

11 April 2023

Dear Professor O’Kane,

We write on behalf of the Research Data Culture Conversation¹, an ongoing activity initiated five years ago by research data and infrastructure managers at Monash University, the University of Melbourne, the University of New South Wales, the University of Queensland and the University of Sydney. As part of articulating the emerging challenges of managing research data at scale we have measured the growth in data in the sector.

We believe the Accord can set in motion activities to address the challenges of research data growth, now and over the coming decades, leading to a shared agenda, aligned planning and investment by governments, universities and other research bodies. Improved data management does lead to knowledge creation and collaboration. However, the current unmanaged research data growth creates a long term resource challenge for universities and Governments, including the Government’s NCRIS program.

Last year we invited a wide range of research sector institutions to answer the following question:

“What volume of unique data is being intentionally managed by your institution for the purpose of future access”

The work took approximately six months. Notwithstanding that sixty seven institutions, including universities, medical research institutions, CSIRO and national research facilities, were involved in meetings and discussion, not one was able to answer that question. However, we were able to collect information on the total corpus of digital research content produced through research project activity that is being managed for future access.

As of December 2021 we estimate that the total content managed is at least three hundred petabytes, and we estimate it to be growing at about twenty five percent per year and doubling every three years.

The creation of this corpus was funded by research schemes but at the end of the schemes, and after operational and legal retention requirements are met, our Universities believe they are obliged to continue to retain it, in case the corpus might be valuable. *The result is an unbounded unfunded unending liability.*

We can also report that our survey respondents were unable to respond with the fraction of their holdings that is in fact valuable, what fraction is sensitive, or what part of the whole is original or copies. The observation we make is that while research data is known to be a valuable asset, we are currently unable to report very much at all about it and its properties.

Given that research data is a key component of our national and global stock of knowledge, a far better understanding of it is needed. This understanding belongs in the Accord between universities and the Australian Government because of the role research data plays in the value creation of universities, a role that is expected to continue and grow. Further, the stewardship of a culture and practice that enables the cost efficient value of research data to be realised, for Australia, requires a systemic response.

Yours Faithfully

Ai-Lin Soo, Coordinator RDCC, (contact: ai.lin.soo@unsw.edu.au)

Luc Betbeder-Matibet, Chair RDCC

Rhys Francis, Facilitator RDCC

¹ <https://www.researchdataculture.org/>