

Online Supplemental Table 3. Framework as at May 2020

Category	Dimension	Definition	Bronze	Silver	Gold	Platinum
Data Documentation	Documentation Quality	Weighted Data Documentation Score	< 66% of metadata specification fields meet format and content requirements	< 76% of metadata specification fields meet format and content requirements	< 86% of metadata specification fields meet format and content requirements	< 96% of metadata specification fields meet format and content requirements
	Availability of additional documentation and support	Available dataset documentation in addition to the data dictionary	Past journal articles demonstrate that knowledge of the data exists	Comprehensive ReadMe describing extracting and use of data, Dataset FAQs available, Visual data model provided	Dataset publication was supported with a journal article explaining the dataset in detail, or dataset training materials	Support personnel available to answer any questions
	Data Model	Availability of clear, documented data model	Known and accepted data model but some key field uncoded or free text	Key fields codified using a local standard	Key fields codified using a national or international standard	Data Model conforms to a national standard and key fields codified using a national / international standard
	Data Dictionary	Provided documented data dictionary and terminologies	Data definitions available	Definitions compiled into local data dictionary which is available online	Dictionary relates to national definitions	Dictionary is based on international standards and includes mapping(?)
	Provenance	Clear description of source and history of the dataset, providing a "transparent data pipeline"	Source of the dataset is documented	Source of the dataset and any transformations, rules and exclusions documented	All original data items listed, all transformations, rules and exclusion listed and impact of these	Ability to view earlier versions, including "raw" or "source" dataset, and review the impact of each stage/step
Technical Quality	Data Quality Management Process	The level of maturity of the data quality management processes	A documented data management plan covering collection, auditing, and management is available for the dataset	Evidence that the data management plan has been implemented is available	Demonstrated compliance with the data management plan	Externally verified compliance with the data management plan
	DAMA Data Quality Element - Completeness	The proportion of stored data against the potential of "100% complete"	<i>Indicators are under development by another project</i>			
	DAMA Data Quality Element - Uniqueness	No thing will be recorded more than once based upon how that thing is identified.				

	DAMA Data Quality Element - Timeliness	The degree to which data represent reality from the required point in time.				
	DAMA Data Quality Element - Validity	Data are valid if it conforms to the syntax (format, type, range) of its definition.				
	DAMA Data Quality Element - Accuracy	The degree to which data correctly describes the "real world" object or event being described.				
	DAMA Data Quality Element - Consistency	The absence of difference, when comparing two or more representations of a thing against a definition.				
Coverage	Pathway coverage	Representation of Multi disciplinary healthcare data	Contains data from a single speciality or area	Contains data from multiple specialties or services within a single tier of care	Contains multimodal data or data that is linked across two tiers (e.g. primary and secondary care)	Contains data across the whole pathway of care
	Length of follow up	Average timeframe in which a patient appears in a dataset (follow up period)	Between 1 - 6 months	Between 6 - 12 months	Between 1 - 10 years	More than 10 years
Access & Provision	Allowable uses	Allowable dataset usages as per the licencing agreement		Non-consented, aggregate data for specific academic uses	Aggregate data, for specific commercial use	Fully consented for commercial uses
	Research environment	Access, tooling and environment (once approved)	Requested analysis can be undertaken by internal teams and provided back in anonymised format to data requestors	The dataset can be used in a trusted research environment	The dataset can be used in a trusted research environment, but other data and tools can be brought in as required	The dataset can be used in requesting companies environment
	Format	The technical presentation of the data format (e.g. DICOM images vs PNG)	Format is explicitly defined	Format is explicitly defined and in widely readable (non-proprietary) format		Format is explicitly defined, in open and referenced format

	Access and approvals	IG Process for gaining approval to access data	Contact details for the relevant authority to request data access	Indicative timeframes for processing data access applications detailed	Details of application process and requirements	Detailed Information Governance Process described, outlining requirements for applications, basis of decisions, and anticipated timeframes
	Timeliness	Data time lag + Average data access request timeframe	More than 12 months	Less than 12 months	Less than 6 months	Less than 3 months
Value & Interest	Access Request	Number of data access inquires received	Some interest. Access inquires/requests >1	Some interest. Access inquires/requests >5	Some interest. Access inquires/requests >20	Some interest. Access inquires/requests >50
		Provenance of data access inquires received	Interest and inquires from colleagues and known associates	Interest and inquires from other national organisations	Interests and inquires from national commercial organisations	Interests and inquires from international bodies
	Current Usage	Number of active projects and tools using the dataset	> 1 active projects or tools using the dataset	> 5 active projects or tools using the dataset	> 10 active projects or tools using the dataset	> 30 active projects or tools using the dataset
	Linkages	Ability to link with other datasets	Identifiers to demonstrate ability to link to other datasets	Available linkages outlined and/or List of datasets previously successfully linked provided	List of restrictions on the type of linkages detailed. List of previously successful dataset linkages performed, with navigable links to linked datasets via at DOI/URL	Existing linkage with reusable or downstream approvals
	Data Enrichments	Data sources enriched with annotations, image labels, phenomes, derivations, NLP derived data labels	The data include additional derived fields, or enriched data.	The data include additional derived fields, or enriched data used by other available data sources.	The derived fields or enriched data were generated from, or used by, a peer reviewed algorithm.	The data includes derived fields or enriched data from a national report.