

FAIR self assessment for D&S/TD96 Discovery Project. University of Sydney. Enhancing the data, tools and policy translation resources on CARDAT

Completed 8/10/2019

BACKGROUND

CARDAT hosts OPEN, GENERAL and RESTRICTED data. OPEN data is in public domain, GENERAL is only accessible to member of CAR and RESTRICTED is only available on request.

Questions for each FAIR component ↓		Answer options: Increasingly FAIR -->				
FINDABLE						
Q1	Does the dataset have any identifiers assigned?	No identifier	Local identifier	Web address (URL)	Globally unique, citable and persistent identifier (e.g. DOI, PURL, or Handle)	
	Start of project			GENERAL and RESTRICTED data have persistent web URL through Cloudstor (e.g. e.g. https://cloudstor.aarnet.edu.au/plus/f/365776). The GENERAL data is accessible to CAR employees and associates on request. The RESTRICTED data is accessible only with permission of the data user.	OPEN data has PURLs and is accessible on OSF (e.g. https://osf.io/5wq3f/).	
	End of project				More GENERAL data to be made available as OPEN data.	
	Two years time					
Q2	Is the identifier included in all metadata records or metadata files describing the data?	No	Yes			
	Start of project		Yes, there is a metadata file in the folder and also in a data-inventory for GENERAL, RESTRICTED and OPEN data. All metadata records include identifiers, but some are non-persistent URLs.			
	End of project		All dataset records to have their metadata to be updated with persistent URLs.			
	Two years time					
Q3	Is the data described by a metadata record?	The data is not described	Brief title and description	Brief title and description, and multiple other fields filled out, albeit briefly.	Comprehensively (a min metadata template will be provided) using a formal machine-readable metadata schema.	
	Start of project				Yes for all data types, however the metadata is limited to information about the dataset as a whole and doesn't include tabular metadata for each of the variables.	
	End of project				Investigate adding tabular metadata information to records where applicable (e.g. for CSV files, not shape files).	
	Two years time					
Q4	What type of repository or registry is the metadata record in?	The data is not described in any registry or repository	Local institutional repository	Domain-specific repository	Generalist public repository	Data is in one place but discoverable through several places (i.e. other registries, RDA, Google Data Search)
	Start of project			All metadata records are stored in our Data Inventory. This is available to CARDAT members (a joint project involving multiple universities and CSIRO).		

	End of project			Expand the availability of the Data Inventory to include CAR affiliates.	Investigate publishing Data Inventory in public domain.	
	Two years time					
ACCESSIBLE						
Q5	How accessible is the data? Note: The access method(s) must be explicitly stated in the metadata record, e.g. if any authentication is needed, or there are any restrictions to access.	No metadata record	Access to metadata only	Unspecified access conditions e.g. "contact the data custodian to discuss access"	Embargoed access after a specified date; or A deidentified version of the data is publicly accessible	Fully accessible public, or to persons who meet and follow explicitly stated conditions and processes, e.g. ethics approval for sensitive data
A5	Start of project			RESTRICTED data is available to CARDAT members on request to the data owner. GENERAL data is available to all CARDAT members.		OPEN data is available to the public under CC-BY licences.
	End of project			No change		Make more GENERAL data available as PUBLIC data.
	Two years time					
Q6	Is the data available online without requiring specialised protocols or tools once access has been approved?	No access to data	By individual arrangement	File download from online location	Non-standard web service (e.g. OpenAPI/Swagger/informal API)	Standard web service API (e.g. OGC)
A6	Start of project			All data is available through Cloudstor, and CoESRA. e.g. https://cloudstor.aarnet.edu.au/plus/f/3657763955		
	End of project			No change		
	Two years time					
Q7	Does the repository/registry agree to maintain the persistence of the metadata record, even if the data product is no longer available?	No (or not applicable, if no metadata record exists)	Unsure	Yes		
A7	Start of project			The Data Inventory is maintained separately to the data storage on CloudStor (although a copy is stored with the data). This can be published to GitHub as required.		
	End of project			No change		
	Two years time					
INTEROPERABLE						
Q8	Are the data available in (an) open (file) format(s)?	Data are mostly available only in a proprietary format	Data are available in an open format	Data are available in an open, documented, widely-used standard format (i.e. NetCDF, CSV, JSON, XML, etc)		
A8	Start of project			Yes available as CSV or shapefile both open formats.		
	End of project			No change		
	Two years time					
Q9	Are the data machine readable?	The data are unstructured	The data are structured and machine-readable (i.e. csv, JSON, XML, RDF, database files, etc)			
A9	Start of project		Yes. Datasets are pre-processed in R and saved as CSV			
	End of project		No change			
	Two years time					
Q10	What best describes the types of vocabularies/ontologies/tagging schemas used to define the data elements?	Data elements are not described (i.e. fields or objects are labelled with codes or not at all)	Data elements are described (so that a human user can correctly interpret the data), but no standards have been used in the description	Recognised standards have been used in the description of data elements, but no published vocabularies with resolvable URIs are used	Published vocabularies using resolvable identifiers linking to explanations are used, so that the data can be read and understood by machines as well as humans.	Published vocabularies using persistent resolvable identifiers linking to explanations are used, so that the data can be read and understood by machines as well as humans.

A10	Start of project			Yes we use the EML international metadata standard		
	End of project				Link metadata fields to EML documentation	
	Two years time				Our confidence is high that the EML standard is being conformed to and we would like to implement an automated software tool to check the validity of metadata and therefore increase our level of sureity that we conform to the standard a very high degree	HARVESTED INTO A RIF/CS BASED METADATA PORTAL SUCH AS RDA
Q11	How is the relationship to other data and resources (e.g. related datasets, services, publications, etc) described in the metadata, to provide context around the data?	There are no links to other metadata or data	The metadata record includes URI links to related metadata, data and definitions	Qualified links to other resources are recorded in a machine readable format, e.g. a linked data format such as RDF		
A11	Start of project		Some metadata records includes information on external resources relevant to the dataset (e.g. publications explaining data collection protocols).			
	End of project		More comprehensive information on related data and links to the relevant resource (e.g. link to the DOI for the relevant publication).			
	Two years time					
REUSABLE						
Q12	Which of the following best describes the license (usage rights) attached to the data?	No license is applied	Non-standard license applied, without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Non-standard license applied, WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record
A12	Start of project				Yes not applied consistently though. Some published with no licence.	
	End of project				Have all datasets published with standard licences.	
	Two years time					
Q13	How much provenance information has been captured to facilitate data reuse? i.e. project objectives, data generation/collection (including from external sources) and processing workflows.	No provenance information is recorded	Partially recorded	Comprehensively recorded in a text format (i.e. TXT or PDF)	Comprehensively recorded in a machine readable format (i.e. in metadata record's schema or PROV, or in RDF, JSON, NetCDF, XML, etc)	
A13	Start of project		Yes, we have included in the metadata explanation and links to provenance for some records.			
	End of project			Provide comprehensive information to all records.		
	Two years time					