Implementing GBIF Pipelines in the Atlas of Living Australia: The first step towards alignment and further collaboration

Javier Molina[‡], Peggy Newman[§], David Martin[§], Vicente J Ruiz Jurado[|]

‡ Atlas of Living Australia, Melbourne, Australia
§ Atlas of Living Australia, Canberra, Australia
| Living Atlases Community - GBIF.es, Madrid, Spain

Corresponding author: Javier Molina (javier.molina@csiro.au)

Abstract

The Global Biodiversity Information Facility (GBIF) and the Atlas of Living Australia (ALA) are two leading infrastructures serving the biodiversity community.

In 2020, the ALA's occurrence records management systems reached end of life after more than 10 years of operation, and the ALA embarked on a project to replace them.

Significant overlap exists in the function of the ALA and GBIF data ingestion pipeline systems. Instead of the ALA developing new systems from scratch, we initiated a project to better align the two infrastructures. The collaboration brings benefits such as the improved reuse of modules and an overall reduction in development and operation costs.

The ALA recently replaced its occurrence ingestion system with GBIF pipelines infrastructure and shared code. This is the first milestone of the broader ALA's Core Infrastructure Project and some of the benefits from it are a more reliable, performant and scalable system, proven by the ability to ingest more and larger datasets while at the same time reducing infrastructure operational costs by more than 40% compared to the previous system. The new system is a key building block for an improved ingestion framework that is being developed within the ALA. The collaboration between the ALA and GBIF development teams will result in more consistent outputs from their respective processing pipelines. It will also allow the broader collective expertise of both infrastructure communities to inform future development and direction. The ALA's adoption of GBIF pipelines will pave the way for the Living Atlases community to adopt GBIF systems and also contribute to them.

In this talk we will introduce the project, share insights on how both the teams from the GBIF and the ALA worked together and finally we will delve into details about the technical implementation and benefits.

Keywords

ALA, data ingestion, biodiversity infrastructure, Living Atlases

Presenting author

Javier Molina

Presented at

TDWG 2021