Preserving Data
More data available increases research productivity. You cannot reduce the time it takes to collect new data, but if the data from previous studies were available, you may not need to gather new data at all.

Increases the impact of your research as you will be cited by everyone who reuses your data.

Raise your research profile which may affect your employability and the success of future research funding bids.

You can also rest assured that your data will be safe in the long term so your future self can also reuse your data.
NOT EVERYTHING

Select data that is
• difficult to reproduce, if it is even reproducible,
• tools that took months, perhaps years, to produce,
• any supporting data,
• electronic versions of results.

What data and tools would have saved you time if they were available to use when you started your project?

What do you need to redo your work and/or expand on it?
It is important to retain quality data and tools of independent merit.

- your raw data,
- your reduced data, if it is at risk from software obsolescence,
- tools and scripts,
- design files,
- finalised measurements,
- supporting documentation and metadata.

Make sure there’s enough information in your publication to reuse these data and therefore verify your work and use it for a concept study or comparison set.

If they can compare easily, they will and credit your research.
Research Data Management
Preserving Data

Data in support...

- The data used to make the plots

- Full catalogues in digital form for easy reuse – abstracts of the catalogue may be printed, but excessive data sets are expensive to print in full.

- Electronic versions of images which contain additional 3D information.

- Videos of models – compressed to limit file size but retaining the essence.
If you have information that is not vital to the study, then don’t share it, anonymise it.

- Remove direct identifiers (e.g., personal information such as addresses)
- Aggregate or reduce the precision of variables that might be identifiable (such as postcode).
- Generalise text variables to reduce identifiability
- Restrict continuous variables to reduce outliers
- Pay particular attention to anonymising relational data - some anonymised variables may become identifiable when considered in combination.

Whenever editing is done, researchers need to be aware of the potential for distorting the data. For example, deleting all possible identifiers from text or sound recordings is a simple but blunt tool that creates data that are confidential but may be unusable.