

The FAIR paradigm as a key to Open Astronomical Data

Fabio Pasian and Marco Molinaro

INAF Osservatorio Astronomico di Trieste, Italy

ASTERICS and EOSCpilot projects



20 Nov 2017

The FAIR paradigm as a key to Open Astronomical Data

1



How to implement the Open Universe tools?

To implement a one-stop-shop for the Open Universe two approaches are possible:

- ~~Creation of a archive containing all data to be accessed through OU~~
 - ~~Data preservation and curation problems ! (expertise in data quality, updates)~~
 - ~~Cost !!~~
- Creation of a mechanism allowing access to a distributed set of archives containing data to be shared through the OU
 - Such a mechanism needs to be compliant to existing standards and best practices accepted worldwide

Open Universe and FAIR data

- Experts Meeting (Rome, 11-12 April 2017) recommended:
 - “For the researcher group – the Open Universe Initiative should work towards: (a) disseminating and promoting existing standards and best practices”
 - “promote among data providers the adoption of the FAIR (Findable, Accessible, Interoperable, Reusable) guiding principles”
- These recommendations apply particularly to the practical implementation of the OU
 - to allow the data to be fetched from the participating archives
 - to allow data from different sources to be comparable within the OU apps

In other science communities ...

... there is a political, legal and technological debate around re-use and reproducibility of research data. Funding agencies, research institutions, coordinators of research infrastructures, individual researchers, data scientists, data librarians / curators and computing scientists are striving to identify which of the FAIR guidelines can be effectively supported by the current technologies, which national policies are supporting their implementation and what are the challenges still open towards a concrete implementation of the FAIR principles.

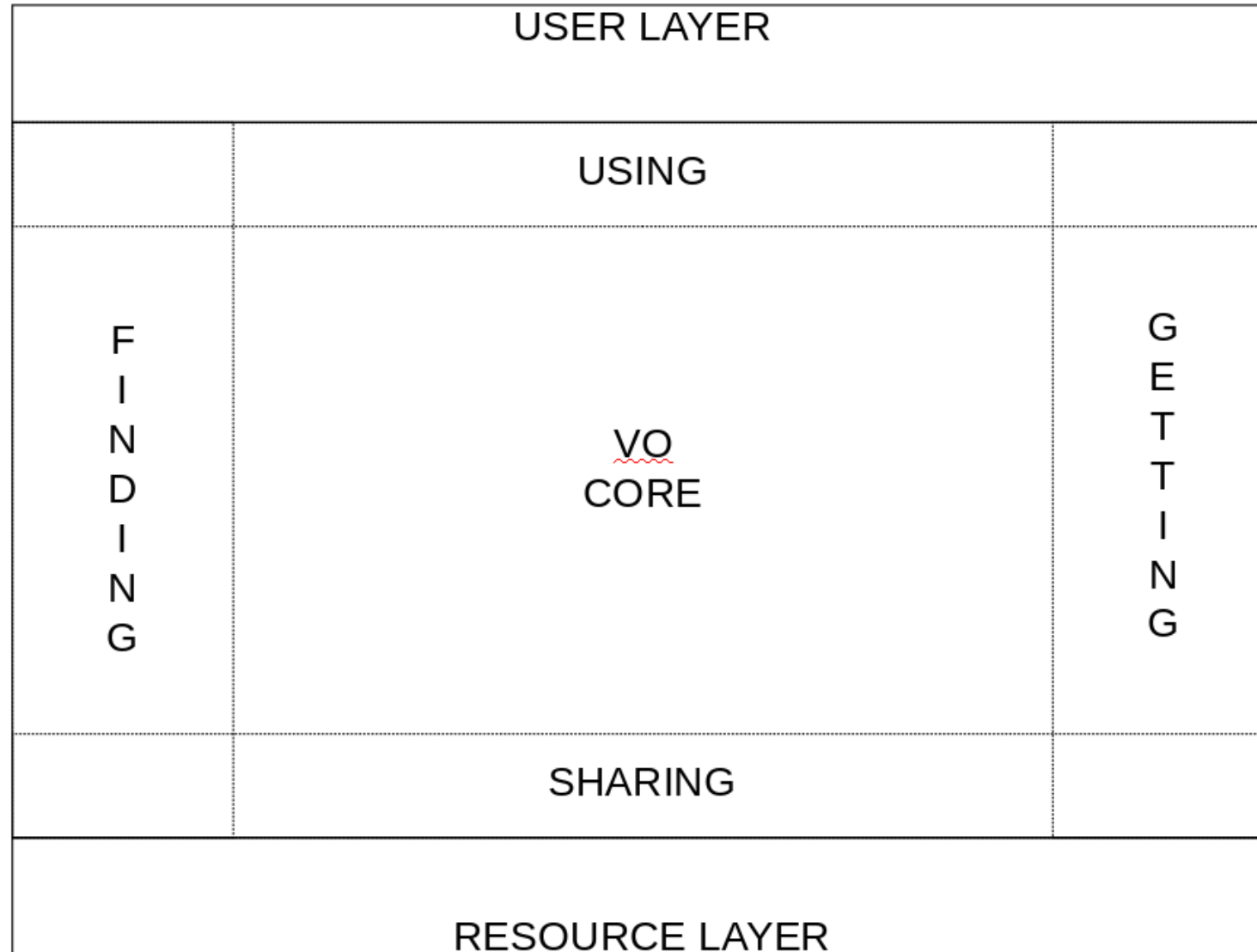
*(from the introduction to the RDA Workshop “FAIR data management: best practices and open issues”)
Firenze, 14-15 November 2016*

LEVEL 0

USERS



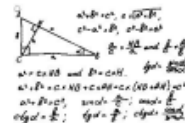
COMPUTERS



IVOA Architecture



PROVIDERS



FAIR principles in Astronomy: the VO (I)



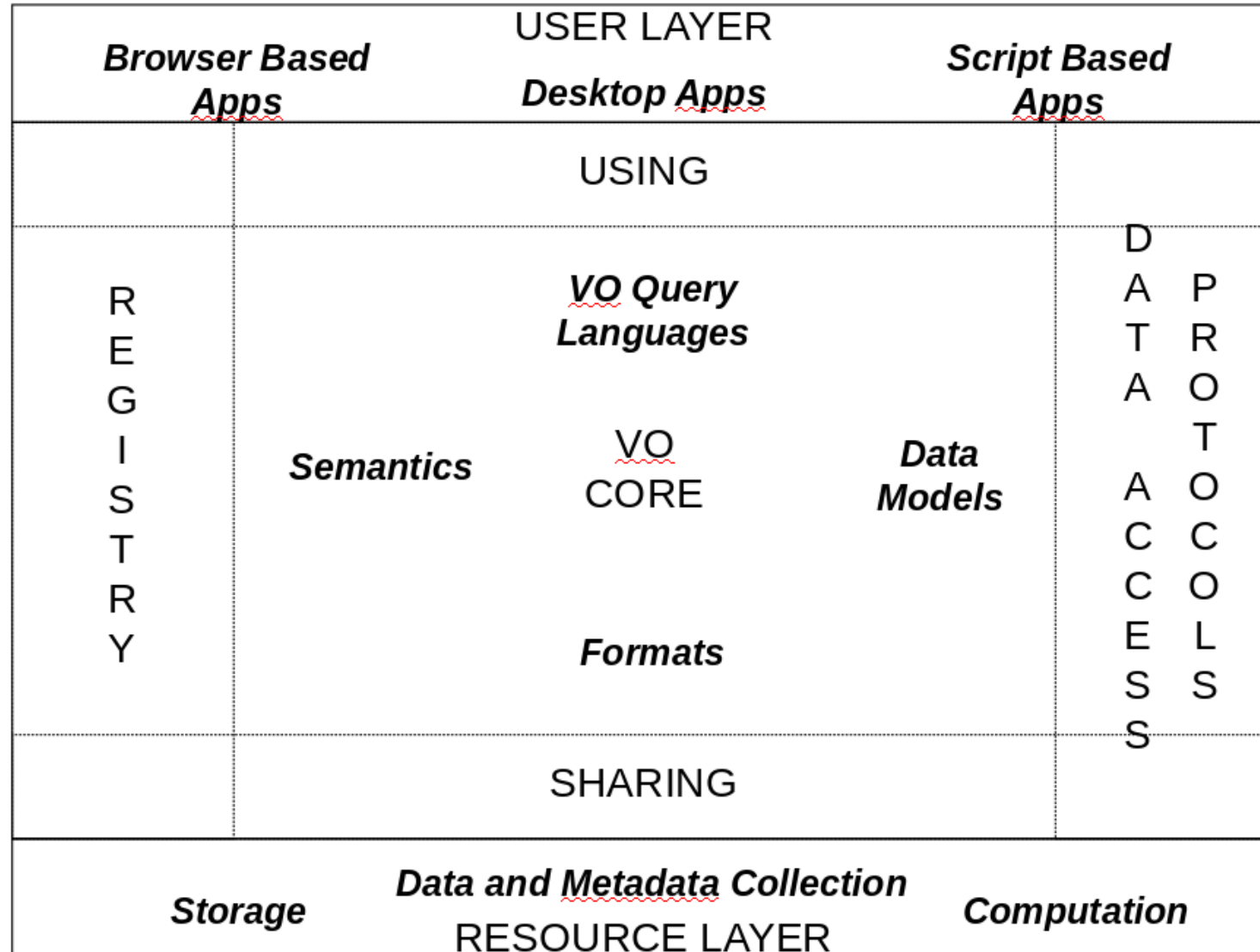
LEVEL 1

USERS

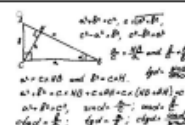


COMPUTERS

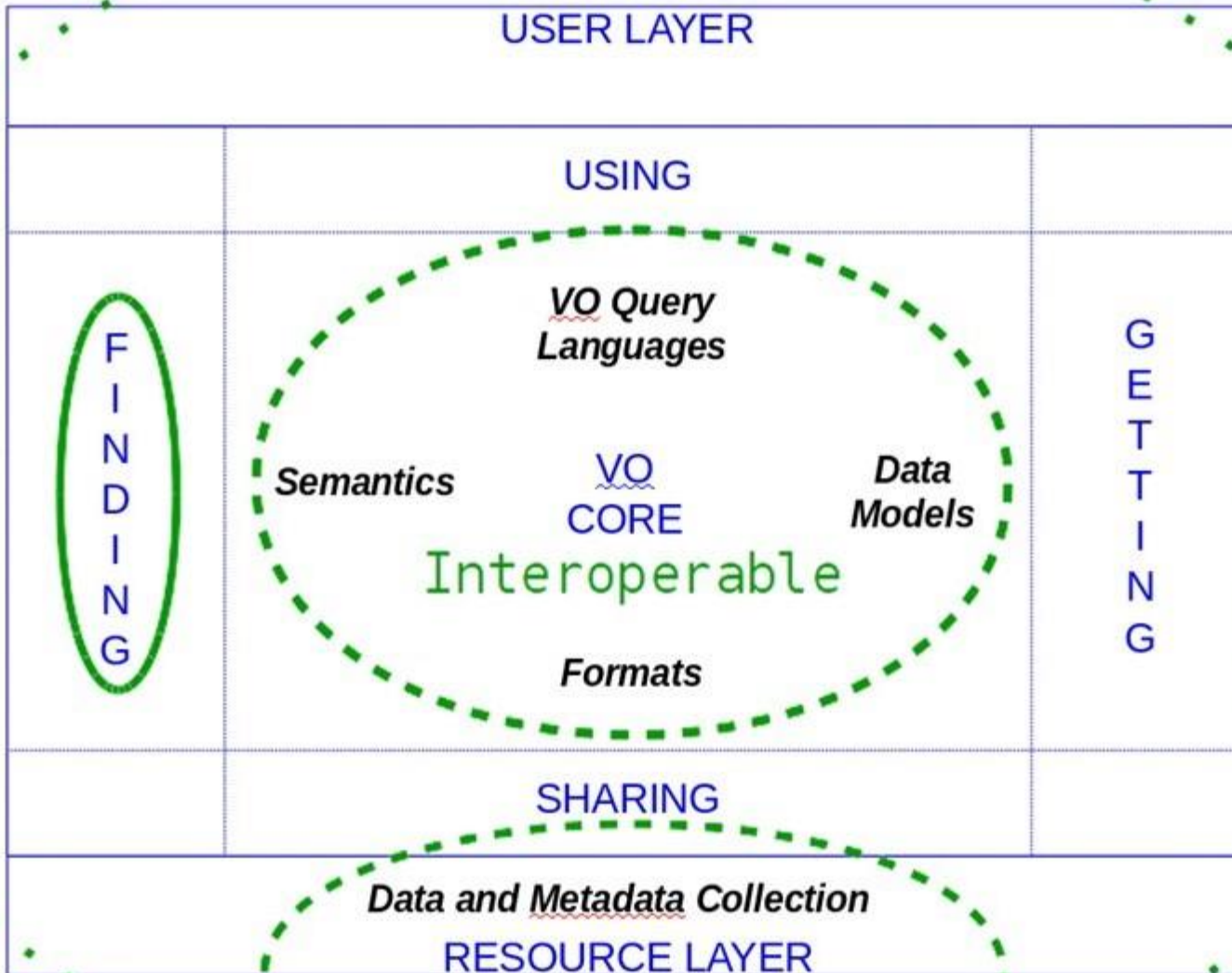
FAIR principles in Astronomy: the VO (II)



PROVIDERS



LEVEL 0
LEVEL 1



FAIR principles in Astronomy: the VO (III)

LEVEL 2
All standards

USERS

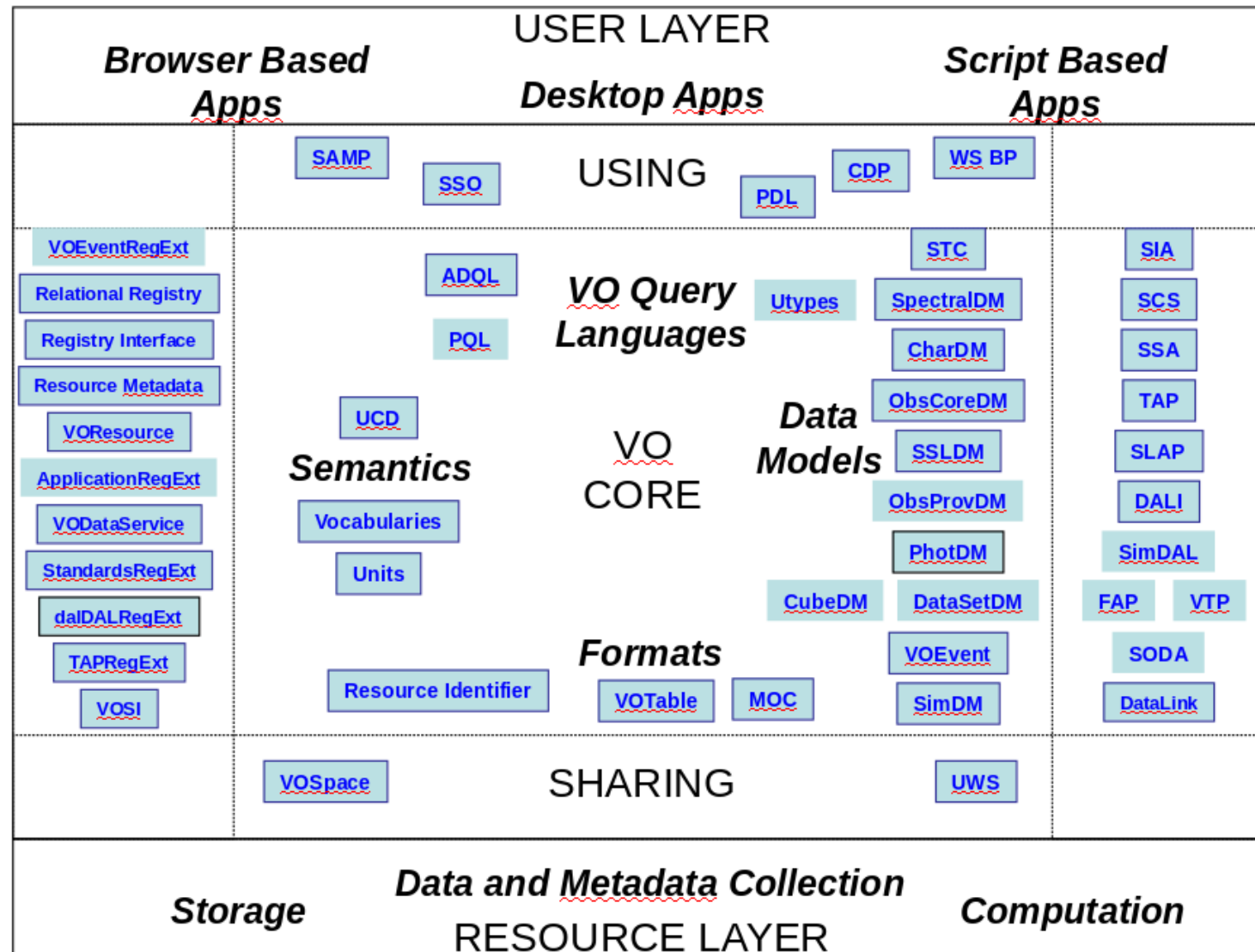


COMPUTERS

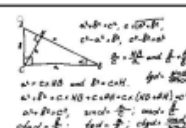
FAIR principles in Astronomy: the VO (IV)

R
E
G
I
S
T
R
Y

D
A
T
A
P
R
O
T
O
C
O
L
S
A
C
C
E
S
S



PROVIDERS



| | | |
|------|---|---|
| F1 | (meta)data are assigned a globally unique and eternally persistent identifier | Resource Identifiers, Registry |
| F2 | data are described with rich metadata | VOResource & extensions, Semantics, Data Model |
| F3 | (meta)data are registered or indexed in a searchable resource | Registry Interfaces, Relational Registry, Resource Identifiers |
| F4 | metadata specify the data identifier | VOResource, Resource Identifiers, Registry |
| A1 | (meta)data are retrievable by their identifier using a standardized communications protocol | All the Data Access Layer Protocols |
| A1.1 | the protocol is open, free, and universally implementable | VO standards and protocols are public, open and implementation agnostic |
| A1.2 | the protocol allows for an authentication and authorization procedure, where necessary | Single Sign-On, Credential delegation Protocol |
| A2 | metadata are accessible, even when the data are no longer available | VO Standard Interfaces, Registry |
| I1 | (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation | VO-DML, Data Models |
| I2 | (meta)data use vocabularies that follow FAIR principles | VO Vocabularies, Semantics, Standards Registry Extension |
| I3 | (meta)data include qualified references to other (meta)data | Data Models, Semantics, Unified Content Descriptors |
| R1 | meta(data) have a plurality of accurate and relevant attributes | VOResource, Data Access Layer Interface, VO Standard Interfaces, VOTable format |
| R1.1 | (meta)data are released with a clear and accessible data usage license | Registry annotation, generally public |
| R1.2 | (meta)data are associated with their provenance | Provenance Data Model (and protocols) |
| R1.3 | (meta)data meet domain-relevant community standards | IVOA enabled data providers |

Are all issues solved ?

Data compliant with VO standards are FAIR

However:

- Not all of the data are VO-compliant
- FAIR compliance is not enough!

Not all of the data are VO-compliant

- To achieve data FAIRness, conversion layers need to be built for every non-VO-compliant data archive (**BIG development effort for OU**)
1. IVOA and data providers shall make efforts to **jointly develop new VO standards** to deal with new classes of data

This already happening with most of the new large projects (in Europe through ASTERICS) and the IVOA Committee on Science Priorities

FAIR compliance is not enough! (I)

- To be effective for OU purposes (for many science and EPO activities) some **computing services** need to be provided by data centers
 - From simple things like image cutouts → catalogue building → machine learning → ...
 - Need for a change: from data providers to science platform providers
- 2. Data providers shall make efforts to **provide computing services** and, with IVOA, **jointly develop standards** to access these new combined services (included in IVOA Registry)

FAIR compliance is not enough! (II)

- Astronomy is an observational science → any area of the sky not observed in a certain moment in time is “missing data”
 - “Time-dependent data should be kept forever” (G.Rossi, ESFRI Chair, 15 Nov.2017)
 - This is not required by FAIR principles (A2 is about metadata)
3. Data providers shall make efforts to **guarantee preservation** of their data over the **longest possible** period of time (include this in their Data Management Plans)

Conclusions – What can UNOOSA do?

To allow the Open Universe initiative to be technically and scientifically feasible, UNOOSA shall

- A. promote among data providers
 - a) long-term preservation policies
 - b) the adoption of IVOA standards
- B. foster common work between IVOA and data providers for the development of new standards wherever needed (including computing as well), with the support of other interested parties



Thank you for your attention

The European Commission is supporting ASTERICS (grant no. 653477) and EOSCpilot (grant no. 739563) in the framework of the Horizon2020 research and innovation programme.