FAIR self assessment for project: TD16 - Standardisation of protocols for collecting linked images and electrophysiological data from in vitro and in vivo studies on neural cells and tissues

Completed 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	Yes	No (User defined)	https://physionet.org/about/database/	None	Notes: Data storage is typically in local repositories, dependent on individual researchers (UNSW example: https://research.unsw.edu.au/research-data-management-unsw). Limited image data and slightly more extensive electrophysiology data can be found at https://physionet.org/
	End of project	Yes	Yes, non-standard local identifiers used describing datasets in some databases. For instance "RRID:SCR_003129" https://neuinfo.org/Resources/search?q=Electrophysiology&l=Electrophysiology	yes, some database provide a URL link for specific datasets. For example: http://crcns.org/data-sets/ac/ac-1	None	Notes: In some databases such as Mendeley Data or the Collaborative Research in Computational Neuroscience there are links to DOI. A review of existing datasets in open platforms has been completed. These have been updated as shown below and links to each database have been included in the report. Neuroelectro.org CRCNS - Collaborative Research in Computational Neuroscience ABA Mouse Brain: CellType EphysData CONP OpenNeuro Physionet.org Neurosynth.org Neuromorph.org Hippocampome.org Neurovault Neurodatabase.org Neuroscience Information Framework Mendeley Data Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC)
	Two years time	Yes	yes	yes	Local identifiers linked to unique standardised identifiers.	Notes: Standardising all data formats from different equipment has been regarded as impractical but it is expected that through collaborative efforts standard identification of datasets can be maximised
Q2	Is the identifier included in all metadata records or metadata files describing the data?					
A2	Start of project	No	No			

	End of project	No	*No*		Example of the DOI linked to a dataset from Mendeley data: Dataset: https://data.mendeley.com/datasets/hw5p35hc9g/1 DOI: http://dx.doi.org/10.17632/hw5p35hc9g.1	Notes: DOIs are often included at the database webpage where the link to access/download the dataset can be found. An example is included at the globally unique column. But most metadata describes only the experimental parameters and data acquisition conditions.
	Two years time	yes	yes		DOI linked to local identifier and to the metadata	Notes: it is expected that for future datasets international identification standards such as those proposed by Neurodata without borders, nature publishing group (ISI standard) or CellPress (STAR Method) can be implemented
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		Notes: Metadata is required in local repositories and data management plans should describe the metadata (UNSW example: https://resdata.unsw.edu.au/pages/help-rdmp.faces
	End of project	No	Yes, all datasets include a brief title and description	Yes in most databases	Yes in some datasets. For example; http://dx.doi.org/10.6080/K00C4T06	Datasets from some databases such as CRCNS contain a brief description of the aims of the study and interpretation of the results. Also, some datasets contain programming codes that are required to retrieve and plot the data from structures contained in Hierarchical Data Format (HDF files). Such datasets sometimes are accompanied by a "readme" text file that describes how to run the codes. but this is not the case for most datasets and require programming skills and understanding on how HDF files are structured.
	Two years time	No	yes	yes	yes	Ongoing efforts are in place to standardised current datasets. However, the metadata varies among datasets and the information provided relies o the author's input.
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		Yes			
	End of project	Research reports estimate that at least 50% o research data is not shared in any registry or repository. Source: https://doi.org/10.1038/nn.3838	Institutions such as UNSW provide a research data management platform for archiving purposes: https://resdata.unsw.edu.au/pages/authenticate.faces	Yes most databases. For instance; http://crcns.org/data-sets/ear/ear-1/about-ear-1	Some search engines that provide access to databases such as Mendeley Data, Harvard Dataverse, and the open science network are generalist public repositories.	*most databases included in the report were found through google search or google scholar*
	Two years time	the amount of data not described in any registry or repository is expected to remain the same unless research journals implement a change in the policy requiring data along with the publication and institutions implement policies to ensure data archiving practices	it is envisioned to promote the use if internal repositories across institutions	Most databases are expected to remain domain-specific	Search engines such as Mendeley data, NIF and NITRC are expected to remain as tools for accessing and distributing shared data	Access to databases and search engines is expected to be discoverable through most common search engines
	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			Yes		
	End of project		The background research did not show a dataset granting access only to metadata	Yes, processed data is available but it is requested to contact the author to access raw data		
	Two years time		To protect data ownership and IP - datasets can be separated through data management and present only metadata		Yes in some databases	yes. Additional notes are added to metadata requesting acknowledgments, citation to linked papers
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	Yes	*Yes*			
	End of project	Few datasets are restricted	Yes, processed data is available but it is requested to contact the author to access raw data. For instance; https://figshare.com/articles/Paradiplozoon_exposure/3474476/1	*yes*	yes, databases such as Neurodata without borders provides with openAPI that allows to communicate with programming interfaces such as MatLab to access the data sets. For instance; https://neurodatawithoutborders.github.io/matnwb/doc/index.html https://neurodatawithoutborders.github.io/matnwb/#tutorials	yes, databases such as NIF or Mendeley Data
	Two years time	Some datasets are expected to be restricted due to embargo or Intellectual properties issues	Requesting to contact the author to access raw data is a condition expected to remain	yes	yes	*yes*
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		
	End of project			*yes*		
	Two years time			yes		
	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		*yes*		Notes: Imaging and electrophysiology data is typically collected in formats that are "machine-dependent"
	End of project	datasets containing data in proprietary format are common; for instance, trace recordings from Multichannel systems or images from Zeiss, Olympus or Leica microscopes	file types such as MAT files or HDF files are common to storage large datasets. These can be opened with software such as MatLab, Python, Julia, R, Fortran and C++	yes		Notes: most data available in open widely-used formats is processed data; in som cases it includes raw data for instance, https://datadryad.org/stash/dataset/doi:10.5061/dryad.2rb57
	Two years time	is expected to reduce the use of proprietary format to share data through standardisation of data	file types such as HDF are expected to be key to maintain a structure to share large dataset	yes		
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				
	End of project	yes, some	yes , most			
	Two years time	yes, some	yes , most			

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*			
	End of project	yes, some	yes	yes	yes, at some extent, identifiers refer to the research article and some datasets provide a written explanation to open the files and a brief description of the study. One limitation is that vocabularies are not subject to a standard and may change depending on the user's input or on the proprietary software used to create the file	Mostly no, common vocabularies are in place but there are no standards. Most datasets require programming codes that often require human input to retrieve and analyse the data from the files.
	Two years time	it is expected to be minimised with through implementation of standards and data sharing policies	Expected to be minimal. Standardising all data formats and file types is impractical. Thus, is expected that some datasets with no standards in the description will persist	Expected to be reduced significantly	yes, although the use of human input for retrieval and use of data may still be required	yes, although the use of human input for retrieval and use of data may still be required
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	Yes				
	End of project	most datasets	Some datasets refer to other data, indicating in their description statements like "A collection of behavioural testing data is also available. (SCR_003129)". For instance, http: //memory.psych.upenn. edu/Electrophysiological_Data	some datasets		
	Two years time	Expected to be minimised	this is expected to be implemented and promoted through collaborative networks in the pilot project	this is expected to be implemented and promoted through collaborative networks in the pilot project		
	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				
	End of project	no	no	no (datasets from public databases contain a link (URL) that describes the license deed.)	yes	yes
	Two years time	Expected to be minimised	Expected to be minimised	Expected to be minimised	yes	yes
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			Notes: Local data archives require Research Data Management Plans completed by individual researchers
	End of project	yes	yes	yes	yes, but not common	Notes: some datasets include documents that describe the context of the data, the aims of the study ad include a "readme" file that states step by step how to access the files and help interpreting metadata. For example: https://gin.g-node. org/doi/multielectrode_grasp
	Two years time	Expected to be minimised	Expected to be minimised	yes	Expected to increase	