Research data explained

The primary purpose of research data is to provide the information necessary to support or validate a research project's observations, findings or outputs. This resource will introduce you to the concept of research data, and how data differ from other types of information. After completion you will:

• Recognise the importance of managing research data

• Be able to distinguish between various types of research data

# What is research data?

## Why is research data and its management important?

Research data is one of the fundamental building blocks of research, because it is the evidence that substantiates research findings. As such, research data is a crucial asset that must be managed. The University believes that research data management is important because:

“Whilst good data management is fundamental for high quality research data and therefore research excellence, it is crucial for facilitating data sharing and ensuring the sustainability and accessibility of data in the long-term and therefore their re-use” (Van den Eynden, V., Corti, L., Woolard, M., Bishop, L. and Horton, L. (2011). Managing and sharing data: Best practice for researchers. UK Data Archive, University of Essex: Essex.)

## Research data management enables:

* Good research practice,
* Effective and efficient research processes,
* The University’s strategic goals of quality, impact and integrity,
* Effective governance and risk management,
* Compliance with University, funders’ and publishers’ policies on research data and data sharing.

## Research data in context

Research data is the evidence that underpins the answer to the research question and can be used to validate findings regardless of its form (e.g. print, digital, or physical).

Research data takes many forms across different subject areas and can mean different things to different people in different contexts.

Within the arts, research data can be evidence of an identified research activity and can include preparatory, unfinished and supportive work.

## Digital and non-digital data

Not all research data is digital. It is also important to consider the management of non-digital data such as hand-drawn sketches, hand-written laboratory notebooks, artefacts, geological samples, medical specimens, etc. To manage the risk of loss, one option is to digitise such data. Guidance on digitising analogue data is available from Jisc Digital Media. If the data or physical object absolutely cannot be digitised, consider other options for protection, such as a fireproof safe in a secure location.

## When does something become research data?

Research data can also be situational - the same digital information may be data for some research questions but not others. For example:

* A photographic image of an old municipal building in a historical archive is an archived image in an image bank. But when used by a researcher to study the history of a city, the photographic image becomes data for that researcher.
* CCTV footage may be archived (or destroyed) by a security firm. However, when used by a researcher to study human behaviour or 21st century surveillance methods, the video footage becomes data for that researcher.

Data can be used by others for a completely different purpose than originally intended; this facilitates innovation.

# Research data types

Research data can comprise of many objects, such as:

* Documents (text, Word), spreadsheets
* Laboratory notebooks, field notebooks, diaries
* Questionnaires, transcripts, codebooks
* Audiotapes, videotapes
* Photographs, films
* Test responses
* Slides, artefacts, specimens, samples
* Artwork, an interpretation, archives, published texts or a manuscript
* Collection of digital objects acquired and generated during the process of research
* Statistical or other 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

## Classifications of research data

Research data can be classified into five categories. Think about different examples of data that would fit within each classification.

* Observational

Observational data results from monitoring events, usually at a specific time and place. This data, therefore, can be unique and irreplaceable. In an observational study, a researcher collects data based only on what is observed and does not interfere with the subjects or variables.

**Examples include** weather measurements, species counts, ethnographies

1. Experimental

Experimental data is generated in controlled environments. This data, therefore, is often reproducible, but it can be expensive. In an experimental study, a researcher applies treatments to the subjects and collects data based only on the effect of the treatments.

**Examples include** measurements from particle accelerators; measurement of crop yields for plants that have been subjected to different levels of ozone exposure.

* Simulation

Simulation data is generated from test models where the model and its associated metadata may be more important than the output data arising from the model.

**Examples include** climate models; economic models.

* Derived

Derived data results from processing data or combining data from multiple sources. This data, therefore, is often reproducible, but it can be expensive. Examples include: compiled databases; compiling outputs from text mining; aggregating census data; combining data from existing data sources with new data.

* Reference

Reference data are collections of datasets that have most probably been published and curated.

**Examples include** gene sequence databanks; chemical structure databases

## Research data formats

Research data come in a variety of forms and physical file formats. Below are some formats with specific examples.

* **Text:** Plain text files, MS Word, Portable Document Format (PDF), Rich Text Format (RTF), Hyper-Text Markup Language (HTML), Extensible Markup Language (XML), flat files such as EMBL
* **Numerical:** MS Excel, SPSS, Stata, SAS, flat files (fixed field format files, delimited files), hierarchical files
* **Multimedia:** JPEG, TIFF, GIF, Dicom, MPEG, Quicktime, Bitmap, PNG
* **Models:** 3D, statistical, similitude, macroeconomic, causal
* **Software:** Binary and code files written in a variety of programming/scripting languages such as: Java, C, Perl, Python, RubyPHP
* **Discipline specific:** Astronomy: Flexible Image Transport System (FITS); Chemistry: Crystallographic Information File (CIF); Meteorology: GRIdded Binary (GRIB);
* **Instrument specific:** Zeiss Vision Image (ZVI), Olympus Image Binary (OIB)

## Primary and secondary data

Primary data (sometimes referred to as raw data) is original evidence collected by the investigator conducting the research.

Secondary data is derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger dataset) and is therefore one step removed from the original data source.

Take a minute to watch [this video](https://youtu.be/G8lonzDwYsQ?feature=shared) of Professor John MacInnes describe in more detail the difference between the two.

### Limitations of secondary data

There are limitations of secondary data:

* You don’t have control over the questions that have been asked - they may not be quite the questions you really wanted to ask.
* It can be time-consuming to understand and organise secondary data before you begin your own analysis

In [this video](https://youtu.be/xlQMVV7VJtA?feature=shared), Professor John MacInnes talks about the limitations of secondary data

# Research Records

This section explores what research records are and why you need to keep them.

Research data is an essential component of the research record. Research records contain recorded information about research activity, and are retained as evidence of that activity, or because they have other informational value. Research records can be organised into the following categories:

* **The Research Process:** research protocols, technical appendices, ethics applications, approvals granted consent forms.
* **Research outcomes or progress:** technical reports, research publications, social media communications such as blogs, wikis, tweets
* **Management of the research process / projects:** grant applications, contracts, financial records, staff records, correspondence (email and paper-based)
* **Research data in both primary and secondary formats:** written notes, completed questionnaires, audio/video recordings, images, instrument readings, databases, samples.

## Why are research records important?

It is important to manage research records both during and beyond the life of a project.

Complete, authentic and reliable research records are required to:

* demonstrate good research practice and strengthen the reliability of research evidence
* demonstrate compliance with legislation, regulations and other requirements
* protect individual and institutional intellectual property rights
* safeguard researchers and institutions from allegations of research misconduct
* demonstrate effective stewardship of resources to auditors and to research sponsors
* The University’s Records Management Office provides guidance on records retention.
* Maintaining an effective record of your research data also ensures that your data can be discovered, accessed, understood, and re-used in the future, either by yourself or others.

# Summary

Research data is the evidence that underpins the answer to the research question and can be used to validate findings regardless of its form (e.g. print, digital, or physical). By understanding research data and how to manage it, you will:

* Improve the efficiency of your research process
* Make your research more reproducible
* Increase the impact of your research by enabling it to be used and built upon, by yourself and others

You have reached the end of this resource. You can find more help and support on the University’s Research Data Management pages on the Library website. If you have any questions about research data management, you can contact:

**Email:** researchdata@manchester.ac.uk

[UK Data Service](https://ukdataservice.ac.uk/): provides information and guidance on all aspects of data management and sharing.

[Research data bootcamp](https://data.blogs.bristol.ac.uk/bootcamp/): an online tutorial that offers an elementary introduction to the key facets of research data management. The tutorial should take about 30 minutes to complete. This resource is provided by the University of Bristol.

[RDMRose](https://orda.shef.ac.uk/collections/RDMRose/5859834): Jisc funded project to produce learning materials in research data management tailored for information professionals