Research Data Alliance activities of interest for ESCAPE

Françoise Genova, François Bonnarel (CDS), Marco Molinaro (INAF), Mark Allen (CDS)

CEVO Technology Forum 1

4 February 2020, Strasbourg

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.

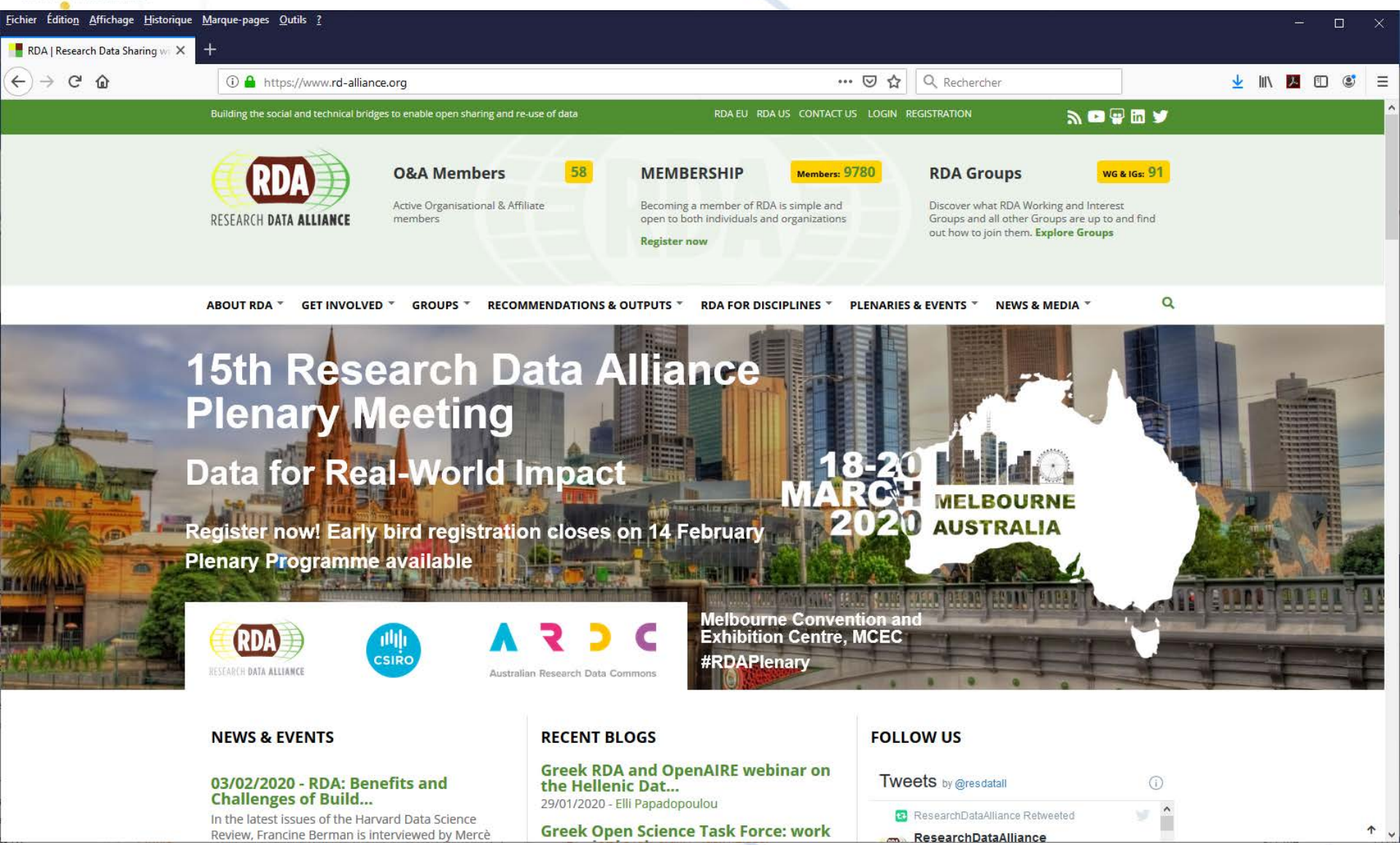
The Research Data Alliance

- Created in 2013 by the European Commission, NSF (USA) and Australian government
- A neutral, open, international forum to discuss all the aspects of scientific data sharing and produce recommendations and other outputs
- Today 9778 members from 137 countries
- 91 Working Groups and Interest Groups tackling many different subjects – technological and sociological
- RDA value for EOSC

https://www.rd-alliance.org/value-research-data-alliance-european-open-science-cloud-eosc

Funded by the European Union's Horizon 2020 - Grant N° 824064

https://www.rd-alliance.org/

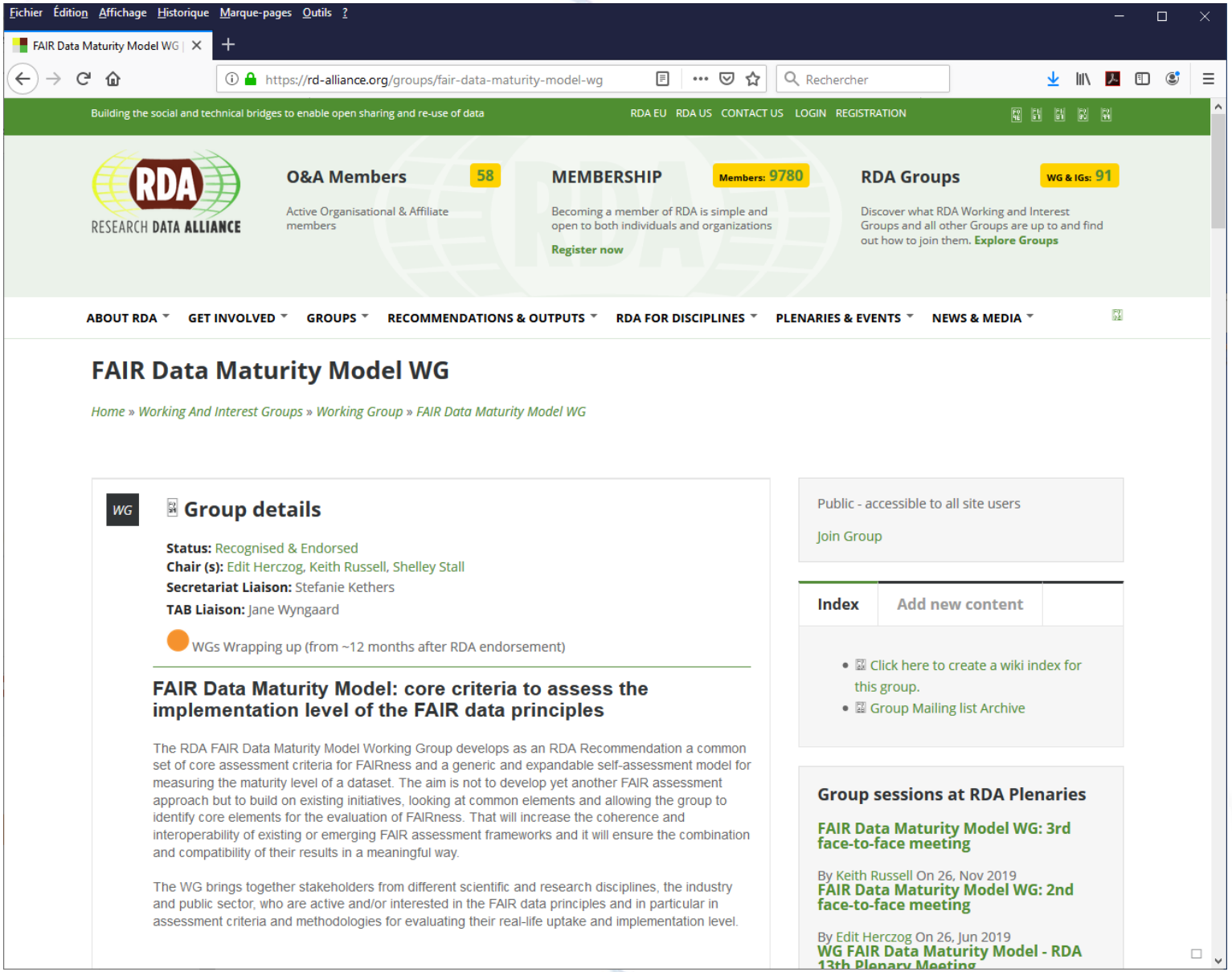


Funded by the European Union's Horizon 2020 - Grant N° 824064

Many relevant activities, among which…

- Global Open Research Commons Interest Group
  - EOSC and other similar initiatives
- Data Repository Audit and Certification Recommendation >>> Core Trust Seal
- FAIR Data Maturity Model Working Group
  - Core criteria for FAIR data

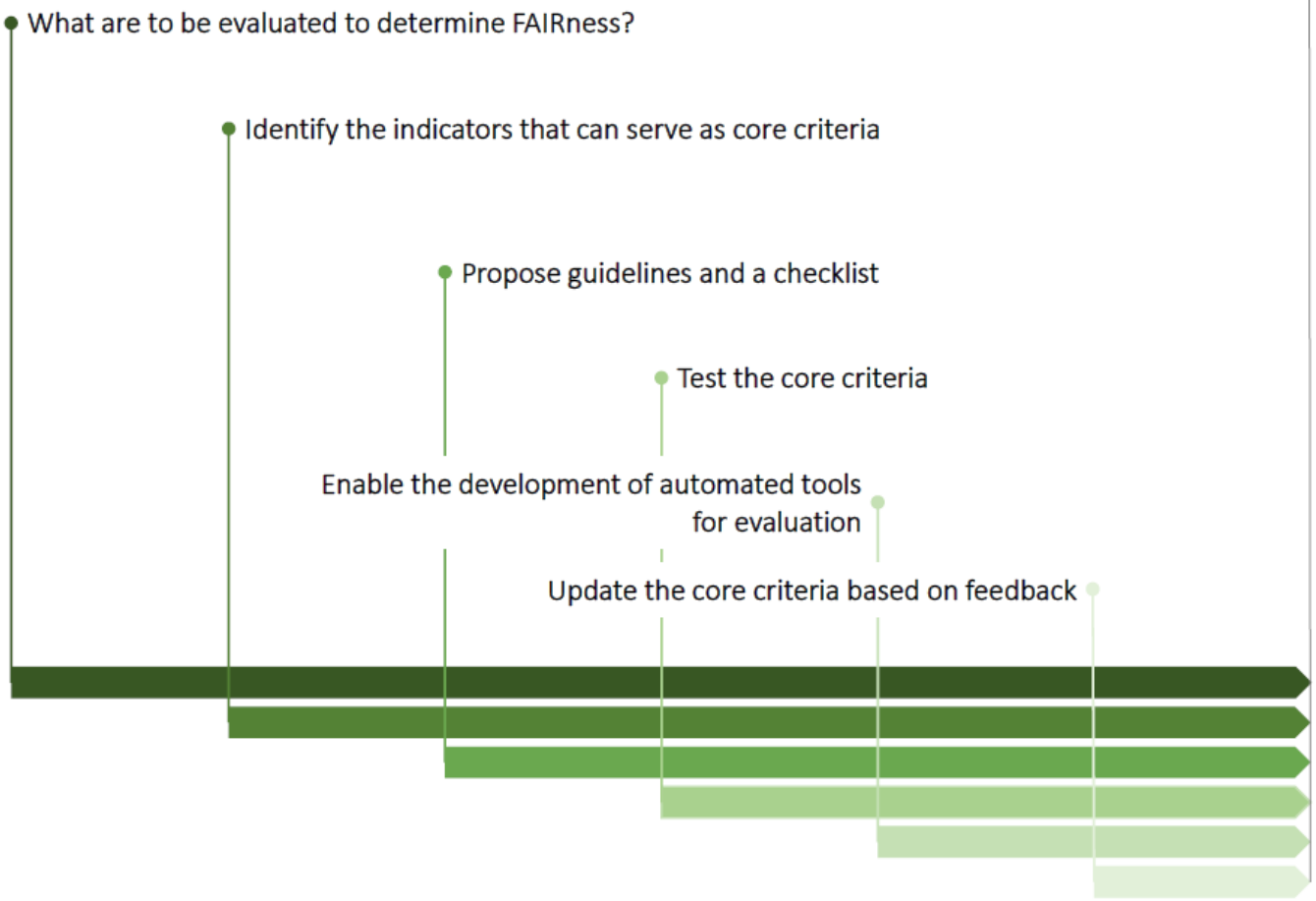
Funded by the European Union's Horizon 2020 - Grant N° 824064



Funded by the European Union's Horizon 2020 - Grant N° 824064

Objectives

FAIR data maturity model



Funded by the European Union's Horizon 2020 - Grant N° 824064

Criteria in November 2029

Successive versions of the criteria

https://docs.google.com/spreadsheets/d/1mkjElFrTBPBH0QViODexNur0xNGhJqau0zkL4w8RRAw/edit#gid=1210743571

	PRINCIPLE	INDICATOR_ID	INDICATORS	PRIORITY
F	F1	F1-01M	Metadata is identified by a persistent identifier	Recommended
	F1	F1-01D	Data is identified by a persistent identifier	Mandatory
	F1	F1-02M	Metadata is identified by a universally unique identifier	Recommended
	F1	F1-02D	Data is identified by a universally unique identifier	Mandatory
	F2	F2-01M	standard	Recommended
	F2	F2-02M	Metadata is provided for the discovery-related elements defined by the RDA Metadata IG, as much as possible and relevant, if no domain/discipline-specific metadata standard is available	Recommended
	F3	F3-01M	Metadata includes the identifier for the data	Mandatory
	F4	F4-01M	Metadata is offered/published/exposed in such a way that it can be harvested and indexed	Recommended
A	A1	A1-01M	Metadata includes information about access conditions	Optional
	A1	A1-01D	Data can be accessed manually (i.e. with human intervention)	Recommended
	A1	A1-02D	Data can be accessed automatically (i.e. by a computer program)	Recommended
	A1	A1-02M	Metadata identifier resolves to a metadata record	Optional
	A1	A1-03D	Data identifier resolves to a digital object	Mandatory
	A1	A1-03M	Metadata is accessed through standardised protocol	Recommended
	A1	A1-04D	Data is accessible through standardised protocol	Recommended
	A1.1	A1.1-01M	Metadata is accessible through a free access protocol	Mandatory
	A1.1	A1.1-01D	Data is accessible through a free access protocol	Mandatory
	A1.1	A1.1-02M	Metadata is accessible through an open-source access protocol	Recommended
	A1.1	A1.1-02D	Data is accessible through an open-source access protocol	Recommended
	A1.1	A1.1-03D	Actions to be taken by a reuser to get access to the data are well documented	Recommended
	A1.2	A1.2-01M	Metadata includes information relevant for access control	Mandatory
	A1.2	A1.2-01D	Data is accessible through an access protocol that supports authentication	Recommended
	A1.2	A1.2-02D	Data is accessible through an access protocol that supports authorisation	Recommended
	A2	A2-01M	Metadata is guaranteed to remain available after data is no longer available	Mandatory
I	I1	I1-01M	Metadata uses knowledge representation expressed in standardised format	Recommended
	I1	I1-01D	Data uses knowledge representation expressed in standardised format	Recommended
	I1	I1-02M	Metadata uses machine-understandable knowledge representation	Optional
	I1	I1-02D	Data uses machine-understandable knowledge representation	Optional
	I1	I1-03M	Metadata uses self-describing knowledge representation	Optional
	I1	I1-03D	Data uses self-describing knowledge representation	Optional
	I2	I2-01M	Metadata uses standard vocabularies	Recommended
	I2	I2-01D	Data uses standard vocabularies	Recommended
	I2	I2-02M	Metadata uses FAIR-compliant vocabularies	Optional
	I2	I2-02D	Data uses FAIR-compliant vocabularies	Optional
	I3	I3-01M	Metadata includes references to other metadata	Recommended
	I3	I3-01D	Data includes references to other data	Recommended
	I3	I3-02M	Metadata includes references to other data	Recommended
	I3	I3-02D	Data includes sufficiently qualified references to other data	Optional
	I3	I3-03M	Metadata includes sufficiently qualified references to other metadata	Recommended
	I3	I3-04M	Metadata include sufficiently qualified references to other data	Optional
R	R1	R1-01M	standard	Recommended
	R1	R1-02M	Metadata is provided for the reuse-related elements defined by the RDA Metadata IG, as much as possible and relevant, if no domain/discipline-specific metadata standard is available	Recommended
	R1.1	R1.1-01M	Metadata includes information about the licence under which the data can be reused	Mandatory
	R1.1	R1.1-02M	Metadata refers to a standard reuse licence	Recommended
	R1.1	R1.1-03M	Metadata includes licence information in the appropriate element of the metadata standard used	Mandatory
	R1.1	R1.1-04M	Metadata refers to a machine-understandable reuse licence	Optional
	R1.1	R1.1-05M	Metadata includes information about consent for reuse (e.g. for personal data)	Recommended
	R1.2	R1.2-01M	Metadata includes provenance information according to community-specific guidelines	Recommended
	R1.2	R1.2-02M	Metadata includes provenance information according to a cross-domain language	Optional
	R1.3	R1.3-01M	Metadata complies with a community standard	Mandatory
	R1.3	R1.3-01D	Data complies with a community standard	Mandatory
	R1.3	R1.3-02M	Metadata is expressed in compliance with a machine-understandable community standard	Optional
	R1.3	R1.3-02D	Data is expressed in compliance with a machine-understandable community standard	Optional

Funded by the European Union's Horizon 2020 - Grant N° 824064

Test with FAIR practices in astronomy

- Essential to test how disciplinary practices fit with the proposed criteria
- Reusing and interoperate data at the core of astronomical research
- We have an operational, international data sharing framework enabling astronomers to find, access, interoperate and reuse data
- Overall test in November: how the criteria fit with our requirements
- Detailed test on-going: CDS/VizieR & Trieste VIALACTEA/IA2

Funded by the European Union's Horizon 2020 - Grant N° 824064

Main results of the overall test

- Our goals are reuse and interoperability, not reproducibility
- We can live with the criteria
- The Virtual Observatory enables some of them ‘for free’ once it is implemented on top of the data holdings
- But issue with the definition of priority criteria
  - Find is a dynamic process for us – the PID is only one among many parameters
  - Access/reuse: at the core of our system but often not compliant – we provide in general open data, in general no licence, usage based on disciplinary ethics (cite what you use). The data is however widey reused
  - At that time no mandatory criteria for interoperability, some added later
- Some metadata are at the collection level, others at the dataset one

Funded by the European Union’s Horizon 2020 - Grant N° 824064

Conclusions of the overall test

- Real critical problem with mandatory (or essential) criteria – we would chose different ones!
- FAIRness is a process: the evaluation method should be inclusive and allow and measure progress (compliance scales)
- Open by default should be considered as acceptable
- There is a cost to adapt a repository to fulfill criteria different from the disciplinary ones, and archives are not supported for that
- Detailed test ongoing

Funded by the European Union's Horizon 2020 - Grant N° 824064

IVOA Note being written

FAIR practices in Astronomy, and how they fit with the FAIR criteria

Status report

Françoise Genova, François Bonnarel (CDS/Observatoire Astronomique de Strasbourg), Marco Molinaro (INAF), Mark Allen (CDS/Observatoire Astronomique de Strasbourg)

ESCAPE WP4 - Connecting ESFRI projects to EOSC through VO framework (CEVO)

V0.4, 30 January 2020

1. Introduction

Disciplinary data sharing frameworks are the pillars of open science. A key element for them to succeed is to be driven by the community science needs, and to take into account community practices (e.g. Genova et al., 2017[1]). 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 (VO) standards, provides a shared 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[2]). The International Virtual Observatory Alliance[3] (IVOA) was created in 2002 to define and maintain the astronomical interoperability standards. Its processes[4] 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[5]). More than 100 "authorities" from all around the world, large agencies as well as smaller teams, have declared at least a service in the VO registry of resources. VO-enabled data services provided by the ground and space-based observatory archives and 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.

Funded by the European Union's Horizon 2020 - Grant N° 824064