

Category	Dimension	Definition	Bronze	Silver	Gold	Platinum
Data Documentation	Documentation Completeness	Proportion of metadata (as in the current HDR UK metadata specification) which is available in the expected format	61%-70%	71%-80%	81%-90%	91%-100%
	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	As Silver, plus dataset publication was supported with a journal article explaining the dataset in detail, or dataset training materials	As Gold, plus support personnel available to answer questions
	Data Model	Availability of clear, documented data model	Known and accepted data model but some keyfield un-coded 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 versions before any transformations have been applied data (in line with deidentification and IG approval) and review the impact of each stage of data cleaning
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		Externally verified compliance with the data management plan, e.g. by ISO, CQC, ICO or other body
	Data Management Association (DAMA) Quality Dimensions	Technical data quality dimensions: Completeness, Uniqueness, Accuracy, Validity, Timeliness and Consistency	These elements will be calculated with data profiling tools, and the category breakdown evaluated following further data collection			
Coverage	Pathway coverage	Representation of multi-disciplinary healthcare data	Contains data from a single specialty 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 more than two tiers
	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 licensing agreement, following ethical and IG approval	Available for specific academic research uses only	Available for academic and non-profit (e.g. charity, public sector) uses only	Available for limited commercial uses (e.g. relating to a specific domain), in addition to academic and other non-commercial uses	Available for wider commercial uses (in line with ethical and IG approval), and addition to academic and other non-commercial uses
	Time Lag	Lag between the data being collected and added to the dataset	Approximately 1 year	Approximately 1 month	Approximately 1 week	Effectively real-time data
	Timeliness	Average data access request timeframe	Less than 6 months	Less than 3 months	Less than 1 month	Less than 2 weeks
Value & Interest	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.

