Biodata Infrastructure within Australia and Beyond: Landscapes and horizons

Jeff Christiansen[‡,§,|,¶], Kathryn Ann Hall[#,|,¤]

‡ Australian BioCommons, Melbourne, Australia
§ QCIF, Brisbane, Australia
| The University of Queensland, Brisbane, Australia
¶ University of Melbourne, Melbourne, Australia
# CSIRO, Altas of Living Australia (ALA), Dutton Park, Australia
¤ Queensland Museum, South Brisbane, Australia

Corresponding author: Jeff Christiansen (jeff@biocommons.org.au)

Abstract

In current life science practice, digital data are associated with all parts of the research lifecycle. Generation and management of data are planned for during project conception; collected from numerous instruments or existing sources; prepared for analysis and analysed to generate new knowledge and information; and then (hopefully) preserved so that the data may be found, shared and re-used by others when appropriate.

This session will begin with a scan of the biodata and biodata infrastructure landscape within Australia. We will explore which organisations fund biodata generation, where data are processed and stored, and how data are made available for reuse by others. Important global and complementary data resources that are hosted offshore will also be discussed. To guarantee reproducibility and integrity for life sciences research, it is critical that each of these infrastructures (whether they are hosted on- or off-shore) are maintained for the long term.

As an example of a resource that utilises a mixture of existing on- and off-shore data infrastructures to underpin a critical research need, the Australian Reference Genome Atlas (ARGA) will be discussed. ARGA is solving the problem of genomics data obscurity for Australian-relevant species by creating an online platform where life sciences researchers can comprehensively and confidently search for genomic data for taxa relevant to Australian research. Publicly available genomics (and genetics) data are aggregated and indexed from multiple sources (both on- and off-shore), and then integrated with occurrence records and the taxonomic frameworks of the Global Biodiversity Information Facility (GBIF) and the Atlas of Living Australia (ALA) to enrich the genomic data and make them searchable using taxonomy, location, ecological characteristics and selected phenotypic data. The presentation sets the scene for a subsequent talk by members of the Global Biodata Coalition (GBC), who will outline the challenges in sustaining the types of disseminated infrastructure discussed and the

GBC's work with the funders who support many of these resources to ensure long-term funding for existing infrastructure, while also channelling support to underpin future growth in data volumes and new technologies.

Keywords

data resources, indexing service, genomics data, genetic data, sustainability

Presenting author

Jeff Christiansen and Kathryn Hall

Presented at

TDWG 2023

Acknowledgements

Australian BioCommons is supported by NCRIS funding via Bioplatforms Australia.

The Australian Reference Genome Atlas (ARGA) is a joint project of Australian BioCommons, Bioplatforms Australia, Atlas of Living Australia and Australian Research Data Commons and is supported by NCRIS funding through these organisations.

Conflicts of interest

The authors have declared that no competing interests exist.