



Contribution ID: 195

Type: **Poster**

## Measuring Data Matters!!

*Monday 13 October 2025 19:10 (20 minutes)*

The Macro View, reported at IDW2023, set out to estimate the national scale of research data that is under management for the purpose of future access in Australia and New Zealand. Two key observations can be made:

1. The participating institutions lacked internal reports on data as an asset, from which a total could be easily aggregated. Instead one off measurement tasks were undertaken.
2. While data was definitely counted, non-data digital content was also being counted.

Further work with a small subset of the institutions, revealed that an expected rising and falling of data volumes as a research project proceeds was never detected in practice. The events that might cause reduction of data towards a small set of refined outputs, either didn't occur in practice, or did not result in the deletion of the intermediate data. Instead, the operative policy could be summarised as "if researchers don't delete it - keep it".

We therefore postulate the existence of a significant volume of content held in institutional archives that is not research or scientific data. However, no measure of its extent is available.

We propose to label this content as 'the digital debris of research', given it arises from the day to day practice of performing research. Some of the digital debris is inherent such as copies of downloaded material, intermediate error filled software versions and their test outputs, and redundant faulty data superseded by correctly gathered data. Some examples of debris are more difficult to evaluate. For example, older data can lose its ability to influence the advancement of knowledge as that knowledge does in fact improve. This might involve prior versions of data falling into disuse as instruments and analysis evolve, creating simply better data (eg. The human reference genome and its downstream by-products is at version 38).

Data creating and supplying entities such as terrestrial observing platforms or population scale genomic libraries, know the digital objects they hold are data, and can measure their data and its use and reuse patterns directly. It appears the research performing institutions, by direct observation during the effort to establish the Macro View, could not. They counted data and debris together because as data volumes have grown, extensive homogeneous file systems have been developed to underpin their research activity. This means that the way we understand data from the experience of our formal data collections, is an incomplete narrative when applied to the management of digital content in research performing institutions. For instance, the digital debris of research retained within the digital corpus under management in an institution, should, most likely, not be made FAIR.

Our poster will highlight results from initial investigations into institutional data practice that support the following two provocations:

1. The digital debris of research is real, accumulates endlessly and over time uselessly and by doing so, renders curating valuable data held with it, increasingly inefficient and impractical.
2. Research performing institutions need to develop a debris policy to enhance their data policy if the desire to maximise the reuse of valuable data is to be realised.

Data and debris intermingle in the day to day research process within institutions, and therefore in institutional archives. Because they should be treated very differently, this is a major unresolved complexity in our research data management practice.

A possible response would be to articulate the life cycle of research debris as distinct to the life cycle of research data, develop policies and guidelines that can separate the pathways for data and debris, and measure all aspects of the journey they each take.

**Primary authors:** SOO, Ai Lin (UNSW Sydney); Dr RYE, Claire (The University of Auckland); Dr CERON, Isabel; BETBEDER-MATIBET, Luc (UNSW Sydney); JONES, Nick (The University of Auckland); Dr FRANCIS, Rhys

**Presenters:** SOO, Ai Lin (UNSW Sydney); Dr RYE, Claire (The University of Auckland); Dr FRANCIS, Rhys

**Session Classification:** Poster Session

**Track Classification:** SciDataCon Persistent Themes: Policy and Practice of Data in Research