Addressing FAIR data challenges in astronomy -
CDS, EOSC, IVOA and ESCAPE

Mark Allen *(CDS)*

CENTRE DE DONNÉES
ASTRONOMIQUES DE STRASBOURG

ESCAPE
European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

IVOA

German Astronomical Society meeting
September 22, 2020

CDS services for reference data

Reference Database of Astronomical Objects :
~11 million objects, ~36 million IDs, 22 million citation links

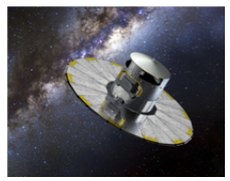
Reference service of astronomical catalogues :
~20000 Catalogues, 38000+ published tables, ~30 billion rows

Visualisation and image database :
900+ HiPS surveys: images and cubes (>350 TB)
Aladin Desktop Client, Aladin Lite - embeddable web app

+ Cross-match service, CDS Portal, and many APIs…

Made interoperable by standards (IVOA, CDS, +)

Contributing to the global astronomy data infrastructure



Ground and Space Observatories, Instruments and missions

ESO esa cnes NASA

CDS

Journals

Astronomy & Astrophysics AAS MONTHLY NOTICES of the Royal Astronomical Society + …

Astronomy Data Centres

astrophysics data system

NED CADC, MAST, HEASARC, IPAC, + …

Virtual Observatory

EURO VO France IVOA

H2020 projects: ESCAPE European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Data e-Infrastructures

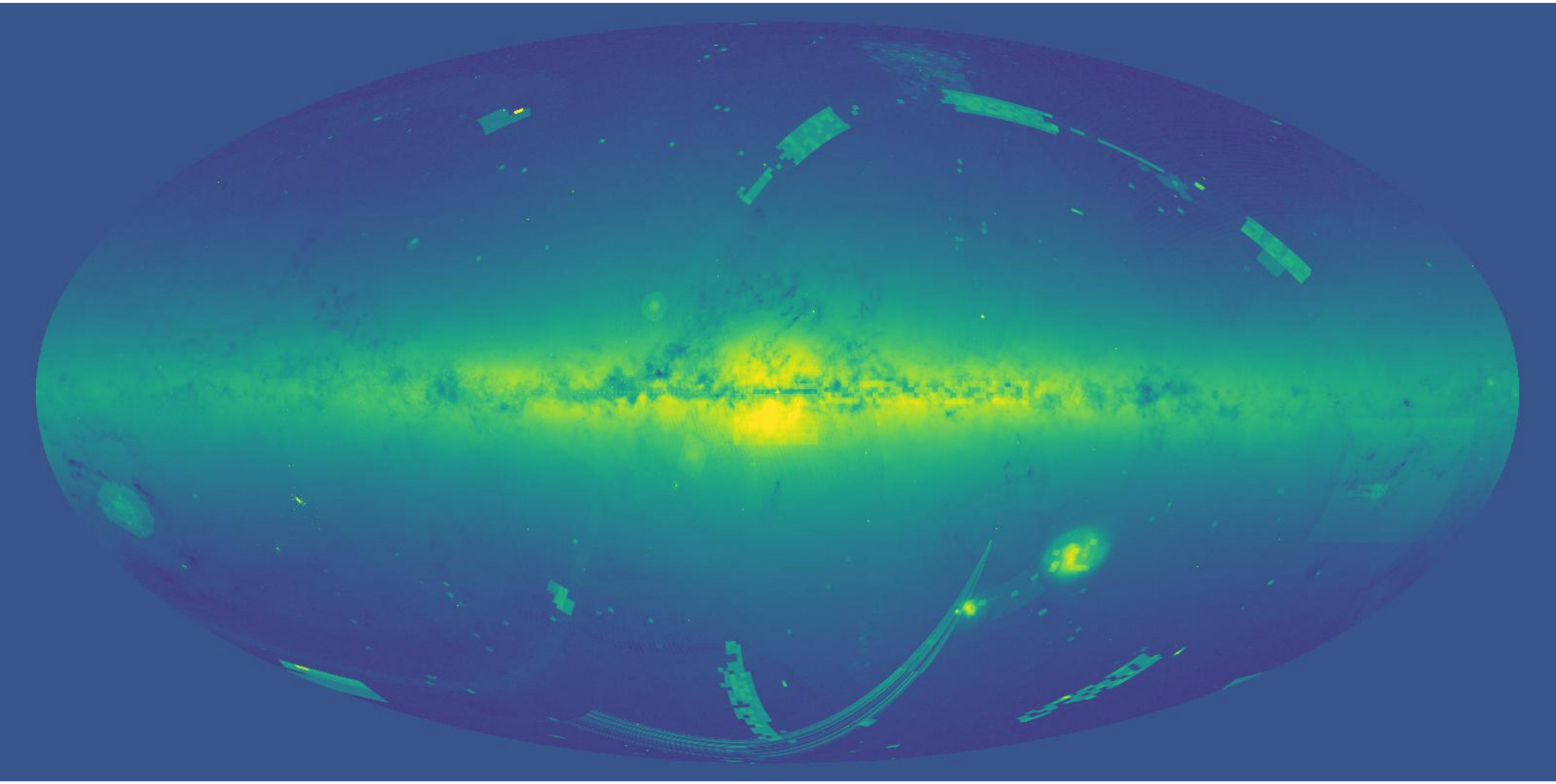
RDA RESEARCH DATA ALLIANCE + RDA Europe

CORE TRUST SEAL

ICSU WORLD DATA SYSTEM

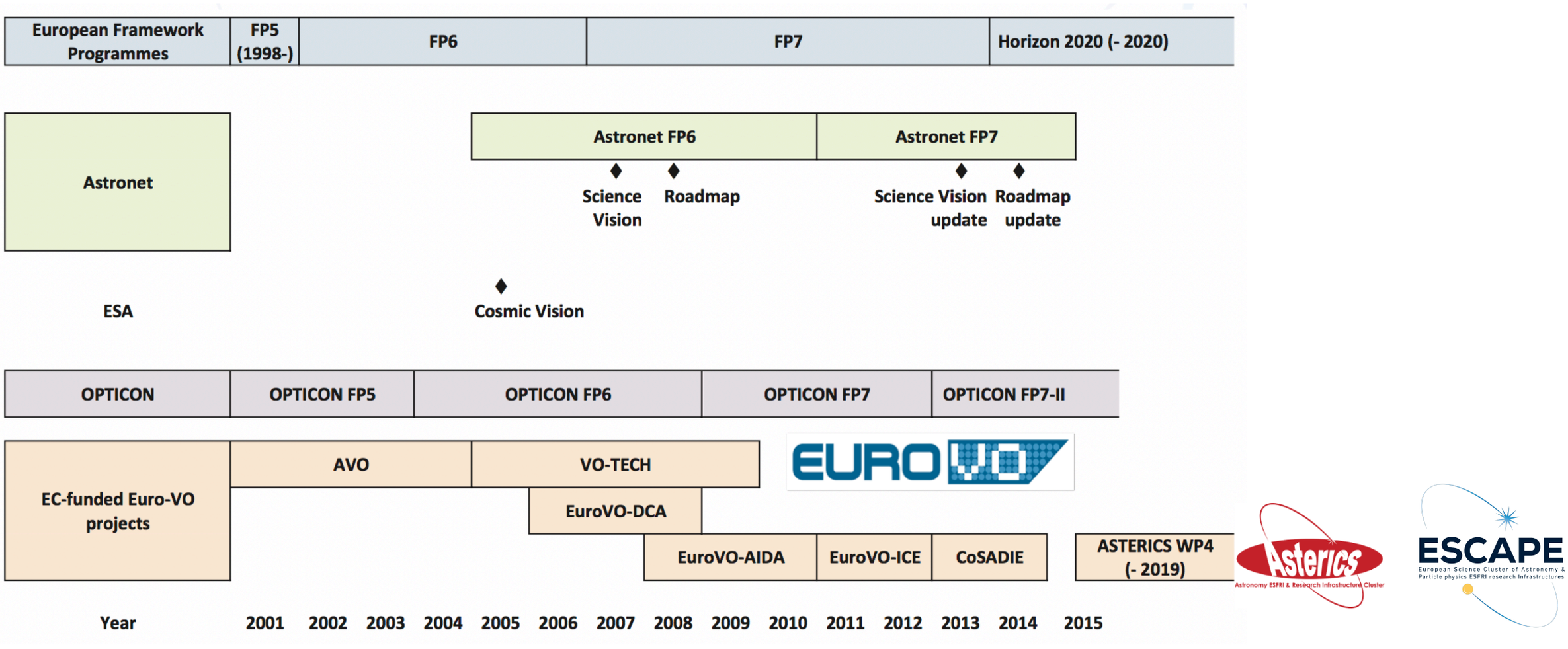
Interoperability enables new views of the data

e.g. Published catalogues (A&A) - density map



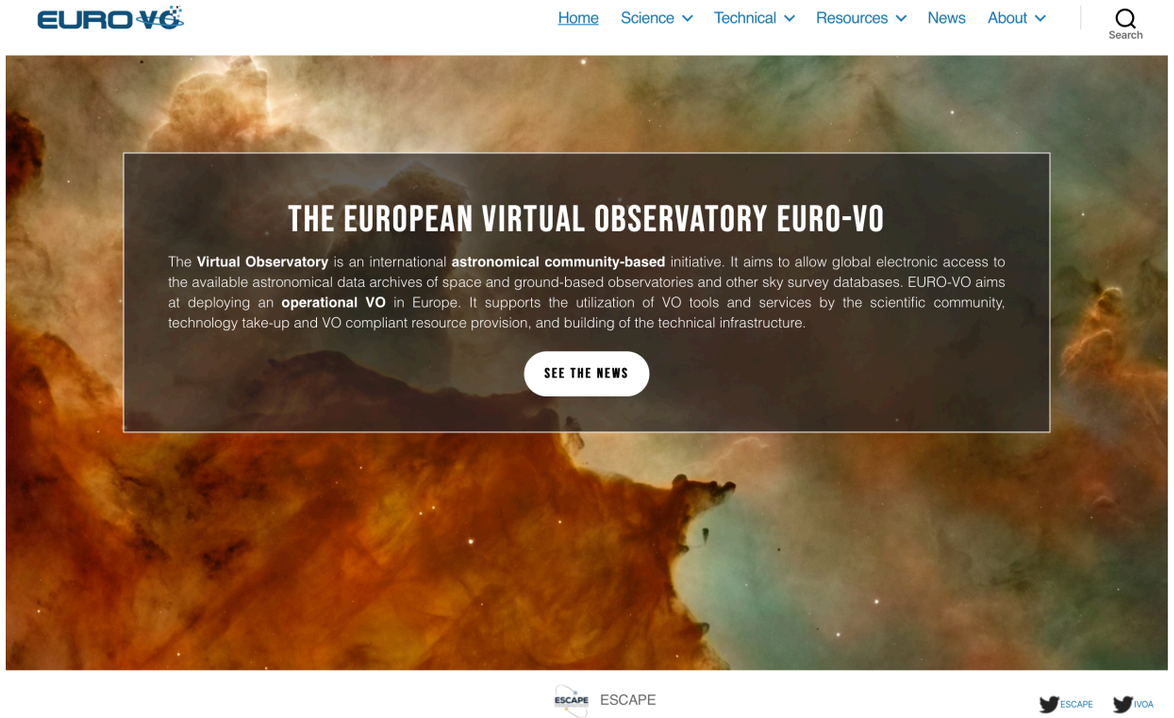
VO projects in Europe… since ~2001

DE, UK, ES, FR, IT, (ESA, ESO)



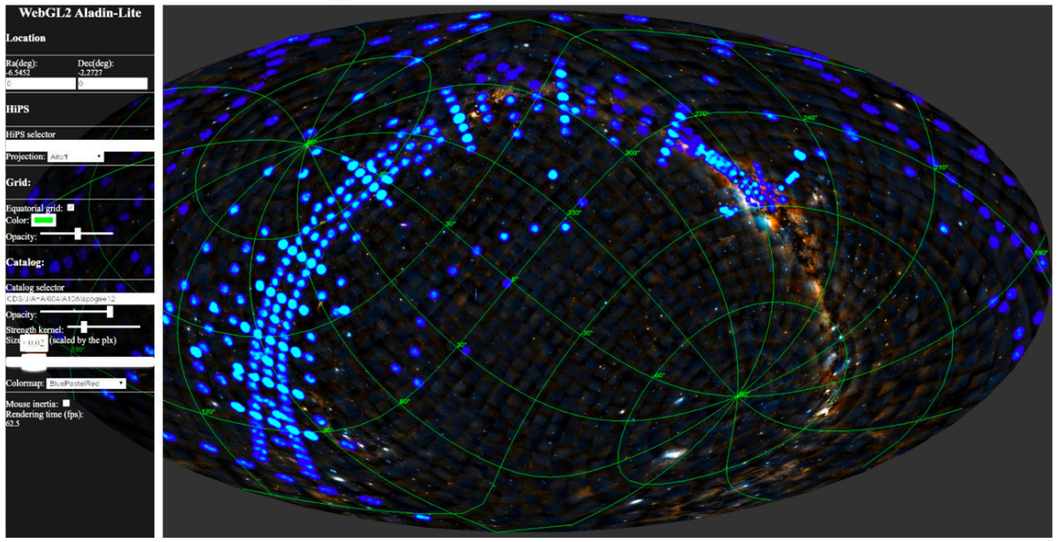
Actions towards different groups:

- **VO teams:** Building VO - standards and tools
- **Data Publishers:** implementation and requirements
- **Science users:** requirements, training, feedback



Euro-VO new web pages - to be launched in October

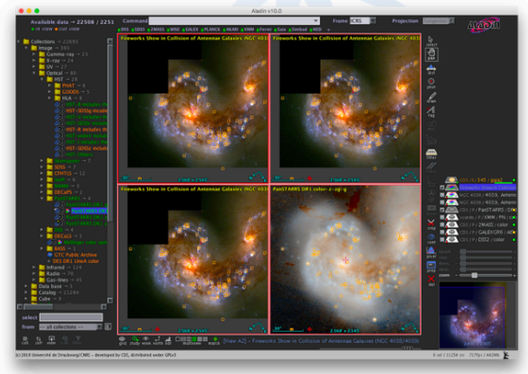
Some science tools in action:

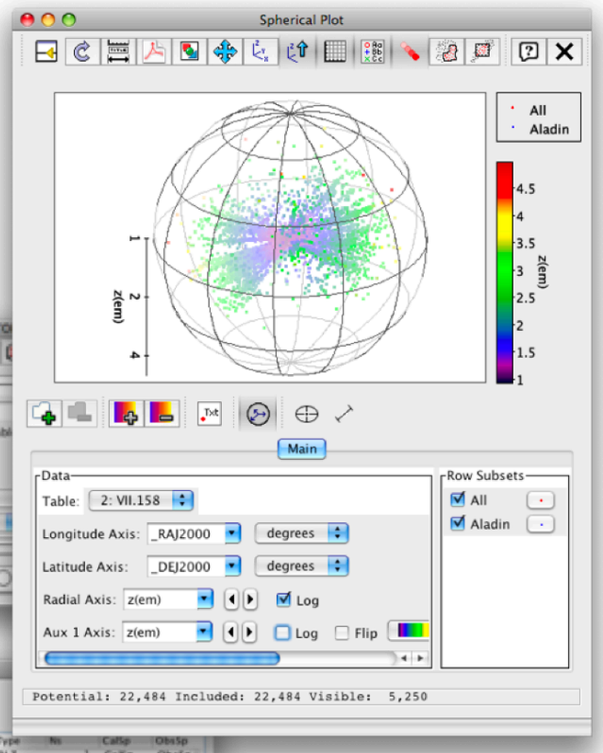


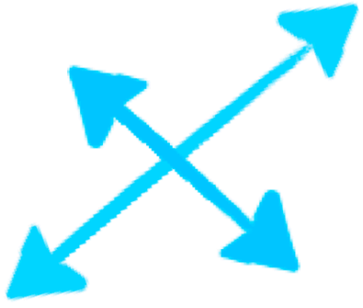
New improved WebGL Aladin Lite

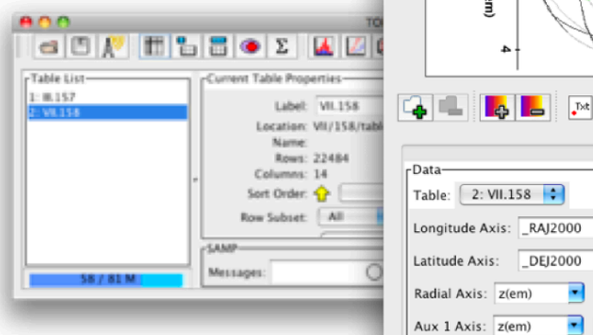
```
In [ ]:
from ipyaladin import Aladin
a = Aladin(target='18 55 24.508 +04 29 46.72', survey='P/Mellinger/color', fov=180)
a

In [ ]:
a.add_moc_from_URL('http://alasky.u-strasbg.fr/footprints/tables/vizier/VI_158_gaia2/MOC', {'color':

In [ ]:
a.survey = 'P/GALEXGR6/AIS/color';a.target = 'M101';a.fov = 0.3

In [ ]:
nloadTable&outputFormat=vot&filename=vizier_M101_II_328_allwise_20190322', {'color': 'red', 'onClick': 'showTable'})
```

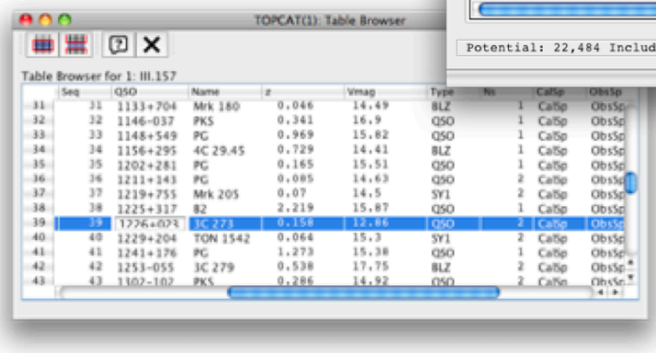
ed10.ipynb#











Notebooks & Platforms

TOPCAT

Broadcast

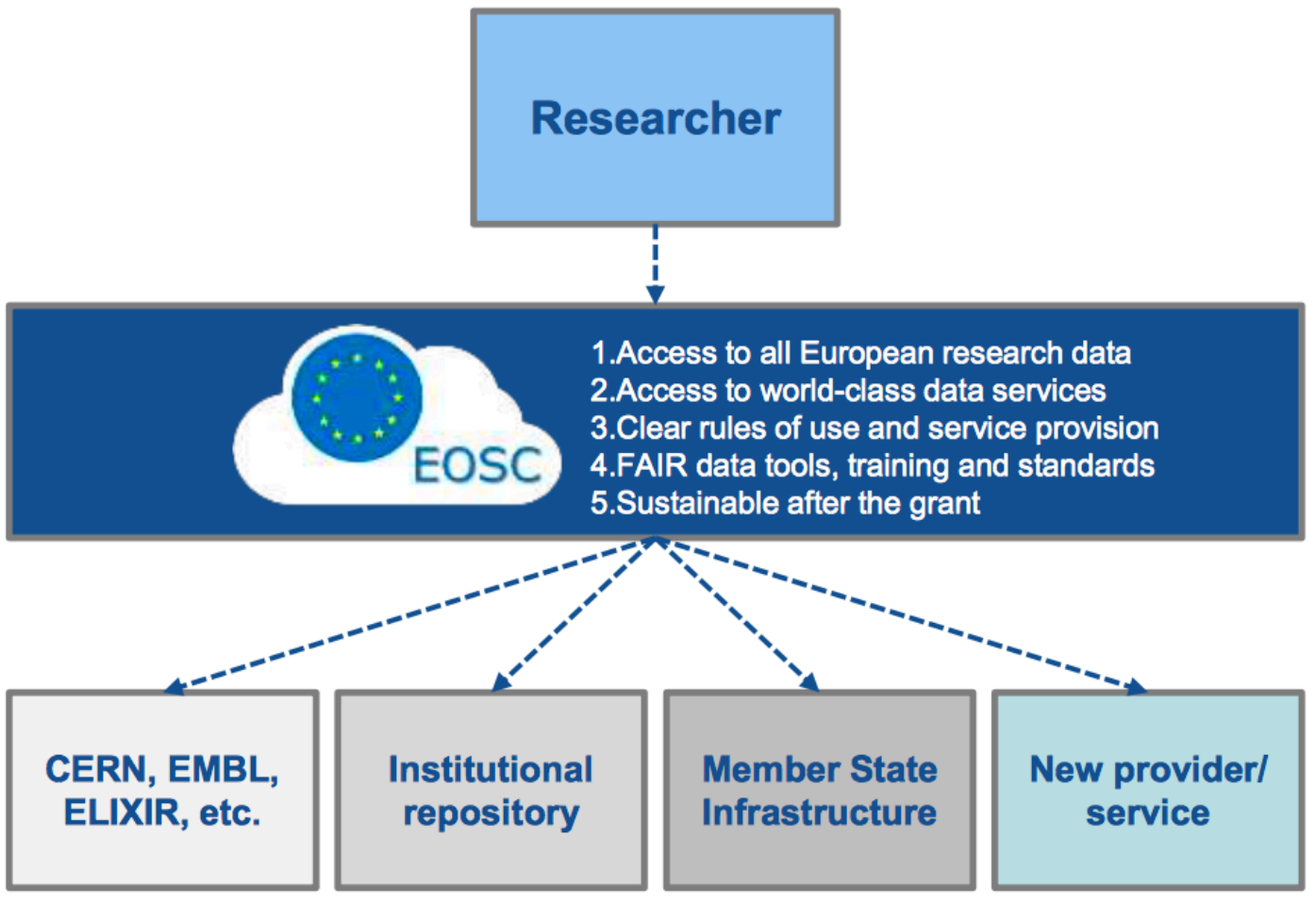
GERMAN ASTROPHYSICAL VIRTUAL OBSERVATORY
GAVO

Registry Services, ADQL, TAP, Semantics, +++

New context: **European Open Science Cloud**

European Commission

A. The EOSC will allow for universal access to data and a new level playing field for EU researchers

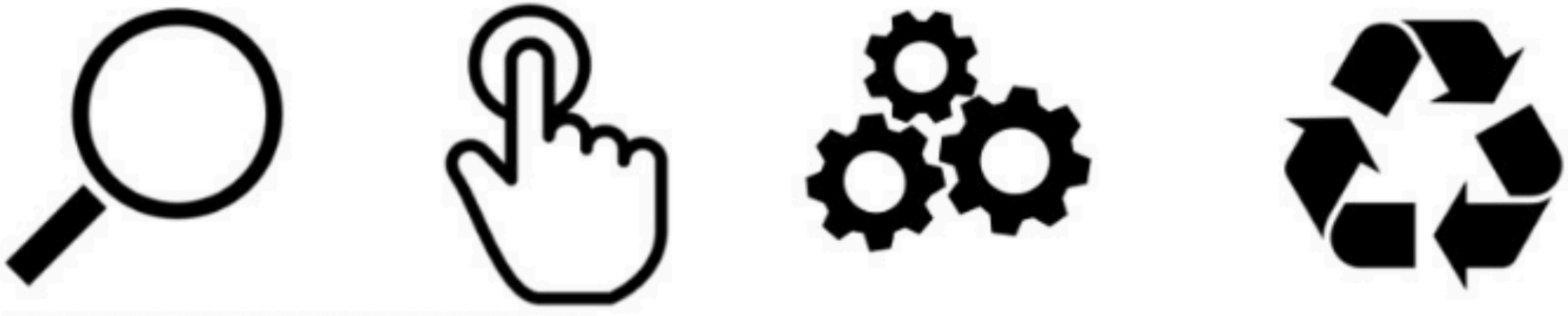


- Easy access through a universal access point for ALL European researchers
- Cross-disciplinary access to data unleashes potential of interdisciplinary research
- Services and data are interoperable (FAIR data)
- Data funded with public money is in principle open (as open as possible, as closed as necessary)
- EOSC will help increase recognition of data intensive research and data science

Seamless environment, enabling interdisciplinary research

Open Science

Findable Accessible Interoperable Reusable



Common language in the changing landscape of data sharing

FAIR

- **F**indable, **A**ccessible, **I**nteroperable, **R**eusable

Open Science

- Data sharing with open and seamless services to analyse and reuse research data to improve science

Stewardship

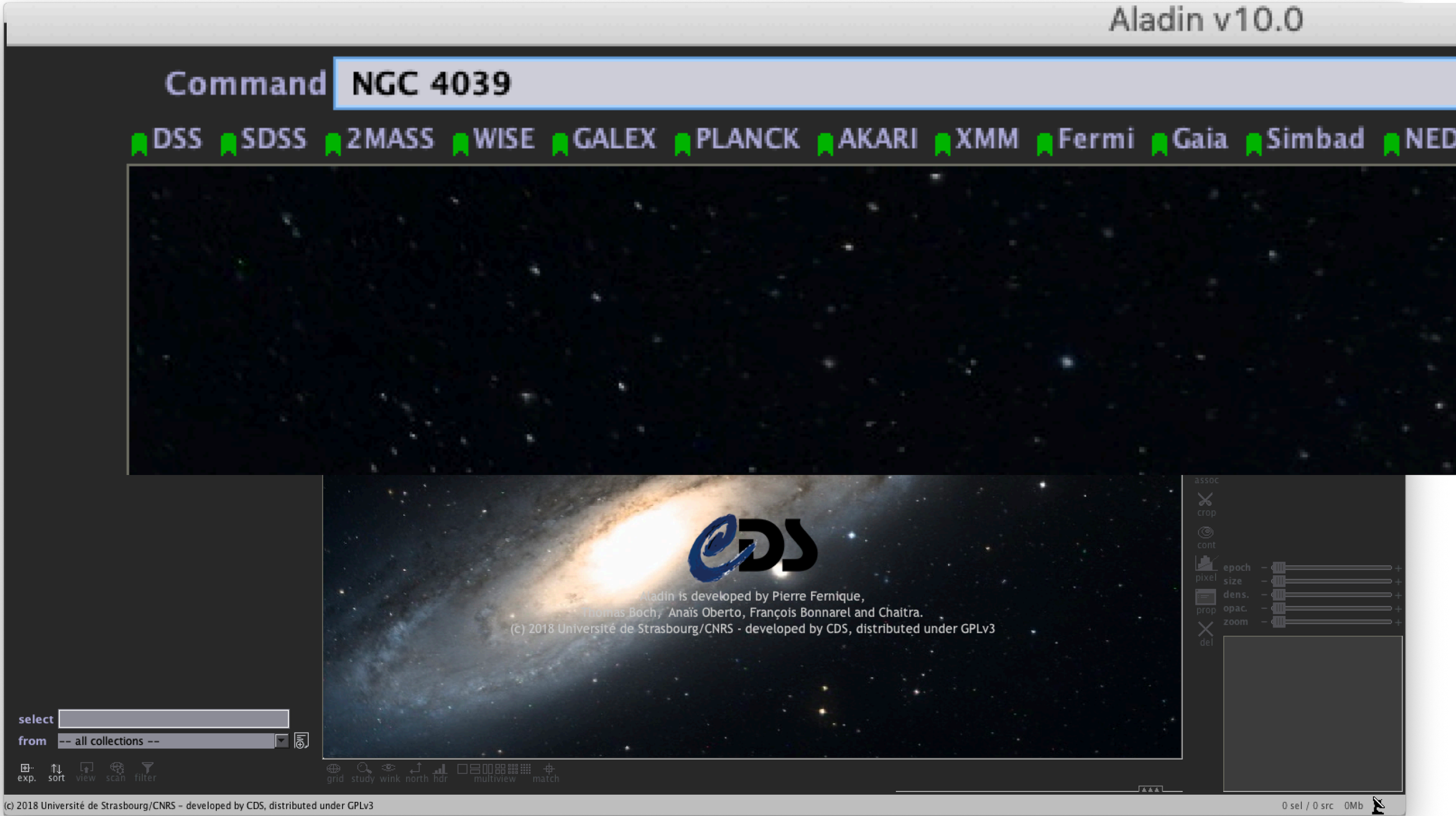
- Human skills for curation, quality content, data management, services

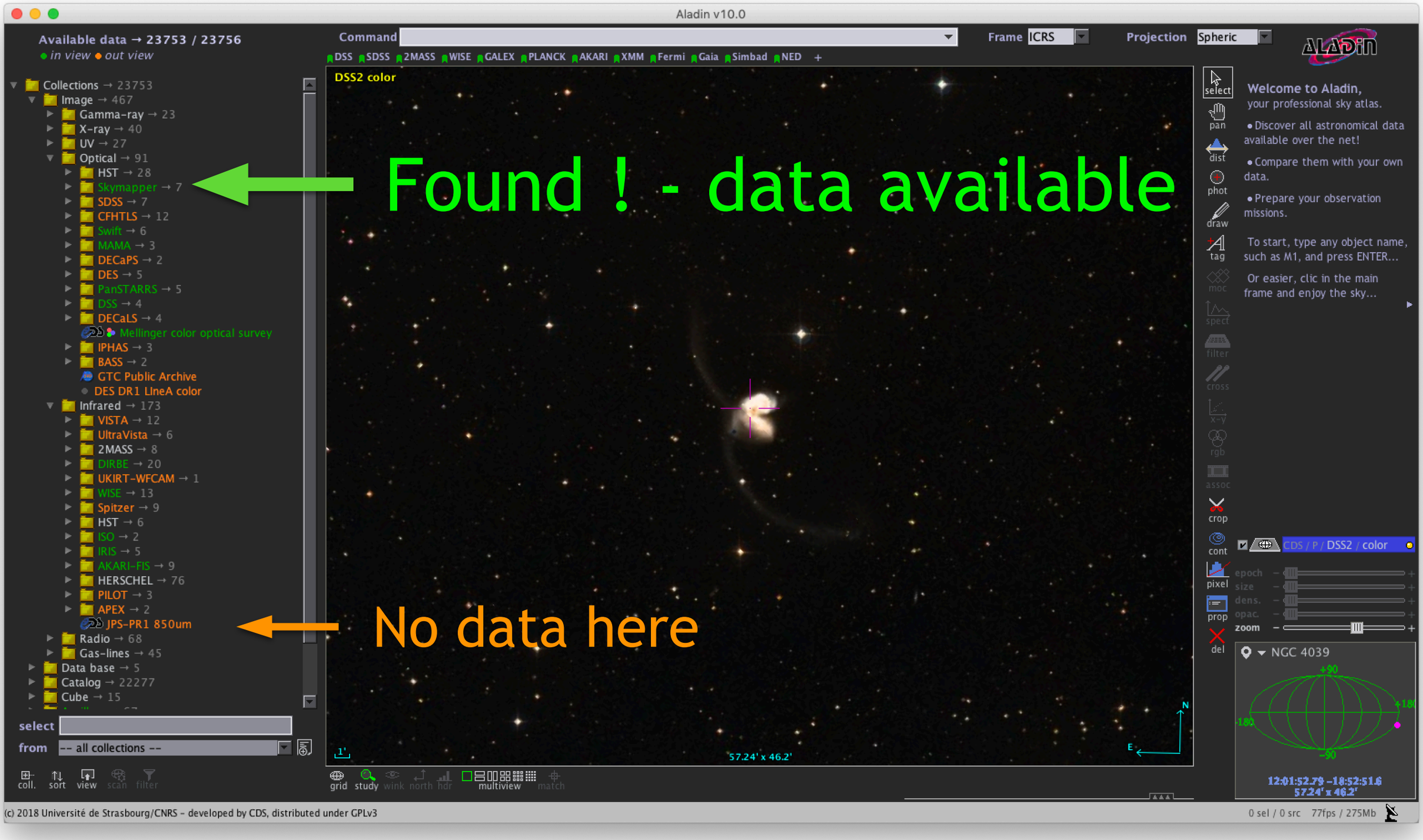
An example:

NGC 4039 - an interacting galaxy

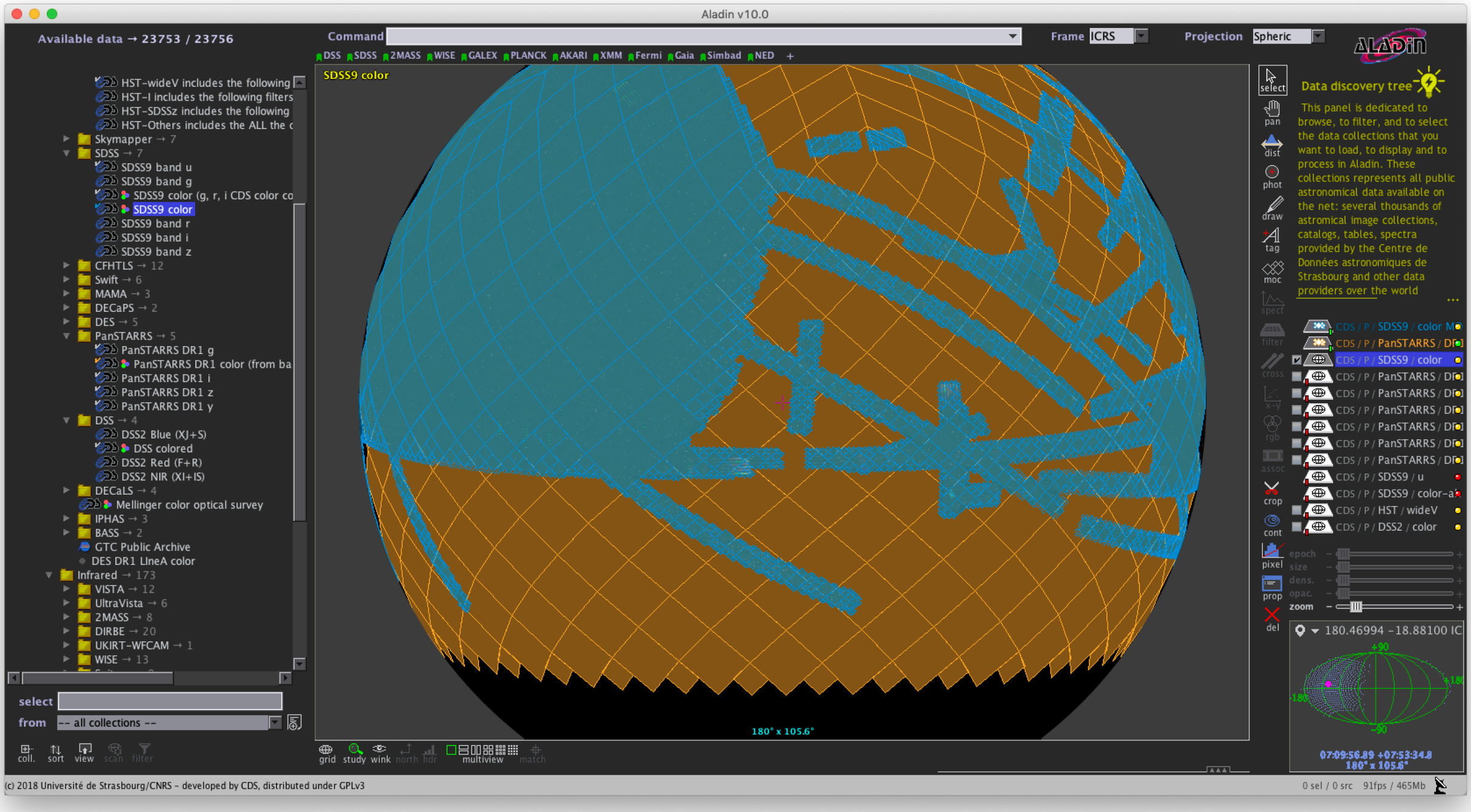
- **F**ind the data available
  - **A**ccess the data
    - **I**nteroperable use of the data with other data
      - **R**e-use the data

Findable

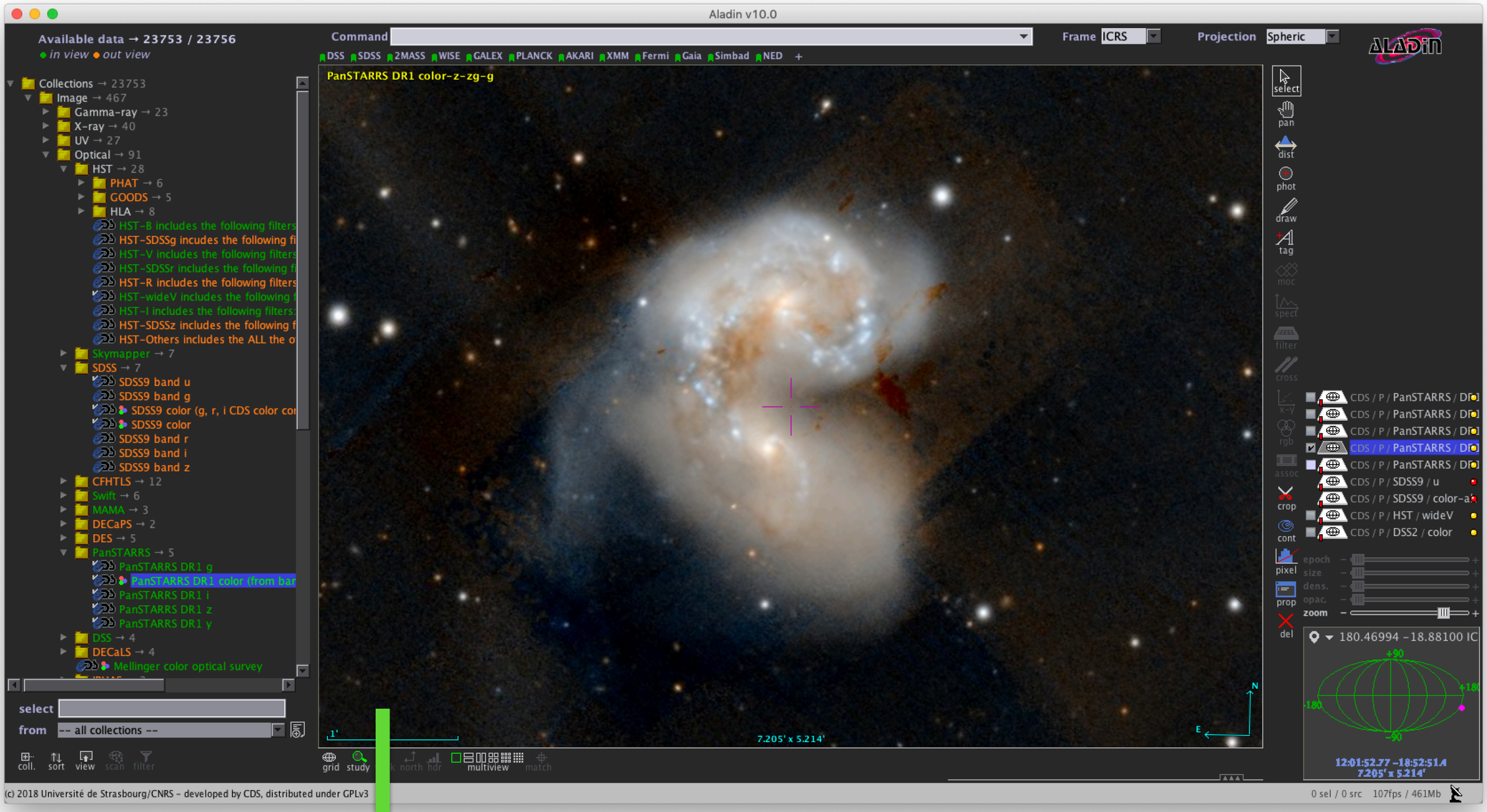

Aladin v10.0
Command
NGC 4039
DSS
SDSS
2MASS
WISE
GALEX
PLANCK
AKARI
XMM
Fermi
Gaia
Simbad
NED
Aladin is developed by Pierre Fernique,
Thomas Boch, Anaïs Oberto, François Bonnarel and Chaitra.
(c) 2018 Université de Strasbourg/CNRS - developed by CDS, distributed under GPLv3
assoc
crop
cont
pixel
prop
del
epoch
size
dens.
opac.
zoom
select
from
-- all collections --
exp.
sort
view
scan
filter
grid
study
wink
north
hdr
multiview
match
(c) 2018 Université de Strasbourg/CNRS - developed by CDS, distributed under GPLv3
0 sel / 0 src
0Mb


Aladin v10.0
Available data → 23753 / 23756
in view out view
Command
Frame ICRS
Projection Spheric
DSS SDSS 2MASS WISE GALEX PLANCK AKARI XMM Fermi Gaia Simbad NED +
DSS2 color
Collections → 23753
Image → 467
Gamma-ray → 23
X-ray → 40
UV → 27
Optical → 91
HST → 28
Skymapper → 7
SDSS → 7
CFHTLS → 12
Swift → 6
MAMA → 3
DECaPS → 2
DES → 5
PanSTARRS → 5
DSS → 4
DECaLS → 4
Mellinger color optical survey
IPHAS → 3
BASS → 2
GTC Public Archive
DES DR1 LIneA color
Infrared → 173
VISTA → 12
UltraVista → 6
2MASS → 8
DIRBE → 20
UKIRT-WFCAM → 1
WISE → 13
Spitzer → 9
HST → 6
ISO → 2
IRIS → 5
AKARI-FIS → 9
HERSCHEL → 76
PILOT → 3
APEX → 2
JPS-PR1 850um
Radio → 68
Gas-lines → 45
Data base → 5
Catalog → 22277
Cube → 15
select
from -- all collections --
coll. sort view scan filter
Found ! - data available
No data here
1'
57.24' x 46.2'
N
E
grid study wink north hdr multiview match
select pan dist phot draw tag moc spect filter cross x-y rgb assoc crop cont pixel prop del
Welcome to Aladin,
your professional sky atlas.
• Discover all astronomical data available over the net!
• Compare them with your own data.
• Prepare your observation missions.
To start, type any object name, such as M1, and press ENTER...
Or easier, clic in the main frame and enjoy the sky...
CDS / P / DSS2 / color
epoch
size
dens.
opac.
zoom
NGC 4039
+90
-90
12:01:52.79 -18:52:51.6
57.24' x 46.2'
(c) 2018 Université de Strasbourg/CNRS - developed by CDS, distributed under GPLv3
0 sel / 0 src 77fps / 275Mb

...based on standardised coverage map indexing

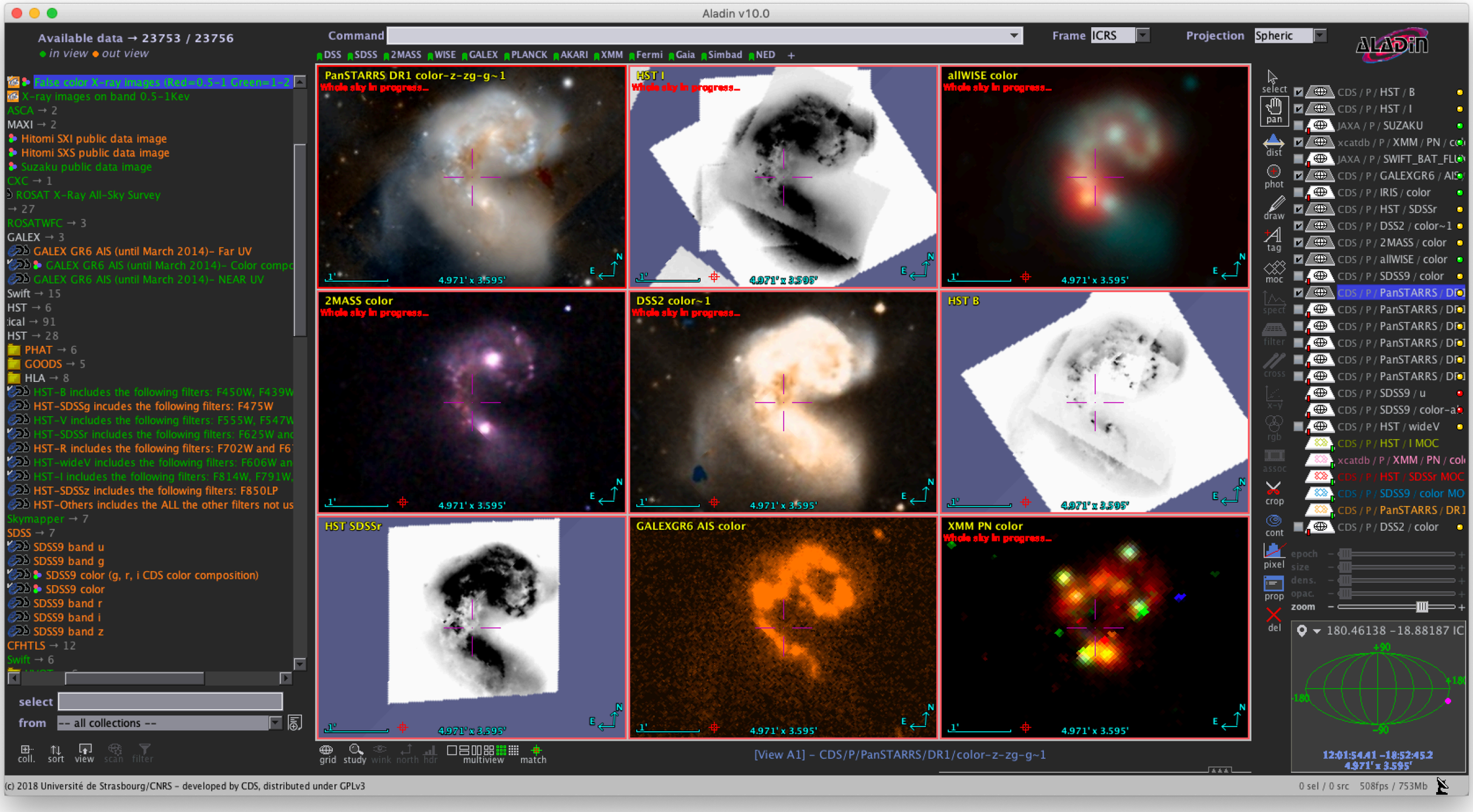


Accessible

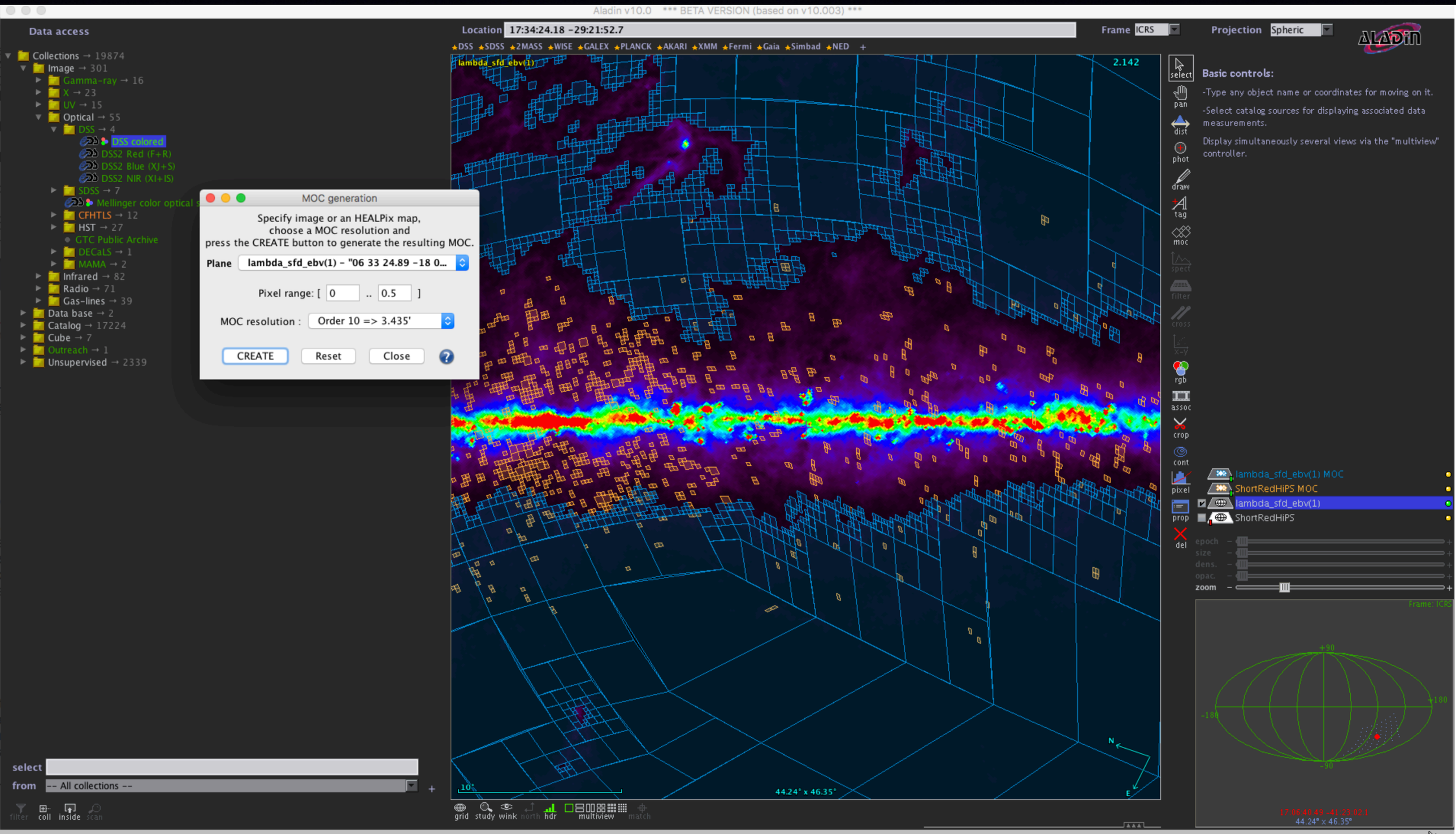


Download in science/visualisation formats

Interoperable



Interoperable: complex catalog/image queries



Reusable

Services for extracting :cut-outs of the data for re-use

hips2fits

Fast generation of FITS cutouts from HiPS datasets

The hips2fits service enables generation of FITS images cutouts of arbitrary size and resolution from a given HiPS.

Try it now!

Use the form above to test the service

By parameters | **By WCS**

HiPS survey	eg: DSS2/red, PanSTARRS/DR1/g
Dimension	1200 x 900 pixels
Target	NGC 4039
Projection	TAN - tangential
Image size	0.5 degrees
Projection frame	ICRS / Galactic
Rotation angle	0.0 degrees

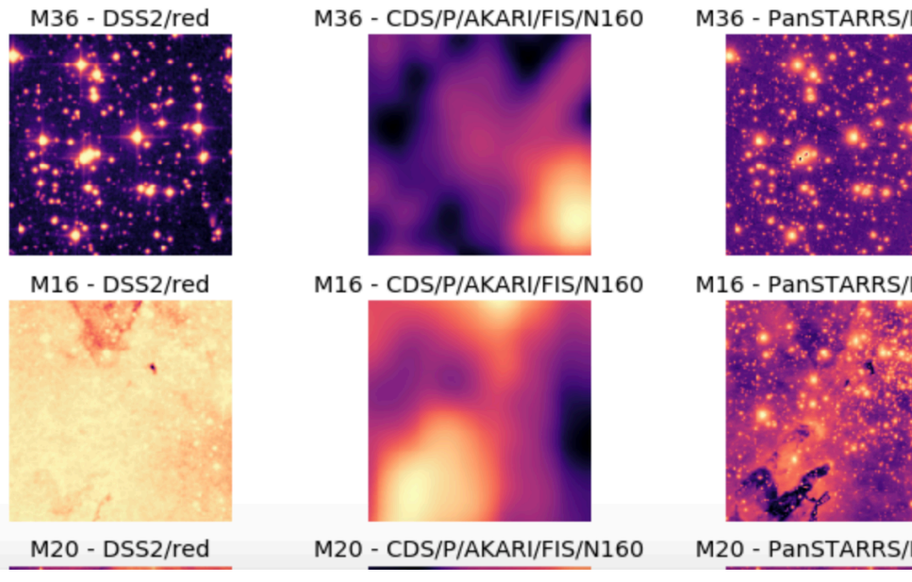
Download FITS | Preview JPG | Open JPG in new tab

jupyter hips2fits (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

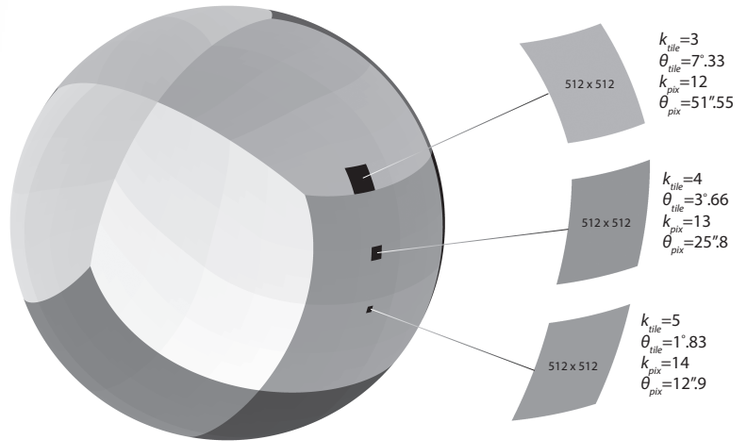
Run Code

```
i = 0
for obj in objects:
    for hips in hips_list:
        axs[i].set_axis_off()

        axs[i].set_title('{} - {}'.format(obj, hips))
        sc = SkyCoord.from_name(obj)
        ra = sc.icrs.ra.deg
        dec = sc.icrs.dec.deg
        url = 'http://alasky.u-strasbg.fr/hips-image-services/hips2fits
        hdu = fits.open(url)

        hdu.writeto('{}-{}.fits'.format(obj, hips.replace('/', '_')), o
        im = hdu[0].data
        norm = ImageNormalize(im, interval=MinMaxInterval(),
                              stretch=AsinhStretch())
        axs[i].imshow(im, cmap='magma', norm=norm, origin='lower')

        i += 1
```



k	N_{side} =		θ_{pix}	$k_{tile,512}$	$N_{tile,512}$	$\theta_{tile,512}$
0						
1						
2						
3						
4						
5						
6						
7						
8						
9					12	58.°6
10					48	29.°3
11					192	14.°7
12					768	7.°33
13					3072	3.°66
14					12288	1.°83
15					49152	55.′0
16					196608	27.′5
17					786432	13.′7
18					3,145,728	6.′87
19					12,582,912	3.′44
20					50,331,648	1.′72
21					201,326,592	51.″5
22					805,306,368	25.″8
23					3.22×10^9	12.″9
24	2^{24}			5	1.29×10^{10}	6.″44
25	2^{25}					

Tiles --------



- WMAP
- PLANCK HFI
- IRAS
- NVSS
- SCUBA
- DSS
- SDSS
- CFHTLS
- HST ACS

International Virtual Observatory Alliance

HiPS – Hierarchical Progressive Survey

Version 1.0

IVOA Recommendation 19th May 2017

This version:
1.0: Recommendation 2017-05-19

Previous version(s):
1.0: Proposed Recommendation 2017-04-06
1.0: Proposed Recommendation 2017-04-03
1.0: Proposed Recommendation 2017-02-07
1.0: Working Draft 2016-06-23

Interest/Working Group:
Applications: http://www.ivoa.net/twiki/bin/view/IVOA/IvoaApplications

Editor:
Pierre Fernique

Authors:
Pierre Fernique, Mark Allen, Thomas Boch, Tom Donaldson, Daniel Durand, Ken Ebisawa, Laurent Michel, Jesus Salgado, Felix Stoehr

Abstract

This document presents HiPS, a hierarchical scheme for the description, storage and access of sky survey data. The system is based on hierarchical tiling of sky regions at finer and finer spatial resolution which facilitates a progressive view of a survey, and supports multi-resolution zooming and

Interoperability

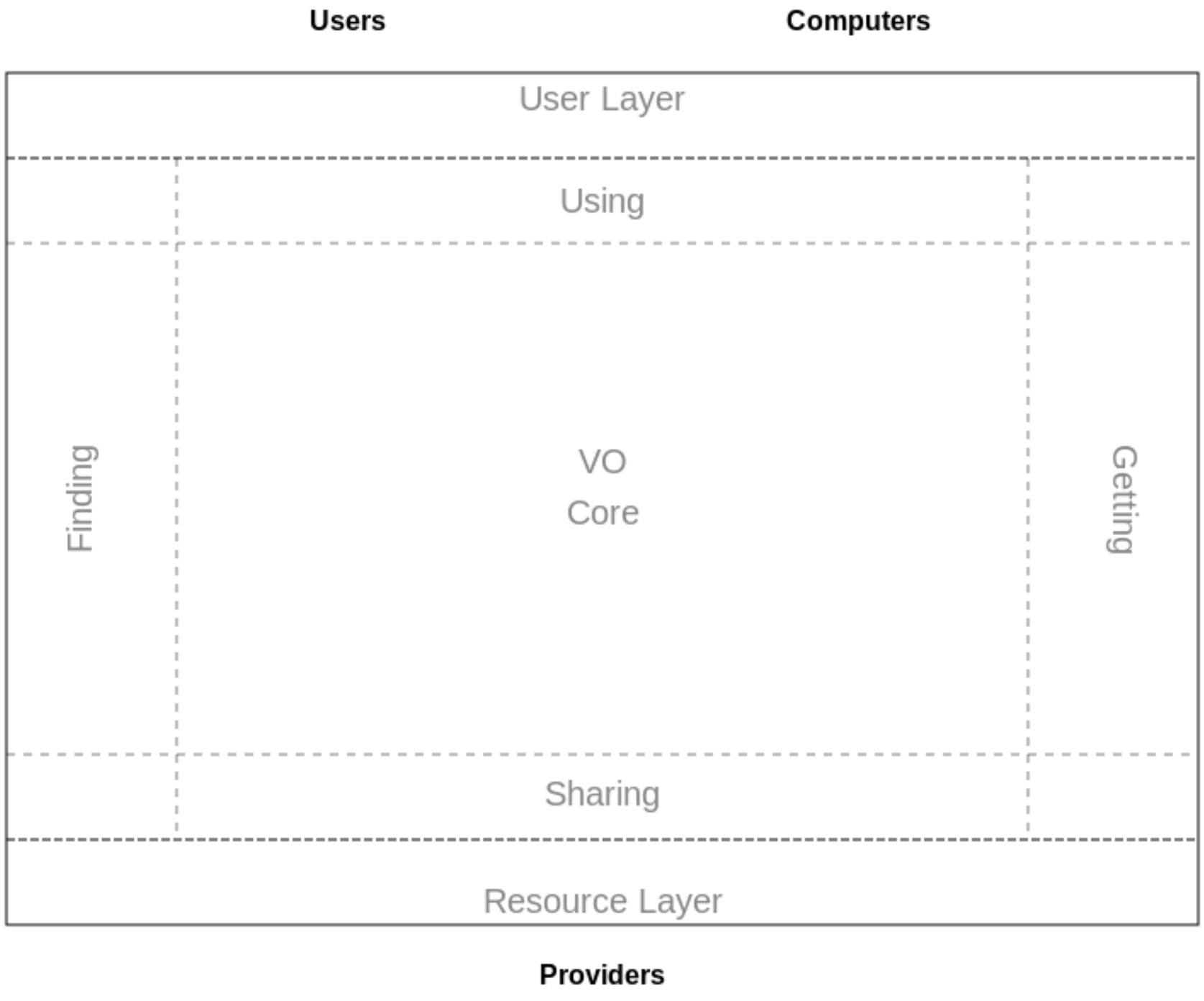
Astronomy - long experience of sharing data

- FITS standard since ~1981
- bibcode for journals

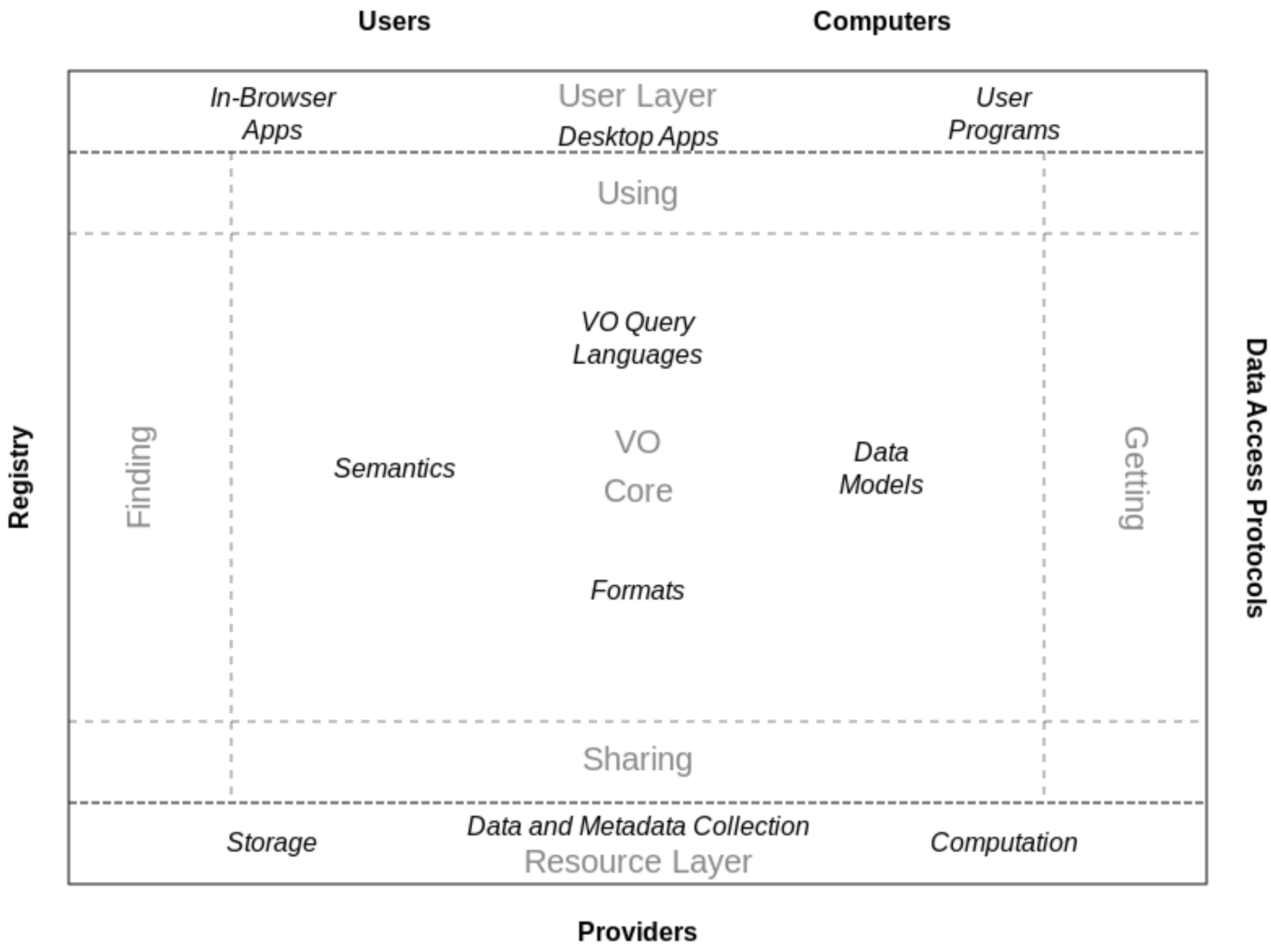
International Virtual Observatory Alliance

- Since 2001
- Standards for astronomy “interoperability”
- 21 international member projects
- 2 Interoperability meetings per year

Architecture diagrams of standards

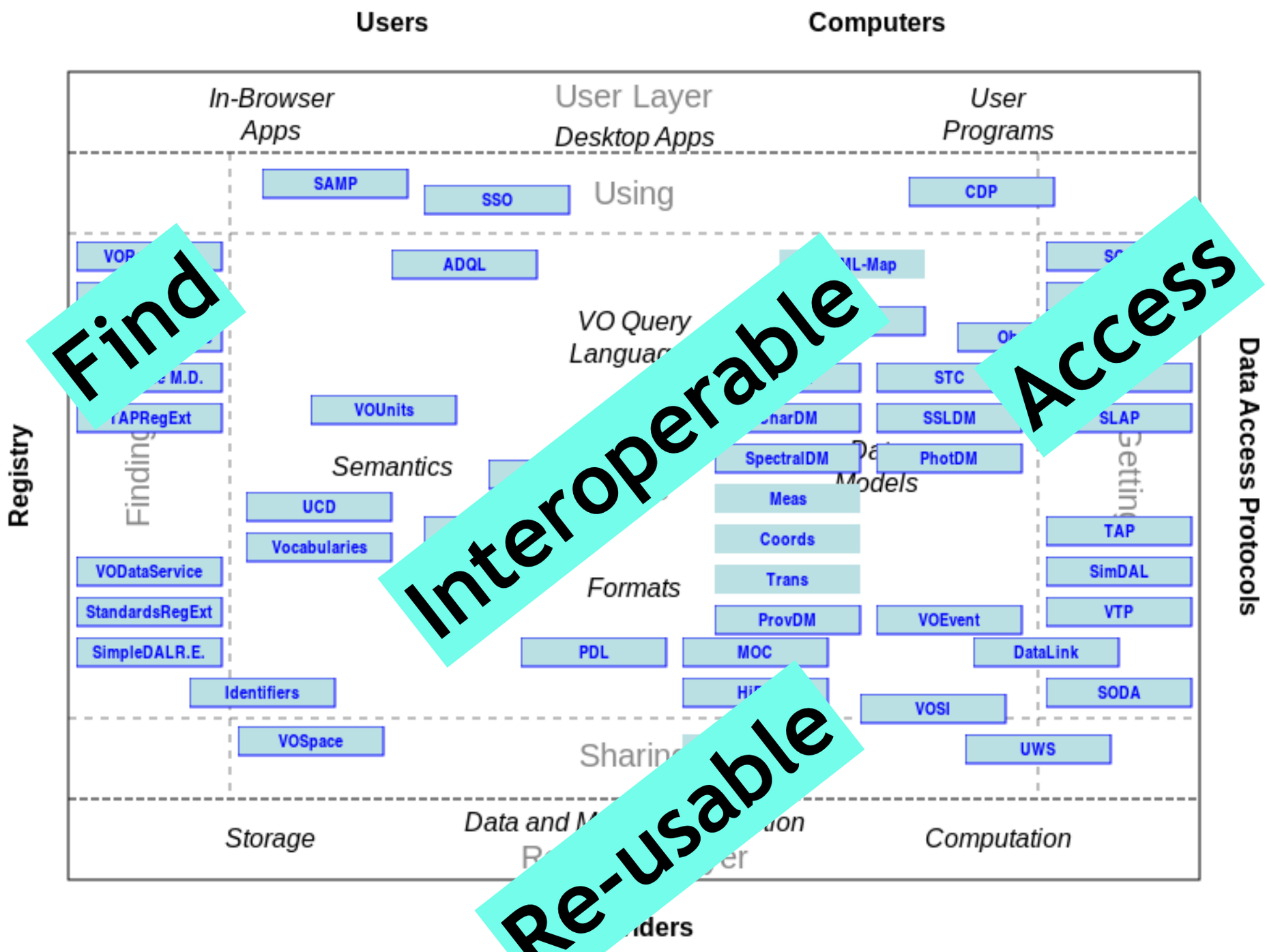


Level 0 - concepts



Level 1 - VO terminology

VO is FAIR



Level 2 - Mapping of VO standards, and their status

Working Draft

IVOA Recommendation

Documents & Standards

DOCUMENTS XML SCHEMA TEMPLATES DOC SUBMISSION

- *Technical Specifications*
- *Notes*
- *Promotion process*
- *IVOA Technical Assessment and Roadmap Documents*
- *Submission Log*

Technical Specifications

>

Group	Title	Most stable	In progress	Version history
App	SAMP - Simple Application Messaging Protocol	1.3		1.3 1.3 1.3 1.3 1.3 1.2 1.2 1.2 1.11 1.11 1.10 1.00
	VOTable - VOTable Format Definition	1.4		1.4 1.4 1.4 1.4 1.4 1.4 1.3 1.3 1.3 1.2 1.2 1.2 1.20 1.20 1.10 1.00
	MOC - HEALPix Multi-Order Coverage Map	1.1		1.1 1.1 1.1 1.1 1.1 1.0 1.0 1.0 1.0 1.0
	HiPS - Hierarchical Progressive Survey	1.0		1.0 1.0 1.0 1.0 1.0 1.0
DAL	DALI - Data Access Layer Interface	1.1		1.1 1.1 1.1 1.1 1.1 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
	DataLink	1.0		1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
	Simple Cone Search	1.03	1.1	1.1 1.03 1.02 1.01 1.00
	SIA - Simple Image Access	2.0		2.0 2.0 2.0 2.0 2.0 2.0 2.0 1.0 1.0 1.0 1.01 1.00
	SLAP - Simple Line Access	1.0	2.0	2.0 2.0 1.0 1.0 1.0 1.0 1.0 1.0
	SSA - Simple Spectral Access	1.1		1.1 1.1 1.1 1.1 1.04 1.03 1.02 1.01 1.01 1.00
	STC-S: Space-Time Coordinate Metadata Linear String Implementation	1.0		1.0
	TAP - Table Access Protocol	1.1		1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1.0 1.0 1.0 1.0 1.0 1.00
	TAPRegExt - A VOResource Schema Extension for Describing TAP Services	1.0		1.1 1.0 1.0 1.0 1.0 1.0 1.0 1.0
	ADQL - Astronomical Data Query Language	2.00	2.1	2.1 2.1 2.1 2.00 2.00 2.00 1.01 1.00
	SimDAL - Simulation Data Access Layer	1.0		1.0 1.00 1.00 1.00 1.00 1.00 1.00
	VOEvent Transport Protocol	2.00		2.00 2.00 2.00 1.00

http://ivoa.net/documents

astrophysics data system

Available via ADS, with DOI

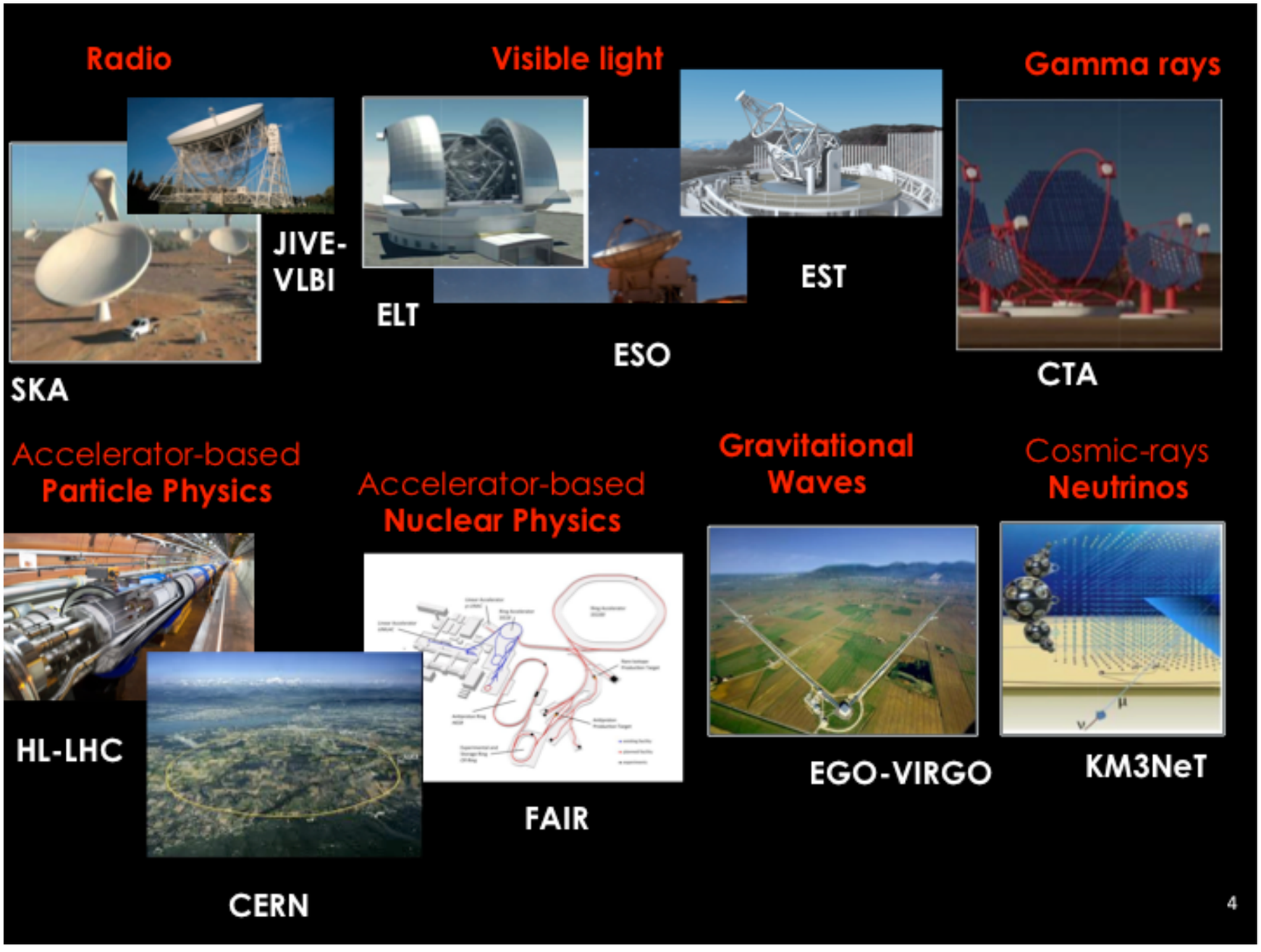
FAIRsharing.org standards, databases, policies

Listed in FAIRsharing

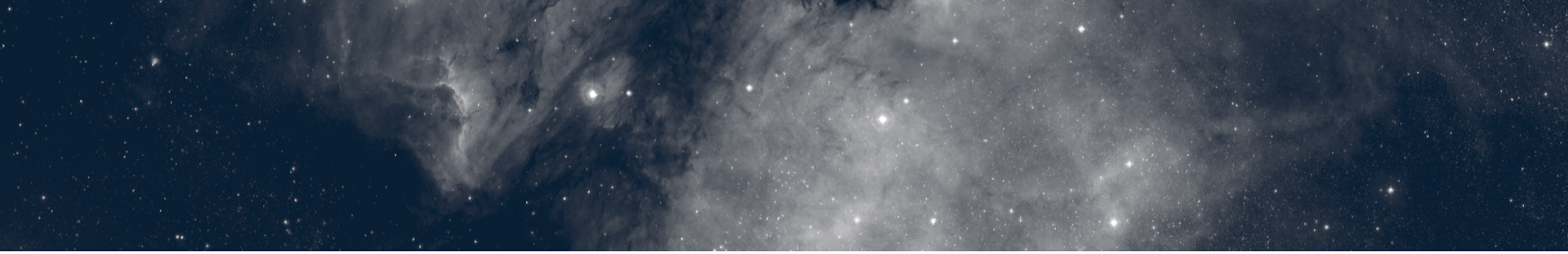
46 standards !

Astronomy (+) and EOSC

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures



- Connecting the Astronomy / Astroparticle / Accelerator Particle Physics ESFRI and there Research Infrastructures to EOSC



ESCAPE in a nutshell

ESCAPE convenes a large scientific community

- **31** partners : **7** ESFRI & landmarks: CTA, ELT, EST, FAIR, HL-LHC, KM3NeT, SKA
- **2** pan-European International Organizations: CERN, ESO (with their world-class established infrastructures, experiments and observatories).
- **4** supporting ERA-NET initiatives: HEP (CERN), NuPECC, ASTRONET, APPEC
- **1 involved initiative/infrastructure: EURO-VO (Virtual Observatory)**
- **2** European research infrastructures: EGO and JIVE-ERIC
- Budget: **16 M€,** Started: **Feb 2019**, Duration: **42** months
- Coordinator: **CNRS (**Centre National de la Recherche Scientifique**)**

Home page: https://projectescape.eu

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

Goals towards a Virtual Research Environment

- Implementing **Science Analysis Platforms** for EOSC researchers to stage data collections, analyse them, access ESFRIs' software tools, bring their own custom workflows.

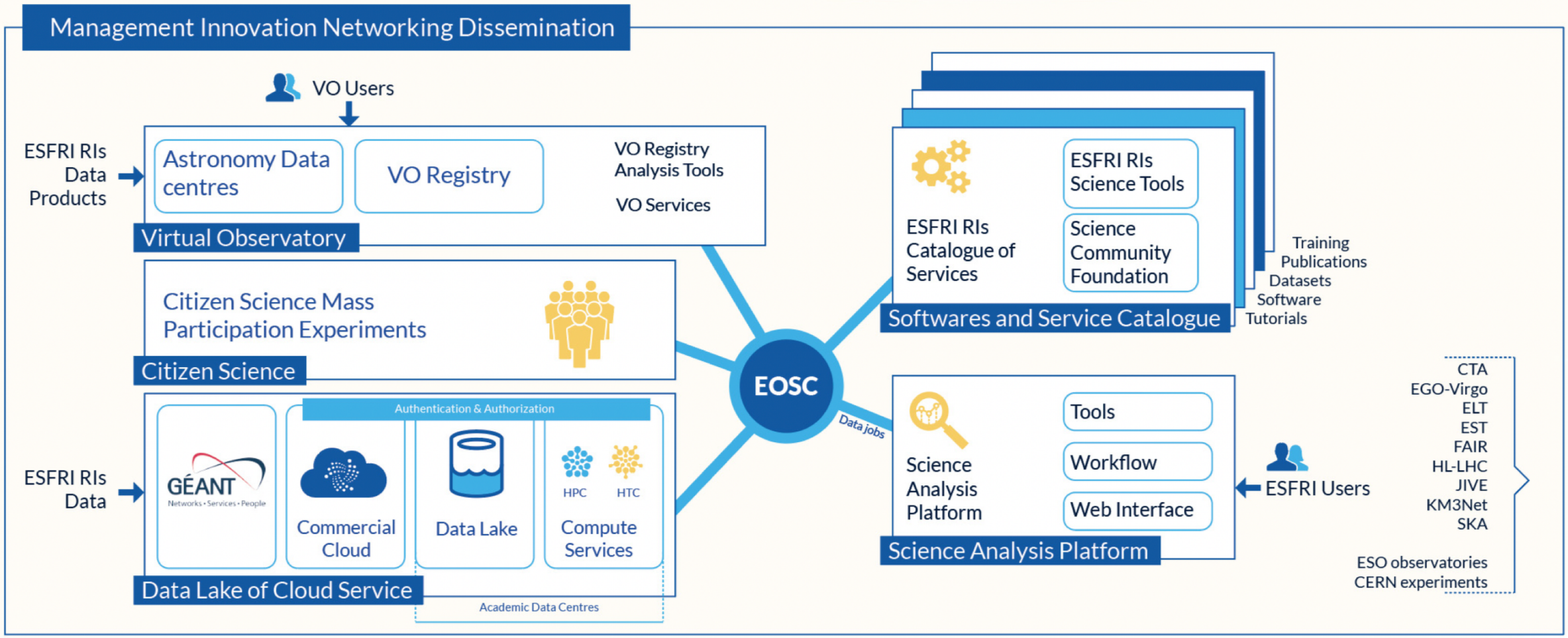
- Contributing to the EOSC global resources federation through a **Data-Lake** concept implementation to manage extremely large data volumes at the multi-Exabyte level.

- Supporting "**scientific software**" as a major component of ESFRI data to be preserved and accessed in EOSC through dedicated catalogues. Implementing a community foundation approach for continuous software shared development and training new generation researchers.

- Extending the **Virtual Observatory standards and methods** according to ***FAIR*** principles to a larger scientific context; demonstrating EOSC capacity to include existing frameworks.

- Further **involving society** in knowledge discovery.

ESCAPE goals: building a domain-based EOSC cell



VO partners and RI partners in WP4

Partners from ESFRIs and astronomy Research Infrastructures

Partners bringing experience from European Virtual Observatory projects

3 ‘FAIR’ related activities in ESCAPE WP4

1. **Integration of astronomy VO data and services into the EOSC**

- e.g. VO registry into the EOSC framework, participation in RDA, EOSC

2. **FAIR principles for data through the Virtual Observatory**

- Interoperability standards based on needs
- ***Support of science community - training schools***
- ***Forum event for data providers (2021)***
- VO data readiness for use in Science Platforms

3 . Adding value to trusted content in astronomy archives

- ML added-value to archive products
  - exploring: ‘search for similar data’

summary

- Integration of astronomy VO data and services into the EOSC

  -**Interaction with EOSC bodies, VO registry in B2FIND, tests of service publishing**

- Implementation of FAIR principles for ESFRI data through the Virtual Observatory

  -**Milestones – representation of ESCAPE priorities at IVOA level** ***- e.g. Radio Astronomy***

  -**ESFRI/RI partners requirements defined, results on tools and VO publishing**

  - e.g. New version of Aladin Lite in webGL - see ADASS in November

  - Training activities delayed due to covid-19, but coming in 2021 !!

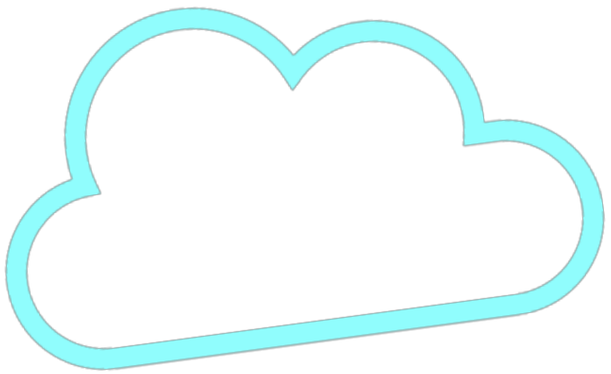
- Adding value to content in astronomy archives

  - First results of machine learning applied to archive data sets ***- see talk by A. D'Isanto***

More about EOSC… it's ambitious

The role of the European Open Science Cloud (EOSC) is to ensure that **European scientists reap the full benefits of data-driven science**, by offering:

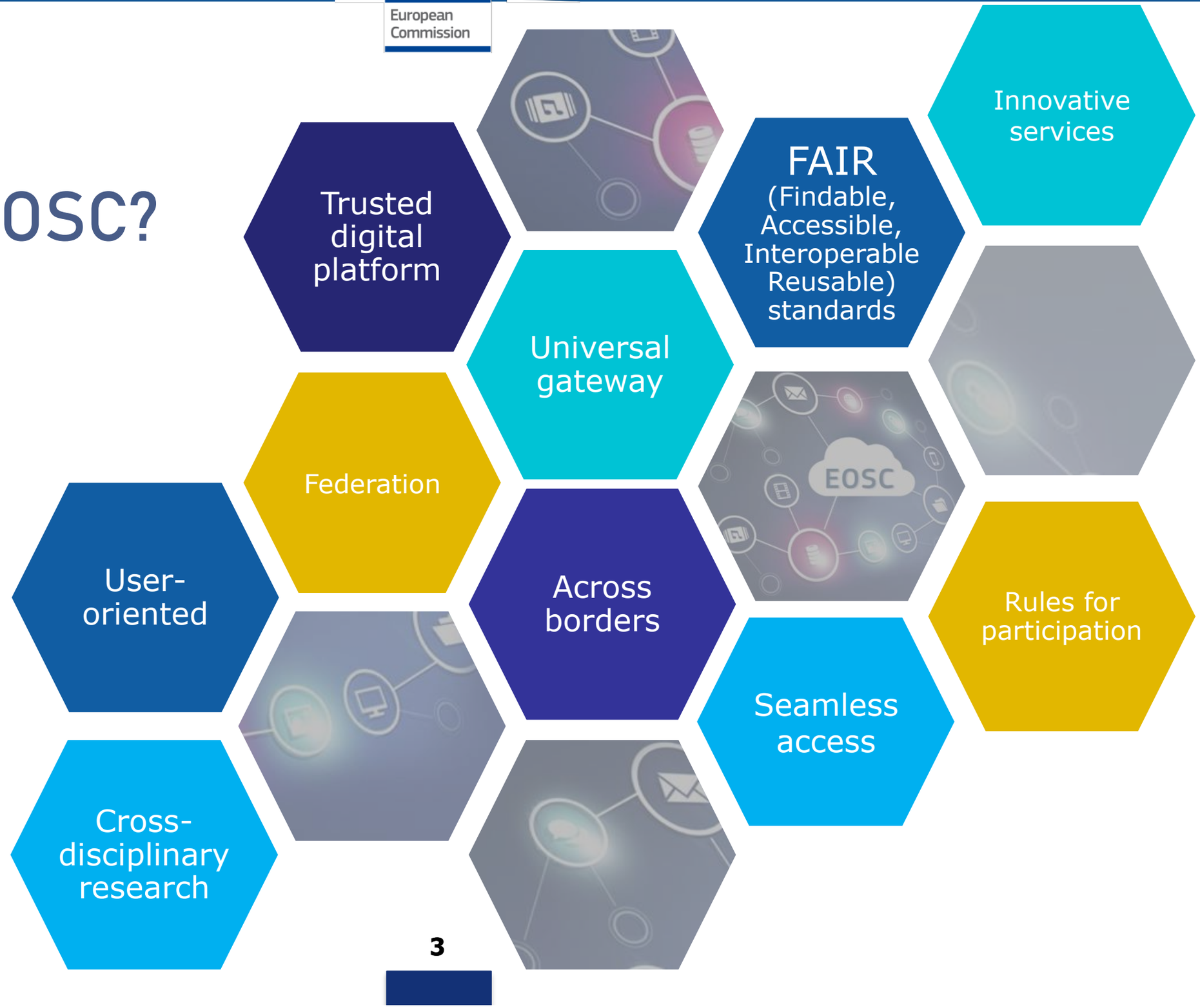
***1.9 million** European researchers and **70 million** professionals in science and technology a **trusted open distributed environment** providing seamless access to data and interoperable services addressing the whole research data life cycle.*



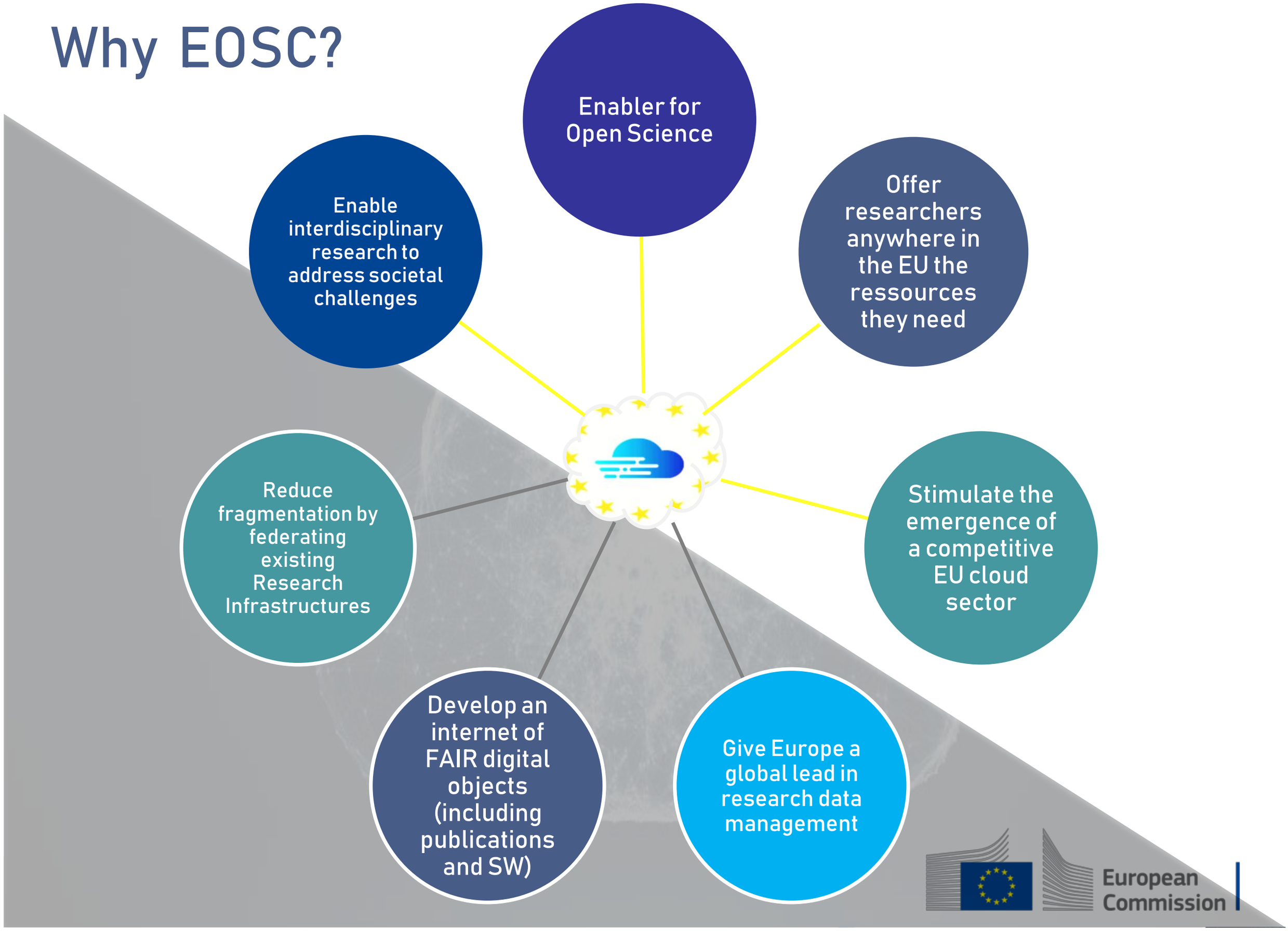
The development of the EOSC realises EU policy objectives including Open Science, FAIR data implementation and the Digital Single Market

European Commission

What is EOSC?


Trusted digital platform
Innovative services
FAIR (Findable, Accessible, Interoperable Reusable) standards
Universal gateway
Federation
EOSC
User-oriented
Across borders
Rules for participation
Seamless access
Cross-disciplinary research

Why EOSC?


Enabler for Open Science
Enable interdisciplinary research to address societal challenges
Offer researchers anywhere in the EU the ressources they need
Reduce fragmentation by federating existing Research Infrastructures
Stimulate the emergence of a competitive EU cloud sector
Develop an internet of FAIR digital objects (including publications and SW)
Give Europe a global lead in research data management
European Commission

EOSC today

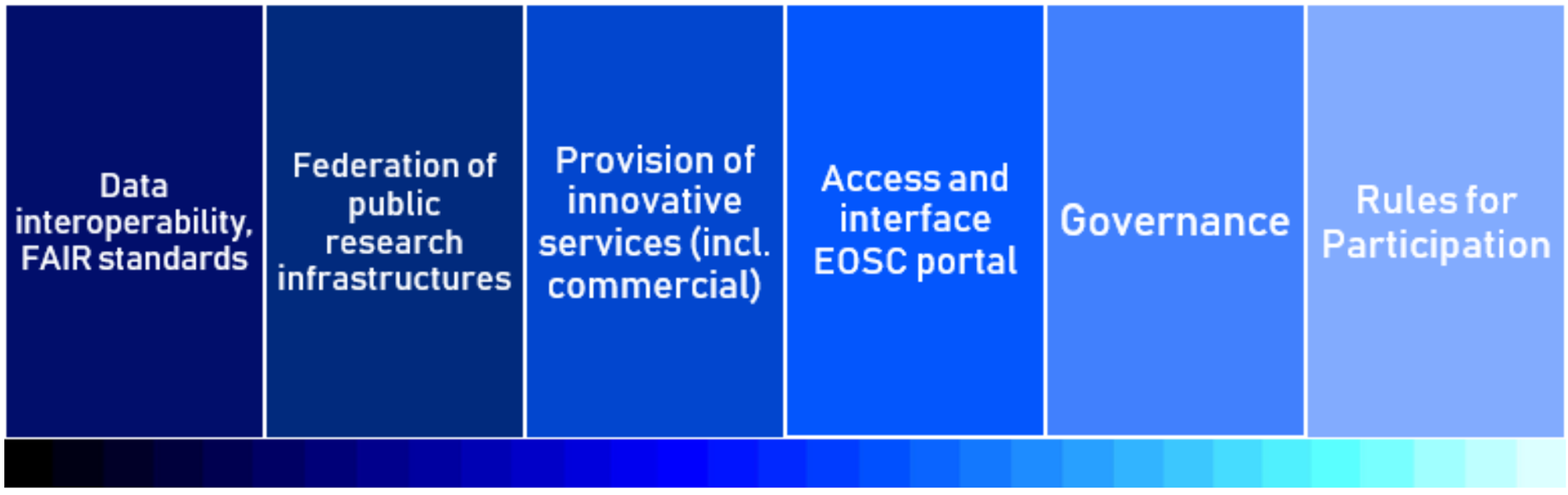
EOSC in numbers

1.9 million researchers and **70 million** professionals will benefit from EOSC

€ 250 million in 2018-2020 (Horizon 2020)

>30 core EU-funded projects implementing EOSC

The [EOSC implementation roadmap], published in 2018 by the Commission, addresses the implementation of EOSC under six action lines:



FAIRSFAIR
Fostering Fair Data Practices in Europe

(5a)

EOSC secretariat eu

Setup and management of the EOSC Secretariat supporting the EOSC Governance

Regional Nodes / Thematic Projects

NI4OS
National Initiatives for Open Science in Europe

EOSC SYNERGY

EOSC-Pillar
Coordination and Harmonisation of National &Thematic Initiatives to support EOSC

EOSC NORDIC

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

FAIR (5C)

Synchronisation Force

FAIRSFAIR
Fostering Fair Data Practices in Europe

EGFC
European Group of FAIR Champions

EOSC Governance Board

EOSC Executive Board

EOSC Working Groups

- Landscape WG
- Sustenaibility WG
- Rules of Participation WG
- Architecture WG
- FAIR WG
- Skills and Training WG

FAIR WG Task Groups

- FAIR practice
- Interoperability
- PIDs
- Metrics and certification

INFRAEOSC-5 Cross Project Collaboration Board (CPCB)

INFRAEOSC-5 Task Forces:

- Landscaping
- FAIR data and Infrastructures
- Services onboarding
- National policies and governance
- Training and skills
- Dissemination and events

EOSC Interest Groups

- Researcher engagement and use cases
- Service and research product catalogue
- Federating core
- Glossary

"Horizontals"

ARCHIVER
ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS

eInfra Central

FREYA

RDA
RESEARCH DATA ALLIANCE

OpenAIRE

OCRE Open Clouds for Research Environments

EOSC-hub

ESFRI Clusters

ENVRI FAIR

panosc
photon and neutron open science cloud

SSHOC
social sciences & humanities open cloud

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

EOSC-Life

Other FAIR Initiatives

FAIR4Health

STM Research Data

GO FAIR

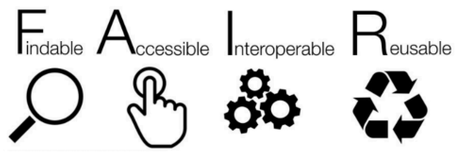
FAIRsharing.org
standards, databases, policies

FAIRplus

Conclusions

FAIR - a good expression for Open Science

- Astronomy is a pioneer of data sharing - we are well placed to make the most of the new opportunities - EOSC



Common challenges include:

- Standardisation at the right level
- Meeting needs of RIs and users - innovations, “*code to data*”
- Critical mass of implementation, and sustainability

Astronomy data sharing - part of a much wider landscape

- benefits from generic components, but need to make sure requirements of astronomy are taken into account

Links

- CDS: https://cds.unistra.fr
- Euro-VO: http://euro-vo.net (coming soon in October 2020)
- IVOA: http://ivoa.net
- ESCAPE: https://projectescape.eu
- EOSC: https://www.eosc-portal.eu/about/eosc