

Data Engineer

**Are you an experienced data engineer seeking a new and challenging opportunity?
Contact the CSIRO's Atlas of Living Australia today!**

Role Overview

- Do you want to apply your data skills to be part of science and research at CSIRO?
- Are you passionate about open-source software and open data?
- Would you like to work on international collaborations?

We are offering two exciting data roles working at CSIRO in biodiversity science, using modern open-source technology and collaborating with stakeholders all around the country and the world.

The Atlas of Living Australia (ALA) is Australia's national biodiversity data aggregator funded under the National Collaborative Research Infrastructure Strategy (NCRIS) and hosted by CSIRO. The ALA is the Australian node of the Global Biodiversity Information Facility (GBIF). Our digital infrastructure is developed in-house to support research activities, government decision-making and community events.

The ALA uses and produces open-source software and tools to aggregate Australian biodiversity data from a variety of providers and make it discoverable and reusable online. Our technology stack is reused by over 25 countries.

The ALA Data Management team is seeking a data engineer for a **two-year contract opportunity** to work on data acquisition, transformation, loading and quality assurance. Our team is technically oriented and uses multiple technologies and platforms to explore and manipulate large datasets into a standardised format, which we then ingest into our processing pipeline.

What will you be doing?

As a successful candidate you will develop and support automated jobs to harvest data from a series of data providers, ensuring data currency and quality is consistent with expectations. You will need to be effective both as a team member and as a reliable point of contact for data providers. We're looking for strong collaboration and communication skills, and the ability to develop great rapport with stakeholders.

Suitable candidates located in either Canberra or Melbourne are encouraged to apply. Our main office is in Canberra, but we have a Melbourne cohort who work completely remotely. Our workplace culture facilitates remote and flexible work.

Duties and Key Result Areas:

- Work to the Data Manager in the Data Management team to build and manage both automated and manual data loading processes
- Map datasets to the Darwin Core standard
- Implement, deploy, schedule and maintain data load processes
- Implement quality assurance and verification on datasets to ensure loaded records meet expectation
- Engage professionally with external stakeholders offering technical guidance on data management issues such as data mapping, automation, and loading
- Contribute to team meetings and planning and review activities

CSIRO Values:

- Communicate effectively and respectfully with all staff, clients and suppliers in the interests of good business practice, collaboration and enhancement of CSIRO's reputation
- Work collaboratively with colleagues within your team, the broader CSIRO and across partner institutions to reach objectives
- Adhere to the spirit and practice of CSIRO's Values, Health, Safety and Environment plans and policies, Diversity initiatives and Zero Harm goals

Who are we looking for?

Essential Criteria:

We understand that women and other marginalised groups don't tend to apply for these roles unless they meet all of the criteria, and we recognise that there can be other things that make a candidate a great fit. If you're enthusiastic about working in biodiversity science, have strengths in just some of these areas and a willingness to learn fast, please get in touch.

- 2+ years demonstrated operations experience in a data driven production system
- Strong knowledge of scripting languages – Python, Spark, Scala, SQL, bash, R, JavaScript
- Strong ETL skills with large datasets with a focus on efficiency and scale
- Experience with Linux OS
- Experience with a variety of open source relational and non-relational databases
- Experience in both delivering and consuming REST services
- Source code management using git, svn or Bitbucket
- Knowledge of SOLR and/or Elasticsearch administration and queries
- Effective stakeholder engagement and technical liaison skills

Desirable Criteria:

- Background or strong interest in biodiversity/ecology/taxonomy
- Enthusiasm and knowledge of open data standards, procedures and policy
- Experience with Apache Beam/Spark/AVRO, Jenkins, ELK, Zabbix, Ansible
- Experience with geospatial data systems and development
- Experience with Darwin Core standard

Eligibility:

The successful applicant will be required to obtain and provide a National Police Check or equivalent.

To be eligible for this position you must be willing and able to travel interstate occasionally.

How to apply

If you would like to be considered for this role, please provide a cover letter of no more than 2 pages with your attached CV that best describes your interest in this role and addresses the above selection criteria. This is a great opportunity to tell the Australian Atlas of Living about you. Please send your resume and covering letter directly to Peggy Newman via email at Peggy.Newman@csiro.au.