FAIR self assessment for project: Supporting FAIR machine maintenance data for Industry and Researchers in the Centre for Transforming Maintenance through Data Science

Completed 1/10/2019

	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)	
A1	Start of project	Data did not exist in a collection prior to the project				
	End of project			All datasets in the repository will have a unique URL		
	Two years time				DOI's will be implemented in the coming months via a CKAN plugin	
Q2	Is the identifier included in all metadata records or metadata files describing the data?	No	Yes			
A2	Start of project					
	End of project		All dataset ID's are referenced in the metadata record			
	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.	
A3	Start of project	WIll vary but much of the data has no structured meta data				
	End of project			Meta data standard is still evolving, but early data collection will meet this as a minimum		
	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)
A4	Start of project					
	End of project			https://data-ittc-app.it.csiro.au/		
	Two years time				Cleaned datasets will be published to Zenodo which supports harvesting and should be combatible with RDA and other aggregators	
	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	WIll vary but much of the data has no structured meta data				
	End of project			Data that isn't accessible provides a simple access request form.		
	Two years time				Data published to Zenodo will be publically accessible or will provide clear embagos ot terms for	
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					
	End of project				CKAN provides a API for data access, standard interfaces exist but were not implemented by project end	
	Two years time					Zendo supports OAI-PMH in addition to an API

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	Data is subject to each companies data policies			
	End of project		The CKAN repository may not exist indefinatly		
	Two years time			Collections published to Zenodo will have their meta retained where possible even whendata needs to be removed.	

Q8	Is 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	This will vary between companies and data types			
	End of project			Data, whenever possible will be coverted to an open format prior to being added to a	
	Two years time				
Q9	Is the data machine readable?	The data is unstructured	The data are structured and machine-readable (i.e. csv, JSON, XML, RDF, database files, etc)		
A9	Start of project				
	End of project		*Majority of the data follows some kind of identifiable structure*		
	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	This varies between companies and datasets			
	End of project			Will vary however meta data is being captued using standardised codes and published descriptions. Functionality exists to provide a full data dictionary for tabular data.	
	Two years time				
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	Data will vary			
	End of project			Metadata is using linked csv's and lookup services	
	Two years time				

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	Most data is internal to companies and won't include a specific license			
	End of project				All data will have an associated license from a list
	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	Most data will be raw or may have a partial record of applied transformations.			
	End of project				

	Two years time			Final collections will aim to have some form of text provenance record of all changes made by the projects.		