

FAIR self assessment for project: TD42 - A pilot database of Industrial Internet of Things Networks for cyber security applications

Completed 18/09/2019

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			Metadata on UNSW Canberra website https://www.unsw.adfa.edu.au/unsw-canberra-		
	End of project			cloudstor: https://cloudstor.aarnet.edu.au/plus/s/ds5zW91vdgjEj9i	ResData metadata record with dataset DOI has been published at https://researchdata.andcs.org.au/new-generations-internet-toniot-datasets , and https://doi.org/10.26190/5d7ac9bfe8487 , respectively	
	Two years time				An intention of creating DOI per each future dataset resolving to RDA record harvested from UNSW ResData	
Q2	Is the identifier included in all metadata records or metadata files describing the data?	No	Yes			
A2	Start of project			url to dataset from website page		
	End of project				ResData metadata record with dataset DOI has been published at https://researchdata.andcs.org.au/new-generations-internet-toniot-datasets , and https://doi.org/10.26190/5d7ac9bfe8487 , respectively	
	Two years time				An intention of creating DOI per each future dataset included within dataset metadata record	
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, within the website, but not metadata records within repository		
	End of project				ResData metadata record with datasets DOI has been published at https://researchdata.andcs.org.au/new-generations-internet-toniot-datasets , and https://doi.org/10.26190/5d7ac9bfe8487 , respectively	
	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	Published on UNSW Canberra website				

	End of project		Metadata record in UNSW ResData			Comprehensive metadata record is published in UNSW ResData and harvested into RDA at https://researchdata.and.s.org.au/new-generations-internet-toniot-datasets with https://doi.org/10.26190/5d7ac9bfe8487 that resolves to this page in RDA
	Two years time		Metadata record in UNSW ResData			Dataset metadata record per each future dataset, registered in UNSW ResData and harvested to RDA
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	There was metadata for UNSW-Nb15 but no metadata record for TON IoT				
	End of project					All publicly accessible in persistent storage - metadata in ResData and data in cloudstor
	Two years time					As above, for all future datasets
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					Some datasets: HTTP. Any tool required for understanding or reading data is made available for use - open source only
	End of project				* Raw datasets were stored at log files (readable by txt editor) and other formats (readable by open source tools as described in the website) * Processed datasets were logged in a CSV format * Metadata and description of datasets were stored in pdf files	All datasets: HTTP. Any tool required for understanding or reading data is made available for use - open source only
	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				Some persistent in local school website https://www.unsw.adfa.edu.au/uns-w-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/ and some in cloudstor	
	End of project					RESData and data owner will do this in accordance with UNSW policy.
	Two years time					UNSW will maintain the persistence of metadata record at Resdata
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, and where a tool is required, it is provided and is open source		
	End of project			As above		
	Two years time			As above		

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, as csv or tool available for reading			
	End of project		As above			
	Two years time		formats if required by users			
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	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	Published vocabularies using persistent resolvable identifiers linking to explanations are used, so that the data can
A10	Start of project			Recognised standards are used, but no resolvable URIs to published terms have been used (see links to standard managers)		
	End of project				data elements/features were described and links of open sources tools were provided in 'ReadMe.PDF' at the cloudstor https://cloudstor.aarnet.edu.au/plus/s/ds5zW91vdgjEj9i	
	Two years time					Open source tools and description of their terms used for generating datasets have been published at at the cloudstor https://cloudstor.aarnet.edu.au/plus/s/ds5zW91vdgjEj9i
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		Existing metadata contains link to related resources.			
	End of project			ResData record in RDA at https://researchdata.andis.org.au/new-generations-internet-toniot-datasets contains link to grant "A pilot database of industrial internet of things networks for cyber security applications" at https://researchdata.andis.org.au/pilot-database-industrial-security-		
	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	Rights statement on some existing metadata				
	End of project				Standard Licence applied without URL is indicated in resdata record https://researchdata.andis.org.au/unswnb15-dataset "GPL" - determined as 'open' in RDA	
	Two years time					All future published datasets will have licence applied when metadata constructed for ResData

Q13	How much provenance information has been captured to facilitate data reuse? i.e. project objectives, data generation/collection (including from external sources)	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)	
A13	Start of project			Comprehensively recorded in a text format (i.e. TXT or PDF)		
	End of project				Comprehensively recorded in a text format (i.e. TXT, CSV or PDF)	
	Two years time				All generated datasets will have machine-readable provenance information, for assisting reuse and regeneration	