

FAIR self assessment for project: ARDC clinical and research dataset discovery in the MACH network

Completed for the grant funding period to 08/10/2019

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	Identifiers for datasets were unknown at the start of the project. The project had no visibility of datasets prior to the project start				
	End of project		Most datasets described have been given identifiers by their custodians.	Some datasets have been assigned a URL	Very few datasets have been given a unique, citable and persistent identifier.	
	Two years time		It is not possible at this time to determine	It is not possible at this time to determine		
Q2	Is the identifier included in all metadata records or metadata files describing the data?	No	Yes			
A2	Start of project	The uses for Identifiers for datasets were unknown at the start of the project. The project had no visibility of datasets prior to the project start				
	End of project		This project aimed to discover metadata about datasets. In most cases an identifier for the dataset was included in the metadata record. It was not possible to discover where or if this identifier was used elsewhere in the MACH institution in the description of the dataset.			
	Two years time		It is not possible at this time to determine			
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	It was not known by the project if a metadata record existed already for the dataset				
	End of project		This project aimed to discover datasets and create a metadata record or catalogue entry for the dataset. All datasets recorded during the project each were given a brief title and description by the data custodian	All datasets recorded had multiple other fields recorded albeit briefly, including key words, collection status and number of participants.		
	Two years time				Some sections of the survey were structured and potentially machine readable, for example the 'study type' or 'availability for research'. Other sections were free entry. Further work on the survey and control of responses is required to make the results of the survey machine readable.	

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) Mutually exclusive: Data can be in multiple locations and (all) discoverable via several registries. Sometimes it's good to be in multiple places, as long as it's done right. This answer is also not in contrast to the other options.
A4	Start of project	The project was a dataset discovery project. The MACH members who participated had little to no visibility of where data custodians recorded the existence of their datasets and what metadata records were kept.				
	End of project		The results of the survey responses were a set of metadata records for datasets hosted in the MACH network. The metadata records collected are hosted on the University of Melbourne's REDCap survey instance			
	Two years time			We aim in two years time to have at minimum a clinical and health research data catalogue made available to members of the MACH network.	We aim to make this metadata collection available via ARDC Research Data Ausatralia and to promote significant collections to obtain their own listings	We aim to make this metadata collection available via ARDC Research Data Ausatralia and to promote significant collections to obtain their own listings
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	The current accessibility of datasets was not known at the start of the project				
	End of project		Due to time constraints and the method of engagement with MACH institutions, it was not possible to discover the current accessibility of the datasets recorded.			
	Two years time					It is hoped that all dataset metadata records will contain information on how to access the dataset including how to obtain the required permissions.
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	The current accessibility of datasets was not known at the start of the project		No	No	
	End of project		Due to time constraints and the method of engagement with MACH institutions, it was not possible to discover the technical requirements for access to the datasets recorded.	The metadata collection will be available via the REDCap survey databases in the first instance	No	

Q7	Two years time			No	No	We aim to make this metadata collection available via ARDC Research Data Ausatralia and to promote significant collections to obtain their own listings
	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		A number of MACH members have not yet been approached about maintaining the survey and metadata records as the first pass of the survey has either not started or is not completed. The response from the institution to the survey will influence the desire to maintain the discovery.			
	End of project			Three institutions, so far, have agreed to continue with the survey and maintain the metadata records of discovered datasets.		
	Two years time				It is hoped that other MACH members will agree to maintain the dataset metadata records.	
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					
	End of project	Due to time constraints and the method of engagement with MACH institutions, it was not possible to discover the inter-operability of the datasets recorded				
Q9	Two years time	The project is not able to help enforce interoperability in any contributing dataset although having the datasets advertised will we hope increase the research use of such data that may lead to better file format conformance				
	Are the data is 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	The machine readability of the datasets was unknown at the start of the project				
	End of project	Due to time constraints and the method of engagement with MACH institutions, it was not possible to discover the machine readability of the datasets recorded				
	Two years time	The project is not able to help enforce interoperability in any contributing dataset although having the datasets advertised will we hope increase the research use of such data that may lead to better conformance with regard to machine readability				

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	Data element descriptions provided locally for the datasets were unknown at the start of the project			
		End of project	Due to time constraints and the method of engagement with MACH institutions, it was not possible to discover data elements nor data element descriptions for the datasets recorded			
Q11	Two years time	The project is not able to help enforce interoperability in any contributing dataset although having the datasets advertised will we hope increase the research use of such data that may lead to better conformance to ontological standards				
	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	Links and other relationships to related resources were unknown at the start of the project.			
Q12	End of project		Due to time constraints and the method of engagement with MACH institutions, it was not possible to request information about other resources, however some responses included information on collaborating institutions.			
	Two years time			It is not possible at this time to determine		
	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	Licensing and usage rights applying to the datasets was unknown as the dataset was unknown			
		End of project		Some responses included information on licensing but this information was not specifically sought due to time constraints. What was captured though was whether the dataset was available for researchers to use.		
Q13	Two years time			It is not possible at this time to determine	It is not possible at this time to determine	It is not possible at this time to determine
	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 information was unknown as the dataset was unknown			
	End of project		Many of the descriptions collected in the meta data record included information on the purpose of the dataset, its content and collection status. In most cases there was sufficient information for a researcher to determine if the dataset could be of interest. No technical information was collected on processing workflow.		
	Two years time			It is likely that with persistent surveying and potentially face to face interview that comprehensive information could be obtained about provenance.	Achieving machine readability is dependent on the hosting, structure and format of the dataset. It is not possible at this time to determine if machine readability of the provenance of the record is achievable.