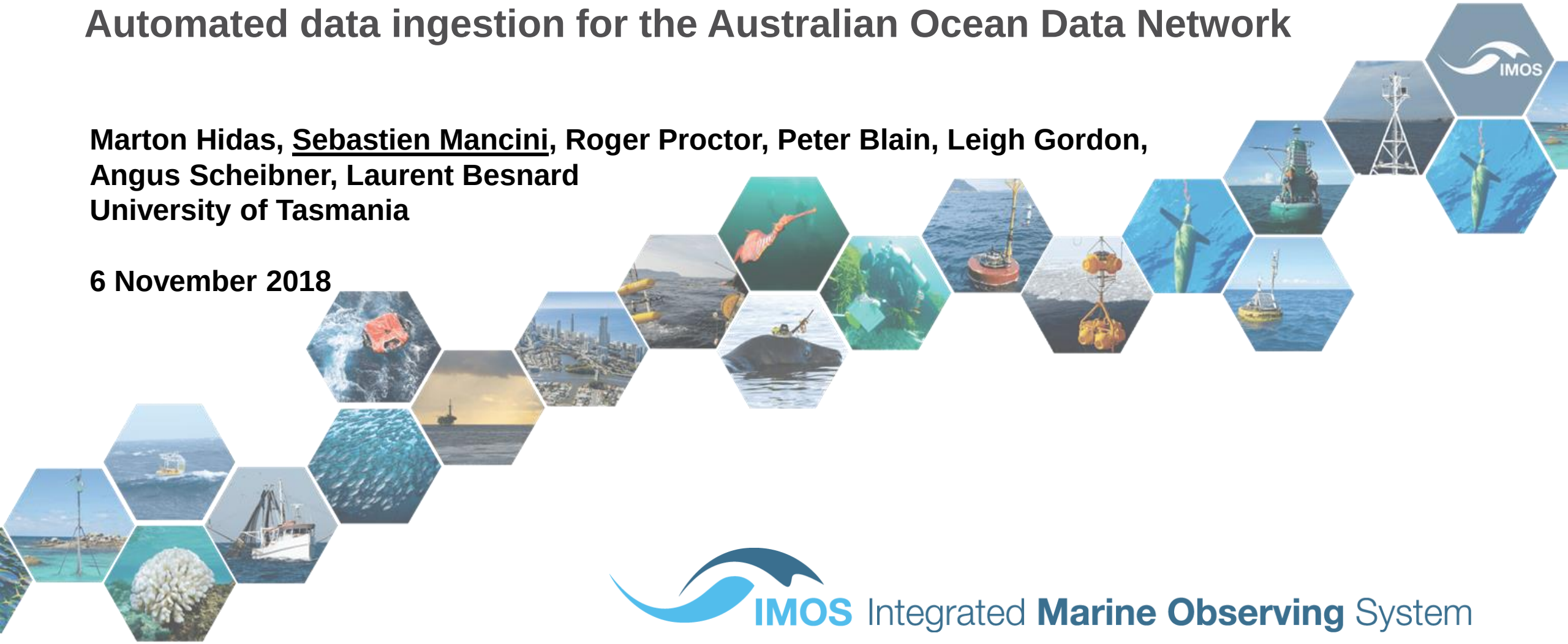
Australia's Integrated Marine Observing System (IMOS)

Automated data ingestion for the Australian Ocean Data Network

Marton Hidas, Sebastien Mancini, Roger Proctor, Peter Blain, Leigh Gordon, Angus Scheibner, Laurent Besnard
University of Tasmania

6 November 2018



IMOS Integrated **Marine Observing** System

Outline of the talk

12 minutes, plus 3 minutes for questions

1. About IMOS
2. Data providers versus data users
3. Automated pipeline
4. Future improvements

IMOS Facilities

AUSTRALIAN OCEAN DATA NETWORK

SATELLITE REMOTE SENSING

WIRELESS SENSOR NETWORKS

OCEAN RADAR

SHIPS OF OPPORTUNITY

AUTONOMOUS UNDERWATER VEHICLES

NATIONAL MOORING NETWORK

ANIMAL TRACKING #2 (ACOUSTIC TELEMETRY)

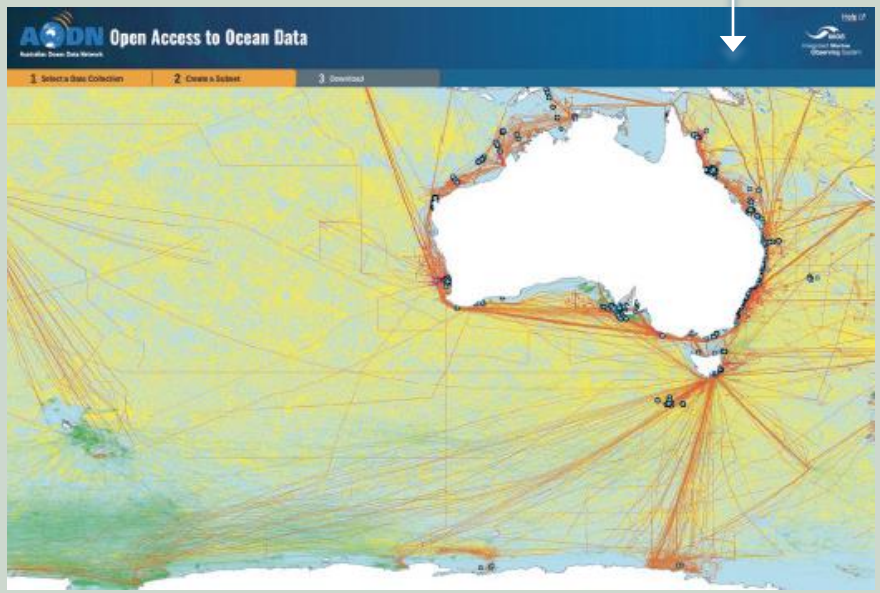
DEEP WATER MOORINGS

ARGO FLOATS

OCEAN GLIDERS

ANIMAL TRACKING #1 (TAGGING)

All data discoverable, accessible, usable and reusable



https://portal.aodn.org.au

Why AODN?

Data providers

Diversity in

- Technologies
- Instruments
- Platforms
- Organisations
- People

Data users/use cases

Diversity in

- Data products required
- Data volume required
- Method of access
- Preferred formats
- Tools used

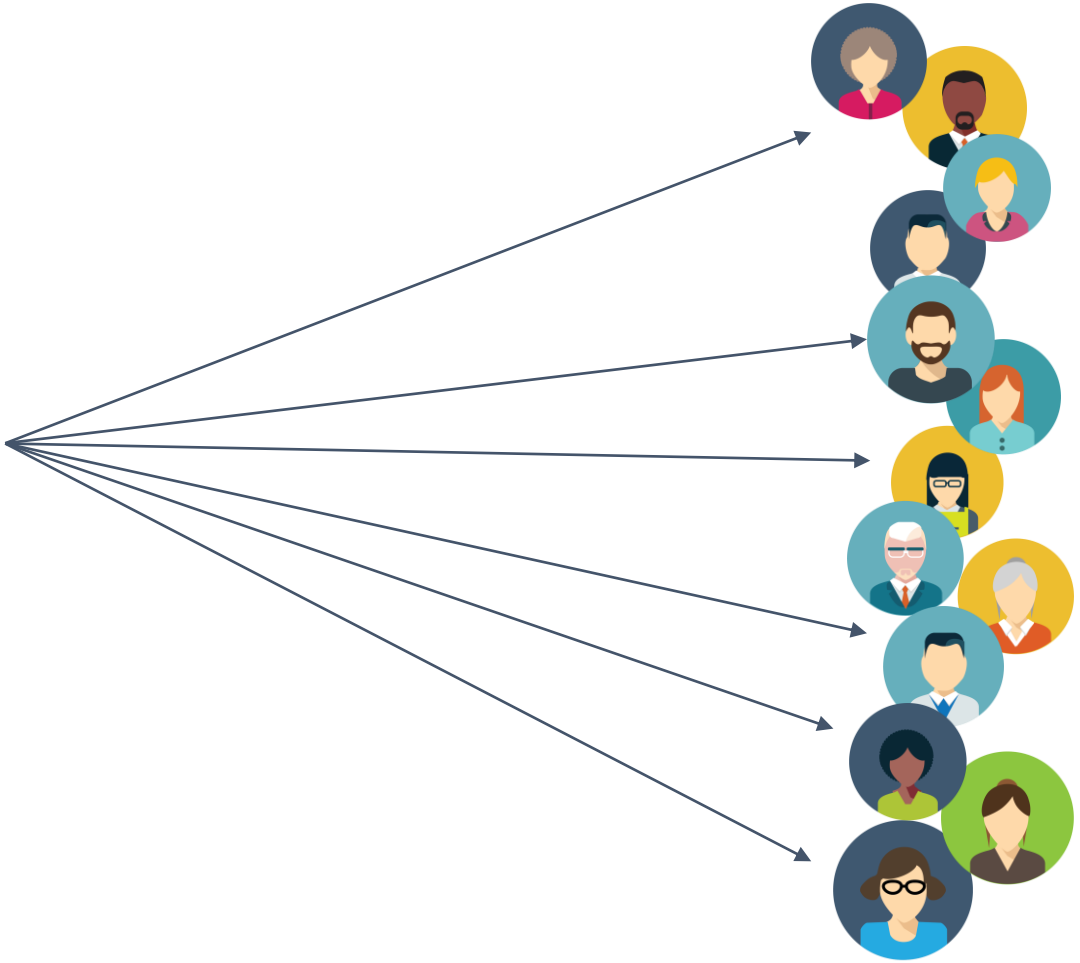
- Research scientists
- Government departments
- Private industry
- Fisheries
- Managers
- Policy makers
- Shipping
- Recreational fishing/boating
- Etc…

Why AODN?

Data providers

Each provider has to cater for many users

Data users/use cases



Why AODN?

Data providers

UNIVERSITY of TASMANIA | IMAS INSTITUTE FOR MARINE & ANTARCTIC STUDIES

Australian Government
Geoscience Australia

CSIRO

EPA VICTORIA

Australian Government
Bureau of Meteorology

Australian Government

AUSTRALIAN INSTITUTE OF MARINE SCIENCE

NSW GOVERNMENT Office of Environment & Heritage

Australian Government
Department of the Environment and Energy
Australian Antarctic Division

Government of South Australia

SARDI
SOUTH AUSTRALIAN RESEARCH AND DEVELOPMENT INSTITUTE

THE UNIVERSITY OF WESTERN AUSTRALIA

Data users/use cases

Each user may need to obtain data from multiple sources

Why AODN?

Data providers

Each provider has to cater for many users

That's a lot of unique connections! Diversity^2

Data users/use cases

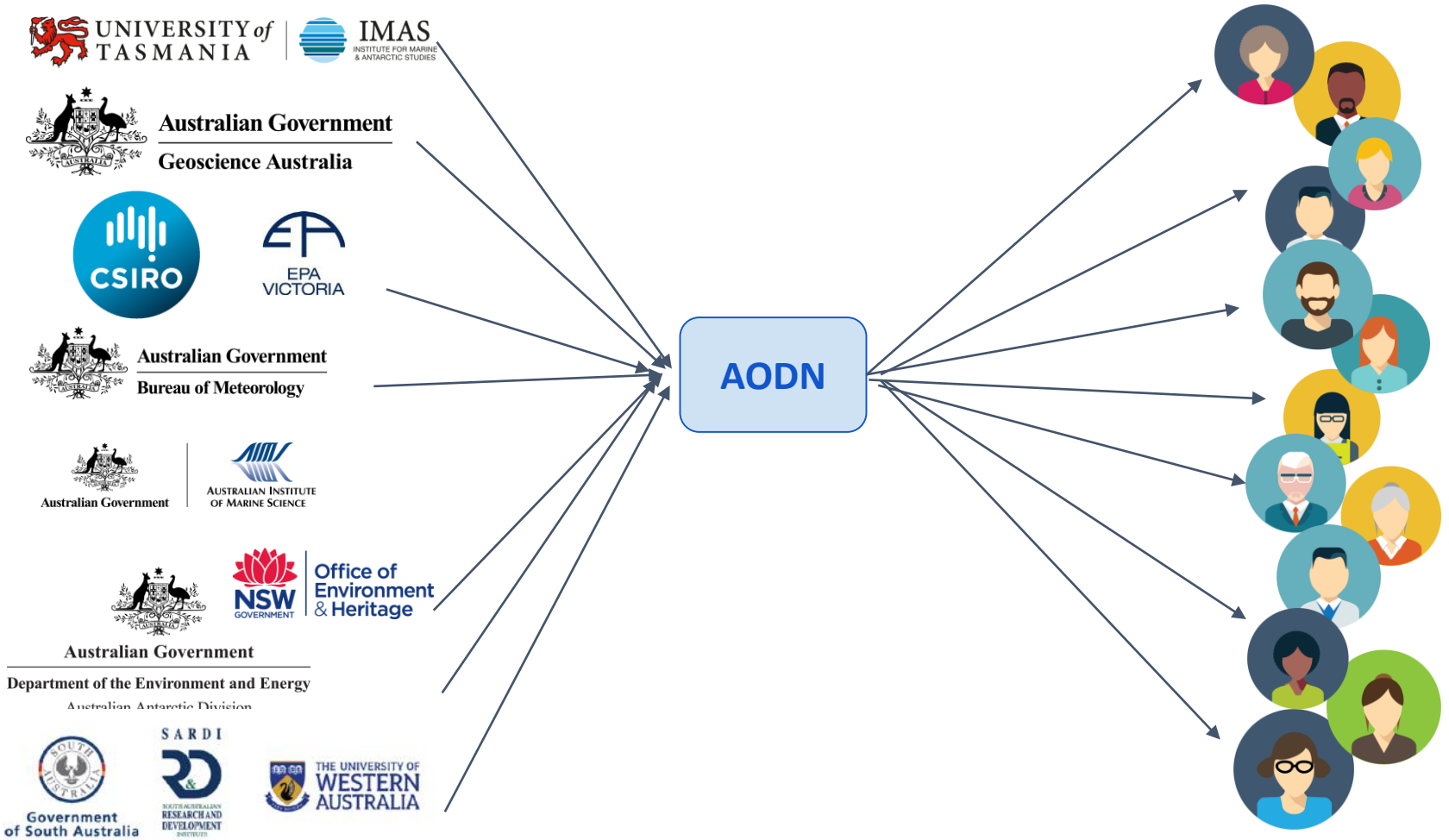
Each user may need to obtain data from multiple sources

Why AODN?

Data providers

Data users/use cases

Standard ingestion

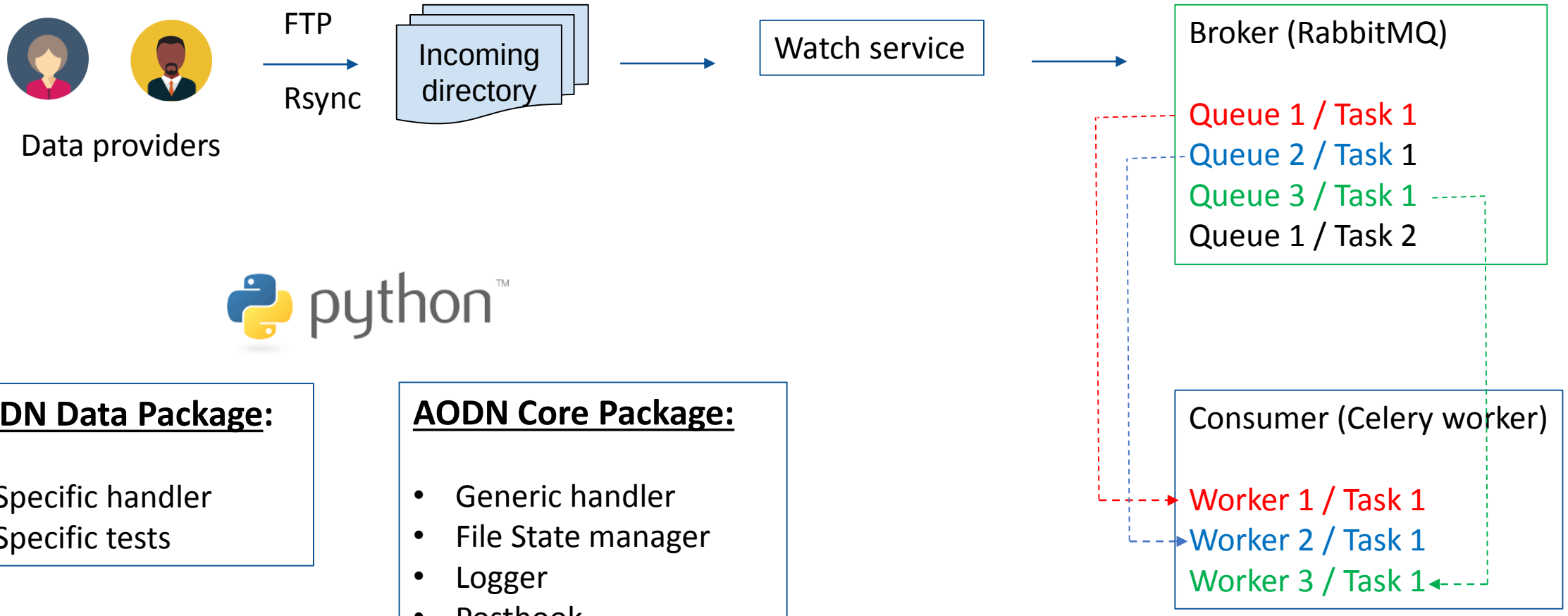


Standard access

IMOS standard ingestion: Key design criteria

- Keep data safe
- Make only high quality data for users
- Do not increase load on front facing systems
- Make data available as quickly as possible
- Robustness
- Transparency

Architecture overview of the pipeline



AODN Data Package:

- Specific handler
- Specific tests

AODN Core Package:

- Generic handler
- File State manager
- Logger
- Posthook
- Tests

Pipeline workflow: the state machine

Data providers

Incoming data files

Resolve

Pre-process

Validate

Process

Publish

Post-process

Notify

Handle error

Notify error

Error files

IOOS

Compliance checker

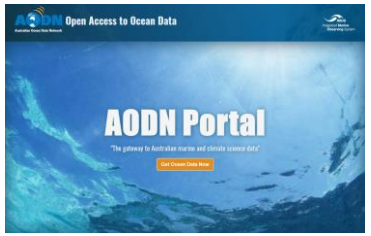
Storage (AWS S3)

ETL (Talend)

DB (Postgres)

Metadata (GeoNetwork)

Data & Maps (GeoServer)



Data providers (and AODN staff) are notified via a detailed email report.

Pipeline workflow: the state machine

Data providers

Incoming data files

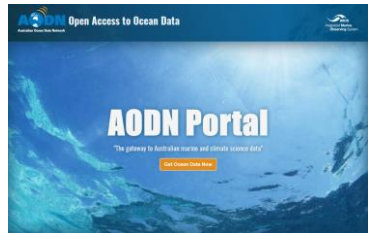
Resolve

Some Numbers:

32 Organisations providing data
60 Data uploaders (+ a handful that we pull from)
24 Handler classes
43 Ingestion pipelines
86 Data/metadata “Harvesters” (95 Harvest jobs)
204 Data collections in the AODN Portal

DB (Postgres)

Data & Maps (GeoServer)



Error files

Handle error

Notify error

Post-process

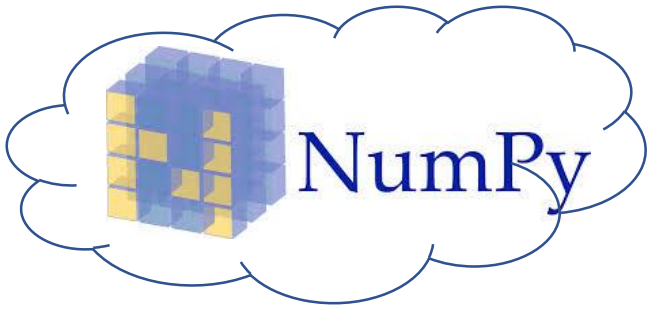
Notify

Data providers (and AODN staff) are notified via a detailed email report.

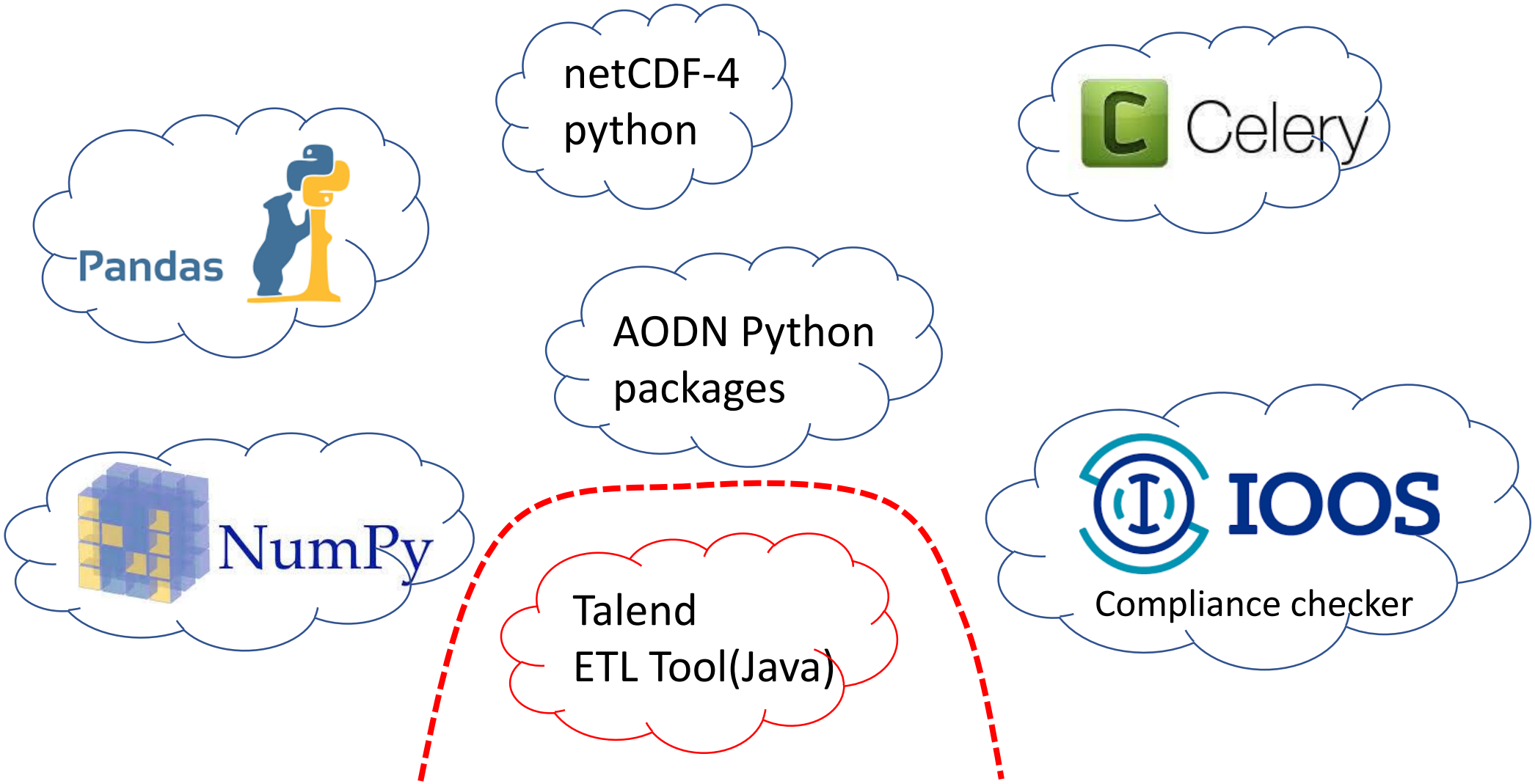
Benefits: Python ecosystem

netCDF-4
python

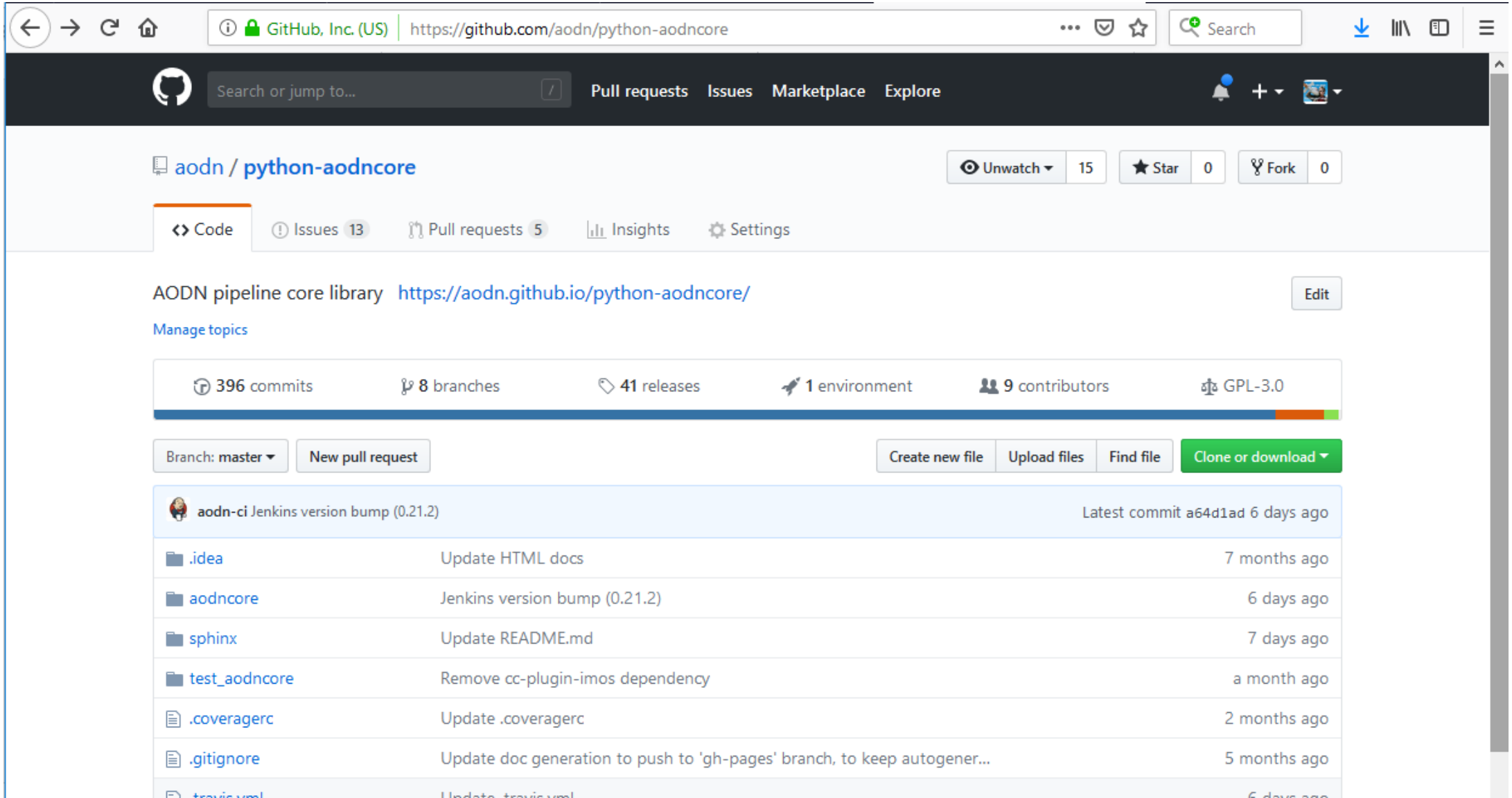
AODN Python
packages



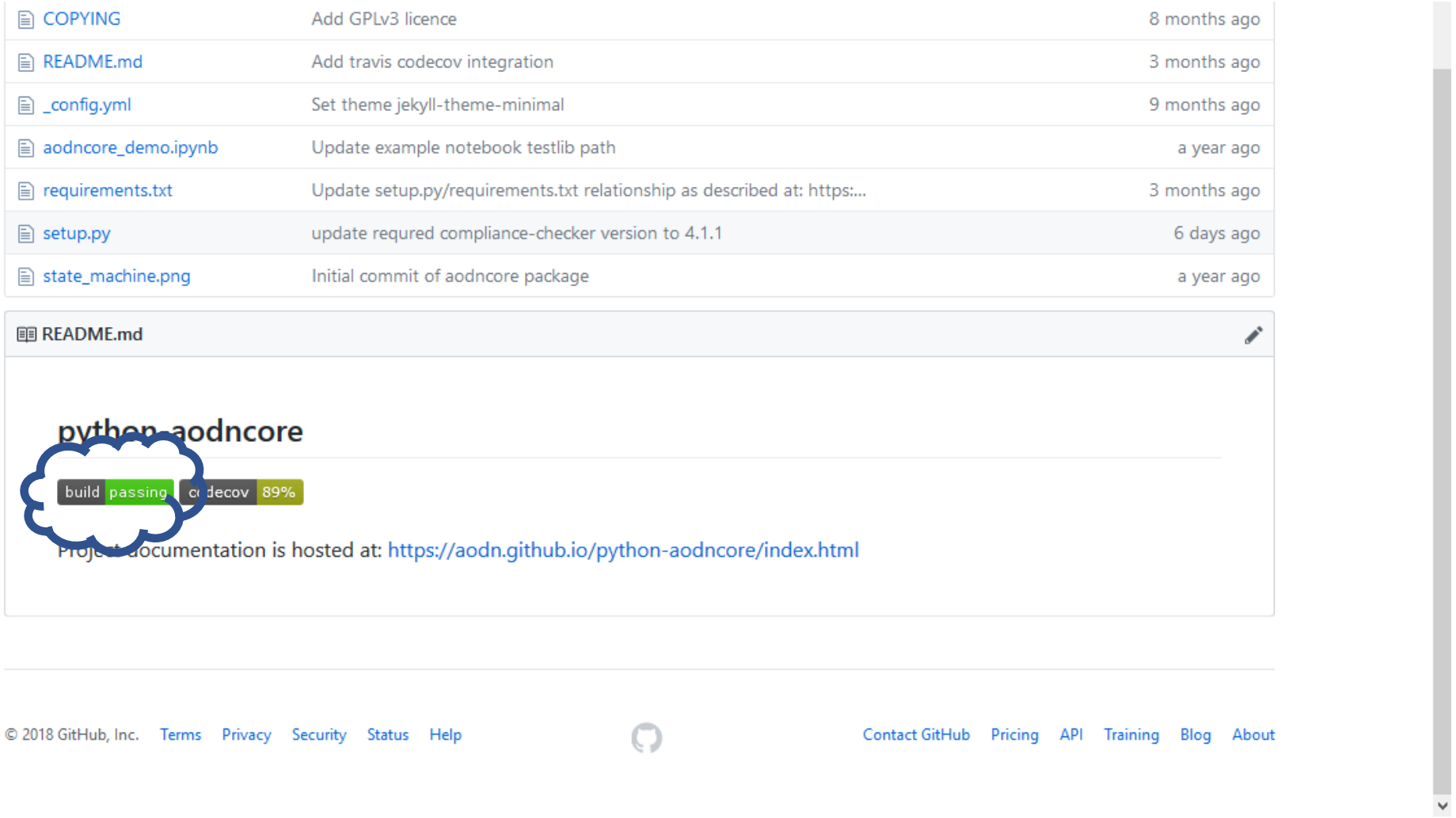
Benefits: Python ecosystem – One outlier



Benefits: Test and deploy code in confidence



Benefits: Test and deploy code in confidence



COPYING	Add GPLv3 licence	8 months ago
README.md	Add travis codecov integration	3 months ago
_config.yml	Set theme jekyll-theme-minimal	9 months ago
aodncore_demo.ipynb	Update example notebook testlib path	a year ago
requirements.txt	Update setup.py/requirements.txt relationship as described at: https:...	3 months ago
setup.py	update requred compliance-checker version to 4.1.1	6 days ago
state_machine.png	Initial commit of aodncore package	a year ago

README.md

python-aodncore

build passing codecov 89%

Project documentation is hosted at: https://aodn.github.io/python-aodncore/index.html

© 2018 GitHub, Inc. Terms Privacy Security Status Help

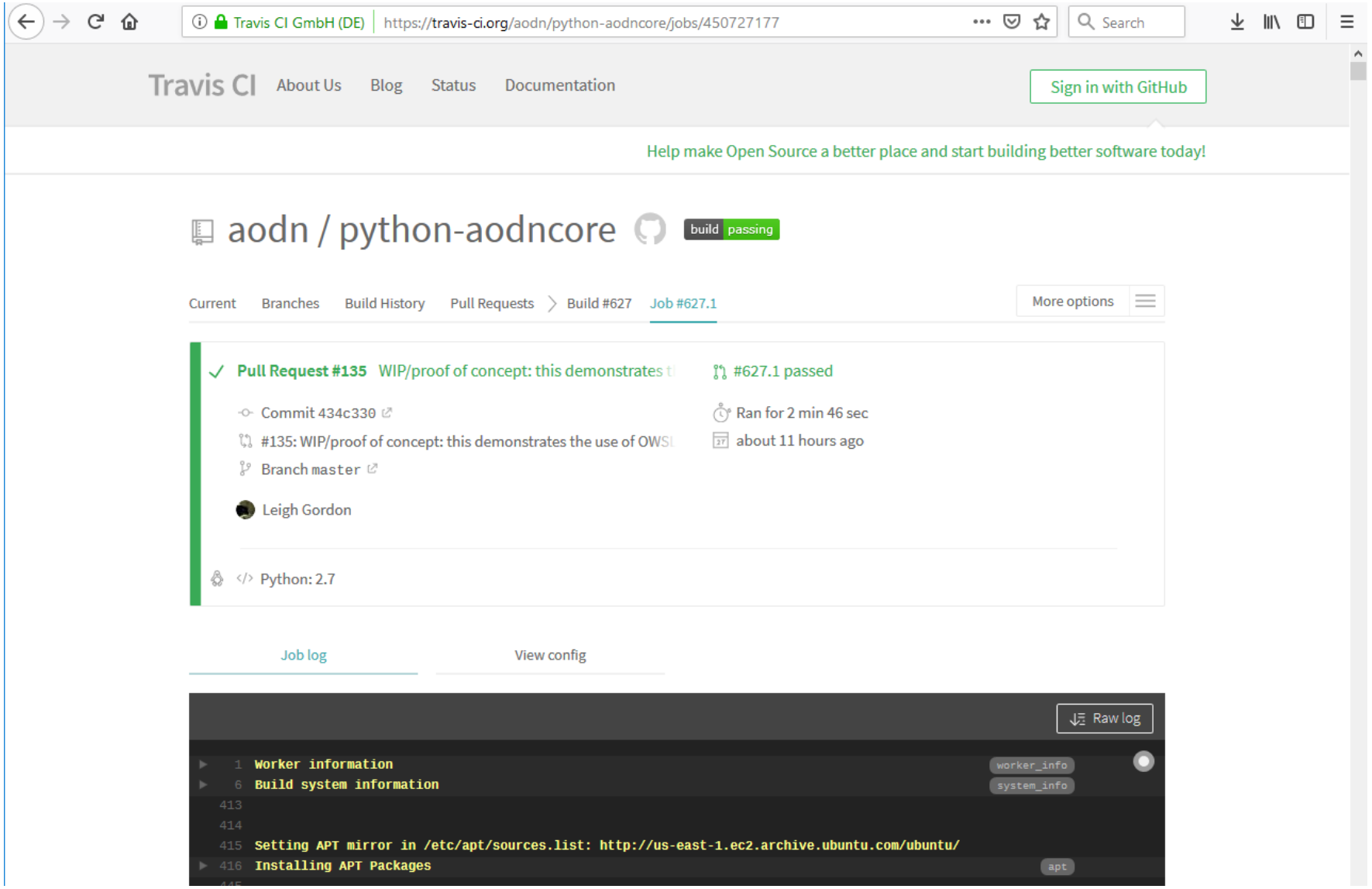
Contact GitHub Pricing API Training Blog About

Travis CI Integration

Benefits: Test and deploy code in confidence

Travis CI

627 Tests passed !!!



Benefits:

- Generic code
- Increased confidence:
  - All features can be tested from one version to the next
- Easier to debug
- Better control of the different version
  - Workflow to build, package and deploy
  - What version is deployed on which environment (Release candidate or Production)
- Faster publication of new data
- Improved consistency within data collections

Future improvements

- Reporting, creation of dashboard in Sumologic
- Improve code:
  - Replace Talend (Java) by a similar tool in Python for a consistent environment
- Use of AWS Batch to improve scalability:
  - No conflict between pipelines
  - Multiple queues for different data streams
  - Only running when needed

GitHub

https://github.com/aodn/python-aodncore
https://github.com/aodn/python-aodndata

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

IMOS is a national collaborative research infrastructure, supported by Australian Government. It is operated by a consortium of institutions as an unincorporated joint venture, with the University of Tasmania as Lead Agent. **www.imos.org.au**

PRINCIPAL PARTICIPANTS

University of Tasmania
(Lead Agent)

Australian Government
Australian Institute of Marine Science
Australian Government
Bureau of Meteorology
CSIRO

Government of South Australia
SARDI
The University of Western Australia
sims sydney institute of marine science
UTS
The University of Sydney
Macquarie University
UNSW Sydney

SIMS is a partnership involving four Universities.

ASSOCIATE PARTICIPANTS

Curtin University

Australian Government
Department of the Environment and Energy
Australian Antarctic Division