

Progress Report CF 2021-045

BIO data collation program

BCS Biodiversity Information Office

Project Core Team

X X **Supervising Scientist** Helen Ensikat
Data Custodian Helen Ensikat

Project status as of July 4, 2023, 1:24 p.m.

X X Update requested

Document endorsements and approvals as of July 4, 2023, 1:24 p.m.

X X
Project Team granted
Program Leader granted
Directorate required

BIO data collation program

A Barker, H Ensikat, K Grogan, Z Huq, S Marcus, D Murphy, H Pichette, C Piper, N Panine, R Cechner

Context

The data collation program of the Biodiversity Information Office (BIO) brings together data from a range of data sources and providers across industry, government, the public, research sectors and community organisations. BIO is initially focusing on high volume, high quality datasets, particularly those that are not easily accessible at present, driving a step-change in the availability of biodiversity data in Western Australia. The data is ingested into the platform after passing automated quality assurance checks, where it is mapped to the Darwin Core biodiversity data standard and undergoes a human-mediated quality control process before being released to users.

The initial datasets ingested into BIO's Dandjoo biodiversity data sharing platform have been sourced directly from industry, DBCA repositories, the Western Australian Museum, and regulators such as the Department of Water and the Environment (DWER). New data will continually be ingested, through future automated delivery from regulators and BIO's rolling program to identify and source other valuable datasets. This data collation program will involve outreach to data custodians across all sectors, exploration of technologies to unlock data in pre-digital documents and ongoing monitoring of existing data to identify temporal and spatial gaps.

Aims

- Provide users with access to a rich collection of high-quality datasets.
- Ingest new and up-to-date data over time.
- Enable access to previously undiscoverable and inaccessible datasets held by government and other sectors.

Progress

- Ingestion of over 1.8 million biodiversity records into Dandjoo; BIO's curatorial work on these records is driving a State-wide uplift in data quality, as many have not undergone curation since they were first collected.
- Progress on the acquisition of Western Australian biodiversity records from a range of international, State and Territory collections so they can be made available to Western Australian data users.
- Commencement of a joint initiative with the Western Australian Museum to accelerate the databasing of a substantial backlog of collections records so these can be ingested into Dandjoo and also support the Museum's research work.
- Continuing collaboration with DWER on the one-off ingestion of historical environmental assessment data and re-engineering regulatory processes to allow for the automated ingestion of new data into the BIO platform.

Management implications

- Access to a greater range of high-quality biodiversity data, including datasets that have not been available in the past, will support better-informed research and decision-making.
- Availability of up-to-date environmental assessment data, mapped to common standards and validated via the curation process, will ensure that environmental decisions are based on current and accurate information.

Future directions

- Continue to ingest high quality datasets available within DBCA, and from the Western Australian Museum and DWER.

- Progress mapping of systematic survey data to the ABIS standard, enhancing the ways in which it can be ingested into, and visualised in, Dandjoo.
- Continue to co-design environmental assessment processes with DWER to embed data ingestion into the BIO platform including through pre-planning and application workflows for industry proponents.
- Further outreach to potential data providers across all sectors to secure and schedule the delivery of new data sets over time.