replication data for: Plutniak,Sébastien, "The Professionalisation of Science – Claim and Refusal
- Dryad The implicitome: a resource for rationalizing gene-disease associations
- Nanopublication The implicitome: a resource for rationalizing gene-disease associations nanopublications
- Zenodo xenon escience center zenodo
- Yale ISPS "Dafoe, Allan (2014). Replication Materials for: 'Science Deserves Better: The Imperative to Share Complete Replication Files,' ISPS Data Archive.
- Figshare I was not able to write a query that would return this resource
- Broad's SCP   fair (more formally: ?search_terms=fair) (Note that the discovery here is of the RESOURCE's container – a study in the Single Cell Portal. A study can have multiple files; one such file in this case is the analysis.json. Our search functionality does not surface study files at this time, only studies.)
- SeaDataNet's CDI free search: 2626292
- Wikidata   https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples
- Euretos

Question 7: FM-A1.1, Access Protocol:

Please provide:

(A) A URL to the description of the protocol:

- Dataverse   https://tools.ietf.org/html/rfc7230
- Dryad https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol
- Nanopublication https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol
- Zenodo https://tools.ietf.org/html/rfc7230
- Yale ISPS https://www.w3.org/wiki/UriSchemes/http
- Figshare https://www.w3.org/wiki/UriSchemes/http
- Broad's SCP   https://tools.ietf.org/html/rfc6749
- SeaDataNet's CDI http://vocab.nerc.ac.uk/collection/L07/current/
- Wikidata http://www.wikidata.org/entity/Q44484
- Euretos

(B) Is the protocol is open (technical details are provided)? (answer true/false):

- Dataverse true
- Dryad true
- Nanopublication true
- Zenodo true
- Yale ISPS true
- Figshare true
- Broad's SCP true
- SeaDataNet's CDI true
- Wikidata true
- Euretos

(C) Is the protocol is (royalty) free? (answer true/false):

- Dataverse true
- Dryad true
- Nanopublication true
- Zenodo true
- Yale ISPS true
- Figshare true
- Broad's SCP true
- SeaDataNet's CDI true
- Wikidata true
- Euretos

Question 8: FM-A1.2, Access authorization:

Please answer:

(A) Authorization is required to access the content of my RESOURCE ID: (answer true/false)

- Dataverse false
- Dryad false
- Nanopublication false
- Zenodo false
- Yale ISPS false
- Figshare false
- Broad's SCP false (It requires authentication, but not authorization)
- SeaDataNet's CDI true
- Wikidata   false

- Euretos

(B) If "true" above, please provide a IRI that resolves to a description of the process to obtain access to restricted content:
- Dataverse N/A
- Dryad N/A
- Nanopublication N/A
- Zenodo N/A
- Yale ISPS NRP
- Figshare NRP
- Broad's SCP NRP
- SeaDataNet's CDI http://www.seadatanet.org/urnurl/SDN:L08::RS
- Wikidata N/A
- Euretos

Question 9: FM-A2, Metadata Longevity:

Please provide the URL to a metadata longevity plan:
- Dataverse https://dataverse.org/best-practices/harvard-dataverse-preservation-policy
- Dryad https://datadryad.org/pages/repository
- Nanopublication The nanopubs do not have a link to such a policy document. So I assume as an implicit default, these nanopubs linked to here in Dryad will use the Dryad persistence policy.
- Zenodo http://about.zenodo.org/policies/
- Yale ISPS none
- Figshare https://knowledge.figshare.com/articles/item/preservation-and-continuity-of-access-policy
- Broad's SCP Note that metadata is user-controlled, and thus subject to the user's desired persistence. Users can delete metadata.
- SeaDataNet's CDI https://www.seadatanet.org/Data-Access/Data-policy (Part of the data policy)
- Wikidata NRP
- Euretos

Question 10: FM-I1, Use a Knowledge Representation Language:

Please provide the URL to the specification of the language (please see the FAIR Metrics for a description of how this is evaluated):
- Dataverse http://www.jelks.nu/XML/xmlebnf.html
- Dryad   https://datadryad.org//resource/doi:10.5061/dryad.gn219?show=full (Metadata hsa XML serialization)
- Nanopublication https://www.w3.org/TR/n-quads/   (We used nquad serialization for the implicitome data. The nanopub schema is defined based on rdfg)
- Zenodo http://nlesc.github.io/Xenon/versions/2.6.0/javadoc/, http://xenonrse2017.readthedocs.io/en/latest/, https://docs.oracle.com/javase/specs/jls/se7/html/index.html (not sure which to use)
- Yale ISPS none
- Figshare none
- Broad's SCP NRP
- SeaDataNet's CDI http://vocab.nerc.ac.uk/isoCodelists/sdnCodelists/gmxCodeLists.xml#LanguageCode" codeListValue="eng"
- Wikidata   http://shex.io
- Euretos

Question 11: FM-I2, Use FAIR Vocabularies:

Please provide one or more (max 3) IRIs representing the vocabularies used within the (meta)data that is returned by resolving the RESOURCE ID (e.g. http://purl.obolibrary.org/obo/GO The prefixes for OBO vocabularies can be found here: http://wiki.geneontology.org/index.php/Identifiers

- Dataverse http://id.loc.gov/authorities/subjects/sh85006507, http://id.loc.gov/authorities/subjects/sh2008111330
- Dryad http://dublincore.org/documents/dcmi-terms/, http://www.prismstandard.org/resources/mod_prism.html
- Nanopublication   http://www.w3.org/2002/07/owl#, http://www.nanopub.org/nschema#, http://www.w3.org/2000/01/rdf-schema#
- Zenodo none
- Yale ISPS none
- Figshare none
- Broad's SCP https://raw.githubusercontent.com/HumanCellAtlas/metadata-schema/4.6.1/json_schema/analysis.json (This analysis.json does not use ontological vocabularies like GO, but it does use interoperable metadata standards like the linked HCA analysis.json. The FAIR publication at https://www.nature.com/articles/sdata201618 mentions "vocabularies or other (meta)data standards" in relation to I2. (More context below.) . Are there examples of FAIR-Interoperable metadata standards that are not vocabularies? From the FAIR publication: "First, when community-endorsed vocabularies or other (meta)data standards do not include the attributes necessary to achieve rich annotation, there are two possible solutions: either publish an extension of an existing, closely related vocabulary, or—in the extreme case—create and explicitly publish a new vocabulary resource, following FAIR principles ('I2'). ")
- SeaDataNet's CDI http://vocab.nerc.ac.uk/collection/V22/current/
- Wikidata https://www.wikidata.org/wiki/Wikidata:List_of_properties
- Euretos

Question 12: FM-I3, Use Qualified References:

Please provide the URL to a formal Linkset (defined at: https://www.w3.org/TR/void/#linkset) or copy/paste the content of a formal linkset that describes at least a portion of the content at RESOURCE ID (you may need to create this, or edit the example below):

> *e.g.   :DBpedia2DBLP a void:Linkset;*
> *void:target :DBpedia;*
> *void:target :DBLP;*
> *void:linkPredicate owl:sameAs;*
> *void:triples 10000;*

- Dataverse NRP
- Dryad https://datadryad.org//resource/doi:10.5061/dryad.gn219?show=full (are there linksets here?)
- Nanopublication http://136.243.4.200:8890/sparql?default-graph-uri=&query=select+distinct+%3Fp+%3Fo+where+%7B%3Chttp%3A%2F%2Frdf.biosemantics.org%2Femco%2Fv1.5%2Fconcepts%2FC3065064%3E+%3Fp+%3Fo%7D&format=text%2Fhtml&timeout=0&debug=on (We did provide linkset for genes and diseases mentioned in this dataset, but we didn't follow the proper linked data principles. For example, look at the gene FGD3 the notation predicate of this subject URI provides information about different identifier for this gene. Instead of using exiting URIs we defined our own datatype to represent these identifiers. For example

<http://rdf.biosemantics.org/emco/datatypes/eg-notation-datatype> datatype is used to represent the Entrez Gene ID.)
- Zenodo Since Zenodo does not use control vocabularies, it cannot use linksets.
- Yale ISPS none
- Figshare none
- Broad's SCP NRP
- SeaDataNet's CDI none RDF/Linkeddata in development
- Wikidata https://www.bridgedb.org/mapping-databases/hmdb-metabolite-mappings/
- Euretos

Question 13: FM-R1.1, Accessible usage license:

Please provide the IRI for your usage license regarding *the content returned from* RESOURCE ID (be that data, or metadata):
- Dataverse https://dataverse.org/best-practices/harvard-dataverse-general-terms-use
- Dryad https://datadryad.org/pages/policies#usage →
  http://creativecommons.org/publicdomain/zero/1.0/
- Nanopublication http://creativecommons.org/licenses/by/3.0/ (Note: the nanopubs carry their own license. I wonder which has legal priority… the nanopubs or Dryad ?)
- Zenodo zenod entry license: http://about.zenodo.org/terms/,
  https://creativecommons.org/publicdomain/zero/1.0/ software license:
  https://www.apache.org/licenses/LICENSE-2.0
- Yale ISPS http://creativecommons.org/licenses/by-nc-nd/3.0/us/
- Figshare https://www.gnu.org/licenses/gpl- 3.0.html
- Broad's SCP NRP
- SeaDataNet's CDI For restricted data, as available in the metadata
  http://www.seadatanet.org/urnurl/SDN:L08::RS
  In broader sense the user has to comply with the SDN license of unrestricted/SDN licensed data
  https://www.seadatanet.org/Data-Access/License/1.0
- Wikidata https://wikimediafoundation.org/wiki/Terms_of_Use/en
- Euretos

Question 14: FM-R1.2, Detailed Provenance:

(A) Please provide the IRIs (maximum 3) for the vocabularies being used to describe the **provenance** of the **content resolved from RESOURCE ID** (be that data, or metadata; e.g. http://purl.org/dc/terms/):
- Dataverse
  https://www.ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/codebook.xsd
- Dryad http://dublincore.org/documents/dcmi-terms/
  http://www.prismstandard.org/resources/mod_prism.html
- Nanopublication Unfortunately the provenance part of this dataset is not captured in rdf. We did add some provenance to the nanopubs but those information is not complete. However, there is a very non machine-readable drawing here
  https://datadryad.org/bitstream/handle/10255/dryad.90809/Copy%20of%20Gene-disease%20paper%20-%20code%20overview.pdf?sequence=1.
- Zenodo http://dublincore.org/documents/2012/06/14/dces/
- Yale ISPS none
- Figshare none
- Broad's SCP Generally: https://github.com/broadinstitute/wdl/blob/develop/SPEC.md
  Specifically: https://api.firecloud.org/#/Submissions/workflowMetadata
- SeaDataNet's CDI NRP
- Wikidata NRP

- Euretos

(B) Please provide the IRIs (maximum 3) for the vocabularies being used to describe the **domain information** of the **content resolved from RESOURCE ID** (be that data, or metadata; e.g. http://edamontology.org/data_ will be used for many bioinformatics data/metadata):

- Dataverse http://id.loc.gov/authorities/subjects/sh85006507, http://id.loc.gov/authorities/subjects/sh2008111330
- Dryad none
- Nanopublication https://bioportal.bioontology.org/ontologies/SIO https://bioportal.bioontology.org/ontologies/STATO https://bioportal.bioontology.org/ontologies/PROV (We used our own vocabulary for the genes and diseases. For the nanopubs we used SIO, STATO, PROV.)
- Zenodo none
- Yale ISPS none
- Figshare none
- Broad's SCP https://raw.githubusercontent.com/HumanCellAtlas/metadata-schema/4.6.1/json_schema/analysis.json (This SCP resources uses interoperable metadata standards like the linked analysis.json, but not ontological vocabularies like GO. The HCA metadata standard incorporates ontological vocabularies, e.g. UBERON as referenced in https://github.com/HumanCellAtlas/metadata-schema/blob/290696a0fa77a777b085f94d534f19c79de4d628/json_schema/module/ontology/organ_ontology.json. But restricting "interoperability" to only ontologies, to the exclusion of metadata standards, might be unduly narrow.)
- SeaDataNet's CDI The vocabularies used in the metadata http://seadatanet.maris2.nl/v_cdi_v3/print_xml.asp?n_code are described here: http://vocab.nerc.ac.uk/collection/V22/current/ (I indicated the vocabs used in the metadata for the second point. But for provenance we have nothing sofar)
- Wikidata   prefix prov: <http://www.w3.org/ns/prov#>, prefix schema: <http://schema.org/>, wikibase: <http://wikiba.se/ontology-beta#>
- Euretos

Question 15: FM-R1.3, Meets Community Standards:
Please provide the IRI that represents the certification from a recognized authority in your community or domain, indicating that the content of RESOURCE ID is compliant with the standards of your community: _____ (unlikely to exist at this time….) _____

- Dataverse NRP
- Dryad does not exist
- Nanopublication there nanopublication schema's created by the community http://www.nanopub.org/nschema however, there are no IRIs yet that can satisfy this metric.
- Zenodo There are community accepted ways of measuring quality of software. See https://guide.esciencecenter.nl/software/testing.html. Build: https://travis-ci.org/NLeSC/Xenon, code coverage: https://codecov.io/github/NLeSC/Xenon?branch=master, quality gage: https://www.sonarqube.org/?id=nlesc%253AXenon
- Yale ISPS none
- Figshare not provided
- Broad's SCP NRP
- SeaDataNet's CDI SeaDataNet is the standardising body for marine metadata and so therefore has the opportunity (and perhaps the obligation) to build the FAIR machine-readable data and metadata models that others can aspire to.
- Wikidata NRP

- Euretos