

# How and where are you preserving your raw machine observable data for future science?

## [PRESENTATIONS FROM SESSION ARE HERE](#)

Pertinent in a wide range of scientific disciplines is the advent of 'next generation' instrumentation routinely collecting data at up to 16 times current data volumes. These data stem from a wide array of data types (including images), acquisition methods, analytical scales, sample or observed materials, and research applications. While many disciplines are developing best practice reporting schema and vocabularies for the data outputs, these are typically restricted to processed data most suitable for end user interaction, application and publication. Primary Observable Datasets (PODs), the fundamental observations, analyses and models underpinning the published data, generally remain inaccessible and unlinked to the published outcome, limiting opportunities for future repurposing or recalculation with evolving techniques and knowledge. This session will explore how frequently PODs remain uncaptured across disciplines, what forms of PODs are generated, and if existing systems for capturing PODs may inform or be translated to other domains. We invite a diverse audience from across the research ecosystem, from researchers who utilise the final data products (are there instances that PODs data has been required, yet unavailable/lost?), analytical facilities which generate data, databases and repositories, and research or data communities and infrastructure to attend and contribute in discussions. The session will be conducted as an open-floor discussion with attendees led by representatives of CODATA, AuScope, NCI and ACCESS-NRI. Key takeaways will inform data infrastructures on the state of PODs storage, necessity and complexity across disciplines, and how this may be addressed to benefit the community.

## 2025 eResearch Australasia Conference – BoF Agenda (60 minutes)

| Time allowed (minutes) | Item/Activity                                                        | Leader/Speaker                              | Format (e.g. Presentation/Panel Discussion/Group Work/Open Forum Discussion/Lightning Talk) |
|------------------------|----------------------------------------------------------------------|---------------------------------------------|---------------------------------------------------------------------------------------------|
| 5                      | Introduction & Welcome                                               | Bryant Ware/Angus Nixon                     | Presentation                                                                                |
| 5                      | The Importance of Variable descriptions and Metadata from initiation | Simon Hodson<br>CODATA                      | Presentation                                                                                |
| 5                      | Discipline PODs Overview - Geochemistry                              | Angus Nixon<br>AuScope Geochemistry Network | Presentation                                                                                |
| 5                      | Discipline PODs Overview - Geophysics                                | Lesley Wyborn<br>NCI                        | Presentation                                                                                |

|    |                                                                           |                             |                       |
|----|---------------------------------------------------------------------------|-----------------------------|-----------------------|
| 5  | Discipline PODs Overview<br>- Climate                                     | Kelsey Druken<br>ACCESS-NRI | Presentation          |
| 35 | Community discussion of<br>state of PODs across<br>scientific disciplines | Bryant Ware/Angus Nixon     | Open Forum Discussion |
| 5  | Wrap-Up                                                                   | Bryant Ware/Angus Nixon     | Open Forum Discussion |

### **Birds of a Feather - Discussion Notes:**

- Are there initiatives to preserve PODs in your community?
  - Auscope is writing into their contracts that is not just the data products be preserved, but also the primary observable data
  - IMOS (complicated of course) but they have a project currently to look at raw/primary source data - same scenario, no system set up but have gotten input from the community of there being a need for this
- What are the challenges for preserving PODs? - Technology
  - What about instrumentation that reduces the data within the 'instrument' - a need to engage with instrument providers
- What are the challenges for preserving PODs? - Social
  - Engaging with editors at journals and what people should be publishing, top down
  - International scientific unions
- What are the challenges for preserving PODs? - Economic
  - Raw data is very big and there is a lot of data reduction that needs to happen, 'can we keep all this data'? 'can new science be down from the raw data with new reduction methods'?
  - Sometimes (if there is plentiful sample material around) cheaper to just measure sample again
- What should we call these concepts?
  - Primary Observation Data (POD)
  - Primary Observable Variable (POV)
  - Primary Observable Property (POP)
  - Raw
  - Native
  - ?
- Desire to continue conversations and think about avenues to continue the discussion (RDA, CODATA, etc.)