FAIR self assessment for project: An Australian historical criminal justice data repository

Completed 16/10/2019

	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	No identifier				
	End of project				http://dx.doi.org/10.26193/U3ZZGY	
	Two years time					
Q2	Is the identifier included in all metadata records or metadata files describing the data?	No	Yes			
A2	Start of project	NA				
	End of project				Yes - eg http://dx.doi.org/10.26193/U3ZZGY	
	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			Yes - but not accessible to public users		
	End of project				Yes - eg http://dx.doi.org/10.26193/U3ZZGY	
	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			metadata record in Prosecution Project repository, only accessible by research team and administrators		
	End of project				Yes - eg Dataverse record http://dx.doi.org/10.26193/U3ZZGY	
	Two years time					Yes - eg Dataverse record http://dx.doi.org/10.26193/U3ZZGY and discoverable via Prosecution Project website (https://prosecutionproject.griffith.edu.au/) and perhaps Trove (trove.nla.gov.au) as well as repositories of external data providers
	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	No publicly available metadata record				
	End of project					*Datasets fully accessible to public users via ADA Prosecution Project Dataverse site (https://dataverse.ada.edu.au/dataverse/australian_historical_criminal_justice_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	Limited access to public users through a searchable web database	Specific requests for datasets through project leader, actioning each request by manually collating and sending files			
	End of project					*Standard web service established via ADA - https://dataverse.ada.edu.au/dataverse/australian_historical_criminal_justice_data*
	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		Yes, at university level			
	End of project					Yes as per ADA policy - eg datasets and their metadata record once published may be 'deaccessioned' but not deleted
	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			data available as csv files		
	End of project			data available as csv files		
	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		data available as csv files			
	End of project					
	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			Columns are labelled with attribute names - some data columns defined by format standards (eg dates, and url identifiers for links to external data sources)		
	End of project			Columns are labelled with attribute names - some data columns defined by format standards (eg dates, and url identifiers for links to external data sources)		
	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	No links to other metadata				

	End of project		metadata record includes URI links to related sources of data (eg Trove, archive collections)			
	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					
	End of project				????	
	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		Provenance recorded in information on Prosecution Project website and in metadata for database records			
	End of project			Metadata record includes URI link to Prosecution Project website with information of project methodologies and sources of data; metadata txt file also includes data collection information		
	Two years time					