"Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results."

University of Edinburgh (http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt/data-mgmt/research-data-definition)

"**Research data** is defined as **recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings**; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created."

Engineering and Physical Sciences Research Council (EPSRC)

http://www.epsrc.ac.uk/about/standards/researchdata/Pages/scope.aspx

---

A detailed description - Dr Jonathan Tedds, Senior Research Liaison Manager (IT Services), University of Leicester:

A typical example from physical sciences (astronomy) distinguishes between broad categories within the research data spectrum:

1. raw/initially processed data produced at a research facility such as an observatory
   a. typically made publically available in this format after an embargo period of e.g. 1 year
   b. in some cases available immediately - e.g. Swift Gamma Ray Burst satellite

2. 'research ready' processed data which has been fully calibrated, combined and cleaned/annotated
   a. often produced by individuals or collaborations
   b. rarely available to anyone outside the collaboration except upon request/collaboration
   c. but needed if you want to reuse for science unless you have detailed sub domain specific knowledge and detailed contextual information to reproduce from raw
   d. considered to enable a competitive advantage for the researchers involved
   e. may well generate future additional samples and papers for the owning collaboration on top of the original published result(s)
   f. in some cases may be produced by dedicated data scientists on behalf of the community for major survey/missions e.g. ESA XMM-Survey Science Centre (Leicester), NASA…

3. published output dataset – following detailed analysis of research ready datasets
   a. forms the data under the graph in a journal publication following analysis of research ready datasets
   b. rarely available to anyone outside the collaboration except upon request/collaboration
   c. may well generate future additional samples and papers for the owning collaboration on top of the original
   d. other researchers may request the data for their own research but may not get it!

4. published catalogue type representation of published output dataset
   a. optional in many cases, mandatory for most major surveys
   b. usually made available via project specific online resource

c. may be provided as table of parameters based on research ready dataset, usually linked from and associated with a journal
d. specifically produced in order for the wider community to reuse (cite!) and repurpose if wanted
e. The well-known Sloan Digital Sky Survey is a classic example or more recently the 2XMMi X-ray catalogue I have a close involvement with (largest X-ray survey of the sky).

Defining research data

Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results.

Classification of research data

Research data can be generated for different purposes and through different processes (Research Information Network classification):

- Observational: data captured in real-time, usually irreplaceable. For example, sensor data, survey data, sample data, neuroimages.
- Experimental: ldata from lab equipment, often reproducible, but can be expensive. For example, gene sequences, chromatograms, toroid magnetic field data.
- Simulation: data generated from test models where model and metadata are more important than output data. For example, climate models, economic models.
- Derived or compiled: data is reproducible but expensive. For example, text and data mining, compiled database, 3D models.
- Reference or canonical: a (static or organic) conglomeration or collection of smaller (peer-reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals.

Research data formats

Research data comes in many varied formats:

- Text - flat text files, Word, Portable Document Format (PDF), Rich Text Format (RTF), Extensible Markup Languague (XML).
- Numerical - Statistical Package for the Social Sciences (SPSS), Stata, Excel.
- Multimedia - jpeg, tiff, dicom, mpeg, quicktime.
- Models - 3D, statistical.
- Software - Java, C.
- Discipline specific - Flexible Image Transport System (FITS) in astronomy, Crystallographic Information File (CIF) in chemistry.
- Instrument specific - Olympus Confocal Microscope Data Format, Carl Zeiss Digital Microscopic Image Format (ZVI).

Research data (traditional and electronic research) may include all of the following:

- Documents (text, Word), spreadsheets
- Laboratory notebooks, field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Photographs, films
- Test responses

- Slides, artefacts, specimens, samples
- Collection of digital objects acquired and generated during the process of research
- Data files
- Database contents (video, audio, text, images)
- Models, algorithms, scripts
- Contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
- Methodologies and workflows
- Standard operating procedures and protocols

The following **research records** may also be important to manage during and beyond the life of a project:

- Correspondence (electronic mail and paper-based correspondence)
- Project files
- Grant applications
- Ethics applications
- Technical reports
- Research reports
- Master lists
- Signed consent forms

University of Edinburgh

http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt/data-mgmt/research-data-definition

That which is collected, observed, or created in a digital form, for purposes of analysing to produce original research results.

University of Edinburgh

http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/data-repository/definitions

Data that are descriptive of the research object, or are the object itself.

University of Bath

http://wiki.bath.ac.uk/display/ERIMterminology/ERIM%20Terminology%20V4

What is research data?

All researchers work with data, but what you call data will depend on your discipline. As a humanities scholar you might talk about your primary sources or texts. If your research is in a social science, you may think in terms of survey results, interviews and statistics. You will probably have different terms again for the outputs of your experiments and observations if you are a scientist.

Research data can be qualitative or quantitative, and comes in print, digital and physical formats. Sometimes research involves using existing data, or you may be collecting or creating new data yourself. In all cases, your research data needs to be cared for so that the results of your research can be validated and built upon.

Researchers in almost all disciplines now create data in digital form. These data can come in many guises: for example, the measurements recorded by environmental monitoring satellites, the products of collisions between fundamental particles, the sequences of entire genomes, the results of social science surveys and interviews, the annotated images of ancient Greek inscriptions or the annotated videos of innovative dance routines.

What Is Research Data?
Data are distinct pieces of information, usually formatted in a special way. Strictly speaking, data is the plural of *datum*, a single piece of information. In practice, however, people use *data* as both the singular and plural form of the word. In database management systems, data files are the files that store the database information.
Research data is data that is collected, observed, or created, for purposes of analysis to produce original research results. The word "data" is used throughout this site to refer to research data.
Research data can be generated for different purposes and through different processes, and can be divided into different **categories**. Each category may require a different type of data management plan.

- **Observational**: data captured in real-time, usually irreplaceable. For example, sensor data, survey data, sample data, neurological images.
- **Experimental**: data from lab equipment, often reproducible, but can be expensive. For example, gene sequences, chromatograms, toroid magnetic field data.
- **Simulation**: data generated from test models where model and metadata are more important than output data. For example, climate models, economic models.
- **Derived or compiled**: data is reproducible but expensive. For example, text and data mining, compiled database, 3D models.
- **Reference or canonical**: a (static or organic) conglomeration or collection of smaller (peer-reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals.

Research data may include all of the following:
- Text or Word documents, spreadsheets
- Laboratory notebooks, field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Photographs, films
- Test responses
- Slides, artifacts, specimens, samples
- Collection of digital objects acquired and generated during the process of research

- Data files
- Database contents including video, audio, text, images
- Models, algorithms, scripts
- Contents of an application such as input, output, log files for analysis software, simulation software, schemas
- Methodologies and workflows
- Standard operating procedures and protocols

The following **research records** may also be important to manage during and beyond the life of a project:
- Correspondence including electronic mail and paper-based correspondence
- Project files
- Grant applications
- Ethics applications
- Technical reports
- Research reports
- Master lists
- Signed consent forms

Boston University

http://www.bu.edu/datamanagement/background/whatisdata/

Definition of Research Data

For the purposes of the KRDS2 study research data is defined as collections of structured digital data from any disciplines or sources which can be used by academic researchers to undertake their research or provides an evidential record of their research. Research data may be created in a number of different contexts: for reasons entirely unrelated to academic research; for academic research or as a by product of (academic) research. It includes a great variety and heterogeneity of data and its accompanying metadata and documentation to make it usable and understood, or the digital representations and records for physical research data. In essence any type of research data already held in data repositories would be in scope. Examples could include: complex data used in climate modelling, aerodynamics, molecular modelling, bioinformatics; video and image archives used in archaeology, art history, anthropology and performance works; digital images/investigatory data of primary physical sources in the humanities; quantitative and qualitative data used in the social sciences; or electronic data and indices for fossils or skin tissue samples.

Neil Beagrie, Brian Lavoie and Matthew Woollard
KEEPING RESEARCH DATA SAFE 2, KRDS2 DATA SURVEY – SELECTION CRITERIA, Review Draft - 31 July 2009

"A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing."

Digital Curation Centre

"Any information you use in your research"

Cambridge Prepare Project

See "What is data" presentation -
http://www.lib.cam.ac.uk/dataman/PrePARe/Whatisdata/PrePARe_Whatisdata.pdf