

# FAIR self assessment for project: A brain imaging database of rare and endangered Australian mammals

Completed 21/09/2019

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	Previously data was kept on the instrument(s) and/or in flat folder structures without explicit identifier. Only personell in the facility or staff related to the specifi research project was able to find specific datasets.				
	End of project			Data has bee tranferred into a publicly accessible and searchable image repository. The database can be accessed via the following web-link: <a href="https://aussiebrain.unsw.edu.au">https://aussiebrain.unsw.edu.au</a> . Some datasets are still pending and more datasets are currently being acquired, which will go into the database in the future		
	Two years time				We envisage to assign digital object identifiers to the data at the specimen level. Individual images will not have an individual doi.	
Q2	Is the identifier included in all metadata records or metadata files describing the data?	No	Yes			
A2	Start of project	No, metadata were generally not included with the datasets				
	End of project		Each data record has an identifier within the XNAT database			
	Two years time				We aim to include global identifiers on the specimen level, i.e. each specimen will have a doi. All imaging sessions within the specimen will be subsumed under the same doi	
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-	
A3	Start of project	No metadata record, but image related metadata included in proprietary imagine data				

	End of project			Data is stored in an instance of the XNAT imaging data platform. Metadata is described on multiple levels. On the specimen level metadata is included in datatypes of the searchable XNAT database. On the image level metadata is included in the dicom headers of the images. Most but not all metadata is searchable.		
	Two years time				We are currently working on a comprehensive XML-schema based metadata description for museum specimens. We will include these descriptors in the database to make more metadata elements available for dat-mining on the datasets.	
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	Data was kept on imaging instruments or in flat and largely unstructured data directories on different servers. There was no explicit metadata record.				
	End of project			Data is now kept in a locally managed, but publicly accessible and searchable image repository. The database can be accessed via the following web-link: <a href="https://dev.xnat.unsw.edu.au">https://dev.xnat.unsw.edu.au</a>		
	Two years time				We envisage that the data will eventually be moved to a different database platform to increase searching capabilities and linking to other dataabases, which is currently limited by restrictions posed by the XNAT platform. We also envisage that the data will be hosted in cloud storage rather than locally.	
<b>ACCESSIBLE</b>						
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 data was generally not publicly accessible. No consistent metadata record existed. If researchers were aware that a specific datasets exist and they had the opportunity to approach the facility manager or researchers that were associated with data acquisition to request access to the data				

Q6	End of project					The imaging datasets are fully accessible without restrictions. Interested parties can apply for a database account via the web-site. They can then search and download data with 'collaborator premissions'. When data is re-used the data origin and data owner has to be referenced and cited.
	Two years time					The imaging datasets are fully accessible without restrictions. Interested parties can apply for a database account via the web-site and search and download data. When data is re-used the data origin and data owner has to be referenced and cited.
	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		Researchers had to approach the facility manager to find out what data is available. Data manager had to action each request by manually converting, collating and sending files.		
A7	End of project				Data is now kept in a locally managed, but publicly accessible and searchable image repository, which is based on the XNAT platform. XNAT provides a non-standard API called REST.	
	Two years time				Data will likely move to an online image repository, but we envisage that a non-standard web-service may prevail.	
	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	NA				
	End of project		Unsure of UNSW repository policy. Our aim is to implment a permanent digital record of the scanned specimens. So deletion of the data product is not planned.			
	Two years time			The repository is planned as a persistent and continuously extending repository for public access. We aim to eventually have a persistent metadata record.		
<b>INTEROPERABLE</b>						
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	The majority of datasets was kept in proprietary image formats of the instrument manufacaturer(s)				

	End of project			Metadata are kept in a database. Part of the metadata are described by XML datatypes in the platform. A general XML-schema for metadata has been outlined to reproducibly document newly generated imaging data. The definition is not comprehensive, yet, so some metadata information that may become important for specific data-mining purposes might still be missing. We are currently working on completeing the description for general museum specimens.		
	Two years time			We aim to contuouously extend and refine the XML meatdata definition to eventually define a standard that allows for comprehensive data-mining in broad variety of imaging data from specimen collections.		
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	The data did not have a consistent metadata record. The image data itself was machine readable with appropriate proprietary or custom software.				
	End of project		Data is stored in an instance of a widely used imaging data platform and metadata is described at multiple levels using an XML-schema based metadata structure. On the specimen level metadata is included in variables of the searchable XNAT database. On the image level metadata is included in the dicom headers of the images.			
	Two years time		Data is stored in an instance of a widely used imaging data platform and metadata is described at multiple levels using an XML-schema based metadata structure. On the specimen level metadata is included in variables of the searchable XNAT database. On the image level metadata is included in the dicom headers of the images.			
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 <b>persistent</b> 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	Metadata was not described consistently and without standard ontologies. Some metadata were free text in lab books, some were tags in images, some information was kept in specimen documentation at within the (museum) collections				
	End of project		The metadata standrad used in this project is described in the 'datatype' manual associated with this project. Some standard ontologies (e.g. for biological taxonomy) were adopted, but there is room for improving standardization and consistency of nomenclature in the future.			
	Two years time				We aim to increase the use of standardized vocabularies and also retrieve specific data from other databases to prevent inconsistencies from archiving of the same information in multiple places.	
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 data were stored with the imaging data				
	End of project		References to other specimen related data sources are included in the database e.g. by referencing the institution, collection, catalog number, etc. of the archives storing the physical counterpart of the digitized imaging data. No direct link to the related museum databases is established, yet. Other resources related to protocols (e.g. for specimen preparation for imaging or data processing) can be (and are partially) stored in the XNAT imaging database established in this project.			
	Two years time			We aim to establish better data links to the specimen related descriptors in museum databases. Specific dicussions are currently ongoing to directly link the EMU specimen database of the Australian Museum to our data. Our XML schema will be extended to include information from other sources, including geneticand biodiversity information from related adatabases (Atlas of Linving Australia, ozMammals).		
<b>REUSABLE</b>						

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	Researchers or public entities re-using the data were expected to clearly name, reference and cite the data origin and data owner. Licences or copyrights were not applied.				
	End of project	Researchers or public entities re-using the data are expected to clearly name, reference and cite the data origin and data owner. Licences or copyrights are not applied.				
	Two years time		Researchers or public entities re-using the data are expected to clearly name, reference and cite the data origin and data owner. Ownership and copyright information will be included in the XML-schema. Currently there is no firm plan to apply specific licenses, but workshops are currently being planned about 3D image data archiving, which will also include discussions around the requirements in this field.			
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		Some provenance information was recorded, but not with the data but in several different places ( lab books of individual researchers, local folders, etc.).			
	End of project		Metadata about processing workflows is now partially included with the data. Some of this information is now in machine readable form (e.g. specific details about acquisition protocols in dicom, etc.). Provenance information is still incomplete and will be extended in the future.			
	Two years time				We aim to establish a comprehensive description of provenance information in our xml-schema based metadata definition. Some parts of this are already implemented, but protocol information is sometimes stored as free text within a single xml tag. To make protocol information truly machine readable the free text information must be broken up into standardized tags and described in the metadata definition.	