



**HAL**  
open science

## Data stewardship of next-generation data products and services related to ESFRI and other astrophysics infrastructures

Mark G. Allen, Stefania Amodeo, François Bonnarel, Marco Molinaro, Françoise Genova, Martino Romaniello, Hendrik Heintz, André Schaaff, Soizick Lesteven

### ► To cite this version:

Mark G. Allen, Stefania Amodeo, François Bonnarel, Marco Molinaro, Françoise Genova, et al.. Data stewardship of next-generation data products and services related to ESFRI and other astrophysics infrastructures. 2021. hal-03372190

**HAL Id: hal-03372190**

**<https://hal.science/hal-03372190>**

Preprint submitted on 9 Oct 2021

**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.

## Data stewardship of next-generation data products and services related to ESFRI and other astrophysics infrastructures

2 MARK G. ALLEN <sup>1</sup>, STEFANIA AMODEO <sup>1</sup>, FRANÇOIS BONNAREL<sup>1</sup>, MARCO MOLINARO <sup>2</sup>, FRANÇOISE GENOVA <sup>1</sup>,  
3 MARTINO ROMANIELLO <sup>3</sup>, HENDRIK HEINL<sup>1</sup>, ANDRÉ SCHAFF <sup>1</sup> AND SOIZICK LESTEVEN <sup>1</sup>

4 <sup>1</sup>*Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbourg, France*

5 <sup>2</sup>*INAF - Osservatorio Astronomico di Trieste, via Tiepolo 11, 34143 Trieste, Italy*

6 <sup>3</sup>*European Southern Observatory (ESO), Karl-Schwarzschild-Str, D-85748 Garching, Germany*

### 7 ABSTRACT

8 The ESCAPE project (European Science Cluster of Astronomy & Particle physics ESFRI research  
9 Infrastructure) is addressing the Open Science challenges shared by the astrophysics ESFRI facilities  
10 (SKA, CTA, KM3NeT, EST) as well as other pan-European research infrastructures (ESO, JIVE)  
11 in the context of the European Open Science Cloud (EOSC). One of the activities of ESCAPE is to  
12 identify the best practices for data stewardship in astronomy and in particular how to use common  
13 IVOA standards to implement the FAIR principles. We present a summary of the interoperability  
14 topics being pursued in ESCAPE in the different scientific domain areas, and relate these activities to  
15 data stewardship.

16 *Keywords:* Astronomy databases (83) — Astronomy web services (1856) — Virtual Observatoires  
17 (1774) — Interdisciplinary astronomy(804)

### 18 1. INTRODUCTION

19 The ESCAPE project (European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures<sup>1</sup>)  
20 is addressing the Open Science challenges shared by the astrophysics and accelerator particle physics facilities that  
21 are labelled as ESFRI<sup>2</sup> by the European Strategy Forum for Research Infrastructures. The astrophysics ESFRI  
22 infrastructures are the CTA (Cherenkov Telescope Array), ELT (ESO Extremely Large Telescope), EST (European  
23 Solar Telescope), KM3NeT (Kilometre Cube Neutrino Telescope), and SKA (Square Kilometre Array). The other  
24 pan-European research infrastructures involved are ESO (European Southern Observatory), JIVE (Joint Institute for  
25 VLBI in ERIC) and EGO/Virgo (European Gravitational Observatory Virgo collaboration). The project is being done  
26 in the context of the European Open Science Cloud (EOSC<sup>3</sup>) which is a major European initiative for Open Science.

27 The CEVO<sup>4</sup> Work Package of the ESCAPE project (Allen et al. 2020) addresses the implementation of FAIR  
28 principles for Astronomy and Astroparticle ESFRI by using IVOA<sup>5</sup> (International Virtual Observatory Alliance)  
29 standards. The FAIR principles (Wilkinson et al. 2016) are guidelines for data resources to be findable, accessible,  
30 interoperable and re-usable. These principles are key to Open Science as described in the EOSC Association Strategic  
31 Research and Innovation Agenda<sup>6</sup>. In order to implement FAIR in practice it requires the scientists, software engineers  
32 and data experts to face the different ‘data stewardship’ challenges across the different facilities and data types. The  
33 work of the project is built around the requirements of the various ESFRI, in particular their needs for standardisation  
34 and interoperable tools. A recent Technology Forum event<sup>7</sup> was held to track the progress of the work. This poster  
35 paper presents a summary of the most important interoperability topics that have been identified in the different

Corresponding author: Mark G. Allen  
mark.allen@astro.unistra.fr

<sup>1</sup> <https://projectescape.eu>

<sup>2</sup> <https://www.esfri.eu>

<sup>3</sup> <https://www.eosc.eu>

<sup>4</sup> CEVO - Connecting ESFRI to EOSC using the Virtual Observatory framework

<sup>5</sup> <https://ivoa.net>

<sup>6</sup> [https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0\\_15Feb2021.pdf](https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf)

<sup>7</sup> <https://indico.in2p3.fr/event/23481/>

scientific domain areas, and the implications for data stewardship of the different types of data and services and how this is evolving with the development of EOSC. All the IVOA standards cited in the text can be found at <https://www.ivoa.net/documents/>.

## 2. RADIO AND MILLIMETRE ASTRONOMY

Radio and millimetre astronomy produce a wide range of data with special characteristics related to the interferometric nature of the observations. These data have typically been difficult to use by non-specialists, but the developments of current facilities (such as ALMA, JIVE, LOFAR) and new facilities (e.g. SKA) are putting much effort into making the data widely usable. The ESCAPE project contributed to a recent review, published as an IVOA Note<sup>8</sup>, of how the Virtual Observatory can be better adapted to work with radio data. The emphasis of the work, based on the ALMA, JIVE and SKA requirements, is on data discovery and visualisation. Firstly to *"Make the data easy to find"* – which requires standards for discovery of interferometric data, and definition of the necessary metadata for interoperability. Secondly, a way to encode *"What happened to these data?"* – for which the IVOA Provenance Data model (Servillat et al. 2020) has been developed and then tested for radio astronomy (Servillat et al. 2021b). In terms of visualisation, there is a strong motivation to implement the all-sky zoomable interfaces that are enabled by the IVOA HiPS standard (Fernique et al. 2015) and also access to VO services for radio data from Python language programs and notebooks.

## 3. HIGH ENERGY ASTROPHYSICS

High energy astrophysics by the future Cherenkov Telescope Array (CTA) will be pursued with an observatory operational model, and many of the interoperability challenges involve the development of archiving and data access services for the complex CTA data. CTA will detect the Cherenkov radiation from particle showers triggered in the Earth atmosphere by high energy particles from astrophysical sources. The data processing chains require the capturing of the 'provenance' information at each step to describe what happened to the data using a common standard for the description. The CTA requirements have been used in the use-cases for the development of the IVOA Provenance Data Model (Servillat et al. 2020). The focus of this effort is to allow users to access quality and reliability information. This work for CTA has been used as an example of how the VO Provenance Scheme can be applied to a large complex project. Its applicability to ESFRI projects and scientific research, has been highlighted in a recent ESCAPE workshop<sup>9</sup>, and also in a review of practical provenance issues for implementation of the provenance standard (Servillat et al. 2021b), and this led to a proposed management system for provenance information (Servillat et al. 2021a).

## 4. NEUTRINO ASTROPHYSICS

The KM3NeT neutrino telescope is involved in ESCAPE CEVO with the work focused on the improvement in the use of common IVOA standards for KM3NeT. Some of the first steps for making neutrino data available in the VO include the establishment of expertise within KM3NeT for the VO publication of tabular data and publication of events (e.g. VOEvents). This is strongly motivated by the need to integrate neutrino data services into multi-messenger astrophysics systems, and to set up complementarity with the services of the KM3NeT Open Data Center<sup>10</sup>. The ESCAPE project work has identified synergies between the requirements for CTA and KM3NeT in terms of provenance standards, as well as for the using of VO applications and tools (e.g. TOPCAT, Aladin).

## 5. OPTICAL/UV/IR ASTRONOMY

The use of VO standards is embedded in the operations of many archives and data centres of optical, ultraviolet and infrared astronomy. IVOA has taken an approach of engaging with the large astronomy projects to develop the necessary standards and tools (e.g. Allen et al. (2020) and Berriman et al. (2020)). The ESCAPE project supports the development of IVOA standards based on the needs, which have been updated recently to include standards for time-domain astrophysics and alerts for multi-messenger astronomy, as well as standards for handling and accessing multi-dimensional data (e.g. cut-outs and visualisation). This is in addition to the maintenance and evolution of the standards that are implemented in services such as the ESO archive (Romaniello et al. 2018). The relevant standards<sup>11</sup> are ADQL 2.0 (Osuna et al. 2008), DataLink v1.0 (Dowler et al. 2015), ObsCore v1.1 (Louys et al. 2017), SSAP v1.1

<sup>8</sup> <https://ivoa.net/documents/Notes/RadioVOImp/index.html>

<sup>9</sup> <https://indico.in2p3.fr/event/21913/>

<sup>10</sup> <https://www.km3net.org/km3net-infradev/open-access-to-km3net-data/>

<sup>11</sup> ADQL - Astronomical Data Query Language, ObsCore - Observational Core Data Model, SSAP - Simple Spectral Access Protocol, TAP - Table Access Protocol

(Tody et al. 2012), TAP v1.1 (Dowler et al. 2019). A major improvement to the Aladin Lite web visualiser is also underway for use in multiple astronomy services, making use of the HiPS 1.0 standard (Fernique et al. 2019). The progress toward all these standards are recorded in Milestone reports of CEVO, which are available in the public library of ESCAPE Milestones and Reports<sup>12</sup>.

One of the innovative activities being pursued in ESCAPE is the prototyping of services for value-added data in astronomy archives. A highlight of this work is the application of Deep Learning techniques to enable searches for 'similar data' within the ESO archive of HARPS instrument spectra - see Sedaghat et al. (2021). Various prototypes have been tested to explore how these types of services can go beyond the traditional data archive interfaces, and offer ways to "let the data speak" in the sense of enabling new views of the data based on characteristics identified by machine learning or other techniques.

## 6. GRAVITATIONAL WAVE ASTRONOMY

The detection of gravitational waves and in particular the example where an electromagnetic counterpart was found by follow-up observations, has marked the beginning a new era of multi-messenger astrophysics (Abbott et al. 2017). Interoperability of multi-wavelength data is essential for these studies which involve measurements from a wide range of instruments. The EGO/Virgo partner in ESCAPE identified the need for tools for visualising and managing the sky regions covered by the credible region probability maps and by the complex footprints of sky survey coverages. The key development supported in the ESCAPE project is the addition of the time-dimension to the IVOA standardised Multi-Order Coverage (MOC) maps. By including the temporal coverage to the spatial sky coverage maps, it enables searches of data for observations that overlap in both space and time. This builds on the work done in the ASTERICS project (Greco et al. 2019), and the new functions have been implemented into Python libraries for the use of the space-time MOCs (Baumann et al. 2020).

## 7. SOLAR PHYSICS

The European Solar Telescope (EST) is included in the astrophysics ESFRI participating in the ESCAPE project. The emphasis of the activities in the CEVO work is the application of IVOA standards for solar physics data, and setting up the basis for interoperability between the astronomy Virtual Observatory and the SOLARNET virtual observatory<sup>13</sup>. This involves a first trial for installation of a IVOA standard TAP service at the Royal Observatory of Belgium, and a study on how to link a TAP service and a solar web browser: JHelioviewer<sup>14</sup>. The ESCAPE supported participation of EST partners in the IVOA interoperability meetings has enabled extensions of the IVOA metadata (Unified Content Descriptors 'UCD') for the semantics of data content for solar data. This is being used to support cross-column, or even cross-catalog search in SOLARNET VO.

## 8. DATA STEWARDSHIP

The ESCAPE project brings the Astronomy and Particle Physics ESFRI infrastructures together to address their common challenges. In doing so, we find that there is a strong motivation to support Open Science via the use of the FAIR principles. One of the concepts reinforced by the project is that it is essential to have disciplinary level standards, and that these standards need to include or allow interfaces with wider interdisciplinary standards and frameworks. Disciplinary level standards are necessary because there are specific data and information characteristics within a given domain, that must be managed by experts in order to maintain scientific integrity of the data and services. In astrophysics, this includes the disciplinary level metadata for semantic content (IVOA standard UCDs), and also the systems for sky coordinates and indexing (e.g. World Coordinate Systems (Greisen & Calabretta 2002), and the IVOA HiPS/MOC standards), as well as for characterising astronomical observations (such as the IVOA Observational Core Data Model) to name just a few examples. In practice, it is critical for the bodies that will implement the standards to be involved in their definition and maintenance, which has implications for the data stewardship roles that are needed to support open science in astronomy infrastructures.

With the creation of Open Science frameworks such as EOSC in Europe, and in particular the inclusion of existing infrastructures into these, the astrophysics standards and data services can be seen as part of the systems that have a wider scope of data sharing across all areas of science. This requires that the astronomy standards be compatible

<sup>12</sup> <https://projectescape.eu/deliverables-and-reports>

<sup>13</sup> <http://solarnet.oma.be/>

<sup>14</sup> <https://www.jhelioviewer.org/>

at some level with more generic standards. This is already the case with the IVOA Resource Registry which uses the Dublin Core metadata and OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), and enables the harvesting of records by EUDAT/B2FIND<sup>15</sup> in readiness for inclusion in EOSC catalogues. Also, the IVOA standards definitions themselves which are included in the FAIRSharing database of standards<sup>16</sup>. Detailed interdisciplinary science will demand higher levels of cross-discipline data compatibility, but this is already a big step forward in terms of findability, access and interoperability of resource descriptions. In terms of the EOSC architecture, the astronomy standards would contribute to the "interoperability layer" and VO services can be considered as "thematic services" alongside of those from other domains (e.g. Environmental Science, Life Science, Social Sciences, Photon and Nuclear Sciences). CEVO is also following the efforts to define the cross-disciplinary semantic standard layer in EOSC, so that we will be ready to interface the IVOA semantic standards. Various services, such as authentication and authorisation may be provided generically via shared systems (e.g. the EOSC Core).

In terms of data stewardship, it requires that scientists, software engineers, data publication experts and librarians be aware of the wider context of their work. For example, in these various roles, it will be important to use interoperable metadata and to benefit from systems/tools/services that can be made common across domains. This is challenging because it requires a level of participation in activities that extend beyond the traditional boundaries of a discipline. The EOSC initiative is promoting widespread uptake of Open Science, and projects such as FAIRsFAIR<sup>17</sup> are fostering FAIR data practices, so there is a rapidly growing set of resources for training which can be of great value. The work on the interoperability topics in the different areas of astrophysics described above is building up data stewardship expertise in the various ESFRI projects in order for common standards to be implemented for data sharing. A longer term aspiration is that the data stewardship skills become more widely recognised and rewarded in career paths, and that it stimulates new levels of innovation to enable new kinds of science.

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 824064.

## REFERENCES

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, *The Astrophysical Journal Letters*, 848, L12, doi: [10.3847/2041-8213/aa91c9](https://doi.org/10.3847/2041-8213/aa91c9)
- Allen, M. G., Dowler, P., Evans, J. D., et al. 2020, in *Astronomical Society of the Pacific Conference Series*, Vol. 527, *Astronomical Society of the Pacific Conference Series*, ed. R. Pizzo, E. R. Deul, J. D. Mol, J. de Plaa, & H. Verkouter, 795. <https://arxiv.org/abs/1912.01481>
- Allen, M. G., Lamanna, G., Espinal, X., et al. 2020, arXiv, arXiv:2012.11534
- Baumann, M., Boch, T., Fernique, P., Nebot, A., & Pineau, F. X. 2020, in *Astronomical Society of the Pacific Conference Series*, Vol. 527, *Astronomical Society of the Pacific Conference Series*, ed. R. Pizzo, E. R. Deul, J. D. Mol, J. de Plaa, & H. Verkouter, 693
- Berriman, G. B., IVOA Executive Committee, T., The IVOA Technical Coordination Group, & IVOA Community, T. 2020, arXiv e-prints, arXiv:2012.05988. <https://arxiv.org/abs/2012.05988>
- Dowler, P., Bonnarel, F., Michel, L., & Demleitner, M. 2015, IVOA DataLink Version 1.0, IVOA Recommendation 17 June 2015, doi: [10.5479/ADS/bib/2015ivoa.spec.0617D](https://doi.org/10.5479/ADS/bib/2015ivoa.spec.0617D)
- Dowler, P., Rixon, G., Tody, D., & Demleitner, M. 2019, Table Access Protocol Version 1.1, IVOA Recommendation 27 September 2019
- Fernique, P., Boch, T., Donaldson, T., et al. 2019, MOC - HEALPix Multi-Order Coverage map Version 1.1, IVOA Recommendation 07 October 2019
- Fernique, P., Allen, M. G., Boch, T., et al. 2015, *A&A*, 578, A114, doi: [10.1051/0004-6361/201526075](https://doi.org/10.1051/0004-6361/201526075)
- Greco, G., Branchesi, M., Chassande-Mottin, E., et al. 2019, *Proceedings of Science, Asterics2019*, 031, doi: [10.22323/1.357.0031](https://doi.org/10.22323/1.357.0031)
- Greisen, E. W., & Calabretta, M. R. 2002, *A&A*, 395, 1061, doi: [10.1051/0004-6361:20021326](https://doi.org/10.1051/0004-6361:20021326)

<sup>15</sup> <https://eudat.eu/services/b2find>

<sup>16</sup> <https://fairsharing.org/standards?q=IVOA>

<sup>17</sup> <https://www.fairsfair.eu>

- 186 Louys, M., Tody, D., Dowler, P., et al. 2017, Observation  
187 Data Model Core Components, its Implementation in the  
188 Table Access Protocol Version 1.1, IVOA  
189 Recommendation 09 May 2017,  
190 doi: [10.5479/ADS/bib/2017ivoa.spec.0509L](https://doi.org/10.5479/ADS/bib/2017ivoa.spec.0509L)
- 191 Osuna, P., Ortiz, I., Lusted, J., et al. 2008, IVOA  
192 Astronomical Data Query Language Version 2.00, IVOA  
193 Recommendation 30 October 2008,  
194 doi: [10.5479/ADS/bib/2008ivoa.spec.1030O](https://doi.org/10.5479/ADS/bib/2008ivoa.spec.1030O)
- 195 Romaniello, M., Zampieri, S., Delmotte, N., et al. 2018,  
196 The Messenger, 172, 2
- 197 Sedaghat, N., Romaniello, M., Carrick, J. E., & Pineau,  
198 F.-X. 2021, Monthly Notices of the Royal Astronomical  
199 Society, 501, 6026
- 200 Servillat, M., Bonnarel, F., Boisson, C., et al. 2021a, in  
201 Provenance and Annotation of Data and Processes, ed.  
202 B. Glavic, V. Braganholo, & D. Koop (Cham: Springer  
203 International Publishing), 244–249
- 204 Servillat, M., Bonnarel, F., Louys, M., & Sanguillon, M.  
205 2021b, arXiv, arXiv:2101.08691
- 206 Servillat, M., Riebe, K., Boisson, C., et al. 2020, IVOA  
207 Recommendation 11 April 2020
- 208 Tody, D., Dolensky, M., McDowell, J., et al. 2012, Simple  
209 Spectral Access Protocol Version 1.1, IVOA  
210 Recommendation 10 February 2012,  
211 doi: [10.5479/ADS/bib/2012ivoa.spec.0210T](https://doi.org/10.5479/ADS/bib/2012ivoa.spec.0210T)
- 212 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al.  
213 2016, Scientific Data, 3, 160018,  
214 doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)