Applying DOI since 2011

RAVE DR3 got the first DOI for all its tables,

- including DR2 and DR1 data in table format

AIP started to publish data

- using IVOA protocols and standards
- providing also DOI for cross community benefits

Subsequent discussions :

- IVOA (ivorn) with focus on services not sufficient
- scientists want to know the data sets, not so much the way it is delivered
- data identifier as *resolvable entities* would be beneficial to use alongside other identifiers

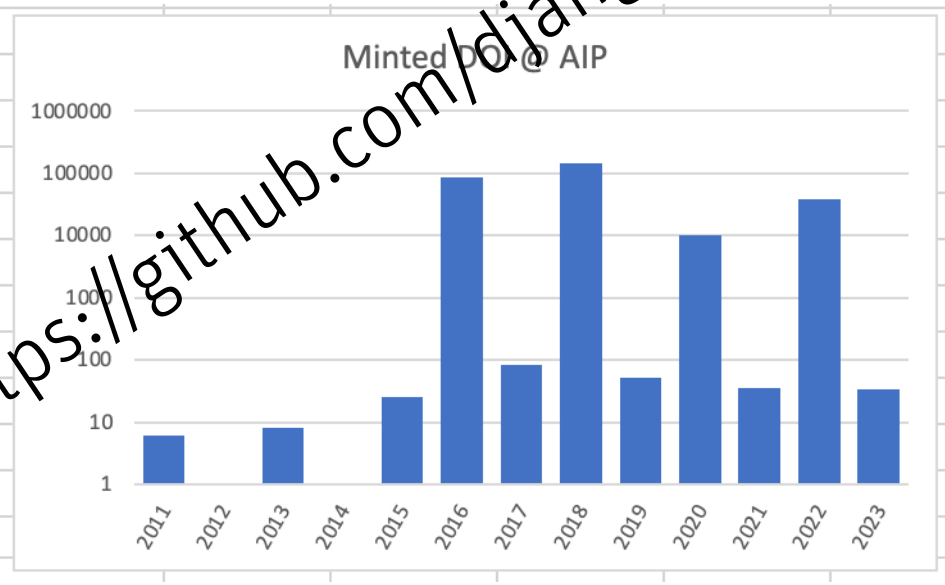
Statistics from DataCite:

Leibniz Institute for Astrophysics Potsdam https://ror.org/03mrbr458

279,286 Works

22,616 Citations

2023	34
2022	38592
2021	36
2020	10160
2019	51
2018	145007
2017	82
2016	85283
2015	25
2013	8
2011	6



DAIQUIRI: https://github.com/django-daiquiri

huge numbers (> 50000) from

APPLAUSE (digitized & processed photographic plates)

www.plate-archive.org

DOI Minting

Policy fundamentals (for AIP)

- DOI is an additional Identifier for *published* data
  - no a substitute for data collection's intrinsic identifiers
  - structure of DOI string should convey basic information about the data collection
  - metadata for DOI should as much as possible come from IVOA metadata
  - landing page of DOI provides IVOA metadata

Missing on IVOA side (initially):

  - default tag for DOI in VOTable Schema
- Registry/Harvesting of data collection or archive
  - via OAI-PMH works for both services

Problems using DOI:

- very rapid development of metadata schema
  - DataCite Kernel was 2.x in 2011
  - DataCite Kernel is 4.x in 2023
  - Kernel 5.x is in preparation

  makes maintenance of data collection DOI difficult for data provider

Benefits of kernel developments:

- improved means to connect data collection to published papers (data table => release paper)
- more standard tags for basic information

Managing Metadata for IVOA and DataCite

Data publication framework:

Daiquiri (Django python based) *https://github.com/aipescience/django-daiquiri*

- metadata for column, table, schema

  standard handlers + models

- metadata for files / objects (non standard)

  tables in science db

  additional adaptors + models

- science data in separate science db
- layered: db (PostgreSQL) / RabbitMQ + Celery / gunicorn
- dockerized
- customizable at data collection level

REST API: OAI-PMH -> IVOA registry and DataCite

REST + TAP API: supports ADQL and Postgres

Datalink API

Using **Datalink** implementation:

- required 'mixing' of metadata
  - in webapp database
  - in science database
- solution :

  add to *tap_schema.datalink*

  #doi (semantics key)

  create table *oai_schema.records*

  with Datalink sematic and ID

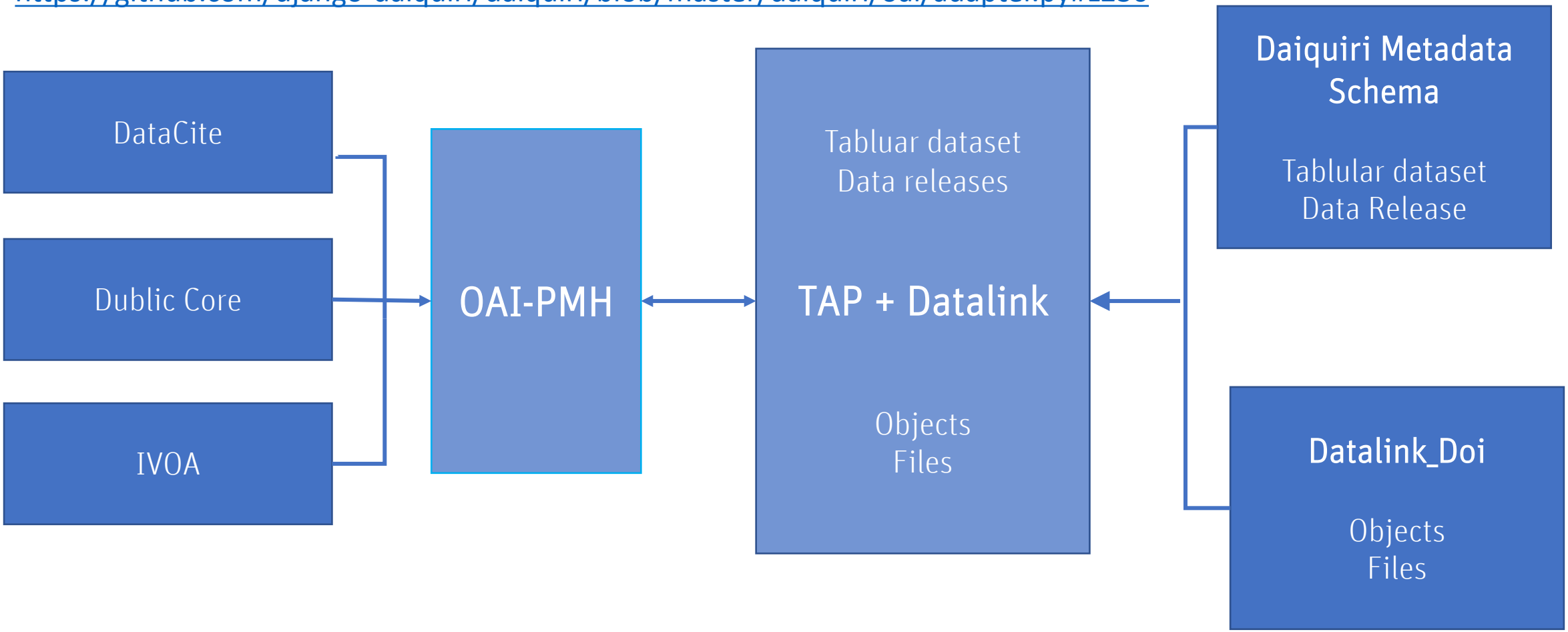
  collect related metadata

  produce all 'relatedIdentifiers' (e.g.) for DOI

  currently static table used, will be based on dynamic linking later

Schematic:

https://django-daiquiri.github.io/docs/administration/#notes-on-datalink
https://github.com/django-daiquiri/daiquiri/blob/master/daiquiri/oai/adapter.py#L230



Use TAP / ADQL, get data+metadata

TAP specification: https://www.ivoa.net/documents/TAP/

URL for TAP protocol queries

https://gaia.aip.de/tap

Python script using pyvo + astropy

```
import pyvo as vo

xurl='https://gaia.aip.de/tap'
tap_session= requests.Session()
tap_service= vo.dal.TAPService(xurl, session=tap_session)

query = "SELECT TOP 10 * from
gaiafpr.interstellar_medium_spectra"

tap_result = tap_service.run_async(query)

# print VOTable
tap_result.to_table().pprint(max_lines=10)
```

```
<?xml version="1.0"?>
<VOTABLE version="1.3"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns="http://www.ivoa.net/xml/VOTable/v1.3"
    xmlns:stc="http://www.ivoa.net/xml/STC/v1.30">
    <RESOURCE type="results">
            <INFO name="QUERY_STATUS" value="OK" />
            <INFO name="QUERY" value="SELECT TOP 10 * from
            gaiafpr.interstellar_medium_spectra" />
            <INFO name="QUERY_LANGUAGE" value="adql-2.0" />
            <INFO name="SOURCE" value="
            gaiafpr.interstellar_medium_spectra" />
        <LINK title="gaiafpr.interstellar_medium_spectra"
        content-role="doc" href="https://doi.org/10.17876/gaia/
        fpr.1/3"/>
        <TABLE name="gaia_user_anonymous.2024-01-16-12-39-21-482003
        ">
            <FIELD name="solution_id" ucd="meta.version" datatype="
            long" />
            <FIELD name="healpix" ucd="pos.healpix" datatype="int"
            />
            <FIELD name="lc" unit="deg" ucd="pos.galactic.lon"
            datatype="float" />
            <FIELD name="bc" unit="deg" ucd="pos.galactic.lat"
            datatype="float" />
            <FIELD name="dc" unit="kpc" ucd="pos.distance" datatype
            ="float" />
            <FIELD name="lambda" unit="angstrom" ucd="em.wl"
            datatype="float" />
            <FIELD name="flux" ucd="phot.flux" datatype="float" />
            <FIELD name="flux_uncertainty" ucd="
            stat.error;phot.flux" datatype="float" />
            <DATA>
```

Use Datalink for multicomponent data sets

Datalink Service:
https://.www.plate-archive.org/datalink/links?ID=plates/101_2
yields available links for all information on **object plate101_2**

Datalinks for plates/101_2

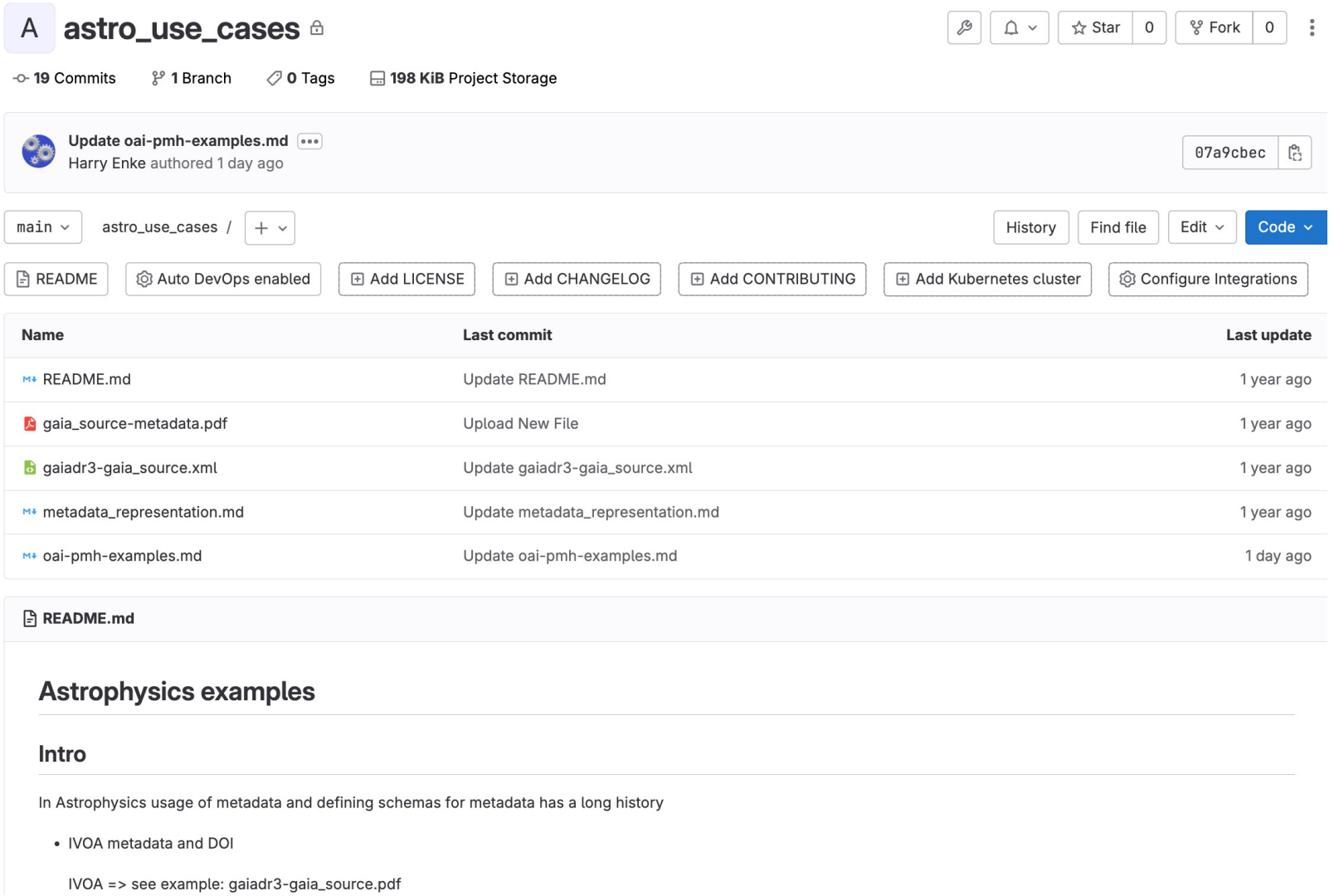
Data Links

access_url	description	semantics	content_type	content_lengt
https://www.plate-archive.org/files/DR3/scans/HAM-LA/LA00010_x.hdr	Header for the scan of the plate 2 in archive 101 from APPLAUSE DR3	#detached-header	text/plain	14.9 KB
https://www.plate-archive.org/files/DR3/scans/HAM-LA/LA00010_y.hdr	Header for the scan of the plate 2 in archive 101 from APPLAUSE DR3	#detached-header	text/plain	14.9 KB
https://doi.org/10.17876/plate/dr.3/plates/101_2	Plate 2 in archive 101 from APPLAUSE DR3	#doi	application/html	None
https://www.plate-archive.org/objects/dr.3/plates/101_2	Object viewer for the plate 2 in archive 101 from APPLAUSE DR3	#preview	application/html	None
https://www.plate-archive.org/files/DR4/scans/HAM-LA/LA00010_x.fits	File for the scan of the plate 2 in archive 101 from APPLAUSE DR3	#this	application/fits	997.0 MB
https://www.plate-archive.org/files/DR4/scans/HAM-LA/LA00010_y.fits	File for the scan of the plate 2 in archive 101 from APPLAUSE DR3	#this	application/fits	998.8 MB

Datalink Viewer: https://.www.plate-archive.org/datalink/plates/101_2

Additional pointers TA4-WP2

https://gitlab-p4n.aip.de/punch/ta4/wp2/astro_use_cases



A **astro_use_cases**

Star 0 Fork 0

19 Commits 1 Branch 0 Tags 198 KiB Project Storage

Update oai-pmh-examples.md
Harry Enke authored 1 day ago

07a9cbec

main astro_use_cases / + History Find file Edit Code

README Auto DevOps enabled Add LICENSE Add CHANGELOG Add CONTRIBUTING Add Kubernetes cluster Configure Integrations

Name	Last commit	Last update
README.md	Update README.md	1 year ago
gaia_source-metadata.pdf	Upload New File	1 year ago
gaiadr3-gaia_source.xml	Update gaiadr3-gaia_source.xml	1 year ago
metadata_representation.md	Update metadata_representation.md	1 year ago
oai-pmh-examples.md	Update oai-pmh-examples.md	1 day ago

README.md

Astrophysics examples

Intro

In Astrophysics usage of metadata and defining schemas for metadata has a long history

- IVOA metadata and DOI

  IVOA => see example: gaiadr3-gaia_source.pdf

Conclusion:

- IVOA and DataCite metadata have considerable overlap
  - especially useful for findability (F in FAIR) in interdisciplinary context
  - Datalink semantics can be used to implement services for both purposes
  - OAI-PMH works for both ‘worlds’
  - VOTable should have an extra tag for DOI for independent data findability