| | |
|---|---|
| Project Title | European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructure |
| Project Acronym | ESCAPE |
| Grant Agreement No | 824064 |
| Instrument | Research and Innovation Action (RIA) |
| Topic | Connecting ESFRI infrastructures through Cluster projects (INFRA-EOSC-4-2018) |
| Start Date of Project | 01.02.2019 |
| Duration of Project | 42 Months |
| Project Website | www.projectescape.eu |

# D4.2 - INTERMEDIATE ANALYSIS REPORT ON USE OF IVOA STANDARDS FOR FAIR ESFRI AND COMMUNITY DATA

| | |
|---|---|
| Work Package | WP4, Connecting ESFRI projects to EOSC through VO |
| Lead Author (Org) | F. GENOVA (CNRS/ObAS) |
| Contributing Author(s) (Org) | F. BONNAREL (CNRS/ObAS), M. ALLEN (CNRS/ObAS) |
| Due Date | 31.03.2020, M14 |
| Date | 31.03.2020 |
| Version | 1.0 |

Dissemination Level

| | |
|---|---|
| X | PU: Public |
| | PP: Restricted to other programme participants (including the Commission) |
| | RE: Restricted to a group specified by the consortium (including the Commission) |
| | CO: Confidential, only for members of the consortium (including the Commission) |

## Versioning and contribution history

| Version | Date | Authors | Notes |
|---|---|---|---|
| 0.1 | 03/03/2020 | F. Genova (CNRS/ObAS) | First draft circulated to co-authors. |
| 0.2 | 04/03/2020 | F. Genova, F. Bonnarel, M. Allen (CNRS/ObAS) | Co-authors' comments included. |
| 0.3 | 23/03/2020 | F. Genova, F. Bonnarel, M. Allen (CNRS/ObAS) | Partners' comments included. |
| 0.4 | 25/03/2020 | F. Genova, F. Bonnarel, M. Allen (CNRS/ObAS) | Text checked by co-authors. |
| 1.0 | 26/03/2020 | F. Genova, F. Bonnarel, M. Allen (CNRS/ObAS) | Version 1.0 submitted to Project Manager. |

## Disclaimer

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 the Grant Agreement n° 824064.

# Table of contents

# Acronym list

AAI: Authentication and Authorisation Infrastructure

ADASS: Astronomical Data Analysis Software and Systems

ADQL: Astronomical Data Query Language

Aladin/AladinLite: Sky atlas and discovery tool

ALMA: Atacama Large Millimetre Array

ASTERICS: Astronomy ESFRI & Research Infrastructure Cluster

ASTRON: the Netherlands Institute for Radio Astronomy

AstroPy: Community Python library for Astronomy

B2FIND: EUDAT metadata service and discovery portal

CERN: European Organization for Nuclear Research

CEVO: Connecting ESFRI projects to EOSC through the Virtual Observatory framework (ESCAPE Work Package 4)

Characterisation: Characterisation Data Model

CNRS: Centre National de la Recherche Scientifique

CNRS-ObAS: CNRS - Observatoire Astronomique de Strasbourg

COAM: Common Observation Archive Model

CTA: Cherenkov Telescope Array

D#: Deliverable number

DADI: Data Access, Discovery and Interoperability (ASTERICS Work Package 4)

ECAP: Erlangen Center for Astroparticle Physics

EGO-Virgo: European Gravitational-Wave Observatory

ELT: Extremely Large Telescope (was E-ELT)

EOSC: European Open Science Cloud

EPN: Europlanet

EPN-TAP: Europlanet Table Access Protocol

ESCAPE: European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

ESFRI: European Strategy Forum on Research Infrastructures

ERIC: European Research Infrastructure Consortium

ESAP: ESFRI Science Analysis Platform (Escape Work Package 5)

ESO: European Southern Observatory

EST: European Solar Telescope

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 the Grant Agreement n° 824064.

EUDAT: European DATa (e-infrastructure)

FAIR: Findable, Accessible, Interoperable, Reusable *or* Facility for Antiproton and Ion Research

FITS: Flexible Image Transport System

FREQSYS: Metadata for frequency/wavelength in VOTable

HDF5: Hierarchical Data Format, version 5

HiPS: Hierarchical Progressive Survey

HL-LHC: High Luminosity Large Hadron Collider

INAF: Istituto Nazionale di Astrofisica

INTA: Instituto Nacional de Tecnica Aeroespacial

IR: Infra-Red

IVOA: International Virtual Observatory Alliance

JIVE: Joint Institute for VLBI ERIC

KIS: Kiepenheuer Institut für SonnenPhysik

KM3NeT: A multi-km3 sized Neutrino Telescope

MS#: Milestone number

MOC: Multi-Order Coverage

mocpy: Python library to create and manipulate MOCs

ObsCore: Observation Core Component Data Model

ObsParis: Observatoire de Paris

ObsTAP: Observation Core Component Data Model using the Table Access Protocol

ProvTAP: Serialization of the Provenance Data Model with TAP

RI: Research Infrastructure

ROB: Royal Observatory of Belgium

SAMP: Simple Application Messaging Protocol

SIA: Simple Image Access

SIMBAD: Set of Identifications, Measurements, Bibliography for Astronomical Data

SKA: Square Kilometre Array

StandardsRegExt: Extension of the resource schema to standards

SODA: Server-side Operations for Data Access

TAP: Table Access Protocol

TAPRegExt: Extension of the resource schema to TAP resources

TIMESYS: Metadata for time in VOTable

UCDs: Unified Content Descriptors

UEDIN: University of Edinburgh

UHEI: Ruprecht-Karls-Universität Heidelberg

UV: Ultraviolet

UWS: Universal Worker Service

VAMDC: Virtual Atomic and Molecular Data Centre

VizieR: Database of astronomical catalogues, published tables and other data

VLBI: Very-long-baseline interferometry

VO: Virtual Observatory

VODataService: Extension of the IVOA Resource Metadata to describe data collections and the services to access them

VOResource: XML encoding schema for resource metadataVOEvent: Virtual Observatory Event

VOTable: Virtual Observatory Table format standard

W3C: World Wide Web Consortium

WG: Working Group

XML: eXtensible Markup Language

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 the Grant Agreement n° 824064.

# Project Summary

ESCAPE (European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures) addresses the Open Science challenges shared by ESFRI facilities (CTA, ELT, EST, FAIR, HL-LHC KM3NeT and SKA) as well as other pan-European research infrastructures (CERN, ESO, JIVE and EGO) in astronomy and particle physics. ESCAPE actions are focused on developing solutions for the FAIRness of large data sets handled by the ESFRI facilities.

These solutions shall: i) connect ESFRI projects to EOSC ensuring integration of data and tools; ii) foster common approaches to implement open-data stewardship; iii) establish interoperability within EOSC as an integrated multi-probe facility for fundamental science.

To accomplish these objectives, ESCAPE aims to unite astrophysics and particle physics communities with proven expertise in computing and data management by setting up a data infrastructure beyond the current state-of-the-art in support of the FAIR principles. These joint efforts are expected result into a data-lake infrastructure as cloud open-science analysis facility linked with the EOSC. ESCAPE supports already existing infrastructure such as astronomy Virtual Observatory to connect with the EOSC. With the commitment from various ESFRI projects in the cluster, ESCAPE will develop and integrate the EOSC catalogue with a dedicated catalogue of open-source analysis software. This catalogue will provide researchers across the disciplines with new software tools and services developed by astronomy and particle physics community. Through this catalogue, ESCAPE will strive to cater researchers with consistent access to an integrated open-science platform for data-analysis workflows. As a result, a large community "foundation" approach for cross-fertilisation and continuous development will be strengthened. ESCAPE has the ambition to be a flagship for scientific and societal impact that the EOSC can deliver.

# Executive Summary

This document reports an initial version of the analysis of the use of IVOA standards for FAIR ESFRI and community data. The analysis is performed in the framework of Task 4.2 "Implementation of the FAIR principles for ESFRI data through the Virtual Observatory" of ESCAPE Work Package 4 "Connecting ESFRI projects to EOSC through VO framework" (CEVO).

The Virtual Observatory (VO) is the international data sharing framework which enables data to be findable, accessible, interoperable and re-usable for the astronomers' needs, together with the use of common data formats. Task 4.2 is gathering requirements from the ESFRIs and other Research Infrastructures of pan-European interest involved in CEVO for their use of the VO framework. The requirements are used to identify the relevant VO standards and we assess whether they need to be updated, or if new standards have to be created to fulfil the ESFRI and RI requirements. The standards can be used by the whole astronomical community world-wide, well beyond the infrastructures participating in CEVO.

FAIR practices in astronomy, their science drivers and how they are implemented through the VO, are briefly described. The method used to perform the assessment, starting from the establishment of WP4 Detailed Work Plan (D4.1) to the current status, is explained. The results detail which VO standards are of interest to enable each of the FAIR principles. A table providing the current priorities of each ESFRI is finally provided.

The next step should have been the May 2020 IVOA Interoperability meeting (ESCAPE Milestone MS22), which was scheduled in Sydney (Australia) in May 2020. The meeting has been cancelled due to the uncertainties with the epidemics situation. IVOA activities will proceed remotely and the Milestone MS22 is still expected to be achieved. The next IVOA Interoperability meeting (Milestone MS23) is planned to be held Granada (Spain) in November 2020. The Granada meeting will be extended to take into account the cancelled meeting in May 2020. The final version of the analysis will be included in D4.8 "Final analysis report on IVOA standards and stewardship best practices", due in May 2022.

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 the Grant Agreement n° 824064.

# 1. Introduction

Work Package 4 of ESCAPE, "Connecting ESFRI projects to EOSC through VO framework" (CEVO), plans to make the seamless connection of ESFRI and other astronomy and astroparticle research infrastructures to the EOSC through the Virtual Observatory (VO) framework. The high level objective of Task 4.2 "Implementation of the FAIR principles[1] for ESFRI data through the Virtual Observatory" is the definition and adoption of common open VO standards for interoperability based on ESFRI requirements, the connection with EOSC being ensured via Task 4.1. The standards can be used by the whole astronomical community world-wide, well beyond the infrastructures participating in CEVO.

The Virtual Observatory is a framework of open standards for making astronomy data FAIR – Findable, Accessible, Interoperable, Reusable. It is an established and operational interoperability framework that has proven to be a great success for many aspects of the interoperability and FAIRness of astronomy data. It is an essential component of the astronomy data landscape, as has been strongly stressed in the ASTRONET Infrastructure Roadmap since its first publication in 2008. International astronomy data providers, in particular ground- and space-based observatories that publish their data using the IVOA standards, and compliant scientific tools and services enable the discovery, access and (re-)use of the data by the whole astronomy research community. A short description of the Virtual Observatory architecture and astronomy FAIR practices will be provided in Section 2.

The inclusion of the data of the ESFRI facilities from astronomy and astroparticle physics in the VO was already well advanced thanks to the collaboration between ESFRI pathfinders and European VO teams in the ASTERICS Data Access, Discovery and Interoperability (DADI) Work Package since 2015. The CEVO objectives build on these developments to make the seamless connection of ESFRI and other astronomy research infrastructures to the EOSC through the VO. Task 4.2 refines and further pursues implementation of FAIR principles for astronomy data via the use and development of common standards for interoperability including the extension of the VO to new communities[2].

D4.1 "Detailed WP4 Project Plan" describes the activities that will be performed under the three WP4 tasks, including Task 4.2, during the project. This deliverable, D4.2, reports an intermediate analysis on the use of IVOA standards for FAIR ESFRI and community data. It is relevant to the two first subtask of Task 4.2 listed in the WP4 description of work:

- Gathering of requirements from ESFRIs for their use of the VO framework and its connection to EOSC.
- Update and definition of standards based on the requirements and priorities, and representation of ESFRI interests in the global VO framework.

The method used to assess the use of VO standards for the ESFRI needs is described in Section 3. The results are described in Section 4, and conclusions and future steps are presented in Section 5.

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016) doi:10.1038/sdata.2016.18 (https://www.nature.com/articles/sdata201618

[2] Neutrino astrophysics (for KM3NeT) and solar physics (for EST) are the main new communities concerned by WP4, but the methodology for defining interoperability standards is expected to be of interest to all partners in ESCAPE.

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 the Grant Agreement n° 824064.

# 2. FAIR practices in astronomy – The astronomical Virtual Observatory

Astronomy has been a pioneer of Open Data Sharing, and remains at the forefront. International agreement on a data format, FITS, and on standards allowing users to find, access and interoperate data, the so-called Virtual Observatory standards, provides a common data sharing framework open to all and enables the development of interoperable tools to access and use data.

FITS was first published in 1981 (Wells, Greisen & Harten[3]). The International Virtual Observatory Alliance[4] (IVOA) was created in 2002 to define and maintain the astronomical interoperability standards. Its processes[5] are adapted from the W3C ones. Data producers have been providing their data in the Virtual Observatory, and the community has been able to find, access, interoperate and reuse astronomical data years before FAIR was defined in the foundational Nature paper in 2016 (Wilkinson et al[6]). More than 100 "*authorities*" from all around the world, large agencies as well as smaller teams, have declared at least one service in the VO registry of resources. VO-enabled data services provided by the ground and space-based observatory archives as well as by value-added data repositories, as well as the VO-enabled interoperable tools providing access to data, are used by the community in their daily research work.

## 2.1. The science drivers of astronomical data sharing

Astronomy research infrastructures are most well known as the ground and space-based telescopes, but data as a whole is also one of the discipline's research infrastructures, which is used by the community 24/7. Data are at the core of the disciplinary scientific approach, which requires to be able to find, access, interoperate and reuse them. Combining observations from different instruments is essential to understanding the physical phenomena at work in astronomical objects. *Multi-wavelength astronomy* where observations across multiple electromagnetic wavebands are used is now the norm and is providing a significant fraction of the publications[7]. *Multi-wavelength astronomy* is now extended to *multi-messenger astronomy*, as spectacularly demonstrated by the first identification of the astronomical object at the origin of a gravitational wave event in 2018 and the identification of blazar TXS 0506+956 as the first source of non-stellar cosmic neutrinos. The fact that many astronomical objects are *variable* on a wide range of time scales, their brightness or other characteristics changing with time, or they have measurable motions on the celestial sphere, is also an important scientific driver to keep and share data on the long term. Of course researchers also want to compare models with observations, like in many other disciplines. Moreover, optimizing the science return of observations by costly instruments is also, at another level, a driver for enabling data reuse for science objectives different from the initial ones. Open access to existing data is also a key factor when planning for new observations or new facilities.

Another of the important characteristics of astronomy is that the data from observatories and missions are almost all made publicly available, usually after a short proprietary period (in general 1 year) for the data obtained in competitive calls for proposals for observation time. The metadata are

---

[3] 1981A&AS...44..363W
[4] http://ivoa.net
[5])
[6] https://doi.org/10.1038/sdata.2016.18
[7] A few years ago, about half of the ESO publications used data from one or more of the other major observatories.

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 the Grant Agreement n° 824064.

generally publicly available during the proprietary period, but data is released when it ends. Practices have been different in the High Energy Physics/astroparticle community, which tends to reserve the data to the people involved in the experiment consortia, but the astroparticle ESFRIs will be Observatories, with data dissemination practices aligned on the astronomy ones.

## 2.2. How FAIR is implemented and put into action in astronomy

The list of the FAIR guiding principles from Williamson et al. is given in Table 1 for reference. Figure 1 is a schema of the Virtual Observatory architecture, showing its main functionalities and how they relate to FAIR.

| |
|---|
| To be Findable:<br>F1. (meta)data are assigned a globally unique and persistent identifier<br>F2. data are described with rich metadata (defined by R1 below)<br>F3. metadata clearly and explicitly include the identifier of the data it describes<br>F4. (meta)data are registered or indexed in a searchable resource<br><br>To be Accessible:<br>A1. (meta)data are retrievable by their identifier using a standardized communications protocol<br>A1.1. the protocol is free, open and universally implementable<br>A1.2. the protocol allows for an authentication and authorization procedure, where necessary<br>A2. metadata are accessible, even when the data are no longer available<br><br>To be Interoperable:<br>I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation<br>I2. (meta)data uses vocabularies that follow FAIR principles<br>I3. (meta)data include qualified references to other (meta)data<br><br>To be reusable:<br>R1. (meta)data are richly described with a plurality of accurate and relevant attributes<br>R1.1. (meta)data are released with a clear and accessible data usage license<br>R1.2. (meta)data are associated with data provenance<br>R1.3. (meta)data meet domain relevant community standards |

**Table 1**: The FAIR Guiding Principles, as defined by Wilkinson et al. 2016

The IVOA standards[8] are defined by Working Groups which tackle different aspects of data sharing (currently Applications, Data Access Layer, Data Models, Grid & Web Services, Registry of Resources, Semantics). In fact, the IVOA standards enable users to find, access and interoperate data. Reusability is mostly in the hands of the data providers. For observational data, FITS integrates data and metadata, in particular about the observation (telescope, instrument, observation field of view, instrument parameters, etc.). Data providers declare their resources in the IVOA Registry of Resources and implement an "*interoperability layer*" on top of their data holdings to interface them with the Virtual Observatory using the VO protocols.

[8] http://www.ivoa.net/documents/

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 the Grant Agreement n° 824064.

As shown in Figure 1, the different facets of FAIR in the astronomy practices can be roughly speaking aligned with the standards defined by the different IVOA Working Groups:

- Findable: Registry of Resources (and Data Access Protocols)
- Accessible: Data Access Protocols
- Interoperable: Data Models, Semantics
- Reusability: Figure 1 refers to the implementation of reusability by the data providers in the resource layer. Going back to the FAIR Guiding Principles, all the IVOA standards ("*community standards*") are relevant, in particular those which define metadata. Provenance is specifically cited in R1.2.

The Applications Working Group activities are relevant to the four facets of FAIR.

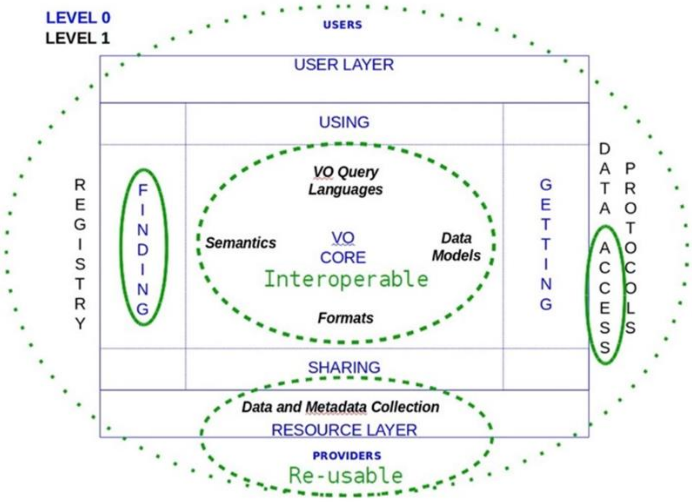


**Figure 1**: The VO architecture, and how it enables users to find, access and interoperate data.

The Virtual Observatory is an open, global data sharing framework, with no central point. Anyone can declare a service in a VO Registry of Resources, and the different registries harvest each other regularly to provide a complete and uniform view of VO-enabled data sources. Anyone can develop and share a tool to access these data, and the VO provides standards to make the tools interoperable. An example of VO tool is displayed in Figure 2: the Aladin interactive sky atlas developed by the Centre de Données astronomiques de Strasbourg[9] (Strasbourg astronomical Data Centre CDS – Bonnarel et al., 2000[10]; Boch & Fernique, 2014[11]). One can visualise all the positions in the sky for which observatory archives or data bases provide data in the image field. If the data is an image, Aladin can visualize the image and compare and combine it with others. If the data is for instance a spectrum, it can seamlessly link by a simple click to a tool which will display the spectrum.

The Virtual Observatory framework was expanded to take into account astroparticle physics in the ASTERICS[12] Cluster, and now in ESCAPE. ESCAPE is also working to reuse some of its elements in solar/heliospheric physics for the European Solar Telescope. Elements of the astronomical Virtual Observatory are reused by planetary sciences, in particular thanks to the Europlanet[13] projects, and

[9] http://cds.unistra.fr/
[10] https://doi.org/10.1051/aas:2000331
[11] https://ui.adsabs.harvard.edu/#abs/2014ASPC..485..277B/abstract
[12] https://www.asterics2020.eu/
[13] http://www.europlanet-2020-ri.eu/

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 the Grant Agreement n° 824064.

also by the Virtual Atomic and Molecular Data Centre VAMDC[14]. Europlanet participants proposed a Solar System Interest Group in the IVOA, which was created in 2017.

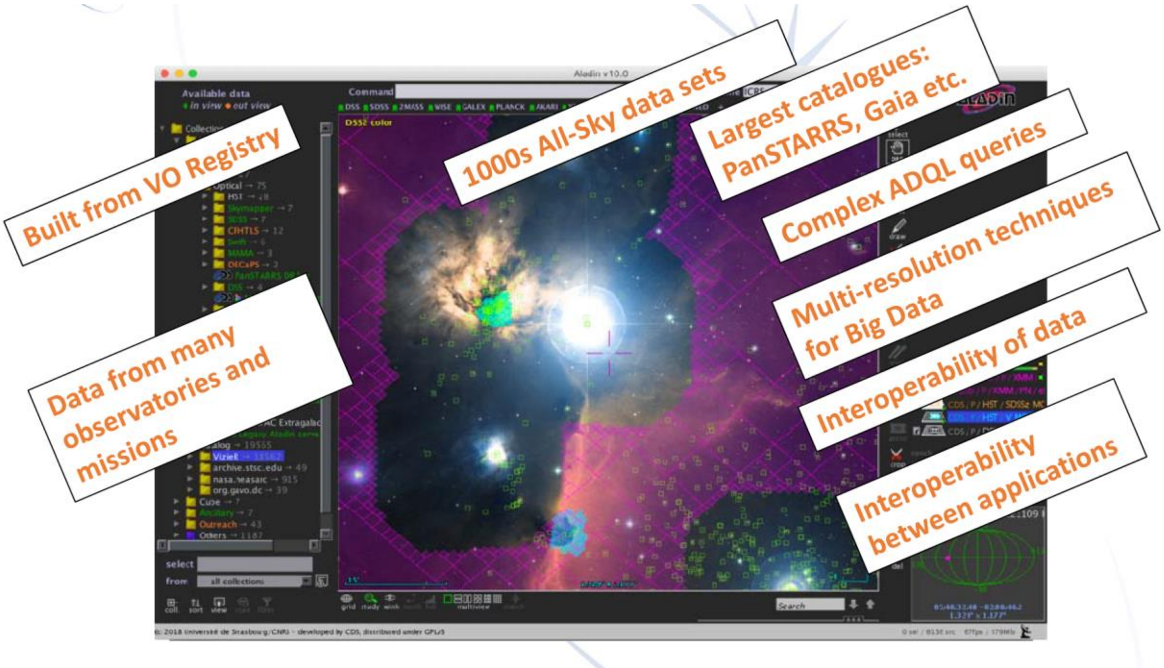


**Figure 2:** Image of the sky with the Aladin interactive sky atlas. The resources (observatory archives, databases) which have data available in the displayed region of the sky are listed in the column on the left. A click on a resource plots the locations for which data is available in the resource, a click on a point displays images on Aladin or other data types with other tools.

## 3. Assessment method

CEVO brings together ESCAPE partners who have expertise in the VO framework, with partners who are connected to the ESFRI projects and other research infrastructures. The VO expertise is provided by the partners: CNRS-ObAS, INAF, INTA, UEDIN, UHEI who have built up the Euro-VO alliance of national European projects since 2002. ObsParis has also become a VO expert and since the ASTERICS project it brings a special link between VO and the Cherenkov Telescope Array. The ESFRI and other research infrastructures who are directly involved in CEVO are: the ESFRI projects: the European Solar Telescope (EST), and the cubic-kilometre-sized Neutrino Telescope (KM3NeT); the ESFRI landmark projects: the Cherenkov Telescope Array (CTA), the Extremely Large Telescope (ELT) and the Square Kilometre Array (SKA). In addition to the ELT, the pan-European International Organization European Southern Observatory (ESO) brings other world-class established astronomical observatories (e.g. ALMA, the La Silla Paranal Observatory). Additionally the research infrastructures (RI) European Gravitational-Wave Observatory (EGO-Virgo) and the Joint Institute for VLBI ERIC (JIVE) are also participating directly in the work package. ASTRON also brings the SKA pathfinder LOFAR, a distributed facility that it coordinates. The ESFRIs and other RIs can be grouped into domain specific areas, Solar Physics, Radio data, Event-based data, UV/Optical/IR data.

[14] http://www.vamdc.org/

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 the Grant Agreement n° 824064.

The Detailed Work Plan D4.1 lists the activities which have been performed to produce this deliverable:

- The first sub-task of Task 4.2, "Gathering of requirements from ESFRIs for their use of the VO framework and its connection to EOSC" is the starting point. It begun with in-depth discussions with the ESFRIs and the other Research Infrastructures (RI) participating in CEVO, in dedicated meetings and teleconferences with each RI or RI group, and also taking advantage of general meetings that relevant people from the VO and ESFRI/RI communities were attending or were invited to attend. The aim was to identify the initial priorities for each ESFRI and RI, and to match these requirements to the priorities set at the international level by IVOA, to establish a plan for implementing VO interoperability for each ESFRI/RI. The current version of the analysis is described in this deliverable. The in-depth discussions and meetings also enabled us to introduce new participants, in particular EST, for interoperability in the VO. The plan will have to be continuously updated to take into account evolving requirements while the ESFRI projects build up their VO layers and feedback from IVOA activities.
- The second sub-task of Task 4.2, "Update and definition of standards based on the requirements and priorities, and representation of ESFRI interests in the global VO framework", is also relevant. It contributes to the update and definition of IVOA standards based on the ESFRI and RI priorities (including the EOSC related aspects when applicable), and include the work performed in the relevant IVOA Working Groups and Interest Groups and participation in the IVOA Interoperability meetings for the planning and definition of the relevant IVOA standards. These meetings are bi-annual interoperability meetings of the global IVOA community. Each of them is an ESCAPE Milestone. A detailed report identifying ESCAPE participation and progress as well as general European contribution to IVOA as been produced for each the two meetings held to date (M20[15], M21[16]).

The list of the meetings and visits organised to gather ESFRI and RI requirements is provided in Table 2. They were completed by teleconferences and by participation in the IVOA Interoperability meetings. The first meetings took advantage of the fact that ASTERICS was still active at the beginning of ESCAPE. Specific meetings were co-located with the last ASTERICS Technology Forum, which was held in Strasbourg 26-28 February 2019[17]. The First ESCAPE WP4 Technology Forum[18], held in Strasbourg 4-6 February 2020, gathered WP4 partners and representatives from WP3 and WP5. It provided a lively view of the relevant WP4 activities. François Bonnarel (CDS) summarized the assessment of the use of IVOA standards by the ESFRIs and RIs in his talk "Status of VO standards versus ESFRI needs in CEVO Task 2"[19], several partners reported on their relevant activities, and the so-called "hack-a-thon" sessions enabled in-depth discussion of on domain/ESFRI-RI needs, VO usage and specific standards.

[15] https://projectescape.eu/Milestone%2020
[16] https://projectescape.eu/deliverables-and-reports/milestone-21-report-progress-and-priorities-ivoa-2
[17] https://www.asterics2020.eu/dokuwiki/doku.php?id=open:wp4:wp4techforum5
[18] https://indico.in2p3.fr/event/20005/
[19] https://indico.in2p3.fr/event/20005/contributions/78637/attachments/57292/76483/CEVO_VO_Standard_Status.pdf

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 the Grant Agreement n° 824064.

| Title | Meeting type | Location | Date | Additional information (specific topic, etc.) |
|---|---|---|---|---|
| EST and the VO | Meeting | Strasbourg | 26 February 2019 | Co-located with the 5th ASTERICS Technology Forum |
| Radioastronomy and the VO | Meeting | Strasbourg | 28 February 2019 | Co-located with the 5th ASTERICS Technology Forum |
| ROB/CDS meeting (EST) | Visit | Brussels | 11-12 March 2019 | |
| KM3Net and the VO | Meeting | Strasbourg | 5 July 2019 | |
| ECAP/UHEI meeting (KM3Net) | Visit | Erlangen | 25-26 July 2019 | Publishing neutrino data in the VO |
| CTA and KM3NeT Provenance meeting[20] | Meeting | Strasbourg | 6 November 2019 | Provenance, CTA/KM3NeT synergies |
| KIS/UHEI meeting (EST) | Visit | Freiburg | 12 November 2019 | Publishing solar data in the VO |
| ASTRON/UHEI meeting (JIVE/radioastronomy) | Visit | Dwingeloo | 6-10 January 2020 | Registration of VO services |
| EGO-Virgo partner visit to CDS | Visit | Strasbourg | 28 January – 14 February 2020 | VO standards and tools for Gravitational Wave localisation and observational follow-up |

**Table 2**: Meetings organised by WP4 to gather ESFRI and RI requirements

The two IVOA Interoperability meetings organised in 2019 were held respectively in Paris, France, (12-17 May)[21] and in Groningen, The Netherlands (11-13 October)[22]. The latter was, co-located with the Annual ADASS conference[23], as usual for the "Northern Fall" IVOA meetings. These meetings were Milestones for CEVO, which have been reported in the standard way with short texts submitted to the EC Portal. In addition, detailed reports (IVOA Paris[24], IVOA Groningen[25]) have been compiled for each of these milestones, including a collection of all the relevant presentations (IVOA Paris[26], IVOA Groningen[27])

[20] https://indico.in2p3.fr/event/20000/
[21] https://wiki.ivoa.net/twiki/bin/view/IVOA/InterOpMay2019
[22] https://wiki.ivoa.net/twiki/bin/view/IVOA/InterOpOct2019
[23] https://www.adass2019.nl/
[24] IVOA Paris report : https://cloud.escape2020.de/index.php/s/vN32b1RO2t4u70v
[25] IVOA Groningen report : https://cloud.escape2020.de/index.php/s/8xQsiLE4oe3h7qH
[26] IVOA Paris presentations (345MB) : https://cloud.escape2020.de/index.php/s/DuKhjzrb9vg2YiQ
[27] IVOA Groningen presentation (203 MB) : https://cloud.escape2020.de/index.php/s/SzX1qJCKelEqKOJ

# 4. Results

This section summarizes the current assessment of the use of VO standards for FAIR ESFRI and community data. The initial in-depth analysis of the ESFRI and RI requirements published in D4.1 provides a strong starting point. The reader is referred to it since the requirement details are not repeated here. D4.1 provided the "ESFRI and RI point of view", this D4.2 is written with the "VO point of view".

The list of VO standards is available and maintained on the IVOA web site at http://www.ivoa.net/documents/. One should note here that the IVOA standards are Recommendations endorsed by the IVOA Executive Board after extensive discussion by the community and following a defined process. The two words "standard" and "recommendation" will be used indifferently in the following to refer to them.

## 4.1. Detailed analysis of the IVOA standards use to enable FAIR data

As explained in Section 2 and shown in Figure 1, the different facets of FAIR in the astronomy practices can be roughly aligned with the standards defined by the different IVOA Working Groups. The Working Groups define standards which are generally applicable, and others aimed at specific kinds of data. A summary of the standards of interest for the ESFRIs and RIs is given below, classified following the FAIR concepts.

**Findability**

The *Registry of Resources* defined by the Working Group with the same name is the core of the VO framework: it lists the available resources and provides a machine readable link to each of them and essential metadata[28]. It is also included in EUDAT B2FIND, providing the Virtual Observatory a first level of visibility in the EOSC. The Registry implements *VOResource*[29], an XML encoding Schema for resource metadata. It uses *IVOA Identifiers*[30] as a globally unique name for a resource within the VO.

Extensions of the resource schema are defined for specific kinds of resources, such as the VO standards themselves (*StandardsRegExt*[31]), collections and services (*VODataService*[32]), or the Table Access Protocol described below (*TAPRegExt*[33]).

All the data resources provided by the ESFRIs and RIs should be listed in the VO Registry. VOResource and the metadata applicability to these new resources has to be checked and extensions defined if necessary.

**Accessibility**

The Data Access Layer Working Group defines remote access protocols for different kinds of data. Most of the protocols are built on the basis of Data Models defined by the Data Modelling WG to ensure interoperability – that WG is described in the Interoperability section below. The *Table Access Protocol TAP*[34] defines a service protocol for accessing general table data, including astronomical

[28] http://www.ivoa.net/documents/RM/index.html
[29] http://www.ivoa.net/Documents/VOResource/index.html
[30] http://www.ivoa.net/Documents/IVOAIdentifiers/index.html
[31] http://www.ivoa.net/documents/StandardsRegExt/
[32] http://www.ivoa.net/documents/VODataService/index.html
[33] http://www.ivoa.net/documents/TAPRegExt/index.html
[34] http://www.ivoa.net/Documents/TAP/

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 the Grant Agreement n° 824064.

catalogues as well as general database tables. It can then be used in multiple different contexts. It can support multiple query languages, including queries specified using the *Astronomical Data Query Language ADQL*[35]. Relational database schema are made interoperable using the "*TAP schema*" formalism. Colleagues from the Solar System community in the Europlanet (EPN) projects have standardized their TAP schema into the so-called *EPN-core*, which is operated as *EPN-TAP*[36]. An early standard, *Simple Cone Search*[37], defines a simple query protocol for retrieving records from a catalog of astronomical sources. The query describes sky position and an angular distance, defining a cone on the sky. This protocol could be updated to provide a very simple access to time domain data by adding constraints on time.

Other recommendations address in particular *Simple Image Access SIA*[38], which enables discovery and access to, and retrieval of datasets including retrieval of 2-D images as well as cubes of three or more dimensions and sparse data such as "event lists". *DataLink*[39] describes the linking of data discovery metadata to access to the data itself, further detailed metadata, related resources, and to services that perform operations on the data. At a lower level, the *Server-side Operations for Data Access*[40] (*SODA*) web service capabilities can act upon data files, performing various kinds of operations and applying functions to the data, including cutout and data extraction.

The on-going developments of the *Observation Locator Table Access Protocol*[41] and of the *Object Visibility Simple Access Protocol*[42] are also of interest for the ESFRIs and RIs. Observation localisation is required for post-alert follow-up observations. The VO of course allows one to define the localisation of an observation but the point is to have an access protocol to a standard description.

The Grid & Web Services Working Group defines the IVOA *Single-Sign-On Profile*[43] and *Credential Delegation Protocol*[44], which have to be assessed with respect to the EOSC work on AAI. The *Universal Worker Service Pattern*[45] UWS defines how to manage asynchronous execution of jobs on a service, also to be examined in the EOSC context. Interfacing these aspects of the VO with the EOSC is under the remit of Task 4.1, although it is also relevant to Task 4.2 aim to enable FAIR data.

**Interoperability**

The Application Working Group defines and maintains the *Simple Application Messaging Protocol*[46] SAMP, which is used to interoperate VO-enabled applications. It also maintains a format, the *Hierarchical Progressive Survey*[47] HiPS, which enables progressive zooming into image data to retrieve the required level of details, and the *Multi-Order Coverage map method*[48] (MOC) to specify

---

[35] http://www.ivoa.net/Documents/ADQL/index.html
[36] Erard, S., Cecconi, B., Le Sidaner, P., et al., The EPN-TAP protocol for the Planetary Science Virtual Observatory, Astronomy and Computing, 7-8, 52, https://doi.org/10.1016/j.ascom.2014.07.008
[37] http://www.ivoa.net/Documents/latest/ConeSearch.html
[38] http://www.ivoa.net/Documents/SIA/
[39] http://www.ivoa.net/documents/DataLink/index.html
[40] http://www.ivoa.net/documents/SODA/index.html
[41] http://www.ivoa.net/documents/ObsLocTAP/index.html
[42] http://www.ivoa.net/documents/ObjVisSAP/index.html
[43] http://www.ivoa.net/Documents/SSO/index.html
[44] http://www.ivoa.net/Documents/CredentialDelegation/
[45] http://www.ivoa.net/documents/UWS/index.html
[46] http://www.ivoa.net/documents/SAMP/index.html
[47] http://www.ivoa.net/documents/HiPS/index.html
[48] http://www.ivoa.net/documents/MOC/index.html

arbitrary sky regions. *VOTable*[49], the first standard defined for the Virtual Observatory, is an XML standard for the interchange of data defined as a set of tables. VOTable is widely used in the VO. Its elements are updated when required to fulfil additional requirements.

The MOC standard is being extended to deal with time coverage (*STMOC*), which will be of high interest for all the ESFRIs and RIs, which have interest in the Time Domain and Alerts. This is a particularly innovative development that provides a practical indexation of astronomy data in space and time, which may also be developed for energy/frequency/velocity axes. The implementation of time into the *mocpy*, the Python library to create and manage MOCs, and its support for the multi-order format provided by LIGO and Virgo is one of the early impacts of this effort, along with the improved MOC visualisation of gravitational wave events in the Virgo-GW skymap[50].

The inclusion of an element to define metadata for time coordinates (TIMESYS) in the last version of VOTable is also relevant for the ESFRI and RI needs. A possible further evolution to fulfil ESFRI and RI needs would be to define wavelength/frequency information (FREQSYS).

The Semantics Working Group is in charge of IVOA vocabularies, with a document proposed on practices related to the use of RDF-based vocabularies in the VO, *Vocabularies in the VO*[51], a recommendation on *Units in the Virtual Observatory*[52], and the IVOA standard to define *Unified Content Descriptors*[53] (UCD), a controlled vocabulary for describing astronomical data quantities. The UCD standard, one of the first defined by the IVOA, has been one of the essential assets to ensure interoperability in the VO. The list of terms[54] evolves following the process defined in the recommendation on *Maintenance of the list of UCD words*[55].

The UCD list was for instance updated to include terms relevant to Solar System studies. It has to be checked to deal properly with the ESFRI and RI data, in particular for EST which brings the new topic of Solar Physics in the Virtual Observatory. Some of the terms defined to describe quantities for the Solar System domain will also be useful for Solar Physics.

The Data Modeling Working Group provides models of the different kinds of entities for which a common description is useful to enable interoperability. Two standards are of particular interest in CEVO context, one being used at the data centre level, the other one at the dataset level. The *ObsCore*[56] recommendation defines the core components of an observation data model that are necessary to perform data discovery when querying data centers for astronomical observations of interest, and how they are implemented in the Table Access Protocol (*ObsTAP*). *EPN-core* (see above) is an extension of the ObsCore table for planetary data. Some ESFRIs and RIs (SKA, JIVE, ALMA) can generate interferometric, or “uv data”, which currently can not be completely described in the ObsCore data model. In the IVOA “radio astronomy into the VO” Interest Group, ESCAPE partners are working together with international observatories and interferometric networks to assess wether an extension of the ObsCore data model to support this type of data is necessary and what parameters

[49] http://www.ivoa.net/Documents/VOTable/
[50] http://www.virgo-gw.eu/skymap.html
[51] http://www.ivoa.net/documents/Vocabularies/20190905/
[52] http://www.ivoa.net/documents/VOUnits/index.html
[53] http://www.ivoa.net/Documents/latest/UCD.html
[54] http://www.ivoa.net/Documents/UCD1+/index.html
[55] http://www.ivoa.net/Documents/UCDlistMaintenance/index.html
[56] http://www.ivoa.net/documents/ObsCore/index.html

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 the Grant Agreement n° 824064.

would be needed to accomodate this. The *Characterisation*[57] data model defines the high level metadata necessary to describe the physical parameter space of observed or simulated astronomical data sets which can be used to derive a structured description of any relevant data and thus to facilitate its discovery and scientific interpretation. These two models have to be assessed to ensure that they will be able to fulfil the requirements of the ESFRIs and Ris, and extended if necessary.

The Provenance data model, which is also under the remit of the Data Modeling WG, will be discussed in the Reusability section.

**Reusability**

For reusability, all the IVOA standards ("community standards", FAIR principle R1.3) are relevant, in particular those which define metadata.

Data formats are also essential components for reusability and interoperability. The widespread usage of the FITS format is a key asset in the domain for astronomy. Other formats, such as HDF5[58], are used by some of the ESFRIs. It is important to assess how to ensure reusability (and interoperability) when these formats are used. The radio-astronomy community has been using FITS files in various flavours for storing and distributing visibility data but is moving away from this. In recent years, the MeasurementSet[59] has established itself as a new standard that is used by many new instruments and supported by a wide variety of data processing software. CTA discusses the usage of HDF5 for low level data and FITS for higher level.

Provenance is specifically cited among the principles (FAIR principle R1.2). Many IVOA standards, as well as the FITS data format, include provenance information. The IVOA *Provenance Data Model*[60] describes how provenance information can be modeled, stored and exchanged within the astronomical community in a standardized way, following the definition of provenance as proposed by the W3C. The model was developed under the leadership of the Paris Observatory CTA team in ASTERICS and ESCAPE. A TAP implementation of the Provenance standard, *ProvTAP*, is being proposed. Provenance is a domain in which CEVO is enabling synergies between CTA and KM3NeT.

FAIR principle R1.1, which requires that the (meta)data is released with a clear and accessible data usage licence, can be an issue for many astronomical data resources: astronomical data is in general openly available, often after a proprietary period (in general 1 year) during which the metadata are in general public. Data and metadata often do not have an explicit usage licence, and usage rights rely mostly on disciplinary ethics: cite the origin of data when data is cited. The IVOA Data Curation and Preservation Interest Group would be an appropriate venue to discuss usage licences for astronomical data.

One can also cite here the *VOEvent*[61] standard, which defines the content and meaning of a standard information packet for representing, transmitting, publishing and archiving information about a transient celestial event, with the implication that timely follow-up is of interest, i.e. of interest for the RIs which deal with alerts, which can be the case for all those involved in CEVO. VOEvent is already used e.g. for gravitational wave alerts and by the neutrino experiments which are CTA pathfinders.

[57] http://www.ivoa.net/Documents/latest/CharacterisationDM.html
[58] https://support.hdfgroup.org/HDF5/whatishdf5.html
[59] https://casa.nrao.edu/casadocs/casa-5.1.0/reference-material/measurement-set
[60] http://www.ivoa.net/documents/ProvenanceDM/index.html
[61] http://www.ivoa.net/documents/VOEvent/index.html

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 the Grant Agreement n° 824064.

## 4.2. ESFRI and RI priorities

Table 3 summarizes the current ESFRI and RI priorities with respect to the use of IVOA standards.

| Infrastructure | Priorities |
| --- | --- |
| EST | • UCD for Solar data<br>• Mapping to ObsCore<br>• Solar data in TAP/EPN-TAP<br>• Visualisation tool: Access via SAMP and ADQL interface<br>• ObsTAP service for solar images |
| SKA | • Explore ObsCore to describe data<br>• Provenance<br>• HiPS for simulated images<br>• Authentication & Authorisation |
| JIVE | • Standards for uv data<br>  ○ Characterisation and metadata for visibilities<br>  ○ Provenance visibility use case<br>  ○ DataLink usage<br>• Service registration<br>• The implementation of VO services with JIVE data will provide feedback on the VO standards |
| LOFAR | • ObsCore/ObsTAP implementation (science-ready and later visibility data)<br>• Advanced discovery and access, incl.<br>  ○ DataLink with additional metadata (uv coverage, frequency amplitude plots, etc), SODA<br>  ○ Datalink to remotely hosted datasets |
| ALMA | • Usage of VO standards<br>  ○ ObsCore<br>  ○ DataLink<br>  ○ SODA and SIA<br>  ○ HiPS via AladinLite<br>• The implementation of VO services with ALMA data will provide feedback on the VO standards |
| CTA | • ObsCore mapping<br>• Discussion of the usage of ADQL for searching bulk and science archive<br>• Extend Provenance implementation<br>• VOEvent implementation<br>• Contribution to data models for multi-observatory and high level data<br>  ○ Characterisation and Provenance extension if needed<br>• IVAO Registry interfaced with EOSC (Task 4.1) |
| ESO | • Time domain alerts<br>  ○ VOEvent, Cone Search with time constraints, STMOC<br>• Authentication<br>• The implementation of VO standards will provide feedback on required evolutions of VO standards and tools |

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 the Grant Agreement n° 824064.

| | |
|---|---|
| | • Machine Learning value-added data (Task 4.3)<br>  ○ Classification, semantics |
| KM3NeT | • VOEvent implementation<br>• Data modelling: TAP + DataLink; VOEvent + VOTable<br>• Provenance implementation<br>• Mapping in HDF5 formats<br>• Feedback on standards to IVOA<br>• Usage of VO services: search for counterparts via VizieR and SIMBAD |
| EGO-VIRGO | • Advanced all-sky Gravitational wave visualisation via HiPS, MOC, SAMP, Aladin/Aladin Lite<br>• STMOC implementation (in mocpy and Aladin tools)<br>• Interoperability of the LIGO multi-order formats with VO tools and standards |

**Table 3**: ESFRI and RI priorities

# 5. Conclusion

The initial assessment of the ESFRI and RI requirements shows that the IVOA standards are already providing an excellent starting point to make their data findable, accessible, interoperable and reusable. CEVO D4.1 Detailed Work Programme identified the requirements and provided an initial list of standards of interest and activities to perform, which is being refined as the assessment progress with work in collaboration between the ESFRI/RI teams and the teams which bring the knowledge of the VO and its tools.

ESCAPE provides essential input to the IVOA. CEVO teams bring the ESFRI and RI requirements to be taken into account in the evolutions of the IVOA data sharing framework. They are instrumental in proposing and leading standard developments. They implement the standards in data services and VO-enabled tools, which provides feedback. The current IVOA priority on the time domain, with an Interest Group which was energized thanks to ASTERICS DADI and is currently led by staff from CEVO partners, fits a critical need of many of the ESFRIs. The creation of a Radio-astronomy Interest Group, which has recently been endorsed by the IVOA Executive Board, was strongly advocated based on CEVO work and input. It will be the vehicle of collaboration with the Working Groups, which develop the standards, to represent the community needs and share the knowledge of the domain science needs and data specificities.

The growing interest of the IVOA for platforms, the liaison established between the IVOA and Astropy[62], a community Python library for astronomy, will provide useful input to WP5 ESFRI Science Analysis Platform (ESAP). More generally, CEVO and ESAP are collaborating to connect ESAP Science Platform to existing VO-enabled astronomical data archives and services. Work is on-going on how to use VO applications in containers, and how to ensure SAMP communication to preserve interoperability enabled by the VO between containerized applications. Liaison with WP2 is also established on AAI. The astronomy and particle physics software repository developed by WP3 is a new resource to develop VO consistent tools and platforms in coordination with WP4 and WP5.

[62] https://www.astropy.org/

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 the Grant Agreement n° 824064.

The next milestone should have been the IVOA Interoperability meeting initially scheduled in Sydney[63] (Australia), 3-8 May 2020 (ESCAPE Milestone MS22). The meeting was cancelled because of the uncertainties linked to the coronavirus epidemics. The next one should be the IVOA Interoperability meeting scheduled in Granada[64] (Spain) 13-15 November 2020 (ESCAPE Milestone MS24), which will be extended to mitigate the May 2020 meeting cancellation. IVOA activities are continuing remotely. The final version of the analysis will be included in D4.8 "Final analysis report on IVOA standards and stewardship best practices", due in May 2022.

[63] https://may2020.asvo.org.au/
[64] https://ivoa2020.cab.inta-csic.es/main/index.php

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 the Grant Agreement n° 824064.