Unlocking the Power of Data:

Streamlined ETL Solutions for the Digital Anthroposphere

eResearch conference, Brisbane

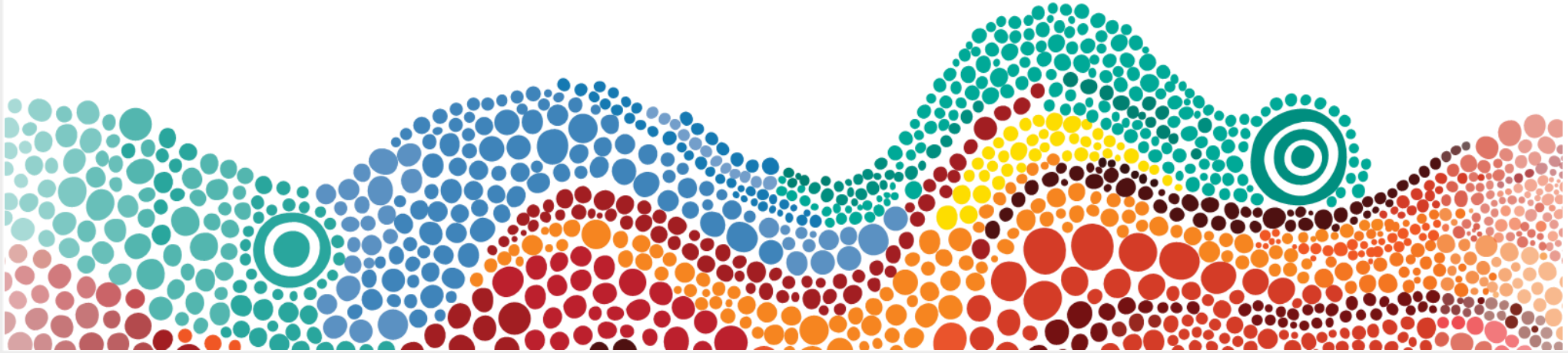
Presenter: Dr. Ali Asghari

AURIN
aurin.org.au

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

Acknowledgement of Country

We acknowledge the Traditional Owners of the land on which this event is taking place and pay respect to their Elders (past and present) and families.



AURIN: An Introduction

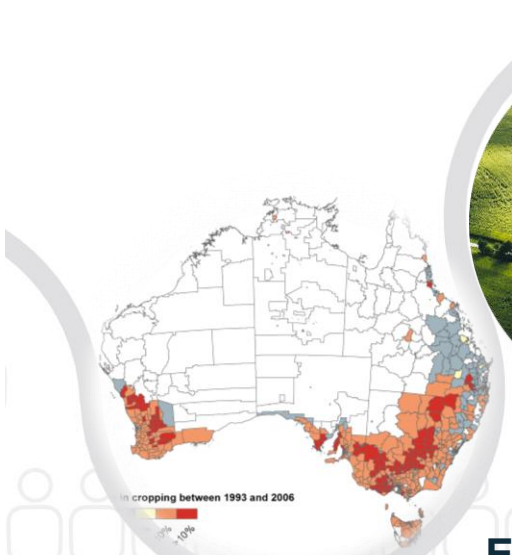
What is AURIN?

Australian Urban *Research*
Infrastructure Network

National Collaborative
Research Infrastructure
Strategy

AURIN: Key Focused Challenges

Key Focused Challenges



Demographic transformation

Energy transition

Climate change

Australian cities, regional centres and communities

AURIN: Key Strategic Areas

Urban Data-as-a-Service

Unlocking, generating, curating, cataloguing, and sharing high value and hard-to-get data for urban and regional research and planning.

Urban Analytics-as-a-Service

Designing, developing and maintaining a modular, cloud-first digital research infrastructure to support urban and regional analytics and modelling.

Foundations for Australian Urban Digital Twins

Understanding the needs of UDTs to inform decisions and positively impact lives. Applying, testing, prototyping and sharing best practice standards, semantics, workflows and cutting-edge technologies for use in UDTs.

Urban Data-as-a-Service: AURIN ETL Pipeline

How does AURIN Manage and publish data?

- AURIN Manages and publishes data using the AURIN ETL pipeline
- Allow to standardise highly valuable and hard-to-get data from a vast range of different data sources (organisations)
- Publish that data in a consistent and machine-readable format and manage it in a modernised way.

Extract

Extracts source data and necessary metadata for an ingestion

Transform

Transforms extracted data into a standardised format for use and storage

Load

Loads standardised data and metadata into software for use and access inside a repository

AURIN ETL Pipeline: Key Characteristics

Why is AURIN ETL Pipeline developed?

To enable the creation of 

data products in a reproducible, atomic, automated, and flexible manner

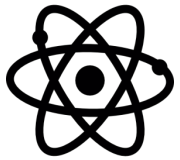
Reproducibility **Atomicity** **Automation** **Flexibility**

AURIN ETL Pipeline: Key Characteristics

Reproducibility

Definition: The ability to recreate any version of a data product at any time, involving documentation and preservation of the data's sources and generation processes.

Importance: Crucial for **data integrity and transparency**, enabling verification, validation, understanding of lineage, and independent result replication.

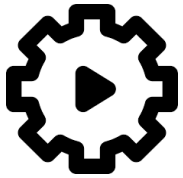


Atomicity

Definition: The concept of treating a sequence of data operations as a single, indivisible unit, where all operations succeed or fail together to prevent partial or inconsistent data states.

Importance: Vital for preserving **data integrity and consistency,** ensuring well-structured data for easier access, analysis, and maintenance.

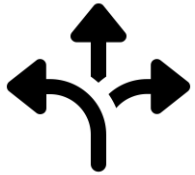
AURIN ETL Pipeline: Key Characteristics



Automation

Definition: Ability to conduct data ingestion processes with minimal manual intervention using scripts, automation tools, or workflows to automate data collection, transformation, and loading tasks.

Importance: Essential for improving **efficiency and scalability** while speeding up data processing and reducing errors.

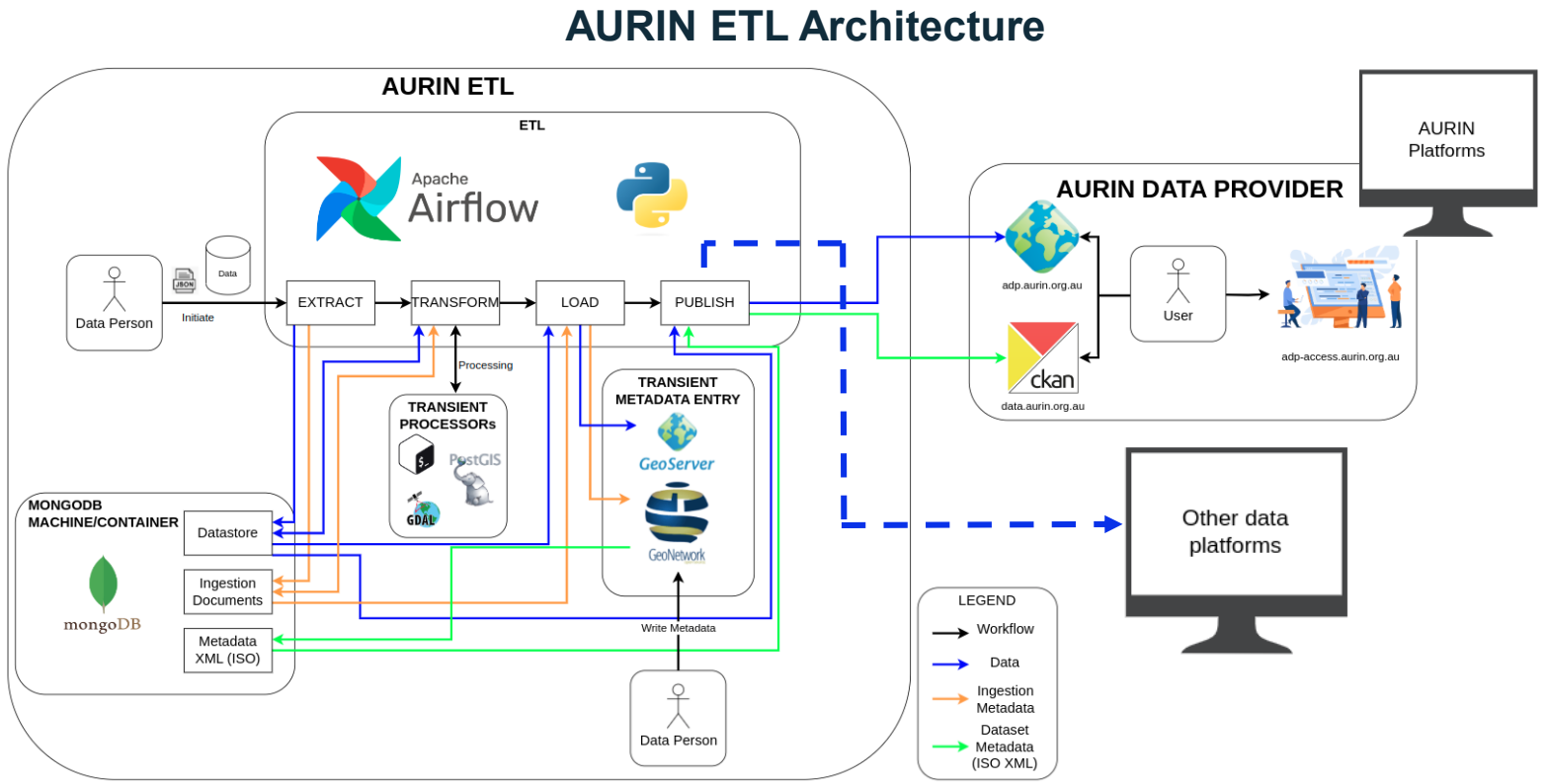


Flexibility

Definition: Ability to adapt to various data formats, schemas, and sources, producing a controlled and standardised output.

Importance: Ensures **interoperability and adaptability** to changing data sources over time.

AURIN ETL Pipeline: Architecture



AURIN ETL Pipeline: Applications and technologies

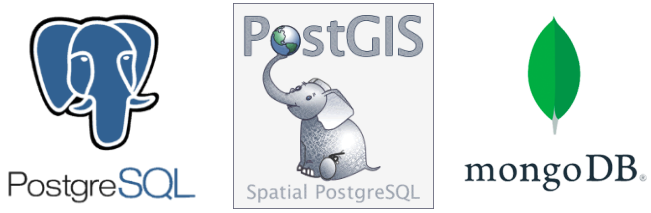
AURIN
aurin.org.au

Infrastructure and DevOps

Geospatial Data Server

Workflow Management

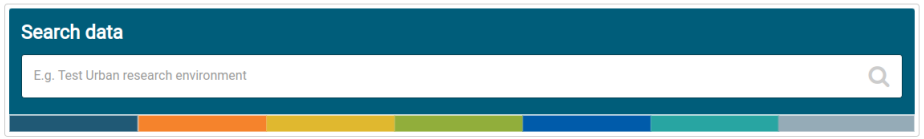
Database Management System

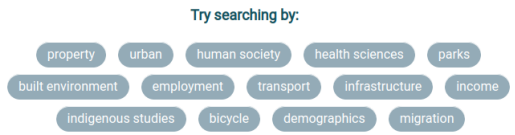


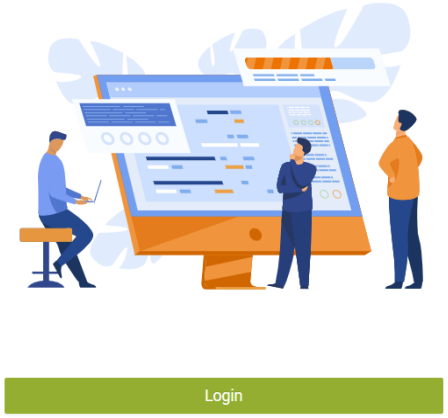
Metadata Management

Urban Data-as-a-Service: AURIN Data Provider (ADP)

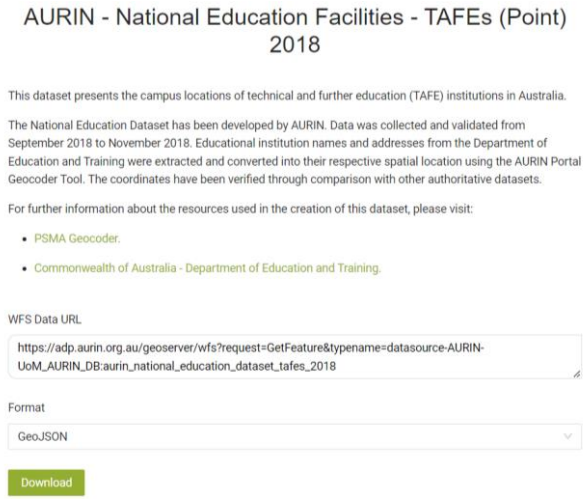
AURIN Data Catalogue







AURIN Data Provider



More information:

- **AURIN Training**

AURIN: Future Directions

What are the future directions for AURIN?

- Focusing on establishing the foundations for Australia's Urban Digital Twins, to address climate change, energy transition, and changing demographics.
- Analysing the requirements of UDT applications and looking to fulfil them wherever possible with data, analytics, workflows, semantic solutions, etc.
- Providing diverse and reliable datasets for researchers will allow them to address more urban challenges and find practical solutions to those challenges.
- Improving the searchability of our AURIN Data Catalogue (metadata enrichments) will help researchers to better find their desired datasets
- Developing data validation methods to make sure the data quality
- Providing new data formats including raster, real-time and 3D datasets for researcher to broaden their horizon into new research ideas and address new challenges

THANK YOU!

COME SEE US AT THE AURIN BOOTH

Dr. Ali Asghari
Data Engineer
E: asgharia@unimelb.edu.au
www.aurin.org.au
@aurin_org_au

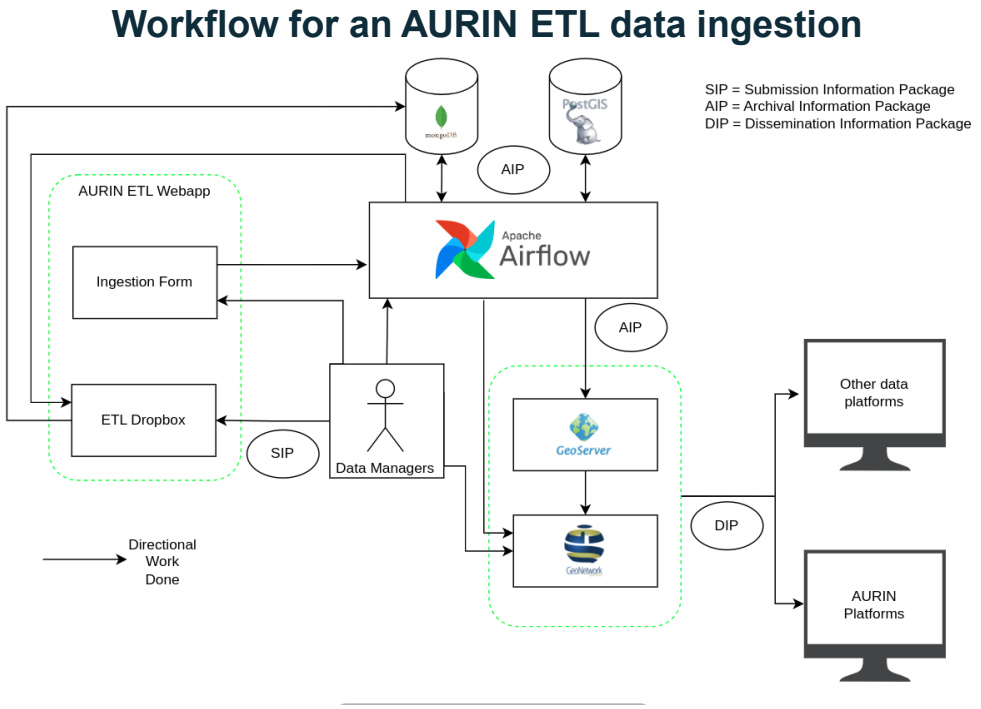
aurin.org.au

AURIN
aurin.org.au

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

AURIN ETL Pipeline: Workflow

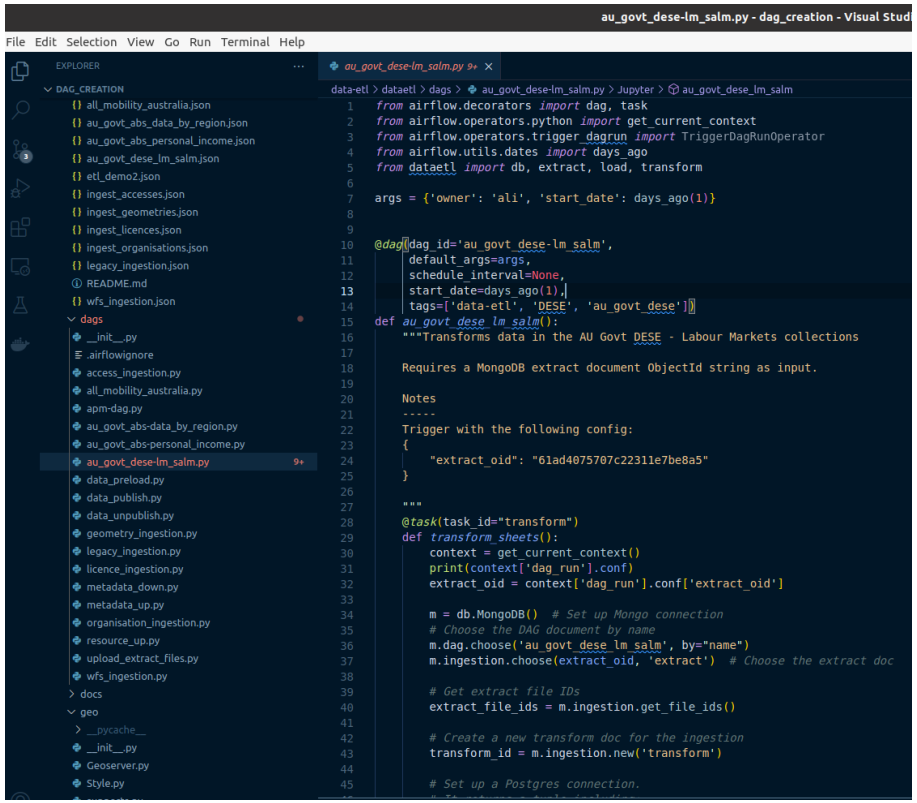
How it works – Workflow



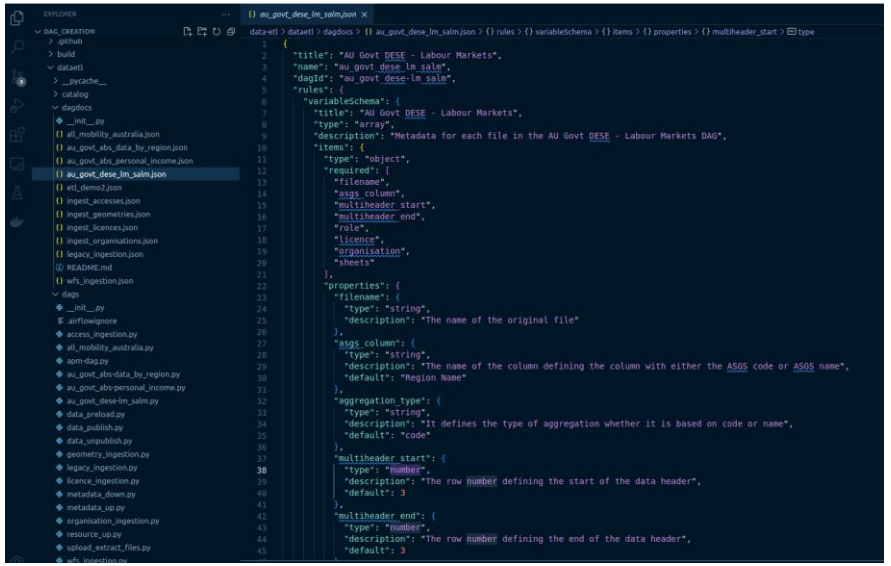
AURIN ETL Pipeline: Workflow

How it works – Develop the workflow

DAG Descriptor

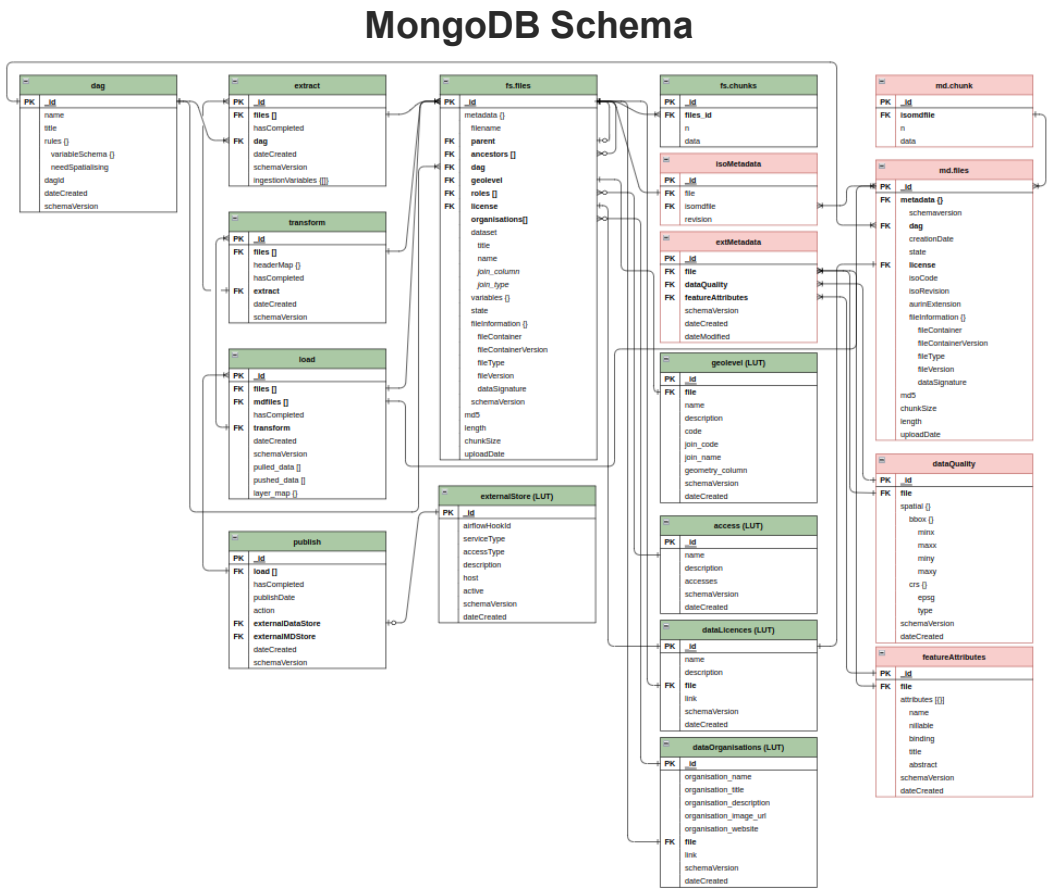


DAG Document



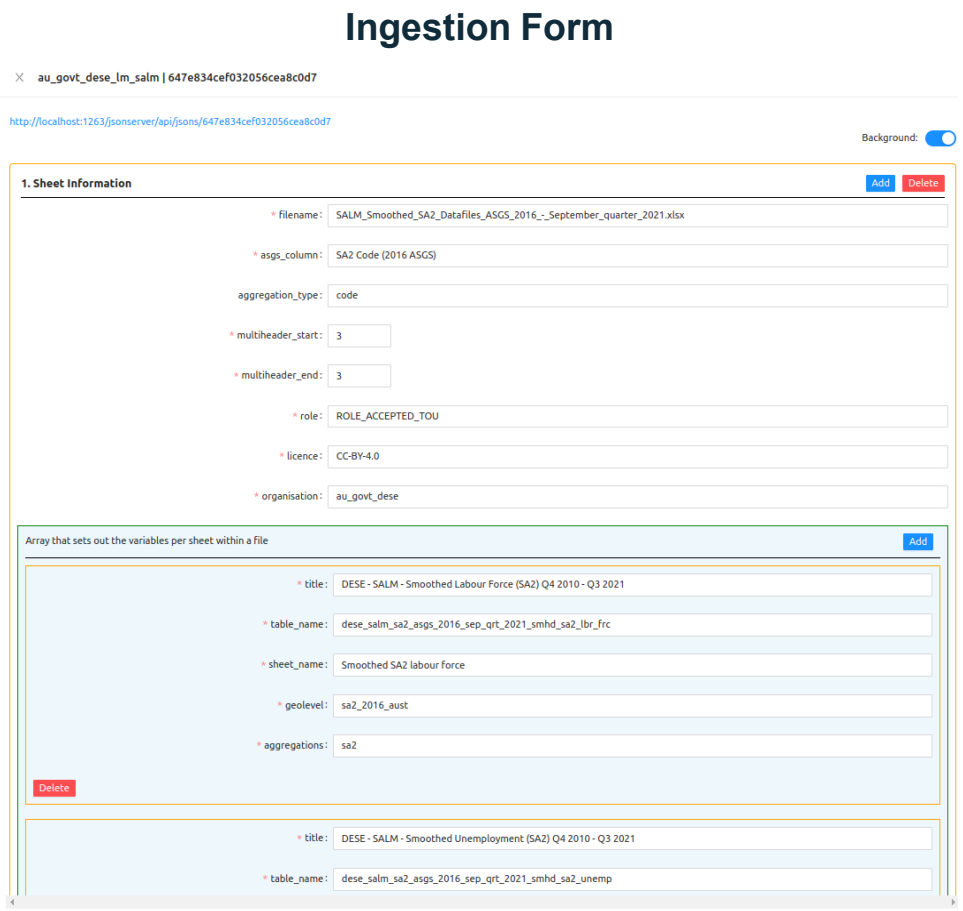
AURIN ETL Pipeline: Workflow

How it works – Register the workflow



AURIN ETL Pipeline: Workflow

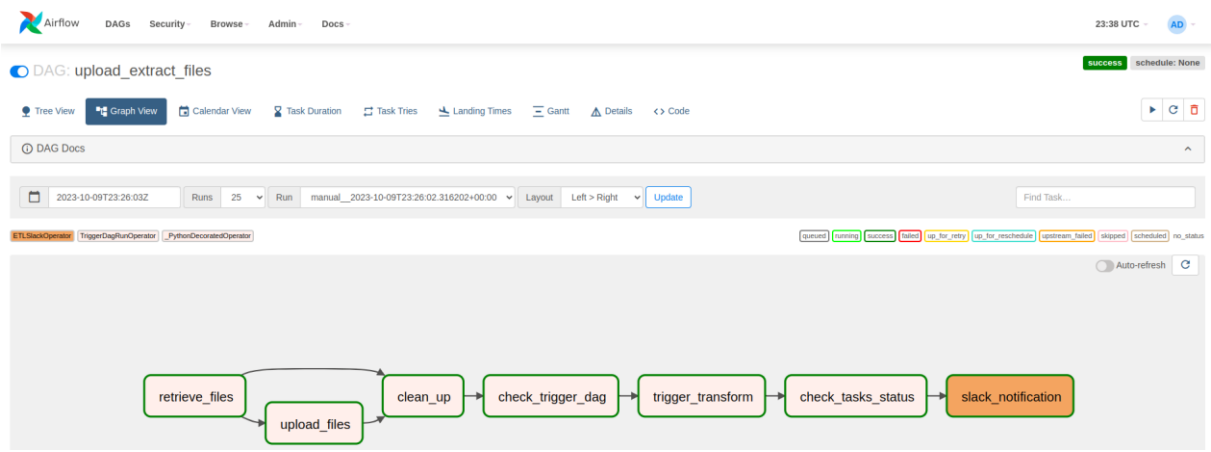
How it works – Populate information required for the workflow



AURIN ETL Pipeline: Workflow

How it works – Run the workflow

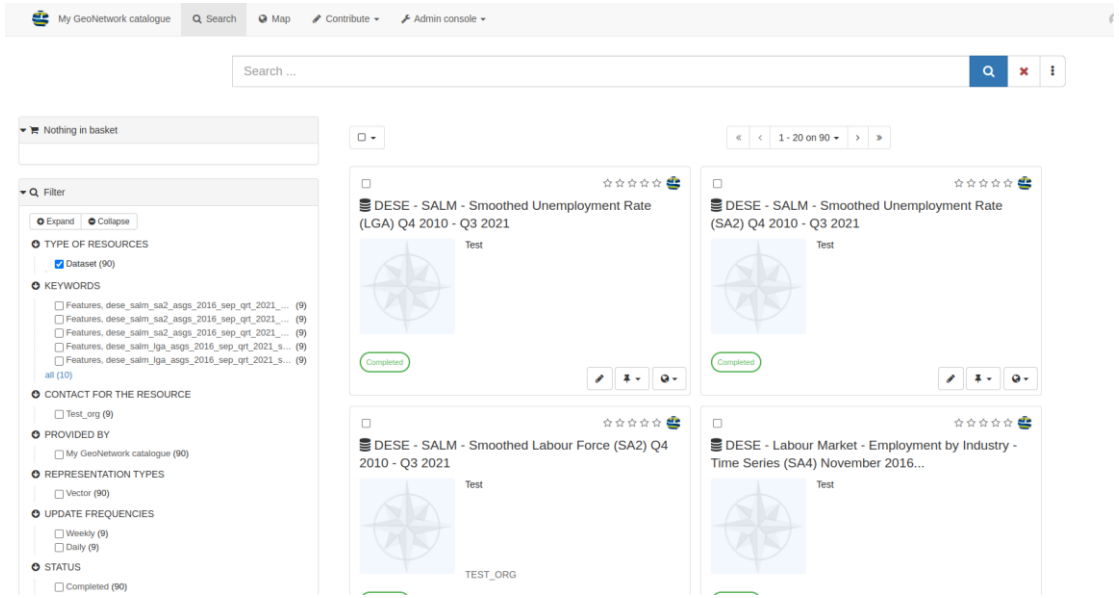
Running DAGs



AURIN ETL Pipeline: Workflow

How it works – Populate and/or enrich the metadata

GeoNetwork standardised template



AURIN ETL Pipeline: Workflow

How it works – Publish the data

Publish Data into GeoServer

Publish Data into CKAN