

## Software tools for data publishing

### A. Galkin, J. Klar, G. Matijevič, H. Enke



#### **Services** @AIP

At the Leibniz-Institute for Astrophysics Potsdam (AIP) we gathered profound experience in publishing astrophysical data for numerous projects and international collaborations. While a project web application requires a dedicated website with a specific look and feel, the process of publishing a data release has steps of data aquisition, data curation and publishing in common. We have build up a software stack, processes and tools for each of the steps.

The Daiquiri web framework offers a dedicated web interface for each of the hosted scientific databases. Scientists all around the world run SQL queries which include specific astrophysical functions.

#### **Daiquiri – a customizable framework for database publication**

#### **The data curation process**

Data curation is a continous process at all stages of the data publication. The to be published scientific data projects are provided on the AIP internal cluster. The data curators develop consitency and integrity checks in close collaboration with the project's data providers. The approved data is then ingested into databases. We use **ASCiilngest**, **DBIngestor** and other custom develped tools and scripts. Again, checks are run to ensure the internal consitency of the database tables, but also to compare the content of the database to the ingested data files. These checks usually are summary checks on how many objects / lines are ingested and also what values are expected. The data curator must have domain specific knowledge about the data to develop and apply the checking and ingesting tools.

The process might iterate several times, as sometimes inconsistencies of the data products are found during the database checks, or the data products are changed by the data providers after problems are identified during the curation process.

Daiquiri allows the publication of data for different roles, thus first the closed collaboration members can access and scrutinise the data. Each table receives a DOI and in the final step the data is released for the public.

![](_page_0_Picture_12.jpeg)

The Daiquiri software stack has now been released in its second incarnation based on the Python Django framework. While the main purpose is to access to data sets stored in SQL databases, also files, images and other data are served by various APIs. The data can be either released to the public or restricted to groups of a collaboration.

The user accesses the data through an interactive SQL query interface, also files and image viewers and a data plotting facility are available. Behind the scienes, the **queryparser** package as part of the Daiquiri framework allows to query the data using the native backend SQL language as well as ADQL (Astrophysical Data Query Language). Currently, MySQL and PostgreSQL are supported. Each user has her/his own private tablespace where the results of all their queries are stored and can be queried, retrieved and downloaded any time. Scripted access is enabled through a REST API and is based on the Table Access Protocol (TAP). Also, an OAI-PMH access is available as a module. The provision of **Digital Object Identifiers (DOI's)** for published data and the automatic generated DOI landing pages is available as a module.

For the project Daiquiri supports user management, contact messages

Data acquisition, consistency and integrity checks

![](_page_0_Picture_17.jpeg)

Last checks by the data providers and curators, published tables are provided with DOI's

Web applications

![](_page_0_Figure_19.jpeg)

and meetings management as well as an integration of Wordpress for documentation and project presentation. Daiquiri consists of a common code base, which is centrally maintained, and site-specific applications, which are highly customizable to the project requirements.

# Consistency checks, checks against the ingested data

Data

publication

#### Links and references:

Services@AIP: www.gaia.aip.de, www.cosmosim.org, www.plate-archive.org, www.rave-survey.org

All software we develop is open source and is avalable on GitHub: Daiquiri, DBIngestor, ASCiiIngest https://github.com/aipescience

 11. Juni 2018 - 7. Workshop des AK Forschungsdaten
 htps://escience.aip.de
 Contact: agalkin@aip.de