RDMT PUBLIC REVIEW 2022

The CODATA RDM Terminology (previously known as the CASRAI RDM Glossary) is open for public review from 15 July until **30 September 2022**. The RDMT has been reviewed and discussed by an international working group of experts as part of its annual review process, and we are now pleased to open the result to community feedback. Please help us make the terminology as useful as possible!

GUIDANCE <= Read this first!

1. First, please read and digest the terminology's Scope Statement.
2. Then, **insert a comment** on any term to which you have feedback, whether it is to **the definition** or to **the term** itself. **Not sure how?** You are also welcome to respond to another comment.
3. **Do not attempt to edit the existing document directly. Such edits will be removed!**
4. **Add any suggestions for additional terms** in the suggestions form here. Suggestions that are in scope, according to the Scope Statement, will be considered in the next annual review. Check your suggestions are not already listed in the Deleted terms list first, though. **Suggestions for new terms will only be considered where a draft definition is also supplied.**
5. Minor suggestions for edits, made via comments, will be applied in this round. Major changes will be considered by the next review cycle.
6. Please note: EDIT = the term was edited in the 2022 review (you can access the prior definition by clicking the hyperlinked term itself); ADD = the term was added in the 2022 review (so there is no prior definition); ACCEPT = the prior definition was reviewed and accepted unchanged; NR = the term was not reviewed in the 2022 review; 'data' = plural; we are using British English spellings. Direct quotes should be followed by the source.
7. There are currently 345 terms. You can feed back on as many or few as you wish, but we don't expect anyone to do them all! Feedback is particularly welcome in response to terms marked NR. May we suggest you pick a term that does not already have comments, and/or those further down the alphabetical list.
8. Definitions are not currently consistently written in line with ISO 704 guidance, but suggestions for such edits are very welcome.
9. Contact the RDMT working group convenor with any questions, and/or if you prefer to provide feedback by email: laura[at]codata.org

Thank you for your input!

TERM	2022 DEFINITION	EXPERT DECISION 2021-22 (ACCEPT / EDIT / ADD / NO REVIEW THIS CYCLE: NR)
Access	Continued availability and ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes the digital material was created and/or acquired for. Users who have access can retrieve, understand, manipulate, and store copies.	Edit
Accessible data	One of the four FAIR principles. Data accessibility is defined by users' ability to gain access to or retrieve data once it has been discovered. This includes data instances where access to the data is limited, such as when user requests need to be authenticated and authorised. Source: https://www.go-fair.org/fair-principles/. RELATED TERMS. FAIR data, FAIR principles.	ADD
Access controls	Access controls define the access relationships between the following metadata: data object name, a user name (or user group, or user role), and access permission(s). The information can be stored as metadata information associated with each data object. The information can be generated dynamically by applying the access controls of the collection that organises the data objects (if a collection sticky bit is turned on). RELATED TERM. Sticky bit.	EDIT
Access workflow	Type of access entity that contains the services and functions which make the data object holdings and their information content and related services visible to data consumers.	NR
Active data	Research data actively accessible and modifiable during the active phase of the research project.	EDIT
Administrative data	Information collected primarily for administrative, and not research purposes. It includes profiles and curriculum vitae of researchers, the scope and impact of research projects, funding, citations, and information about research outcomes. This type of data is collected by government departments and other organisations for the purposes of registration, transaction and record keeping, usually during the delivery of a service. These data are also recognized as having research value.	EDIT
Administrative metadata	Used to manage administrative aspects of the digital objects such as intellectual property rights and acquisition. Administrative metadata also documents information concerning the creation, alteration, and version control of the metadata itself. This is sometimes known as meta-metadata. SYNONYM. Meta-metadata	ACCEPT
Aggregated data	High level data that are expressed in a summary form (e.g. summary statistics).	EDIT
Aggregation	The compiling of elements, often from different sources. Types of aggregation differ by the nature of the processes by which elements are brought together and the intention for aggregating. Aggregations differ in the nature of relations between the constituent parts.	EDIT
Analogue data	Data created and presented in the form of physical materials, e.g. written notes; sketches; maquettes; specimens. RELATED TERM. Analogue materials.	EDIT

Analogue materials	Non-digital materials that have a physical presence (e.g. written and printed materials, maquettes, specimens). RELATED TERM : Analogue data.	EDIT
Analytics	The discovery of meaningful patterns in data through systematic analysis.	EDIT
Anonymity	Ethical principle that is applied in research to maintain the privacy of research participants by keeping their identity unknown through irreversible processes. RELATED TERM: Masking.	EDIT
Archive	(Noun). A curated collection or repository containing physical or digital static records, objects, metadata and data deemed suitable for permanent retention, set up and managed to established standards, (e.g. ISAD(G) https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition or CoreTrustSeal https://www.coretrustseal.org/) that ensure long term integrity, security, authenticity and accessibility of the records, objects, metadata and data. RELATED TERM. Archiving, Archivist, Data Archive.	EDIT
Archiving	Curation activity that ensures that records, objects, metadata and data are properly selected, stored, and can be accessed, and for which logical and physical integrity are maintained over time, including security and authenticity. RELATED TERM. Archive, Archivist.	EDIT
Archivist	Person responsible for appraising, acquiring, arranging, describing, preserving, and providing access to records of enduring value, according to the principles of provenance, original order, and collective control to protect the materials' authenticity and context. Such persons may also have responsibility for management and oversight of an archival repository or of records of enduring value. There is international variance in how this term is used; some archivists primarily interact with inactive records, while others would have responsibility for both inactive and active records.	ACCEPT
At-risk data	Data that are at risk of being lost. At-risk data include data that are not easily accessible, have been dispersed, have been separated from the research output object, are stored on a medium that is obsolete or at risk of deterioration, data that were not recorded in digital form, and digital data that are available but are not useable because they have been detached from supporting data, metadata, and information needed to use and interpret them intelligently.	Accept
Audit	In the context of RDM, an audit involves an evaluation of an organisation, system, group, project or product with respect to its data and processes around this, often in accordance with a standard, guide, or framework used to structure the work. This can involve assessing, describing, and classifying any data held. An audit can be carried out internally by those who have access to the data or participate in related processes regularly, or by an independent, external actor.	EDIT
Authenticity metadata	Type of metadata that conveys information needed to link a data object to its original source. RELATED TERM. Provenance Metadata.	EDIT
Big data	1. An evolving term that describes any voluminous amount of structured, semi-structured and/or unstructured data that have the potential to be mined for information. 2. Extensive datasets/collections/linked data primarily characterised by big volume, extensive variety, high velocity (creation and use), and/or variability that together require a scalable architecture for efficient data storage, manipulation, and analysis. The definition can vary by sector, depending on what kind of software tools are commonly available and what sizes of datasets are common in a particular discipline. With those caveats, big data in many sectors today will range from a few dozen terabytes to multiple petabytes.	EDIT
Bit Sequence	A representation of digital content in an assembly of the fundamental unit of digital bits.	NR

Bit Stream	Unstructured sequence of bits that is identified as a unit (e.g., bits in a communication transmission). It may be stored as a unit or may exist as a pattern and be generated. A digital object may be represented as a bit stream of finite length that encodes its informational content.	NR
Born digital	Digital materials that are not intended to have an analogue equivalent, either as the originating source or as a result of conversion to analogue form. This term is used to differentiate them from 1) digital materials that have been created as a result of digitising analogue originals; and 2) digital materials that may have originated from a digital source but have been printed to paper, e.g. some electronic records.	ACCEPT
Boundary value	Data value that corresponds to a minimum or maximum input or output value specified for a system or component.	ACCEPT
Canonical data collection	Data collection that has been normalised by some established criteria to allow for effective data management. Examples include: data files that belong to a certain experiment, all files that are created by one specific simulation, all files that belong to a specific observation (same day, same place, etc.).	NR
CARE Principles for Indigenous Data Governance	Set of principles for Indigenous data governance. CARE stands for Collective benefit, Authority to control, Responsibility and Ethics. These principles complement the existing FAIR principles.	ADD
Catalogue	(Noun) Index describing, indicating the location of, and recording other details of resources, materials works, etc. Curated and organised using a formal metadata schema (e.g. MARC, ISAD(G), Dublin Core, DataCite etc.) RELATED TERMS: Cataloguing, Metadata, Data catalogue.	EDIT
Cataloguing	Process of describing digital or analogue data in accordance with a formal metadata standard to create a record of that dataset's characteristics, provenance and location. Creates, adds to or edits a catalogue. RELATED TERMS: Catalogue, Metadata.	EDIT
Change log	Document, spreadsheet, or digital tool that tracks the progress of each change in a dataset, code or other research object. RELATED TERMS: Version Control, Documented Data.	EDIT
Checksum	Alphanumeric signature (similar to a fingerprint) calculated from a digital object's content and structure using a mathematical algorithm. The algorithm will always produce the same checksum unless any change, no matter how small, is made to the file. Comparing checksums over time facilitates the management of integrity and authenticity of digital content.	EDIT
Citable data	Standalone dataset that can be cited in a similar manner to other research outputs. The dataset appears in a data repository, data paper or project website, and has a Persistent Identifier. Most current referencing systems provide a format for citing datasets.	EDIT
Cloud computing	Large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms and services are delivered on demand to external customers over the Internet. Key features are that: it is a specialised distributed computing paradigm; it is massively scalable; it can be encapsulated as an abstract entity that delivers different levels of services to customers outside the Cloud; it is driven by economies of scale; and the services can be dynamically configured (via virtualization or other approaches) and delivered on demand.	ACCEPT

Cloud ecosystem	Ecosystem that includes software, infrastructure, consultants, integrators, partners, third parties and anything in their own environments that has a bearing on the other components.	EDIT
Collection management identification	Type of data provenance that adds metadata to identify data collections. The organisation doing the collection management is stated in the metadata along with the provenance of collection management events such as source of data acquisition, conservation, and movement.	NR
Comma separated values	File that contains the values in a table as a series of ASCII text lines organised so that each column value is separated by a comma from the next column's value and each row starts a new line. SYNONYM: CSV.	EDIT
Confidentiality	The duty and practice of ensuring that individuals' personal information only flows from one entity to another according to legislated or otherwise broadly accepted norms and policies. It is the responsibility of the researcher/organisation to ensure that the participants' identities are not disclosed.This can be done by either restricting access to the data or certain variables in the data, and/or by protecting their identities using an anonymisation method.	Edit
Conformance	The state of having satisfied the requirements of some specific standard(s) and/or specification(s). Conformance is used with respect to voluntary standards and specifications, whereas compliance is used with respect to mandatory standards and regulations.	NR
Consumer data	The information trail customers leave behind as a result of their Internet use. This data, which sometimes comprises personal information, comes from such sources and channels as social media networks, marketing campaigns, customer service requests, call centre communications, online browsing data, mobile applications, purchasing history and preferences, and more.	ACCEPT
Container	1. A logical collection of objects, using a standard such as Bag-IT for archival management purposes (https://en.wikipedia.org/wiki/BagIt) 2. A computing technology which allows application code, data, dependencies, and configurations to be packaged into a single object that can be deployed in any environment. Examples include Singularity (https://en.wikipedia.org/wiki/Singularity_(software)) and Docker (https://en.wikipedia.org/wiki/Docker_(software))	EDIT
Content replication	A type of digital migration where there is no change to the Packaging Information, the Content Information, or the PDI. The bits used to represent these Information Objects are preserved in the transfer to the same or new media instance. RELATED TERM. OAI repository.	edit
Controlled vocabulary	A list of standardised terminology, words, or phrases, used for indexing or content analysis and information retrieval, usually in a defined information domain. RELATED TERM. Ontology.	ACCEPT
Corpus	Set of documents that has a scientific meaning. A corpus can be produced by an individual researcher's activity (including its archival materials) or from a laboratory's research, a field campaign, a survey, or any other discrete research activity.	EDIT
Corrupt data	Deterioration of computer data as a result of some external agent such as viruses, hardware or software incompatibility, flaws, or failures, power outages, dust, water, extreme temperatures, etc. RELATED TERM. Noisy data.	ACCEPT
Cross-disciplinary	Explains aspects of one discipline in terms of another (e.g., the physics of music; the politics of literature)	ACCEPT
Curation	The activity of managing and promoting the use of data from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment	EDIT

	or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials. RELATED TERM. Data Curation.	
Curation workflow	A type of workflow that includes active steps to curate data as an aid to on-going management of data through their lifecycle.	EDIT
Dark data	Operational data that are not being used, such as information assets that organisations collect, process and store in the course of their regular business activity, but generally fail to use for other purposes. Such data are seen as an economic opportunity for companies if they can take advantage of it to drive new revenues or reduce internal costs. Examples include server log files that can give clues to website visitor behaviour; client call detail records that can indicate consumer sentiment; and mobile geolocation data that can reveal traffic patterns to aid in business planning. RELATED TERM. At-Risk Data, Legacy Data.	EDIT
Data	Facts, measurements, recordings, records, or observations about the world collected by scientists and others, that are yet to be processed/interpreted/analysed. Data may be in any format or medium taking the form of writings, notes, numbers, symbols, text, images, films, video, sound recordings, pictorial reproductions, drawings, designs or other graphical representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing algorithms, or statistical records.	EDIT
Data access protocol	A system that allows outsiders to be granted access to databases without overloading the system.	NR
Data access statement	A short statement accompanying a research publication that describes whether supporting data is available to access, where this has been made available and under what conditions.	ADD
Data acquisition	The process of acquiring data from some source. For example, data may be acquired by download from a repository, transfer from a data logger, data capture, etc. SYNONYM. Data reception; Data download. RELATED TERM. Data capture	ACCEPT
Data analysis	A data lifecycle stage that involves the techniques that produce synthesised knowledge from organised information. A process of inspecting, cleaning, transforming, and modelling data with the goal of highlighting useful information suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains.	ACCEPT
Data archive	An archival service providing long-term, permanent care and accessibility for digital objects with research value. The standard for such repositories is the Open Archival Information System reference model. RELATED TERM. Repository; Trusted Digital Repository	accept
Data availability	The state when data are available in a timely manner in the place and form as needed by the user.	Edit
Data capture	The process or means of obtaining and storing external data, particularly images or sounds, for use at a later time. In biometric security systems, for example, capture is the acquisition of, or the process of acquiring an identifying characteristic such as a finger image, palm image, facial image, iris print, or voice print. In order to capture the data, a transducer is employed that converts the actual image or sound into a digital file. The file is then stored. At a later time, it can be analysed by a computer, or compared with other files in a database to verify identity or to provide authorization to enter a secured system. Screen capture is the acquisition and storage of an image on a monitor or display exactly as it appears at a specific time. This can sometimes (but not always) be done by hitting the “print screen” key, in which case the	NR

	image appears as a bitmap file in the clipboard. It can also be done by photographing the display screen with a digital camera external to the computer. Electronic signals from scientific instruments, dataloggers, sensors, etc., can also be captured, converted to data, and stored for use at a later time. RELATED TERM. Data acquisition	
Data catalogue	A curated collection of metadata records describing datasets and their data elements. Curated and organised using a formal metadata schema appropriate to data and data sets (e.g. ReCollect, DataCite etc) RELATED TERMS: Cataloguing, Metadata, Catalogue.	Edit
Data centre	1. A facility providing IT services, such as servers, massive storage, and network connectivity. 2. A facility holding large scale data repositories. SYNONYM. Research data centre. RELATED TERM. Digital Infrastructure.	EDIT
Data citation	The process of citing a dataset in a similar manner to other research outputs. The dataset must be a standalone output that appears in a data repository, data paper or project website, and has a Persistent Identifier. Most current referencing systems provide a format for citing datasets.	EDIT
Data cleaning	The process of detecting and correcting corrupt or inaccurate records from a dataset. Data cleaning is a continuous process that requires corrective actions throughout the data lifecycle. Data cleaning involves identifying, replacing, modifying or deleting incomplete, incorrect, inaccurate, inconsistent, irrelevant, and improperly formatted data. Typically, the process involves updating, correcting, standardising, and de-duplicating records to create a single view of the data, even if they are stored in multiple disparate systems. SYNONYM: Data cleansing; Data scrubbing.	EDIT
Data collection	1. A logical grouping of (research) datasets that share a common aspect or concept. A data collection is the highest entity in the hierarchy of data groupings (data collection, dataset, row or record in a dataset). It comprises a grouping of datasets that have a strong connection and it is organised coherently around a single element or concept (e.g., model, instrument). 2. Creating, recording, acquiring, or linking to specified data or type of data. A description of this process includes specifying file formats, naming conventions, data structure, and what provisions have been made for version control, data re-use, sharing, and long-term access to the data.	edit
Data completeness	The degree to which all required measures are known. Values may be designated as “missing” in order not to have empty cells, or missing values may be replaced with default or interpolated values. In the case of default or interpolated values, these must be flagged as such to distinguish them from actual measurements or observations. Missing, default, or interpolated values do not imply that the dataset has been made complete.	NR
Data compliance	Data compliance consists of the ongoing processes to ensure adherence of data to both enterprise business rules (government department, university, industry, or agency), and to legal, regulatory and accreditation requirements. Data compliance includes five areas: controls, audit, legal compliance, regulatory compliance, and accreditation conformance.	accept
Data container	A software stack that chunks digital objects at a physical layer. Typical containers are file systems, database management systems, content management systems, clouds etc. The software stack implies some form of encapsulation of the digital object.	edit
Data curation	A managed process, throughout the data lifecycle, by which data/data collections are cleansed, documented, standardised, formatted and inter-related. This includes versioning data, or forming a new collection from several data sources, annotating with metadata, adding codes to raw data (e.g., classifying a galaxy image with a galaxy type such as “spiral”). Higher levels	accept

	of curation involve maintaining links with annotation and with other published materials. Thus a dataset may include a citation link to publication whose analysis was based on the data. The goal of curation is to manage and promote the use of data from its point of creation to ensure it is fit for contemporary purpose and available for discovery and re-use. For dynamic datasets this may mean continuous enrichment or updating to keep it fit for purpose. Special forms of curation may be available in data repositories. The data curation process itself must be documented as part of curation. Thus curation and provenance are highly related.	
Data custodian	A data custodian is an IT individual or organisation responsible for the IT infrastructure providing and protecting data in conformance with the policies and practices prescribed by data governance. SYNONYM. Technical data steward; Data manager. RELATED TERM. Data stewardship	NR
Data de-noising	Removing noise from data.	NR
Data deletion	Process of destroying data stored on tapes, hard disks and other forms of electronic media so that it is no longer there and it cannot be restored. RELATED TERM. Data destruction, Tombstone record.	ADD
Data destruction	Process of destroying data stored on tapes, hard disks and other forms of electronic media so that it is completely unreadable and cannot be accessed or used. RELATED TERM. Data Deletion, Tombstone Record.	Edit
Data dictionary	A collection of descriptions of the data objects or items in a data model. A first step in analysing a system of objects with which users interact is to identify each object and its relationship to other objects. This process is called data modelling and results in a picture of object relationships. After each data object or item is given a descriptive name, its relationship is described (or it becomes part of some structure that implicitly describes relationship), the type of data (such as text or image or binary value) is described, possible predefined values are listed, and a brief textual description is provided. This collection can be organised for reference into an eBook called a data dictionary.	Accept
Data dredging	A data mining practice in which large volumes of data are analysed seeking any possible relationships between data. The traditional scientific method, in contrast, begins with a hypothesis and follows with an examination of the data. Data dredging often circumvents traditional data mining techniques and may lead to premature conclusions. Uncovered patterns may be presented as statistically significant without any specific hypothesis as to the underlying causality. SYNONYM: Data fishing.	edit
Data driven decision management	Approach to governance that values decisions that can be backed up with data that can be verified. The success of the data-driven approach is reliant upon the quality of the data gathered and the effectiveness of its analysis and interpretation. Errors can creep into data analytics processes at any stage of the endeavour and serious issues can result when they do. SYNONYM. DDDM.	NR
Data driven disaster	Serious problem caused by one or more ineffective data analysis processes. In addition to the financial burden, problems with data quality and analysis can have a serious impact on security, compliance, project management and human resource management, among others. Error can creep into data analytics at any stage. The data quality may be inadequate in the first place. The data could be incomplete, inaccurate, not current, or may not be a reliable indicator of what they are intended to represent. Data analysis and interpretation are prone to a similar number of pitfalls. There can be confounding factors and the mathematical method can be flawed or inappropriate. Correlation can be erroneously considered to suggest causation. Statistical significance may be mistakenly attributed when the data do not support it. Even if the data and	NR

	analytic processes are valid, data may be deliberately presented in a misleading manner to support an agenda. Problems arise when insufficient resources are applied to data processes and too much confidence placed in their validity. To prevent data-driven disasters, it’s crucial to continually examine data quality and analytic processes, and to pay attention to common sense and even intuition. When data seem to be indicating something that does not make logical sense or just seems wrong, it is time to re-examine the source data and the methods of analysis.	
Data element	In a database, an example of a data element is a data field. One also says that a data element is an attribute of a data entity. A unit of data for which the definition, identification, representation (term used to represent it), and permissible values are specified by means of a set of attributes. For example, the data element “age of a person” with values consisting of all combinations of 3 decimal digits; A personnel record may include the data elements “name” and “address”. In the context of the personnel record, “name” and “address” function as an indivisible unit, e.g., the data element “name” and the data element “address” each can be stored and retrieved as an indivisible unit. However, in a different context, “address” itself may be considered a record that contains its own data elements “street address”, “city”, “postal code”, “country”.	NR
Data entity	An object, event, or phenomenon about which data are stored in a database and which has intermediate representation in a Data Model.	NR
Data ethics	"Branch of ethics concerned with the moral implications of practices involving data, including data collection, description, storage, accessibility, rights, ownership, and uses. This also includes any corresponding practices, such as the use of data in algorithms, artificial intelligence, or in policy development, which have the potential to positively or negatively impact individuals, groups, or society. Source: Floridi L, Taddeo M. 2016 What is data ethics? Phil. Trans. R. Soc. A 374: 20160360. http://dx.doi.org/10.1098/rsta.2016.0360 "	ADD
Data exploration	Data exploration involves summarising the main characteristics of a dataset using visualisation and should be the first step in data analysis.	ACCEPT
Data file format	The layout of a file in terms of how the data within the file are organised and encoded for storage. A program that uses the data in a file must be able to recognize and possibly access data within the file. A particular file format is often indicated as part of a file’s name by a filename extension (suffix). Conventionally, the extension is separated by a period from the name and contains three or four letters that identify the format. There are as many different file formats as there are different programs to process the files. Examples include: Word documents (.doc), Web text pages (.htm or .html), Web page images (.gif and .jpg), Adobe Postscript files (.ps), Adobe Acrobat files (.pdf), Executable programs (.exe), Multimedia files (.mp3 and others). Preferred formats are those designated by a data repository for which the digital content is maintained. If a data file is not in a preferred format, a data curator will often convert the file into a preferred format, thus ensuring that the digital content remains readable and usable. Usually, preferred formats are the de facto standard employed by a particular community. SYNONYM. Research data format. RELATED TERM. Data structure.	EDIT
Data governance	The exercise of authority, control and shared decision making (planning, monitoring and enforcement) over the management of data assets. It refers to the overall management of the availability, usability, integrity, and security of the data employed in an organisation. A sound data governance program includes a governing body or council, a defined set of procedures, and a plan to execute those procedures.	EDIT

Data harmonisation	Making data from different sources comparable. The processes involved in producing inferentially equivalent data. The data are interchangeable between different information systems with uniform and unambiguous mutual understanding. RELATED TERM. Data integration.	EDIT
Data hygiene	The collective processes conducted to ensure the cleanliness of data. Data are considered clean when they are relatively error-free. RELATED TERM. Dirty data.	NR
Data ingestion	Obtaining, importing, and processing data for later use or storage in a database. This process often involves altering individual files by editing their content and/or formatting them to fit into a larger document. An effective data ingestion methodology begins by validating the individual files, then prioritises the sources for optimum processing, and finally validates the results. When numerous data sources exist in diverse formats (the sources may number in the hundreds and the formats in the dozens), maintaining reasonable speed and efficiency can become a major challenge. To that end, several vendors offer programs tailored to the task of data ingestion in specific applications or environments.	EDIT
Data integration	Combining diverse datasets from disparate sources into one unified dataset or database. Data are accessed and extracted, moved, validated, cleaned, transformed and loaded. SYNONYM: Data compilation. RELATED TERM. Data linkage; Privacy-preserving data linkage	NR
Data integrity	1. In the context of data and network security: The assurance that information can only be accessed or modified by those authorised to do so. 2. In the context of data quality: The assurance the data are clean, traceable, and fit for purpose.	NR
Data item	A type of data element that expresses a proposition that binds one or more property values to some data entity.	NR
Data lake	A data repository that is centralised, usually at the organisation/institution level, where raw data of any size and format can be stored for any length of time before it is processed, analysed or used.	ADD
Data librarian	Librarian who manages the sharing and publishing of datasets as openly as possible and as closed as necessary, and the management and curation of repositories required to achieve this. Broad role requirements include support for sharing and publishing datasets, finding, accessing, interoperating and re-using these datasets, reviewing and supporting Data Management Plans and training delivery.	Edit
Data linkage	The process of bringing together from two or more different sources, data that relate to the same individual, family, place or event). Example: linkage may be used to bring together information about an individual's health status, prescription drug use, and social media habits. SYNONYM. Linkage. RELATED TERM. Data integration	accept
Data management infrastructure	An infrastructure used to provide data management and enforce data management policies. A data management infrastructure would include resources such as a data repository and an information catalogue.	NR
Data management plan	A formal statement describing how research data will be managed and documented throughout a research project and the terms regarding the subsequent deposit of the data with a data repository for long-term management and preservation.	accept
Data management policy	A written document backed by management describing policy and providing guidance to ensure that appropriate standards, consistent guidelines, and common strategies are used, providing linkages to and consistency with other similar systems, and fostering a true network across an organisation producing data.	accept

Data mart	A repository of data designed to serve a particular community of knowledge workers. The goal of a data mart is to meet the particular demands of a specific group of users. Because data marts are optimised to look at data in a unique way, the design process tends to start with an analysis of user needs. Generally, an organisation's data marts are subsets of the organisation's data warehouse. A data mart tends to be tactical and aimed at meeting an immediate need. Data virtualization software can be used to create virtual data marts, pulling data from disparate sources and combining it with other data as necessary to meet the needs of specific business users. A virtual data mart provides knowledge workers with access to the data they need while preventing data silos and giving the organisation's data management team a level of control over the organisation's data throughout its lifecycle. The difference between a data warehouse and a data mart can be confusing because the two terms are sometimes used incorrectly as synonyms. RELATED TERM. Data warehouse	NR
Data migration	The process of transferring data between storage types, formats, information technologies, or computer systems. A data migration project is usually undertaken to replace or upgrade servers or storage equipment, for a website consolidation, to conduct server maintenance or to relocate a data centre.	EDIT
Data mining	analysing multivariate datasets using pattern recognition or other knowledge discovery techniques to identify potentially unknown and potentially meaningful data content, relationships, classification or trends. Data mining parameters include: Association (looking for patterns where one event is connected to another event); Sequence or path analysis (looking for patterns where one event leads to another later event); Classification (looking for new patterns); Clustering (finding and visually documenting groups of facts not previously known); Forecasting, or predictive analytics (discovering patterns in data that can lead to reasonable predictions about the future.	Accept
Data model	A model that specifies the structure or schema of a dataset. The model provides a documented description of the data and thus is an instance of metadata. It is a logical, relational data model showing an organised dataset as a collection of tables with entity, attributes and relations. RELATED TERM. Data modelling.	NR
Data modelling	Formalising and documenting existing processes and events. It captures and translates complex system designs into easily understood representations of data flows and processes, creating a blueprint for construction and/or re-engineering. Data modellers often use multiple models to view the same data and ensure that all processes, entities, relationships and data flows have been identified. There are several different approaches to data modelling, including: Conceptual Data Modelling (identifies the highest-level relationships between different entities); Enterprise Data Modelling (similar to conceptual data modelling, but addresses the unique requirements of a specific organisation); Logical Data Modelling (illustrates the specific entities, attributes and relationships involved in a business function. Serves as the basis for the creation of the physical data model); Physical Data Modelling (represents an application and database-specific implementation of a logical data model). RELATED TERM. Data Model.	EDIT
Data organisation	Denotes the complexity of measures that are used by a repository to form aggregations of data objects (including collections and metadata) to describe the properties of data objects, to register PIDs, to build the PID records, to link between all components, and to set up the containers (software stack) that are used to store all components.	NR
Data policy	An organisationís stated data/information management processes designed to assist and protect the organisation's data research assets. It is a set of high-level principles that establish a guiding framework for data management. A data policy	accept

	can be used to address strategic aspects such as data access, relevant legal matters, data stewardship issues and custodial duties, data acquisition and other issues.	
Data preprocessing	Any type of processing performed on raw data to prepare it for another processing procedure. Preprocessing may include: data sampling, data transformation, de-noising, data normalisation, data standardisation, or feature extraction.	ACCEPT
Data processing	A generic concept referring to all kinds of procedures being executed on data at any point in the data lifecycle. RELATED TERM. Research data lifecycle.	EDIT
Data product specification	Detailed description of a dataset or dataset series together with additional information that will enable it to be created, supplied to and used by another party. A data product specification provides a description of the universe of discourse and a specification for mapping the universe of discourse to a dataset. It may be used for production, sales, end-use or other purposes. Source: ISO 19131:2007	ADD
Data production	Includes all activities involved in the planning, collecting, processing, analysis and maintenance of data in the original research project. Among these activities are selecting a study design, constructing instruments for data collection, conducting data collection/creation, performing data editing/verification/validation, analysing data, backing up data versions and preparing and tagging metadata.	ACCEPT
Data profiling	Statistical analysis and assessment of the quality of data values within a dataset for consistency, uniqueness and logic. The data profiling process cannot identify inaccurate data; it can only identify business rules violations and anomalies. The insight gained by data profiling can be used to determine how difficult it will be to use existing data for other purposes. It can also be used to provide metrics to assess data quality and help determine whether or not metadata accurately describes the source data. Profiling tools evaluate the actual content, structure and quality of the data by exploring relationships that exist between value collections both within and across datasets. For example, by examining the frequency distribution of different values for each column in a table, an analyst can gain insight into the type and use of each column. Cross-column analysis can be used to expose embedded value dependencies and inter-table analysis allows the analyst to discover overlapping value sets that represent foreign key relationships between entities.	EDIT
Data publication	The release of research data, associated metadata, accompanying documentation, and software code (in cases where the raw data have been processed or manipulated) for re-use and analysis in such a manner that they can be discovered on the Web and referred to in a unique and persistent way. Data publishing occurs via dedicated data repositories and/or (data) journals which ensure that the published research objects are findable, accessible, interoperable and re-usable. RELATED TERMS: Data Sharing, Repository	edit
Data quality	The reliability and application efficiency of data. It is a perception or an assessment of a dataset's fitness to serve its purpose in a given context. Aspects of data quality include: Accuracy, Completeness, Update status, Relevance, Consistency across data sources, Reliability, Appropriate presentation, Accessibility. Within an organisation, acceptable data quality is crucial to operational and transactional processes and to the reliability of analytics, business intelligence, and reporting. Data quality is affected by the way data are entered, stored and managed. Maintaining data quality requires going through the data periodically and scrubbing it. Typically this involves updating, standardising, and de-duplicating records to create a single view of the data, even if it is stored in multiple disparate systems. Data quality assurance (DQA) is the process of verifying the reliability and effectiveness of data. RELATED TERM. Data cleaning	accept

Data recovery	The process of restoring data that have been lost, accidentally deleted, corrupted or made inaccessible for any reason. The data recovery process may vary, depending on the circumstances of the data loss, the data recovery software used to create backups, and backup target media. In some cases, end users may be able to restore lost files themselves. Restoration of a corrupted database from a tape backup is a more complicated process that requires specialised intervention. Data that were not backed up and were accidentally deleted from a computer's file system may sometimes be recovered from file fragments that remain on the disk. An organisation's disaster recovery plan should make known who in the organisation is responsible for recovering data, provide a strategy for how data will be recovered and document acceptable recovery point and recovery time objectives. SYNONYM. Data restoration	NR
Data reduction	The process of reducing the amount or size of stored data. This may be achieved by eliminating redundant copies of data files, deduplicating data files by removing redundant records, or by compressing the data files.	NR
Data registration	A curation process on a data object by which it receives a persistent object identifier (PID) from a trusted registration authority. Registration must be accompanied by the step(s) to upload the data object to a persistent repository. RELATED TERM. Repository; Persistent identifier.	accept
Data repository management	Management of a national, discipline or institutional repository of published datasets. Provision of infrastructure, curation, policy and training that govern the organisation, control, and properties of the repository such as: required file formats, access control restrictions, integrity, replication, retention, disposal, etc.	Edit
Data representation	An object describing the context of the data, including provenance, description, structural, and administrative information.	NR
Data rescue	Recovery and/or transformation and digitization of dark data and at-risk data so that they can be preserved, accessed, shared, and used. Data rescue also involves the addition of rich metadata to make the content understandable and more easily re-usable.	NR
Data residency	The physical or geographic location of an organisation's data or information. Data residency also refers to the legal or regulatory requirements imposed on data based on the country or region in which it resides. Cloud computing, which allows organisations to deliver hosted services over the Internet, can create data residency concerns.	edit
Data retention policy	An organisation's established protocol for retaining information for operational or regulatory compliance needs. The objectives of a data retention policy are to keep important information for future use or reference, to organise information so it can be searched and accessed at a later date, and to dispose of information that is no longer needed. A data retention policy must consider both the value of data over time, and regulations to which the data may be subject.	Accept
Data review	An activity through which the correctness conditions of the data are verified. It also includes the specification of the type of the error or condition not met, and the qualification of the data and its division into the "error-free" and "erroneous" data. Data review consists of both error detection and data analysis, and can be carried out in manual or automated mode. RELATED TERM. Data validation; High quality data; Dirty data; Data cleaning; Data processing; Data integrity	NR
Data sampling	Selection of a statistically representative subset from a large population of data	NR
Data scaling	Techniques used to deal with parameters having different units and scales. SYNONYM. Data rescaling. RELATED TERM. Data standardisation	NR

Data selection	A process that creates a new dataset from an original source. Examples include: creating a subset of the data,querying a database.	NR
Data sharing	The practice of making data available for checking, reproducing or reusing. The mechanisms available for achieving this are: making data available on request, as supplementary information to journal articles, or as published datasets in data repositories. RELATED TERMS: Data Publication, Repository.	Edit
Data splitting	An approach to protecting sensitive data from unauthorised access by encrypting the data and storing different portions of a file on different servers. An unauthorised person would need to know the locations of the servers containing the parts, be able to get access to each server, know what data to combine, and how to decrypt it. Data splitting can be made even more effective by periodically retrieving and recombining the parts, and then splitting the data in a different way among different servers, and using a different encryption key.	NR
Data standard	A technical specification that defines how data should be structured and formatted to ease interoperability across different systems, publishers and users.	ADD
Data standardisation	Conversion of multiple datasets to a single common format and structure. RELATED TERM. Data scaling. Standardisation.	EDIT
Data store	A repository for persistently storing collections of data, such as a database, a file system or a directory. The data stored can be of any type that can be rendered in digital format and placed in electronic media. Examples include text, image, video files and audio files.	NR
Data stream	A sequence of digitally encoded, coherent signals used to send or receive a representation of information content as transmitted.	NR
Data structure	A specialised format for organising and storing data. General data structure types include the array, the file, the record, the table, the tree, and so on. Any data structure is designed to organise data to suit a specific purpose so that it can be accessed and worked with in appropriate ways. In computer programming, a data structure may be selected or designed to store data for the purpose of working on it with various algorithms.	NR
Data structure continuum	The continuum of data structure that includes unstructured data, semi-structured data, and structured data.	NR
Data table attribute	1. A field or column in a database table. It is an abbreviation for ‘physical data attribute’ which is a single data element related to a data object, such as a table in a database. The database schema associates one or more attributes with each database entity (i.e. table). 2. A logical or conceptual attribute such as in an entity-attribute-relationship (EAR) data model.	edit
Data traceability	Data traceability follows the lifecycle of data to track all access and changes to the data. It helps demonstrate transparency, compliance and adherence to regulations. Data traceability, along with data compliance, can be considered part of a data audit process. Data traceability is fundamental to reproducible research.	NR
Data transformation	Manipulation of raw data to produce a single output. RELATED TERM. Data selection; Data processing; Data preprocessing.	NR
Data type registry	1. A registry that links data types of all sorts with the executable data processing functions that can be useful for working with a specific data type. Examples include: complex file types in biology (diagnosis), registering categories that appear in	NR

	PID records to describe data properties. Data types range from complex digital objects to simple categories that occur in digital objects. 2. A type of registry for data types supporting their standardisation, uniqueness and discoverability.	
Data upload database	A collection of interrelated data often with controlled redundancy, organised according to a scheme to serve one or more applications; the data are stored so that they can be used by several programs without concern for data structures or organisation.	NR
Data validation	Provides well-defined guarantees for fitness, accuracy, and consistency for any of various kinds of user input into an application or automated system. Data validation checks that data are valid, sensible, reasonable, clean, usable, and secure before they are processed. Failures or omissions in data validation can lead to data corruption, security vulnerability. Improperly validated data can cause computer code processing the data to crash, generate error messages, behave in an unanticipated manner, or generate incorrect results that may be difficult or impossible to detect. RELATED TERM. High quality data; Dirty data; Data cleaning; Data processing; Data integrity	NR
Data warehouse	A central repository for all or significant parts of the data that an organisation's various business systems collect. A data warehouse tends to be a strategic but somewhat unfinished concept. Data warehousing emphasises the capture of data from diverse sources for useful analysis and access, but does not generally start from the point-of-view of the end user who may need access to specialised data marts. There are two approaches to data warehousing: The top down approach spins off data marts for specific groups of users after the complete data warehouse has been created. The bottom up approach builds the data marts first and then combines them into a single, all-encompassing data warehouse. SYNONYM. Data mart	NR
Data wrangling	The process of manually or semi-automatically converting or mapping data from one form into another format that allows for more convenient consumption of the data with the help of semi-automated tools. Gathering and organising disparate data from different sources, often collected by many different investigators. Activities include developing and supporting search tools that utilise standardised metadata, harmonising the coding of data for specific variables, engineering new methods of combining data. with the help of semi-automated tools. The result of data wrangling is repurposed data. RELATED TERM. Repurposed data	NR
Data z-score scaling	Standardising data so that the mean is centred at zero. Therefore, participants' scores reflect their distance from the mean in standard deviations (i.e., whether their score is higher or lower than the mean).	EDIT
Database	A collection of data that is organised in a according to a conceptual structure/model describing the characteristics of these data and the relationships among their corresponding entities, supporting one or more application areas. A database allows its contents to be easily accessed, managed and updated. The type of database used depends on the requirements of the study. A common type is the relational database, where data are related to each other in a systematic manner so that they can be reorganised and accessed in a number of different ways. A database may house one or many datasets.	ACCEPT
Database administration	The function of managing the physical aspects of data resources, including database design and integrity, backup and recovery, performance and tuning.	NR
Dataset	organised collection of data or objects in a computational format, that are generated or collected by researchers in the course of their investigations, regardless of their form or method, that form the object on which researchers test a hypothesis.	EDIT

	This includes the full range of data: raw, unprocessed datasets, proprietary generated and processed data and secondary data obtained from third parties. The presentation of the data in the application is enabled through metadata.	
Dataset series	A collection of datasets sharing the same product specification. A dataset series is a type of aggregation or collection with some “logical grouping” such as by a topic (specification) with the (product) unit being a dataset series. Example: A series of earth observations. Each year, month or week (depending on the volume) might be a dataset and the series could run from a specified year to the present.	NR
Datetime	A standard way to express a numeric calendar date that eliminates ambiguity, acceptable formats being defined by ISO 8601. ISO 8601 is applicable whenever representation of dates in the Gregorian calendar, times in the 24-hour timekeeping system, time intervals and recurring time intervals or of the formats of these representations are included in information interchange. It includes calendar dates expressed in terms of calendar year, calendar month and calendar day of the month; ordinal dates expressed in terms of calendar year and calendar day of the year; week dates expressed in terms of calendar year, calendar week number and calendar day of the week; local time based upon the 24-hour timekeeping system; Coordinated Universal Time of day; local time and the difference from Coordinated Universal Time; combination of date and time of day; time intervals; recurring time intervals. SYNONYM. ISO date format; ISO time format	NR
De-anonymization	A reverse engineering process in which de-identified data are cross-referenced with other data sources to re-identify the personally identifiable information. This could occur if a de-identification process had not been not successfully performed, or had not been undertaken in the first place. Also known as triangulation or re-identification.	Edit
De-identification	One or more techniques designed to make the risk of identifying a particular individual in a dataset negligible, whilst retaining the re-usability of the dataset. The purpose is to protect the privacy of the individual and comply with legislation, whilst enabling data sharing. Methods include removing direct and indirect identifiers such as names, addresses, social insurance numbers, or dates of birth, or using obfuscation methods such as encryption, hashing, generalisation, pseudonymization, and perturbation.	Edit
Denormalization	In a relational database, denormalization is an approach to speeding up read performance (data retrieval) in which the administrator selectively adds back specific instances of redundant data after the data structure has been normalised. A denormalized database should not be confused with a database that has never been normalised. After data has been duplicated, the database designer must take into account how multiple instances of the data will be maintained. One way to denormalize a database is to allow the database management system (DBMS) to store redundant information on disk. This has the added benefit of ensuring the consistency of redundant copies. Another approach is to denormalize the actual logical data design, but this can quickly lead to inconsistent data. Rules called constraints can be used to specify how redundant copies of information are synchronised, but they increase the complexity of the database design and also run the risk of impacting write performance.	NR
Derived data product	The results of applying a procedure to transform a data object in order to obtain a desired data product that is stored in a repository along with the provenance and descriptive metadata. RELATED TERM. Data transformation	NR
Descriptive metadata	Metadata that describe a dataset or resource in such a way that people can discover and identify it. Contains information that aids with findability, for example, information (metadata elements) on the creator(s), affiliation(s), title, abstract, keywords, persistent identifier, related publications, etc.	EDIT

Digital archiving	1. Often used interchangeably with 'digital preservation' in library and archiving professional communities. 2. In the context of computing, 'Digital archiving' is the process of backup and ongoing maintenance as opposed to strategies for long-term digital preservation. RELATED TERM. Archiving	edit
Digital data	Data in the form of digital materials. RELATED TERM. Digital materials	Accept
Digital infrastructure	Those layers that sit between base technology (a computer science concern) and discipline-specific science. The focus is on value-added systems and services that can be widely shared across scientific domains, both supporting and enabling large increases in multi- and interdisciplinary science while reducing duplication of effort and resources (e.g. including hardware, software, personnel, services and organisations). RELATED TERM. E-Research Infrastructure.	edit
Digital materials	1. Digital surrogates created as a result of converting analogue materials to digital form (digitisation); 2. “Born digital” for which there has never been and is never intended to be an analogue equivalent; 3. Digital records. RELATED TERMS. Born digital; Digital objects; Digital records; Digital data; Electronic records; Analogue Materials.	EDIT
Digital object	A digital object is editable, interactive, accessible and modifiable by means of digital objects other than the one governing its behaviour, and is distributed over information infrastructures. It is a machine-independent data structure consisting of one or more elements in digital form that can be parsed by different information systems; the structure helps to enable interoperability among diverse information systems in the Internet.” A digital object is composed of a structured sequence of bits/bytes. As an object it is named. The bit sequence realising the object can be identified and accessed by a unique and persistent identifier or by use of referencing attributes describing its properties. SYNONYM. Digital entity	NR
Digital Object Identifier	A name (not a location) for an entity on digital networks. It provides a system for persistent and actionable identification and interoperable exchange of managed information on digital networks. A DOI is a type of Persistent Identifier (PID) issued by the International DOI Foundation. This permanent identifier is associated with a digital object that permits it to be referenced reliably even if its location and metadata undergo change over time. SYNONYM. DOI	Accept
Digital preservation	The series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Digital preservation is defined very broadly and refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change. Those materials may be records created during the day-to-day business of an organisation; “”born-digital”” materials created for a specific purpose (e.g. teaching resources); or the products of digitisation projects. This definition specifically excludes the potential use of digital technology to preserve the original artefacts through digitisation. RELATED TERM. Digitisation; Preservation	NR
Digital research data	Research data in digital form. It may have been originally created in digital form, or it may have been converted from paper, or other non-digital form to a digital representation. RELATED TERM. Research data.	Edit
Digital scholarship	Scholarship which is dependent upon digital methods, tools or resources. May include building a digital collection of information for further study and analysis; creating appropriate tools for collection-building; creating appropriate tools for the analysis and study of collections; using digital collections and analytical tools to generate new intellectual products; or creating authoring tools for these new intellectual products, either in traditional forms or in digital form.	Edit

Digitisation	The process of creating digital files by scanning or otherwise converting analogue materials. The resulting digital copy, or digital surrogate, would then be classed as digital material and then subject to the same broad challenges involved in preserving access to it, as “born digital” materials.	NR
Direct identifier	A data variable or value that directly discloses the identity of a single living individual, for example, a name, passport number or fingerprint.	ADD
Dirty data	Data that contain errors. Dirty data can be caused by a number of factors including: inaccurate, incomplete or erroneous data such as spelling or punctuation errors, incorrect data or incorrect data type associated with a field, incomplete or outdated data, duplicate data, inconsistent data, incorrectly ordered data, improper parsing of fields from disparate systems, etc. Errors can be introduced at any stage as data are entered, stored and managed. Using a dirty dataset can lead to spurious associations, false conclusions and misdirected investments. SYNONYM. Dirty dataset	NR
Document type definition	The building blocks of an XML document.	NR
Documented data	Data that are delivered with all associated metadata, data dictionary, description of methods and instruments used to collect and process the data, and other supporting data (e.g., duplicate sample results, replicate analyses, percent recovery, etc.) with the purpose of providing the full context in which the data were created.	Edit
Dublin Core	Widely used metadata element set, formally titled ISO 15836-1:2017, Information and documentation — The Dublin Core metadata element set — Part 1: Core elements.	Edit
Dynamic data	Data the content of which is changing frequently and at asynchronous moments. Examples include: Data streams that are generated by sensors when it is unpredictable when data segments will appear in time (i.e. data streams have gaps); Data streams that are generated by humans in crowdsourcing scenarios where it is not clear when which cell in a database will be filled.	accept
E-Research	(Historical) Computationally intensive, large-scale, networked and collaborative forms of research and scholarship across all disciplines, including all of the natural and physical sciences, related applied and technological disciplines, biomedicine, social science and the digital humanities.	edit
E-Research infrastructure	(Historical) Science supported to a significant degree by digital information-processing and/or computational technologies. E-Science often involves intensive use of cutting edge technologies that are advanced in technique, collaborative or on a large scale (over various possible measures: volumes of information, computational intensity, extent of distribution, variety of information types handled). E-science as a term is becoming a historic one and the term Digital Research Infrastructure may be used in its place.	EDIT
E-Science	(Historic) Science supported to a significant degree by digital information-processing and/or computational technologies. E-Science often involves intensive use of cutting edge technologies that are advanced in technique, collaborative or on a large scale (over various possible measures: volumes of information, computational intensity, extent of distribution, variety of information types handled). E-science as a term is becoming a historic one and the term Digital Research Infrastructure may be used in its place.	EDIT

Ecosystem	The technical infrastructure available, consisting of interoperable systems, within a researcher's workflow, that impacts how research data is handled and by whom. RELATED TERM Landscape.	Edit
Electronic health record	1. A compilation of core electronic health data submitted by various healthcare providers and organisations, accessible by numerous authorised parties from a number of points of care, possibly even from different jurisdictions. 2. An official health record for an individual that is shared among multiple facilities and agencies. 3. Electronic health records typically include: Contact information, Information about visits to health care professionals, Allergies, Insurance information, Family history, Immunisation status, Information about any conditions or diseases, A list of medications, Records of hospitalisation, Information about any surgeries or procedures performed. Digitised health information systems are expected to improve efficiency and quality of care and, ultimately, reduce costs. The benefits of electronic health records include: The ability to automatically share and update information among different offices and organisations, More efficient storage and retrieval, The ability to share multimedia information, such as medical imaging results, among locations, The ability to link records to sources of relevant and current research, Easier standardisation of services and patient care, Provision of decision support systems (DSS) for healthcare professionals, Less redundancy of effort, Lower cost to the medical system once implementation is complete, The governments of many countries are working to ensure that all citizens have standardised electronic health records and that all records include the same types of information. The major barrier for the adoption of electronic health records is cost. SYNONYM. Digital medical record. RELATED TERM. Electronic medical record	NR
Electronic medical record	An electronic version of the paper record that doctors have traditionally maintained for their patients and which is typically only accessible within the facility or office that controls it. RELATED TERM. Electronic health record	NR
Encoding schema	Machine processable specifications which define the structure and syntax of metadata specifications in a formal schema language.	NR
Engineering and scientific support	A technical service involved in the performance, inspection and leadership of skilled technical activities.	NR
EXtensible Markup Language	Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. SYNONYM. XML	ACCEPT
Extensible resource identifier	A defining scheme used for identification of resources (including people and organisations) and the sharing of data across domains, enterprises, and applications. XRI TC will define a Uniform Resource Identifier (URI) scheme and a corresponding Uniform Resource (URN) namespace.	NR
Extract-Transform-Load	ETL involves the following steps: (a) Extract data from homogeneous or heterogeneous data sources which are often managed by different people. An intrinsic part of the extraction involves data validation to confirm whether the data pulled from the sources have the correct/expected values; (b) Transform the data for storing it in proper format or structure for querying and analysis purposes; An important function of transformation is the cleaning of data; and, (c) Load the data into the final target (database, operational data store, data mart, or data warehouse). ETL processes can involve considerable complexity, and significant operational problems can occur. SYNONYM. ETL	NR
FAIR data	Data that is managed in such a way that it is findable, accessible, interoperable, and reusable. RELATED TERM. FAIR data principles. FAIR Guiding Principles for scientific data management and stewardship.	ADD

	Source: https://www.force11.org/group/fairgroup/fairprinciples	
FAIR data principles	Set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. RELATED TERM. FAIR data, FAIR Guiding Principles for scientific data management and stewardship.	ADD
FAIR Guiding Principles for scientific data management and stewardship	Set of foundational principles [...] that all research objects should be Findable, Accessible, Interoperable and Reusable (FAIR) both for machines and for people. Source: https://www.nature.com/articles/sdata201618. RELATED TERM. FAIR principles; FAIR data.	ADD
Fair use	A legal concept in some jurisdictions that allows the reproduction of copyrighted material for certain purposes without obtaining specific permission and without paying a fee or royalty. Purposes permitting the application of fair use generally include review, news reporting, teaching, or scholarly research. When in doubt of whether a 'fair use' condition applies to a resource, the quickest and simplest thing may be to request permission from the copyright owner.	edit
Feature extraction	Selecting specific data that are significant in some particular context.	NR
Field	A data table column name. SYNONYM. Column.	Edit
Findable	One of the four FAIR principles. Findable data and metadata can be located by humans and machines. They should be assigned a globally unique persistent identifier, be described with rich metadata, and ideally registered in a searchable catalogue or index. Source: https://www.go-fair.org/fair-principles/ RELATED TERMS: FAIR data, FAIR data principles	ADD
Fixed data	Data that are not, under normal circumstances, subject to change. Examples of fixed data include results from concluded research, medical records, and historical data. SYNONYM. Reference data; Archival data; Fixed-content data; Permanent data	ACCEPT
Foundational interoperability	Foundational interoperability allows data exchange from one information technology system to be received by another and does not require the ability for the receiving information technology system to interpret the data.	NR
Framework	A real or conceptual structure intended to serve as a support or guide for the building of something that expands the structure into something useful. The ability to make refinements may require that the design is fully known, and this is not necessarily known at the outset. “Framework” is thus sometimes used as a ‘fuzzy’ term. In computer systems, a framework is often a layered structure indicating what kind of programs can or should be built and how they would interrelate. Some computer system frameworks also include actual programs, specify programming interfaces, or offer programming tools for using the frameworks. A framework may be for a set of functions within a system and how they interrelate; the layers of an operating system; the layers of an application subsystem; how communication should be standardised at some level of a network; and so forth. A framework is generally more comprehensive than a protocol and more prescriptive than a structure. Examples of frameworks that are currently used or offered by standards bodies or companies include: Resource Description Framework (a set of rules from the World Wide Web Consortium for how to describe any Internet resource such as a Web site and its content); Internet Business Framework (a group of programs that form the technological basis for the mySAP product from SAP, the German company that markets an enterprise resource management line of products); Sender Policy Framework (a defined approach and programming for making email more secure; Zachman framework (a	NR

	logical structure intended to provide a comprehensive representation of an information technology enterprise that is independent of the tools and methods used in any particular IT business).	
Golden record	1. Single, well-defined version of all the data entities in an organisational ecosystem. In this context, a golden record is sometimes called the “single version of the truth,” where “truth” is understood to mean the reference to which data users can turn when they want to ensure that they have the correct version of a piece of information. The golden record encompasses all the data in every system of record within a particular organisation. A system of record is an information storage and retrieval system that serves as the authoritative source for a particular data element in a system containing multiple sources of the same element. To ensure data integrity, a single system of record must always exist for each and every data element. A well-maintained, current golden record should be a fundamental element of the Master Data Management policy for every enterprise. 2. The word “golden” is sometimes used in information technology to express the importance of some type of source. In the context of virtualization, for example, a golden image is a template for a virtual machine, virtual desktop, servers, or hard disk drive.	NR
Hashing	1. The transformation of a string of characters into a usually shorter fixed-length value or key that represents the original string. Hashing is used to index and retrieve items in a database because it is faster to find the item using the shorter hashed key than to find it using the original value. 2. Used in many encryption algorithms.	NR
Heat map	A two-dimensional representation of data in which values are represented by colours. Heat maps communicate relationships between data values that would be much more difficult to understand if presented numerically in a spreadsheet.	NR
High quality data	High-quality data are complete, timely, accurate, consistent, relevant, reliable, traceable, cleaned, validated, and well documented.	NR
Human-readable format	Data and code that are commented so that humans can understand what it represents, its design, and purpose.	NR
Identity ecosystem	More formally known as the National Strategy for Trusted Identities in Cyberspace, the identity ecosystem is a proposal from the United States federal government to improve identity authentication on the Internet and make online transactions safer. The proposal has four goals: To develop a comprehensive Identity Ecosystem framework; To build and implement an interoperable identity infrastructure aligned with the framework; To enhance confidence and willingness to participate in the Identity Ecosystem; To ensure the long-term success of the Identity Ecosystem; The proposal invites citizens to suggest ideas for creating the ecosystem. “We seek a future where individuals can voluntarily choose to obtain a secure, interoperable, and privacy-enhancing credential from a variety of service providers – both public and private – to authenticate themselves online for different types of transactions.” Although there has already been some resistance to the scheme from privacy advocates, it is likely that to the end user, an Identity Ecosystem will be familiar and work in much the same way as the financial ecosystem that allows people to withdraw cash from an ATM machine, even when the machine belongs to another bank in another city. The Identity Ecosystem will add another layer of security, aimed at reducing identity theft and simplifying the user experience for various other types of electronic transactions. Once the Identity Ecosystem standards are defined, the government hopes to spark adoption by making the federal government an early adopter and offering businesses financial incentives for developing, training and implementing the framework.	NR

Indigenous Data	Data that pertains to (is created or gathered by, or is about) Indigenous peoples. RELATED TERM. Indigenous Data Sovereignty.	ADD
Indigenous Data Sovereignty	The right of Indigenous peoples to own, control, access and possess data that derive from them, and which pertain to their members, knowledge systems, customs or territories. RELATED TERM. Indigenous Data.	ADD
Indirect Identifier	A data variable or value that, when combined with other identifiers, either within the same dataset or combined with another available dataset, directly discloses the identity of a single living individual, for example, age, geographical location or health condition.	ADD
Informaticist	A person who identifies, defines, and solves information problems using information systems, systems integration, and the management of human interactions with machine and data.	Accept
Information	The aggregation of data to make coherent observations about the world, meaningful data, or data arranged or interpreted in a way to provide meaning.	Accept
Information management advisor	A person having a broad knowledge of information management disciplines and who provides guidance and support to program and staff functions on all aspects of managing the information resource.	NR
Information management specialist	A person who is expert in one or more of the information management disciplines that support the effective and efficient management of information.	NR
Information silos	Heterogeneous data sources.	NR
Information technology specialist	Information systems and technology infrastructure manager, expert, or technician.	NR
Instrument	1. A tool or device that is used to do a particular task. 2. A device that is used for making measurements of something.	NR
Instrument output data	Raw electronic data generated by an instrument, analyser, or data logger before any human action on the data and before any processing of the data by automated or semi-automated 3rd-party software or algorithms. RELATED TERM. Raw data	NR
Integrated access management	A combination of business processes, policies and technologies that allows organisations to provide secure access to confidential data. Integrated access management software is used by enterprises to control the flow of sensitive data in and out of a network.	edit
Integration	1. Bringing together smaller components into a single system that functions as one. 2. Stitching together different, often disparate, subsystems so that the data contained in each becomes part of a larger, more comprehensive system that, ideally, quickly and easily shares data when needed. This often requires that organisations build a customised architecture or structure of applications to combine new or existing hardware, software and other communications.	edit
Integrity	In the context of data and network security: The assurance that information can only be accessed or modified by those authorised to do so. Measures taken to ensure integrity include controlling the physical environment of networked terminals and servers, restricting access to data, and maintaining rigorous authentication practices. Data integrity can also be threatened by environmental hazards, such as heat, dust, and electrical surges.	NR

Inter-disciplinary	A study undertaken by combining two or more distinct research disciplines. The research is based upon a conceptual model that links or integrates theoretical frameworks from those disciplines, uses study design and methodology that is not limited to any one field, and requires the use of perspectives and skills of the involved disciplines throughout multiple phases of the research process.	edit
International chemical identifier	A non-proprietary identifier for chemical substances that can be used in printed and electronic data sources thus enabling easier linking of diverse data compilations.	EDIT
International standard	A standard that is used in multiple nations and whose development process is open to representatives from all countries. Some international standards are promulgated by multinational treaty organisations (e.g., the International Telecommunications Union (ITU); the United Nations Food and Agriculture organisation (FAO)). Some international standards are promulgated by multinational non treaty organisations (e.g., the International organisation for Standardization (ISO); the International Electrotechnical Commission (IEC)). Some international standards are promulgated by organisations that originated as national industry associations, professional societies, or standards developers, but over time evolved into a global presence with multinational participation (e.g., ASTM International, SAE International, and NFPA International). Annex 4 of the World Trade organisation (WTO) Committee on Technical Barriers to Trade Report 2000 contains a good discussion of what constitutes an international standard. In short, the WTO suggests that a standard may be considered international if the processes and procedures used to develop it are transparent, open, impartial, and provide meaningful opportunities for WTO members, as a minimum, to contribute to the development of the standard so that the standard does not favour any particular suppliers, countries, or regions. Equally important, the standard must have a global relevance and use.	NR
International organisation for Standardization (ISO)	A non-governmental organisation that promotes the development of standardisation and related activities to facilitate the international exchange of goods and services, and to develop cooperation in intellectual, scientific, technological, and economic activity. A worldwide federation of national standards bodies from 143 countries. The results of ISO technical work are published as international standards. SYNONYM. ISO	EDIT
Interoperable	One of the four FAIR principles. Interoperable data is data and metadata that can be integrated with other data/metadata and can interoperate with applications or workflows for analysis, storage, and processing. Semantic and syntactic interoperability are the two main types of interoperability. Source: https://www.go-fair.org/fair-principles/ RELATED TERMS: FAIR data, FAIR principles	ADD
Knowledge	The rules and organising principles gleaned from aggregated data. The internalised or understood information that can be used to make decisions.	Accept
Landscape	Regarding research data: the broad communities of research data management and related areas that influence researcher incentives and behaviours concerning data. RELATED TERM. Ecosystem.	Edit
Legacy data	Older data that can no longer be accessed or processed easily because they are stored in obsolete formats or systems. RELATED TERMS: At Risk Data, Dark Data.	Edit
Linked open data	Data where relationships/connections between them are available to allow easy data access. A typical case of a large Linked dataset is DBPedia (http://dbpedia.org/), which essentially makes the content of Wikipedia available in RDF. This	NR

	related collection of interrelated datasets is stored on the Web and available via a common format RDF. SOURCE: http://www.w3.org/standards/semanticweb/data#summary	
Long-term preservation	Continued access to digital materials, or at least to the information contained in them, indefinitely.	NR
Machine actionable	A machine readable dataset or file format, that is structured in such a way as to allow machines to take automated programmed actions as a result. RELATED TERM. Machine readable.	ADD
Machine readable	A broad term encompassing: (a) digital surrogates created as a result of converting analogue materials to digital form (digitisation); (b) 'born digital' objects for which there has never been and is never intended to be an analogue equivalent; and, (c) digital records. RELATED TERM. Born digital; Digital objects; Digital records; Digital data; Electronic records, Machine Actionable.	Edit
maDMP	Machine actionable Data Management Plan, that involves moving away from document format, to have actionable functions such as automatic requesting of additional storage, or automatic ingest of dataset publication details.	ADD
Masking	The application of a set of data transformation techniques to de-identify data without any concern for the analytical utility of the data. This is a good approach for fields that are not required to be analysed. Masking is applied to direct identifiers such as name and phone number. Masking techniques include, among others, removal of direct identifiers or replacement of direct identifiers with pseudonyms. RELATED TERM. Anonymity.	Edit
Meaningful use	In the context of health information technology (HIT): defines minimum government standards for using electronic health records (EHR) and for exchanging patient clinical data between healthcare providers, between healthcare providers and insurers, and between healthcare providers and patients.	NR
Medium-term preservation	Continued access to digital materials beyond changes in technology for a defined period of time but not indefinitely.	NR
Metadata	Literally, “data about data”. It is data (or information) that defines and describes the characteristics of other data. It is used to improve the understanding and use of the data.	EDIT
Metadata catalogue	A catalogue containing metadata records in XML-encoded (machine-readable and human-readable) format that enables services to find data and services.	NR
Metadata dataset	The set of metadata describing a specific dataset.	NR
Metadata profile	Document that modifies a metadata standard. A profile may reduce the overall number of metadata elements defined by a standard. A profile may further restrict the optionality of a metadata element, making it mandatory where before it was optional; however, a profile cannot make mandatory elements optional. A profile may further restrict the values allowed in a metadata element. Metadata profiles can be adopted by a standards body, agency, or organisation in place of a metadata standard. Source: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/metadata/metadata-standards-and-styles.htm#ESRI_SECTION1_1F1ECCA678DF4D2EBF8E53817BE1C46F	ADD

Metadata record	A collection of data defined by a theme, category, which reflects what is being measured, observed, monitored at the various sites. The Metadata Record is an information resource of business value.	NR
Metadata standard	A high level, commonly shared representation of the metadata elements related to a dataset, collection, or other digital object. A metadata standard may also provide an XML schema describing the format in which the elements should be stored. Typically, a standard XML format is defined using XML Schema or document type definition (DTD). Standards are typically ratified by national or international standards bodies.	ADD
Migration	A means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration differs from the refreshing of storage media in that it is not always possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the resource with the new generation of technology.	NR
Minimal metadata	A description of a digital object with a limited number of fields, that would include at least a name and persistent identifier.	edit
Missing data	Data that are missing on a variable. The missing data can be missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). The reason why the data are missing informs how it should be curated (e.g. whether to impute the missing data or not).	EDIT
Namespace	Uniquely identifies a set of names so that there is no ambiguity when objects having different origins but the same names are mixed together. Using the Extensible Markup Language (XML), an XML namespace is a collection of element type and attribute names. These element types and attribute names are uniquely identified by the name of the unique XML namespace of which they are a part. In an XML document, any element type or attribute name can thus have a two-part name consisting of the name of its namespace and then its local (functional) name.	NR
Noisy data	Meaningless data, including: Any data that cannot be understood and interpreted correctly by machines, such as unstructured text; any data that has been received, stored, or changed in such a manner that it cannot be read or used by the program that originally created it. RELATED TERM. Corrupt data	NR
Non identifiable data	Data that could not lead to the identification of a specific individual, to distinguishing one person from another, or to personally identifiable information. These may be data that have been de-identified, or that could not lead to personally identifiable information in the first place. RELATED TERM. Non personally identifiable information	NR
Non personally identifiable information	Data that could not lead to the identification of a specific individual, to distinguishing one person from another, or to personally identifiable information. These may be data that have been de-identified, or that could not lead to personally identifiable information in the first place. RELATED TERM. Non identifiable data.	edit
Normalisation	Organising data into tables so that the results of using the database are always unambiguous and as intended. Normalisation is typically a refinement process after the initial exercise of identifying the data objects that should be in the database, identifying their relationships, and defining the tables required and the columns within each table. First normal form (1NF) is the "basic" level of normalisation: Data and information are contained in two-dimensional tables with rows and columns. Each column corresponds to a sub-object or an attribute of the object represented by the entire table. Each row represents a unique instance of that sub-object or attribute and must be different in some way from any other row (that is,	edit

	no duplicate rows are possible). All entries in any column must be of the same kind. For example, in the column labelled "Date," only dates are permitted. In Second normal form (2NF), the tables are in first normal form and, in addition, each column in a table that is not a determiner of the contents of another column must itself be a function of the other columns in the table. At the second normal form, modifications are still possible because a change to one row in a table may affect data that refers to this information from another table. In Third normal form (3NF), the tables are in second normal form and, in addition, there is no transitive functional dependency. For example, if A is functionally dependent on B, and B is functionally dependent on C, then C is transitively dependent on A via B. In Domain/key normal form (DKNF), a key uniquely identifies each row in a table. A domain is the set of permissible values for an attribute. By enforcing key and domain restrictions, the database is assured of being freed from modification anomalies. DKNF is the normalisation level that most designers aim to achieve.	
OAI repository	A type of repository with a network accessible server that can process the 6 OAI-PMH requests in the manner described in the OAI Implementation Guide.	NR
Object attribute	An object model that is the logical attributes or properties associated with a particular object. In a data object this would be the associated properties.	NR
Object model	A collection of descriptions of classes or interfaces, together with their member data, member functions, and class-static operations.	NR
Object property	The characteristics of any digital object can be described by a number of properties which are typically stored in metadata and/or PID records.	NR
Ontology	Shared and standardised list of words, terms and phrases to describe components of a particular discipline or domain, along with a taxonomy of their relations. Compare this to a controlled vocabularies, which tend not to include a structure of relations between their terms. Ontologies are typically developed by domain-specific institutions or communities to aid in the precise referencing of elements. RELATED TERM. Controlled vocabulary	ADD
Open Archives Initiative Protocol for Metadata Harvesting	A low-barrier mechanism for repository interoperability. Data Providers are repositories that expose structured metadata via OAI-PMH. Service Providers then make OAI-PMH service requests to harvest that metadata. OAI-PMH is a set of six verbs or services that are invoked within HTTP. SYNONYM. OAI-PMH	NR
Open data	Data that are accessible, machine-readable, usable, intelligible, and freely shared. Open data can be freely used, re-used, built on, and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike.	edit
Open science	A movement to make scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community. Source: UNESCO Recommendation on Open Science, 2021 https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en	ADD

ORCID	ORCID (Open Researcher and Contributor IDs) is a unique identifier for researchers. This unique identifier disambiguates researchers and their work and allows researchers to connect their ID with additional professional information, including affiliations, grants, and publications.	ADD
Original repository	A type of repository where the original copy of data was stored and probably a data identifier registered.	NR
Persistent identifier	A persistent identifier is a long-lasting reference to a digital object that gives information about that object regardless of what happens to it. Developed to address "link rot," a persistent identifier can be resolved to provide an appropriate representation of an object whether that object changes its online location or goes offline. SYNONYM. PID	ACCEPT
Persistent uniform resource locator	This is a URL. However, instead of pointing directly to the location of an Internet resource, a PURL points to an intermediate resolution service. The PURL resolution service associates the PURL with the actual URL and returns that URL to the client. SYNONYM. PURL	NR
Personal information privacy	The World Wide Web Consortium's Platform for Personal Privacy Project (P3P) offers specific recommendations for practices that will let users define and share personal information with Web sites that they agree to share it with. The P3P incorporates a number of industry proposals, including the Open Profiling Standard (OPS). Using software that adheres to the P3P recommendations, users will be able to create a personal profile, all or parts of which can be made accessible to a Web site as the user directs. A tool that will help a user decide whether to trust a given website with personal information is a Statement of Privacy Policy that a website can post.	NR
Personally identifiable information	1. Data which relate to a living individual who can be identified (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual. 2. Any data that could potentially identify a specific individual. Any information that can be used to distinguish one person from another and can be used for de-anonymizing anonymous data can be considered personally identifiable data. 3. Data are identifiable if the information contains the name of an individual, or other identifying items such as birth date, address or geocoding. Data will be identifiable if the information contains a unique personal identifier and the holder of the information also has the master list linking the identifiers to individuals. Data may also be identifiable because of the number of different pieces of information known about a particular individual. It may also be possible to ascertain the identity of individuals from aggregated data where there are very few individuals in a particular category. Identifiability is dependent on the amount of information held and also on the skills and technology of the holder. SYNONYM. Personal data	NR
Perturbation	A method of data anonymisation, which involves adding 'noise' to the data, so that single living individuals cannot have their identities disclosed.	ADD
PID attribute	A single data element related to a PID and part of its record content.	NR
PID domain	For a single identifier, the class of entity it refers to. For a PID system, the typical class of entities it is intended to be used for. Examples include: digital objects, physical objects, bodies, actors.	NR
PID record	A type of record (and organisation) that stores an instance of an executable/understandable PID. The content of a PID record distinguishes a registered digital or data object from other DOs. A PID record is a type of record that includes property information that characterises the digital object it is identifying. Important parts of a PID record are location and	NR

	checksum. However there is a large variation in usage. In some data models the PID is simply used as a unique label with an empty record. A PID record has a lifecycle including creation, publication, Curation and the destruction.	
PID resolution	The process of resolving a PID to a useful state of information about a digital object by using a globally available system.	NR
PID service	A service that provides a connection between a PID and its target object.	NR
PID system	Consists of at least one PID resolver, a name schema and a defined mechanism for issuing PIDs that conform to the name schema. Examples include: DOI, Handle System, URN, ARK, PURL, etc.	NR
Pipe separated values	Values in a table presented as a series of ASCII text lines organised so that each column value is separated by a pipe (|).	edit
Preservation	An activity within archiving in which specific items of data are maintained over time so that they can still be accessed and understood through changes in technology. SYNONYM. Conservation	NR
Preservation metadata	Documents actions that have been undertaken to preserve a digital resource such as migrations and checks sum calculations. Example: Metadata Encoding and Transmission Standard (METS)	NR
Privacy governance	Monitoring the risk to privacy posed by data requests from researchers, and the practices of data custodians in providing data (information governance) to ensure that confidentiality is protected. Such governance requires specialised knowledge of technology, law, and statistical methods.	edit
Privacy-preserving data linkage	Data linkage where the resulting product has been de-identified. RELATED TERM. De-identification.	NR
Proportionate governance	Keeping the procedural mechanisms that researchers and data custodians must follow when engaged in data sharing and linkage proportional to the degree of risks associated with such practices. Proportionate governance operates in situations that are too variable to be regulated by hard laws (e.g., custom data access requests). It requires that analytical judgments be performed to ensure that the governance mechanisms deployed for a given research proposal correspond to the level of risk it entails. Proportionality is an important cross-cutting consideration across all types of governance that are put in place.	NR
Proprietary	File formats of datasets that are specific to a company, organisation or individual that are not in wider use, and can therefore impact on the re-usability and preservation of datasets.	Edit
Protocols	A formal or official record of scientific experimental observations. SOURCE: https://www.protocols.io/what-is-protocol	EDIT
Provenance	A type of historical information or metadata about the origin, location or the source of something, or the history of the ownership or location of an object or resource including digital objects. For example, information about the Principal Investigator who recorded the data, and the information concerning its storage, handling, and migration.	Accept
Provenance metadata	Information concerning the creation, attribution, or version history of managed data. Provenance metadata that indicates the relationship between two versions of data objects and is generated whenever a new version of a dataset is created. Examples include: (i) the name of the program that generated the new version, (ii) the commit id of the program in a code version control system like GitHub, (iii) the identifiers of any other datasets or data objects that may have been used in creating the new version. Provenance information is gathered along the data lifecycle as part of curation processes. A finer level of provenance metadata would be concerned only with data flowing between various stores such as curated	NR

	databases and managed repositories. Provenance metadata is designed to allow queries over the relationship between versions, and includes either or both fine-grained and coarse-grained provenance data. Different applications may store different provenance data.	
Quality assurance	The process or set of processes used to measure and assure the quality of a product. SYNONYM. QA.	NR
Quality control	The efforts and processes put in place to ensure that the management of research data is of sufficient quality to allow researchers to carry out their work effectively.	EDIT
Raw data	Data that have not been processed for meaningful use. Although raw data have the potential to become information, they require selective extraction, organisation, and sometimes analysis and formatting for presentation. SYNONYM. Source data. RELATED TERM. Instrument output data. Information.	EDIT
Re-use	The re-analysis of a dataset or combination of datasets outside of the original research purpose for which the dataset was created.	Edit
Real-time data	Data that are being received, processed and stored at the time of their occurrence with only small delays. Examples include: stock quotes, manufacturing statistics, Web server loads, data warehouse activity and sensor feeds to data collectors. Real-time data are often used for navigation or tracking. Real-time data are data streams that are typically generated by sensors and received via direct networking connections.	NR
Record	(noun) 1. Collection of data items arranged for processing by a program. Multiple records are contained in a file or dataset. Typically, records can be of fixed-length or be of variable length with the length information contained within the record. 2. (noun) Sometimes called a row, a group of fields (sometimes called columns) within a table that are relevant to a specific entity. For example, in a table called 'Client contact information', a row would likely contain fields such as: ID number, name, street address, city, telephone number, etc. SYNONYM. Data record.	Edit
Record provenance information	Information for a data object that includes: * the person who deposited the data object in the repository, * the source of the data object, * the date when the object was deposited, and * authenticity information needed to link the data object to its original source.	NR
Record standardisation	A process in which files are first parsed (assigned to appropriate fields in a record) and then translated to a common format. Data often lack consistency simply because there are many ways of saying the same thing. Standardising the record ensures that when a query is run for a particular field, accurate results will be returned.	EDIT
Records retention schedule	A policy that depicts how long data items must be kept, as well as the disposal guidelines for these data items.	NR
Referable data	A type of data (digital or not) that is persistently stored and which is referred to by a persistent identifier. Digital data may be accessed by the identifier. Some data object references may access a service on the object.	edit
Reference model	A design covering a class of frameworks with the following characteristics: (1) it can be used to generate more specific models that still belong to the class and (2) it can be used to compare a concrete framework design to identify whether it belongs to the same class.	NR

Reformatting	Copying information content from one storage medium to a different storage medium (media reformatting) or converting from one file format to a different file format (file reformatting).	ACCEPT
Refreshing	Copying information content from one storage media to the same storage media.	NR
Registered data	Data that have gone through a registration process and have been assigned an identifier metadata to aid in their search and retrieval.	NR
Registry	A database containing information about trusted repositories that are provided by the repository managers and are useful for human and machine users. It is a registry information system on which a register is maintained. These registries do not contain information about all metadata descriptions of digital objects, nor do they offer a list of PIDs of all stored digital objects. They do offer information based on standardised types on how to retrieve such information (e.g., the port under which OAI-PMH can be accessed to offer metadata). It is a set of files containing identifiers assigned to items with descriptions of the associated items. It is the assignment of a permanent, unique and unambiguous identifier to an item. SYNONYM. Data Registry.	Edit
Reliability	The likelihood of observing the same result if the data were to be collected in another sample. That is, we can rely on the results to be accurate at a different time or in a different context.	EDIT
Remote data access	The ability to access and download data from a repository.	edit
Replica number	A type of metadata used as part of a replication process or access.	NR
Replication	1. Generation of a copy of a data object that is referenced by the same name, but with a different replica number. When changes are made to the data object, the replica can be updated to track the changes. As part of replication, data may be given a PID for a repository. Enhanced metadata may be stored in a repository as part of replication. A PID should allow replicated objects from different communities to be identified as such. 2. Repeated measurement of the same object. RELATED TERMS: Reproducibility, Reproducible research, Replica Number, PID, Repository.	EDIT
Repository	A physical or digital storage location that can house, preserve, manage, and provide access to many types of digital and physical materials in a variety of formats. Materials in online repositories are curated to enable search, discovery, and reuse. There must be sufficient control for the physical and digital material to be authentic, reliable, accessible and usable on a continuing basis.	edit
Representation	A resource that conveys either the content of a resource (if it is a digital object instance), or provides a digital object that conveys the intention of the resource in a form useful to a user (machine or human).	NR
Representation object	Provides some context for a data object. It contains provenance, description (e.g. format, encoding scheme, algorithm, structural, and administrative information about the object. This is a form of metadata.	NR
Reproducibility	The ability to replicate the results of a study using the same input data and procedures used by the original investigator. RELATED TERM. Reusable, Reproducible research, Replication.	ADD
Reproducible research	Published results that can be replicated using the documented data, code, and methods employed by the author or provider without the need for any additional information or needing to communicate with the author or provider. SYNONYM. Reproducibility.	Edit

Repurposed data	New datasets obtained by combining data appropriately from a variety of existing files, generating new data products that did not previously exist. Repurposed data result from data wrangling. RELATED TERM. Data wrangling.	NR
Requirements analysis	The process of determining user expectations for a program, system, dataset, or product. Requirements analysis is a team effort that must take into account hardware, software, end use, and human factors engineering expertise. Requirements analysis also requires skills in dealing with people. Requirements analysis involves frequent communication with end users to determine specific feature expectations, resolution of conflict or ambiguity in requirements as demanded by the various users or groups of users, avoidance of feature creep and documentation of all aspects of the project development process from start to finish. Energy should be directed towards ensuring that the final system or product conforms to client needs rather than attempting to mould user expectations to fit the requirements.	edit
Research data	Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data. Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.	accept
Research data lifecycle	The entire period of time that research data exists. This lifecycle describes the flow of research data starting from planning, collecting, processing, analysing, preserving, sharing and finally reusing the research data. Research data often have a longer lifespan than the research project.	EDIT
Research data management	Storage, access and preservation of data created or collected in the course of research. Research data management practices cover the entire lifecycle of the data, from planning the investigation to conducting it, and from backing up data as it is created and used to long term preservation of data deliverables after the research investigation has concluded. Specific activities and issues that fall within the category of data management include: File naming (the proper way to name computer files); data quality control and quality assurance; data access; data documentation (including levels of uncertainty); metadata creation and controlled vocabularies; data storage; data archiving and preservation; data sharing and reuse; data integrity; data security; data privacy; data rights; notebook protocols (lab or field) as required. RELATED TERM. Data stewardship.	Edit
Research data management infrastructure	The configuration of staff, services and tools assembled to support data management across the research lifecycle and more specifically to provide comprehensive coverage of the stages making up the data lifecycle. It can be organised locally and/or globally to support research data activities across the research lifecycle.	NR
Research data publication workflow	Activities and processes in a digital environment that lead to the publication of research data, associated metadata and accompanying documentation and software code on the Web. In contrast to interim or final published products, workflows are the means to curate, document, and review, and thus ensure and enhance the value of the published product. Workflows can involve both humans and machines and often humans are supported by technology as they perform steps in the workflow. Similar workflows may vary in the details depending on the research discipline, data publishing product and/or the host institution of the workflow (e.g., individual publisher/journal, institutional repository, discipline-specific repository).	NR
Research governance	Ensures that the benefits to society of research outweigh any risks, from both an ethical and legal perspective.	NR

Research metadata format	Acceptable formats for transmitting and sharing research metadata, e.g. ISO 19115-2:2009	NR
Research organisation Registry	ROR (Research organisation Registry) is a community-led registry of open, sustainable, usable, and unique identifiers for every research organisation in the world. Source: https://ror.org/about/	ADD
Research results	Research results are the journal articles, reports, books, slideshows, or websites that announce the projectís findings and try to convince us that the results are correct.	NR
Research software	Software written to create, generate, analyse or display research data.	ADD
Resource authorization	The process of deciding if a subject (person, program, device, group, role, etc.) is allowed to have access to or take an action against a resource. Authorization relies on a trusted identity (authentication) and the ability to test the privileges held by the subject against the policies or rules governing that resource to determine if an action is permitted for a subject.	NR
Retention period	A metadata operation to create state information for a data object that defines the date when retention of the data object should be evaluated. The retention period must have an associated disposition policy for deciding what to do when the retention period expires.	NR
Reusable	One of the four FAIR principles. Reusable data is data that can be utilised to replicate research findings and/or can be analysed in settings outside of the original context in which it was produced or collected. The reusability of research data can depend on its format, licensing, and the richness of the relevant metadata. Source: https://www.go-fair.org/fair-principles/ RELATED TERM. FAIR data, FAIR principles	ADD
Schema	1. The organisation or structure for a database. The activity of data modelling leads to a schema. (The plural form is schemata.) The term is used in discussing both relational databases and object-oriented databases. The term sometimes seems to refer to a visualisation of a structure and sometimes to a formal text-oriented description. Two common types of database schemata are the star schema and the snowflake schema. 2. A formal expression of an inference rule for artificial intelligence (AI) computing. The expression is a generalised axiom in which specific values or cases are substituted for each symbol in the axiom to derive a specific inference.	NR
Science and technology data	Qualitative or quantitative attributes of a variable or set of variables. Data refers to representations of physical, biological or chemical facts, typically the results of measurements/observations. It also includes related socio-economic and cultural representations. Data are normally in a structured, tabular, numeric, character, geo-referenced, and/or computer-readable format. SYNONYM. Scientific data; Technological data.	NR
Scientific data infrastructure	What is required to enable researchers to create, store and share the data resulting from their experiments, and to find, access and process the data they need. RELATED TERM. Science and technology data; Scientific data services	NR
Scientific data services	Assist organisations in the capture, storage, curation, long-term preservation, discovery, access, retrieval, aggregation, analysis, and/or visualisation of scientific data, as well as in the associated legal frameworks, to support disciplinary and multidisciplinary scientific research. RELATED TERM. Scientific data infrastructure	NR

Scientific method	Series of steps to follow for the systematic discovery of knowledge: 1. Ask a research question which is grounded in existing research and/or theory, 2. Generate a hypothesis, 3. Collect the data, 4. Analyse the data, 5. Interpret the results, 6. Report the results.	Edit
Scientific workflow	A set of chained operations. The simplest computerised scientific workflows are scripts that can involve several ingredients such as data, programs, models and other inputs such as human or sensor observations. Workflows produce outputs that may include, for example, visualisations and analytical results. Preserved workflows are important for reproducible research. They simplify complex sequences of activities and enable researchers to automate and track the provenance of the work in workflow execution. Workflow scripts are digital objects.	NR
Semantic data	Data that are tagged with particular metadata that can be used to derive relationships between data.	NR
Semantic interoperability	The ability of computer systems to transmit data with unambiguous, shared meaning. Semantic interoperability is a requirement to enable machine computable logic, inferencing, knowledge discovery, and data federation between information systems. Semantic interoperability is achieved when the information transferred has, in its communicated form, all of the meaning required for the receiving system to interpret it correctly, even when the algorithms used by the receiving system are unknown to the sending system. Syntactic interoperability is a prerequisite to semantic interoperability. Semantic interoperability ensures that the precise format and meaning of exchanged data and information is preserved and understood throughout exchanges between parties; in other words, what is sent is what is understood. RELATED TERM. Syntactic interoperability.	Edit
Semi-structured data	Data that have not been organised into a specialised repository, such as a database, but that nevertheless have associated information, such as metadata, that makes them more amenable to processing than raw data. Semi-structured data lie somewhere between structured and unstructured data. They are not organised in a complex manner that makes sophisticated access and analysis possible. However, they may have information associated with them, such as metadata tagging that allows elements contained to be addressed. Example: A Word document is generally considered to be unstructured data. However, metadata tags could be added in the form of keywords and other metadata that represent the document content and make it easier for that document to be found when people search for those terms — the data are now semi-structured. Nevertheless, the document still lacks the complex organisation of a database, so falls short of being fully structured data	NR
Sensitive data	Data that must be protected against unintended access or disclosure. This includes information regarding an individual, organisation or other entity. Source: Sensitive Data Toolkit for Researchers Part 1: Glossary of Terms for Sensitive Data used for Research Purposes. https://doi.org/10.5281/zenodo.4088946	ADD
Service object	In the context of reproducible research, a service object is a type of digital object containing executable code, considered as a unit.	NR
Short-term preservation	Short-term preservation. Access to digital materials either for a defined period of time while use is predicted but which does not extend beyond the foreseeable future and/or until it becomes inaccessible because of changes in technology.	NR
Stakeholder	Individuals, groups or organisations that have an interest or share in an undertaking or relationship and its outcome – they may be affected by it, impact or influence it, and in some way be accountable for it.	NR

Standard	A document that applies collectively to codes, specifications, recommended practices, classifications, test methods, and guides, which have been prepared by a standards developing organisation or group, and published in accordance with established procedures.	NR
Standard Operating Procedure for the collection of harmonised or integrated data	Written methods, instructions, and tools that, when applied in different data collection contexts produce data that are ready to be harmonised or integrated without further manipulation. RELATED TERM. Data harmonisation; Data integration	NR
Standardisation	The process of converting data to a common format to ease analysis and facilitate the comparison or collation of different sets of data. RELATED TERM . Data standardisation.	EDIT
Statistical de-identification	Application of data transformation techniques to de-identify data in such a manner that the resulting transformed fields retain a very high analytic value.	edit
Sticky bits	A user ownership access-right flag that can be assigned to digital objects such as directories. When the sticky bit flag is set, files added to the directory will inherit the access permissions associated with the directory.	NR
Storage location	A physical storage location where a data object will be stored upon ingestion into a data repository. This requires identifying the IP address and the physical path name within the storage location where a data object will be stored. The sequence of these chained activities is conceptualised as a workflow object. For retrieval, the data object location is specified by the storage location and the physical path name.	NR
Structural metadata	A type of metadata that indicates how compound objects are put together (e.g., how pages are ordered to form chapters; how data are organised in a table; how datasets are organised in a collection) 2. The underlying structural metadata of digital objects that tells computers how to assemble them.	NR
Structured data	Data whose elements have been organised into a consistent format and data structure within a defined data model such that the elements can be easily addressed, organised and accessed in various combinations to make better use of the information, such as in a relational database. SYNONYM. Structured information	accept
Syntactic interoperability	Syntactic interoperability defines the structure or format of data exchange and is achieved through tools such as XML or SQL Standards. RELATED TERM. Semantic interoperability.	Edit
System metadata	Digital entity properties that are generated by the data management system (e.g., creation time; owner; storage location; data retention period; the length of time a digital entity will be retained).	NR
Tab Separated Values	A file that contains the values in a table as a series of ASCII text lines organised so that each column value is separated by a TAB from the next column's value and each row starts a new line. SYNONYM. TSV	accept
Table	1. organised grouping of columns (i.e. fields). 2. In a relational database, a table (sometimes called a file) organises the information about a single topic into rows and columns. The process of normalisation determines how data will be most effectively organised into tables. 3. A decision table (often called a truth table) contains a list of decisions and the criteria on which they are based. All possible situations for decisions should be listed, and the action to take in each situation should	edit

	be specified. A decision table can be inserted into a computer program to direct its processing according to decisions made in different situations. Changes to the decision table are reflected in the program. 4. An HTML table is used to organise Web page elements spatially or to create a structure for data that is best displayed in tabular form, such as lists or specifications. RELATED TERM. Tabular data.	
Tabular data	Data that are arranged in tabular forms, in rows and columns. RELATED TERM. Table.	ADD
Technical metadata	Information describing the technical processes used to produce, or required to use a digital object	NR
Temporary version	Copy of a data object such as a file during the course of routine operations.	NR
Text file	File that contains the values in a table as a series of ASCII text lines organised so that each column value is separated by a character (e.g., pipe). SYNONYM. Comma separated values; Character separated values; Pipe separated values; TXT	NR
Tombstone record	A metadata record relating to a dataset that has been destroyed, deleted, corrupted, or otherwise made unavailable that acknowledges the past existence of the data and preserves the details of it.	ADD
Topical metadata	Describes the topic or “aboutness” of an information/data object – what are these data about. In order to make sense to an agent or systems, this may include a variety of vocabularies for describing, subjects, topics, categories, etc.	NR
Transdisciplinary	Research efforts conducted by investigators from different disciplines working jointly to create new conceptual, theoretical, methodological, and translational innovations that integrate and move beyond discipline-specific approaches to address a common problem. Transdisciplinary research transcends interdisciplinary research.	NR
Trusted Digital Repository	An infrastructure component that provides reliable, long-term access to managed digital resources. It stores, manages, and curates digital objects and returns their bit streams when a request is issued. Trusted repositories undergo regular assessments according to a set of rules such as defined by CoreTrustSeal or TRAC (ISO 16363). Such an assessment has the potential to increase trust from its depositors and users. Certain quality criteria need to be met to distinguish trusted repositories from other entities that store data, such as notebooks or lab servers.	Edit
Unified data management platform	Centralised computing system for collecting, integrating and managing large sets of structured and unstructured data from disparate sources.	NR
Uniform resource identifier	String of characters used to identify or name a resource on the Internet. Such identification enables interaction with representations of the resource over a network, typically the World Wide Web, using specific protocols. SYNONYM. URI	NR
Uniform resource namespace	Internet resource with a name that, unlike a URL, has persistent significance – that is, the owner of the URN can expect that someone else (or a program) will always be able to find the resource. A frequent problem in using the Web is that Web content is sometimes moved to a new site or a new page on the same site. Since links are made using Uniform Resource Locators (URLs), they no longer work when content is moved. SYNONYM. URN	NR
Universal Numeric Fingerprint	Unique signature of the semantic content of a digital object. It is not simply a checksum of a binary data file. Instead, the UNF algorithm approximates and normalises the data stored within. A cryptographic hash of that normalised (or canonicalized) representation is then computed. The signature is thus independent of the storage format. E.g., the same data object stored in, say, SPSS and Stata, will have the same UNF. SYNONYM. UNF. RELATED TERM. Version control.	NR

Universally Unique Identifier	128-bit number used to guarantee unique identity for different objects on the internet over time. File system partitions. SYNONYM. UUID	NR
Unstructured data	Data that have not been organised into a format and identifiable data structure that makes them easy to access and process. These data can often be searched as long as they are digital, but they are difficult to use for computer analyses. SYNONYM. Unstructured information	ACCEPT
Usable data	Data that can be used: delivered in a form that meets the needs of different end-user audiences, is ready for the tasks that the end-user needs to accomplish, and that has been adapted to the end-user's needs. Usable data have been cleaned, structured, are in machine readable format, fully documented, and ready for analysis and interpretation.	Edit
Use case	Methodology used in system analysis to identify, clarify, and organise system requirements. The use case is made up of a set of possible sequences of interactions between systems and users in a particular environment and related to a particular goal. It consists of a group of elements (e.g., classes and interfaces) that can be used together in a way that will have an effect larger than the sum of the separate elements combined. The use case should contain all system activities that have significance to the users. A use case can be thought of as a collection of possible scenarios related to a particular goal, indeed, the use case and goal are sometimes considered to be synonymous.	Accept
Verify checksum	Generate a unique reduced representation for a data object by applying a procedure and compare the result to the original reduced representation that has been stored as provenance information. Examples include: a checksum, a hash, a digital signature.	NR
Version control	Control over time of data, computer code, software, and documents that allows for the ability to revert to a previous revision, which is critical for data traceability, tracking edits, and correcting mistakes. Version control generates a (changed) copy of a data object that is uniquely labelled with a version number. The intent is to track changes to a data object, by making versioned copies. Note that a version is different from a backup copy, which is typically a copy made at a specific point in time, or a replica. SYNONYM. Source control; Revision control; Versioning. RELATED TERM. Universal numeric fingerprint; Data citation	Edit
Visualisation	The representation of a dataset in visual form, for example, a chart, diagram or picture, used to gain insights that tabular data would not provide.	ADD
Web resource	1. Addressable units of information that are addressed through Uniform Resource Identifiers (URIs). 2. The early notion of static addressable documents or files has evolved to a more generic and abstract definition. Every 'thing' or entity that can be identified, named, addressed or handled in any way whatsoever in the web at large or in any networked information system. Examples include: an electronic document or data stored on the Web, an image, a service (e.g., "a weather report), a collection of other resources. Each resource must have a URI.	accept

SCOPE STATEMENT 2022

The goal of this terminology is to gather the key terms needed for a common understanding of the research data management domain.

Research data management (RDM) refers to the storage, access and preservation of data created or collected in the course of research. Research data management practices cover the entire lifecycle of the data, from planning the investigation to conducting it, and from backing up data as it is created and used to long term preservation of data deliverables after the research investigation has concluded.

Definitions are intended to be clear and unambiguous, and where possible, fit with common usage. Definitions should be apposite across research data management activities of key stakeholders, including those ***working on research data management within the context of*** research, data management, digital curation and preservation, research management, research policy, open data advocacy, computer science, information management, research administration, library, scholarly publishing, digital archiving and research funding roles. Some terms may have more than one definition, in which case the relevant context should be specified.

If a consensus definition can be easily found elsewhere, the term is out of scope. This terminology is limited to the specific concepts necessary for a common understanding of RDM.

DELETED TERMS 2022

Terms were **removed** if they were considered by the expert WG to be beyond the published scope of the RDM Terminology. For inclusion, terms *must be specific to research data management (RDM), or have a specific application in the RDM domain, or be considered a foundational concept within RDM.* There is no obligation to attempt to list every standard, tool and concept in RDM. We aim to disambiguate and clarify, rather than define the set of everything that has been, is, or may be in use. Particularly, tools and standards with authoritative, unambiguous definitions easily found elsewhere need not be included unless they are likely to be often confused with similar terms in natural language.
Many of the removed terms appear to have been from related but distinct domains (particularly research practice, project management and computing science). Removal is no reflection on the quality of the definition writing, but entirely due to scope issues.

- Active archive
- Ad hoc
- Ad hoc testing
- Algorithm
- Analogue signals
- Applied science
- Access control list
- Analytical quality control
- Anomaly
- Architecture
- Authentication
- Behavioural competencies
- Best practice
- Black box
- Blueprint
- Bug
- Causation
- Change management
- Chief data officer
- Chief digital officer
- Chief information officer
- Chief technology officer
- Client
- Commit
- Compute intensive
- Computer code
- Computer intensive
- Computer systems

- Concept
- Confidential information
- Content information
- Correlation
- Creativity
- Darwin information typing architecture
- Data identifier - covered by ‘Digital object identifier’
- Data management - replaced by ‘Research data management’
- Data reference model
- Data munging
- Data tension
- De facto standard
- Demilitarized zone
- Deep archive
- Defect
- Digital
- Dissambuation
- Enhancement
- Error
- Error seeding
- Evaluation
- Executive
- Failure
- Firefighting
- Governance
- Governance and accountability model
- Gremlin
- Health science
- Hypermedia As The Engine Of application State
- Impact
- Import
- Incumbent based
- Indeterminate employment
- Innovation
- Intellectual leadership
- International Standards organisation (term corrected to ‘International organisation for Standardization’)
- ISO 17025
- ISO 19115
- ISO 8000

- ISO 9000
- Interoperability
- Investigation
- Key stakeholder
- Laboratory manager
- Laboratory supervisor
- Laboratory technician
- Laboratory technologist
- Machine-readable format
- Manage datasets in a repository
- Manage metadata catalog
- Manager
- Managing research
- Mandatory standard
- Mashup
- Meets requirements
- Message privacy
- Middleware
- Moof monster
- Murphy’s law
- National standard
- Negative testing
- Open government
- Operational management
- organisational leadership
- Peer review
- Physical science
- Principal Investigator
- Preprint
- Pretty Good Privacy fingerprint
- Pretty Good Privacy ID
- Privacy
- Process
- Productivity
- Professional standard
- Program
- Program governance
- Program manager
- Project

- Project lifecycle
- Project management lifecycle
- Project manager
- Project quality control
- Project team member
- Recognition
- Redundancy
- Related scientific activities
- Relational database
- Relations
- Repeatable process
- Repository access
- Representation and client services
- Requirements
- Requirements creep
- Requirements stability index
- Research
- Research and development
- Research context
- Research data format
- Research, development and analysis
- Research manager
- Research scientist
- Researcher level
- Researcher promotion documentation
- Resistance management
- Resource
- Responsibility
- Result
- Revision control system
- Robustness
- Role
- Science
- Scientist
- Services
- Silver bullet
- SMART
- Specialty [sic]
- Standard Operating Procedure

- Steering committee
- Strategy
- Support service
- SWOT
- System
- Technique
- Technology
- Tool
- Total Quality Management
- University teaching
- Use metadata
- User acceptance testing
- Valued outcome
- View
- Voluntary standard