



HAL
open science

Prototyping access from visualisation tools to SKA science images and cubes stored in a Rucio Data Lake through IVOA discovery and access services

François Bonnarel, Mark G. Allen, Jesus Salgado, Robert Barnsley, Matthieu Baumann, Thomas Boch, Caroline Bot, Robert Butora, James Collinson, Pierre Fernique, et al.

► To cite this version:

François Bonnarel, Mark G. Allen, Jesus Salgado, Robert Barnsley, Matthieu Baumann, et al.. Prototyping access from visualisation tools to SKA science images and cubes stored in a Rucio Data Lake through IVOA discovery and access services. 2024. hal-04738803

HAL Id: hal-04738803

<https://hal.science/hal-04738803v1>

Preprint submitted on 15 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prototyping access from visualisation tools to SKA science images and cubes stored in a Rucio Data Lake through IVOA discovery and access services

François Bonnarel,³ Jesus Salgado,¹ Mark Allen,³ Rob Barnsley,¹ Matthieu Baumann,³ Thomas Boch,³ Caroline Bot,³ Robert Butora,² James Collinson,¹ Pierre Fernique,³ Vincenzo Galluzzi,² Rohini Joshi,¹ Andrea Lorenzani,² Marco Molinaro,² Manuel Parra-Royon,⁴ Jesus Sanchez-Castaneda,⁴ Susana Sanchez-Exposito,⁴ Eva Sciacca,² Giuseppe Tudisco,² Fabio Vitello,² and Alessandra Zanichelli²

¹*SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, Cheshire SK11 9FT, UK*

²*INAF - Istituto Nazionale di Astrofisica, Rome, Italy*

³*Université de Strasbourg, CNRS, Observatoire Astronomique de Strasbourg, UMR 7550, F-67000 Strasbourg, France ; francois.bonnarel@astro.unistra.fr*

⁴*Institute of Astrophysics of Andalusia (IAA-CSIC), Granada, Spain*

Abstract. SKA is the major low frequency radioastronomy project of the future with several major scientific applications: it will upgrade the amount of available science data by several orders of magnitudes reaching eventually more than 700 petabytes of storage per year. The SKA Observatory will proceed to the initial data processing to deliver observatory data products while the SKA Regional Center Network (SRCNet) will provide storage for those and processing capabilities to deliver and store advanced data products for the user community. Within the scope of the SRCNet, Orange (visualisation), Magenta (data management) and Coral (node implementation) teams have prototyped the discovery access and visualisation of science data. Our visualisation tools VisIVO and Aladin discover, access and visualize test science data produced by SKA precursors and pathfinders stored in the Rucio (Barisits et al. 2019) Data Lake. Science metadata functionality has been implemented by the Magenta team to the Rucio Data Lake prototype to demonstrate a means of enabling IVOA-compliant data discovery and server-side processing. VisIVO, Aladin Desktop and Aladin Lite are able to query the Discovery service built on ObsCore (Louys et al. 2017) and SCS (Plante et al. 2008) IVOA protocols. This allows them to load DataLink (Dowler et al. 2015) responses providing links towards a SODA (Bonnarel et al. 2017) cutout service developed by the Orange team able to extract sub-cubes or images directly from the datasets stored in the Rucio Data Lake. The Rucio Storage Element and SODA developments have been deployed and configured on the Spanish SRC node, providing computing and storage resources, managed by the Coral Team members. This prototype paves the way to collaborative development in the SRCNet and shows the possible integration of VO services and visualisation tools in Data Lakes and science platforms.

1. Introduction

Within the scope of the Square Kilometer Array (SKA) project, the SRCNet (SKA Regional Centres Network) project is a worldwide distributed storage and computation system made of collaborating regional nodes. It will be in charge of advanced data processing, data storage and distribution. The prototype described here intends to show how some astronomy oriented visualisation tools (here: Aladin and VisIVO) could discover, access and display data stored in the Data Lake distributed data management system of the SRCNet (here prototyped with Rucio).

2. SRCNet prototyping

The development in the SRCNet project is following the Agile methodology rules. In the current prototyping phase the project is organized in a dozen of Agile prototyping teams. There are three prototyping teams involved in building the prototype presented in this poster: Magenta team (SKAO) for development of data management and metadata services, Orange team (INAF + CNRS) for adaptation of visualisation tools (VisIVO and Aladin) and SODA (IVOA cutout service) development and Coral team (with members from Spain, UK, Sweden and Switzerland) for providing storage and computing resources that are part of the Spanish SRC prototype platform (espSRC).

3. Setting IVOA metadata in the Data Lake metadata storage system

Figure 1 shows how the IVOA metadata are set in the Rucio Metadata Storage System and then interfaced through DaCHS (Demleitner et al. 2014) VO interfaces with visualisation tools.

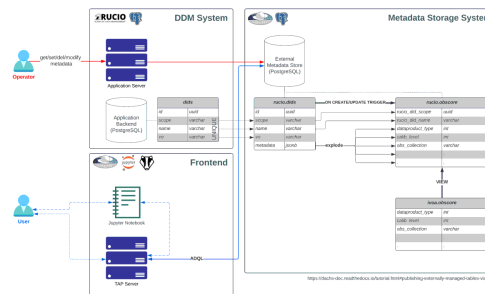


Figure 1. IVOA metadata management and access in the Rucio Distributed Data Management (DDM) system.

For each dataset stored in the Rucio DDM system, we have set IVOA ObsCore-style metadata in the external metadata storage system using scripts of `'rucio set-metadata'` commands. Metadata have been extracted from test data FITS headers and mapped to ObsCore data model in order to build automatically the set-metadata scripts. The database is interfaced by DaCHS to be accessible through VO services such as TAP or SCS.

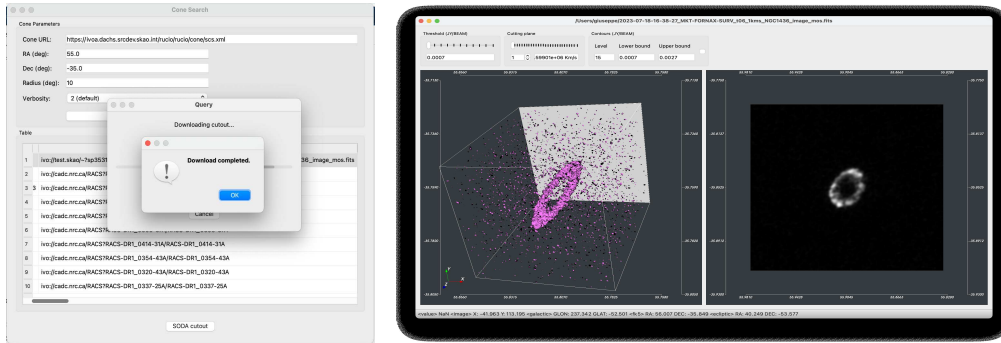


Figure 2. VisIVO SCS interface and ObsCore response (left) and visualisation of a spectral cube extracted via SODA(right)

4. Test data

In order to demonstrate the capabilities of the system, we have integrated test data (images and cubes) in the Data Lake. These datasets are from SKA precursor and pathfinder telescopes, such as MeerKAT, ASKAP, LOFAR and JVLA, considering, in particular, large programmes such as Apertif and VLASS. In addition, SKA simulated images and cubes produced within the scope of the three SKA data challenges have been also integrated. These datasets cover a wide range of frequency bandwidth, spectral resolution and data size (from hundred of megabytes to tens of gigabytes).

5. Nodes installation

This setup process involves configuring essential components for effective data management. The Rucio server and metadata service are deployed at the Science & Technology Facilities Council Cloud (STFC) in the United Kingdom, ensuring centralized and robust data dissemination. Simultaneously, the installation of the Rucio Storage Element (a logical abstraction of a storage system for physical files in one location) and SODA service, both on the espSRC (Garrido et al. 2022), facilitates efficient utilization of computing resources, allowing users to execute tools for visualizing datasets stored within the Rucio data lake. This strategic setup not only supports day-to-day operations but also serves to simulate specific behaviours, contributing to a more comprehensive understanding of the SRCNet.

6. Discovering and accessing datasets from visualisation tools through IVOA standards

VisIVO (figure 2), Aladin Desktop and Aladin Lite (figure 3) have been customized to access the Rucio data discovery service delivering ObsCore metadata through an IVOA SCS interface. From that response, the three tools are able to load the DataLink response in order to discover full retrieval and dataset cutout accesses. Each tool is interfacing with the IVOA SODA cutout service to extract datasets of interest. The SODA service is directly accessing datasets by their Rucio physical file names.

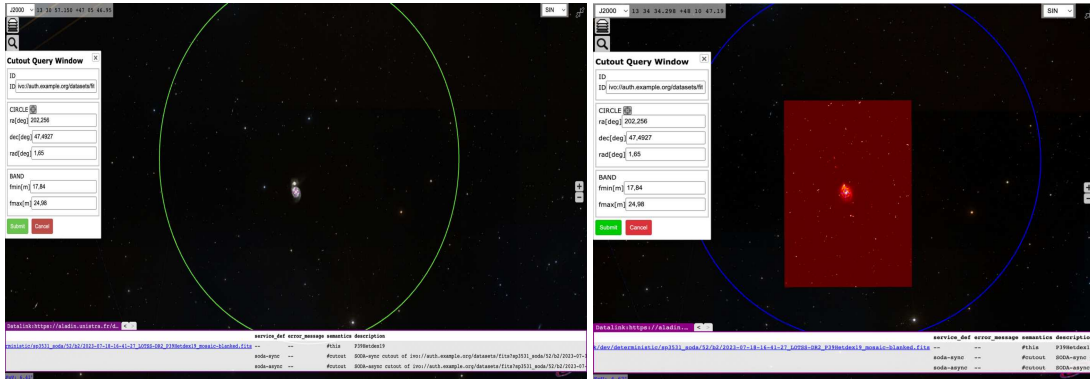


Figure 3. SODA Form (left) and result cutout image from SODA server display (right) within AladinLite

7. Newer developments

We are currently implementing access control using IAM in front of the SODA service. We create additional DataLinks records with derived data such as moment maps or spectral index maps in FITS and HiPS format for advanced data discovery.

References

- Barisits, M., Beermann, T., Berghaus, F., Bockelman, B., Bogado, J., Cameron, D., Christidis, D., Ciangottini, D., Dimitrov, G., Elsing, M., Garonne, V., di Girolamo, A., Goossens, L., Guan, W., Guenther, J., Javurek, T., Kuhn, D., Lassnig, M., Lopez, F., Magini, N., Molfetas, A., Nairz, A., Ould-Saada, F., Prenner, S., Serfon, C., Stewart, G., Vaandering, E., Vasileva, P., Vigne, R., & Wegner, T. 2019, arXiv e-prints, arXiv:1902.09857. 1902.09857
- Bonnarel, F., Dowler, P., Demleitner, M., Tody, D., & Dempsey, J. 2017, IVOA Server-side Operations for Data Access Version 1.0, IVOA Recommendation 17 May 2017. 1710.08791
- Demleitner, M., Neves, M. C., Rothmaier, F., & Wambsganss, J. 2014, Astronomy and Computing, 7, 27. 1408.5733
- Dowler, P., Bonnarel, F., Michel, L., & Demleitner, M. 2015, IVOA DataLink Version 1.0, IVOA Recommendation 17 June 2015. 1509.06152
- Garrido, J., Darriba, L., Sánchez-Expósito, S., Parra-Royón, M., Moldón, J., Mendoza, M. A., Luna-Valero, S., Alberdi, A., Márquez, I., & Verdes-Montenegro, L. 2022
- Louys, M., Tody, D., Dowler, P., Durand, D., Michel, L., Bonnarel, F., Micol, A., & IVOA DataModel Working Group 2017, Observation Data Model Core Components, its Implementation in the Table Access Protocol Version 1.1, IVOA Recommendation 09 May 2017
- Plante, R., Williams, R., Hanisch, R., & Szalay, A. 2008, Simple Cone Search Version 1.03, IVOA Recommendation 22 February 2008. 1110.0498

Acknowledgments. Susana, Manuel and Jesus acknowledge the Spanish Prototype of an SRC (espSRC) service and support funded by the Spanish Ministry of Science, Innovation and Universities, by the Regional Government of Andalusia, by the European Regional Development Funds and by the European Union NextGenerationEU/PRTR. The espSRC acknowledges financial support from the State Agency for Research of the Spanish MCIU through the "Center of Excellence Severo Ochoa" award

to the IAA-CSIC (SEV-2017-0709) and CEX2021-001131-S funded by MCIN/AEI/
10.13039/501100011033.