Leibniz-Institut für Astrophysik Potsdam

Using DOI on astronomical data @ AIP

Harry Enke + Yori Fournier / IVOA Spring Meeting 2023 / 10.05.2023

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
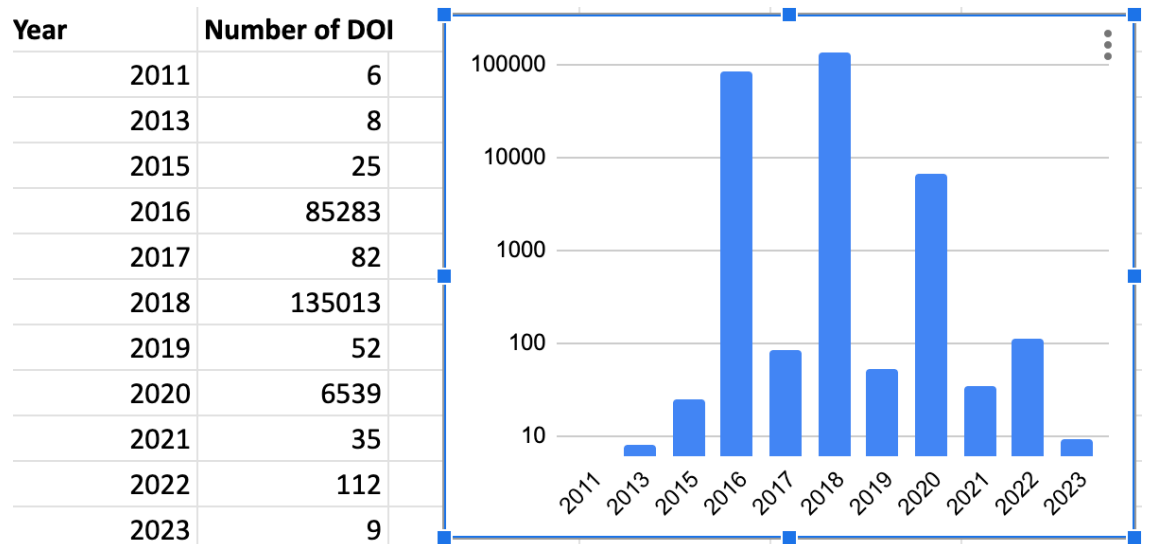
Subsequent discussions :

- 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

227,173
Works

22,609
Citations ⑦

Year	Number of DOI
2011	6
2013	8
2015	25
2016	85283
2017	82
2018	135013
2019	52
2020	6539
2021	35
2022	112
2023	9

huge numbers (> 50000) from

APPLAUSE (digitized & processed photographic plates)

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:
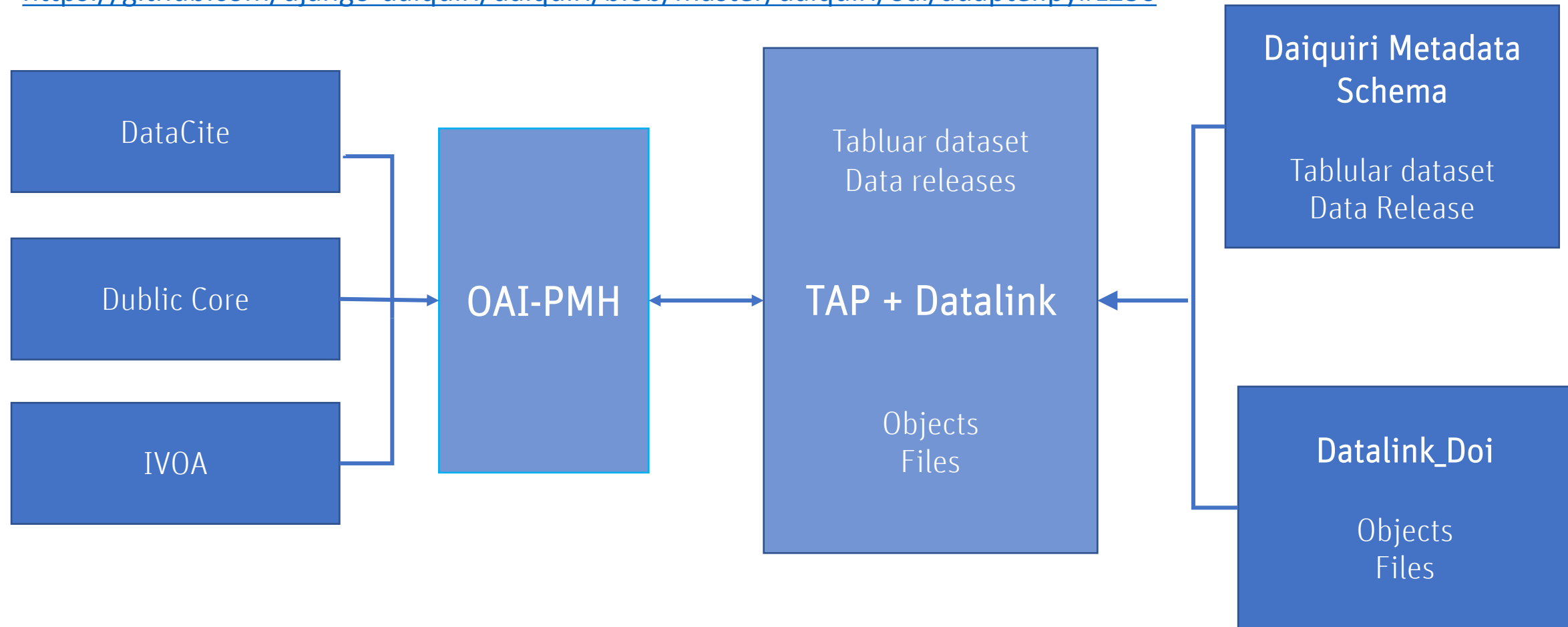
Example:

Datalink Service: https://applause-dev.sirrah.aip.de/datalink/links?ID=plates/101_2

Datalinks for plates/101_2

Data Links

access_url	description	semantics	content_length
https://applause-dev.sirrah.aip.de/files/DR2/scans/HAM-LA/LA00010_x.hdr	FITS header for the scan of the plate 2 in archive 101 from the APPLAUSE DR2	#detached-header	13587
https://applause-dev.sirrah.aip.de/files/DR3/scans/HAM-LA/LA00010_y.hdr	FITS header for the scan of the plate 2 in archive 101 from the APPLAUSE DR3	#detached-header	15288
https://applause-dev.sirrah.aip.de/files/DR2/scans/HAM-LA/LA00010_y.hdr	FITS header for the scan of the plate 2 in archive 101 from the APPLAUSE DR2	#detached-header	13587
https://applause-dev.sirrah.aip.de/files/DR3/scans/HAM-LA/LA00010_x.hdr	FITS header for the scan of the plate 2 in archive 101 from the APPLAUSE DR3	#detached-header	15288
https://doi.org/10.17876/plate/dr.3/plates/101_2	Digital object identifier (DOI) for the plate 2 in archive 101 from the APPLAUSE DR3	#doi	None
https://applause-dev.sirrah.aip.de/objects/dr.3/plates/101_2	Plate 2 in archive 101 from the APPLAUSE DR3	#preview	None
https://applause-dev.sirrah.aip.de/files/DR4/scans/HAM-LA/LA00010_x.fits	FITS file for the scan of the plate 2 in archive 101 from the APPLAUSE DR4	#this	1045405440
https://applause-dev.sirrah.aip.de/files/DR4/scans/HAM-LA/LA00010_y.fits	FITS file for the scan of the plate 2 in archive 101 from the APPLAUSE DR4	#this	1047283200

Datalink Viewer: https://applause-dev.sirrah.aip.de/datalink/plates/101_2

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