

Application of FAIR Principles in Astronomy

by G. Bruce Berriman

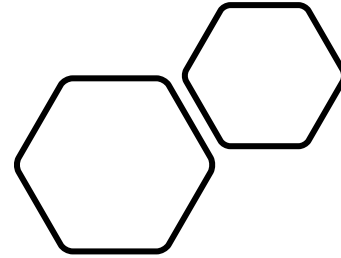
Making the Major Facilities Data Lifecycle FAIR to Provide AI-Ready Data

March 1, 2022

CI Compass Cyberinfrastructure for NSF Major Facilities Workshop

Application of FAIR Principles in Astronomy

NSF Cyberinfrastructure for
Major Facilities Workshop
(March 2022)



G. Bruce Berriman

(Caltech/IPAC-NExSci)

Chair, IVOA Exec Committee, May
2021 – Oct 2022.



Caltech

JPL

Special Thanks

Simon O'Toole

(Australian All-Sky Virtual Observatory)

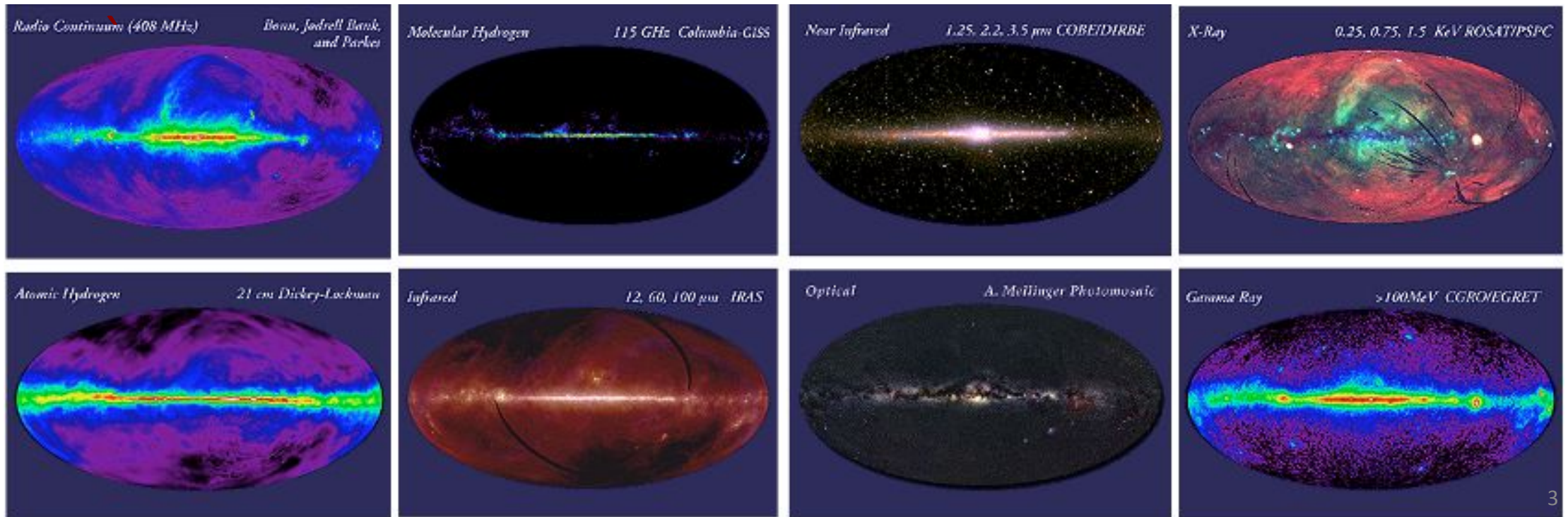
Raffaele D'Abrusco

(Harvard-Smithsonian Center for Astrophysics)

What Is The Virtual Observatory?

“A multi-wavelength digital sky that can be searched, visualized, and analyzed in new and innovative ways.”

- *Pepi Fabbiano*

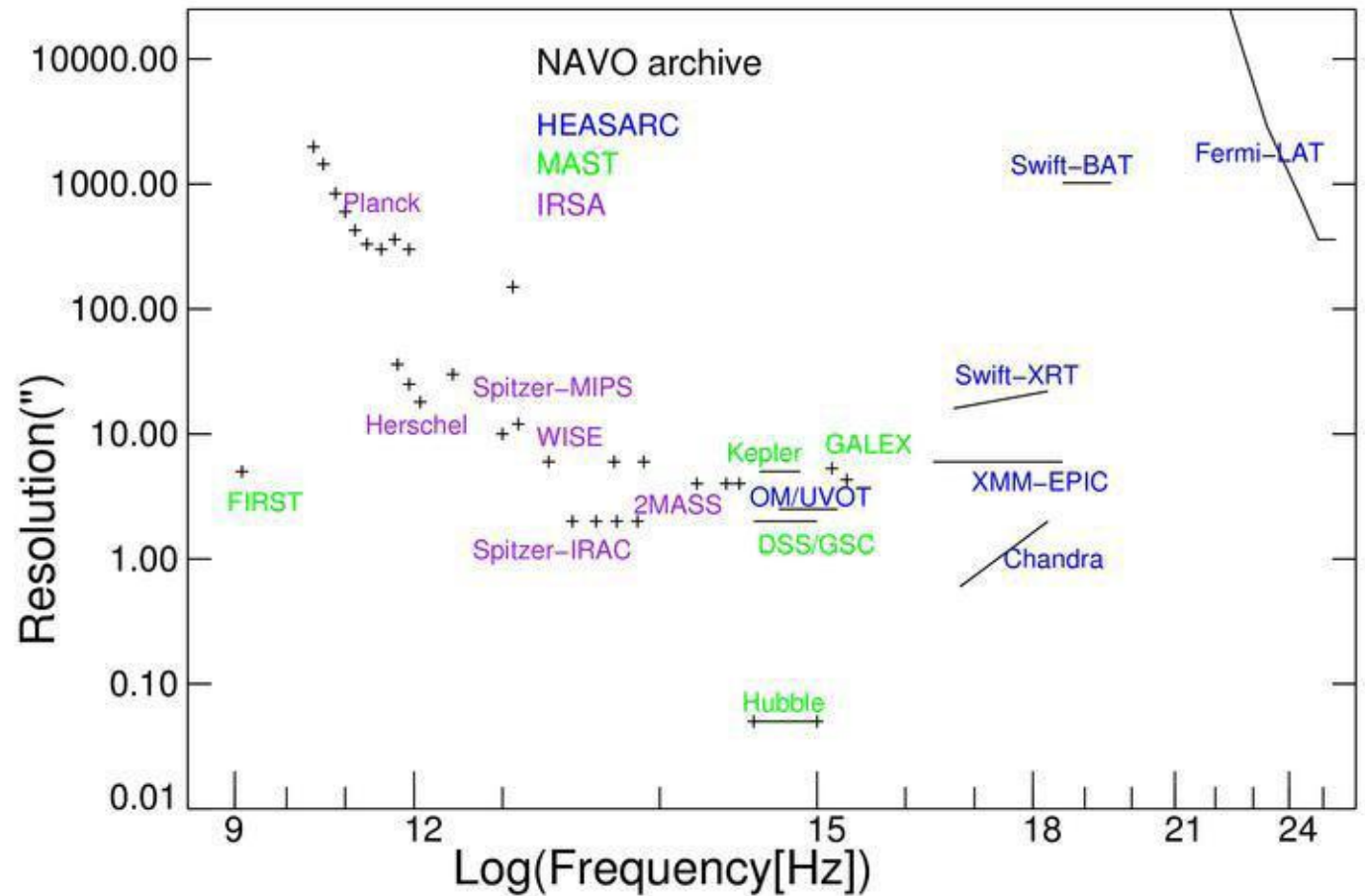


The International Virtual Observatory Alliance: How The VO Gets Built

- Data are in archives distributed worldwide → define standards that enable archives to interoperate seamlessly.
- The International Virtual Observatory Alliance (IVOA) is the international body that defines these standards (<https://ivoa.net/>)
- Founded in 2002, the IVOA today has 22 national VO member projects and one IGO.
- Goals from the outset was to enable seamless **interoperability** of **open** data and services -> **Implementing FAIR principles before they were formalized as such by Hutchinson et al. 2016**
- Implementing IVOA standards makes your data *almost* FAIR.

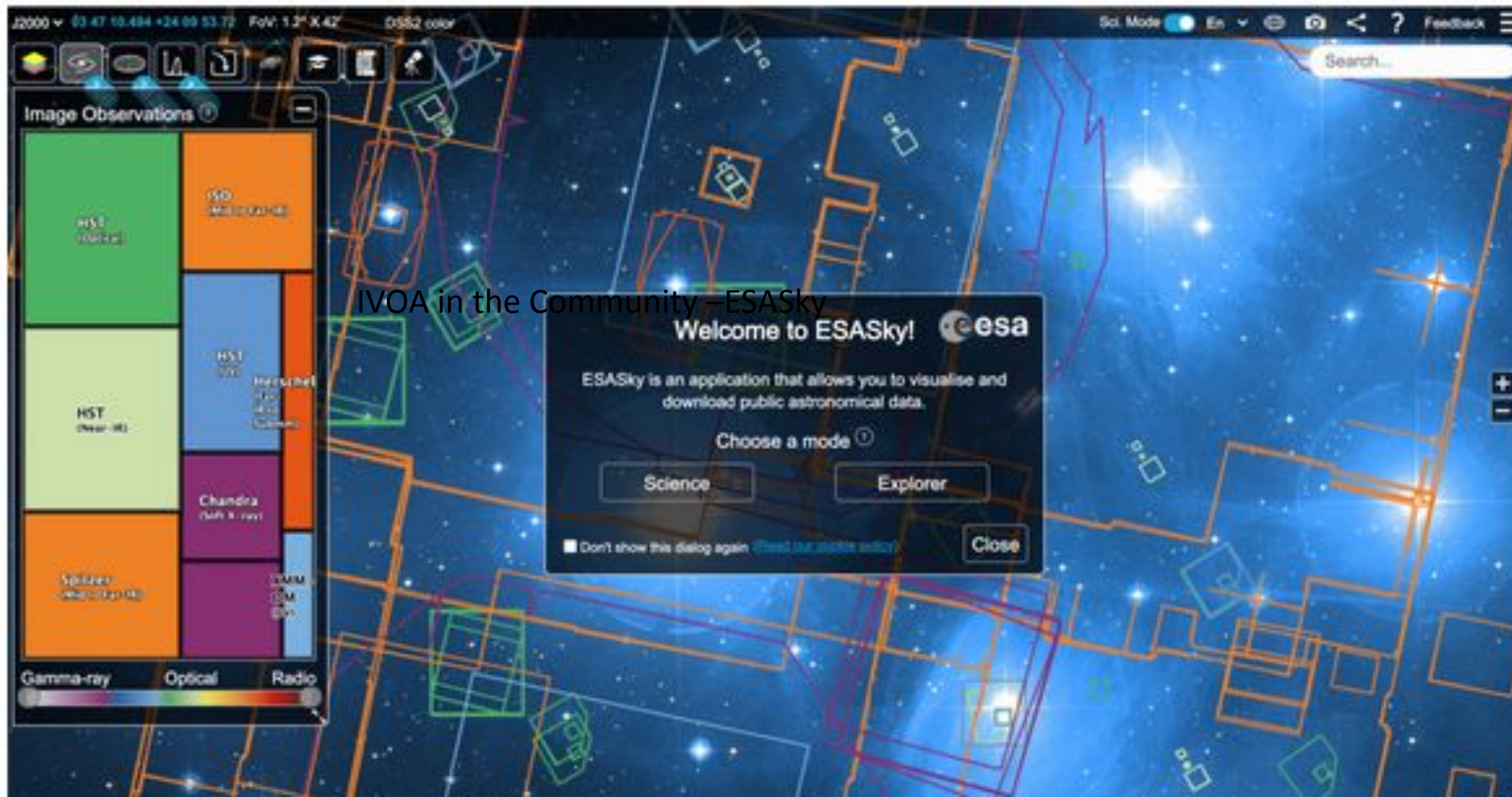
The VO In Action

NASA archives have implemented IVOA standards across distributed archives of heterogeneous data over 15 decades of frequency



- Access to all of these data through common set of machine-based APIs.
- First such large-scale *distributed* implementation of VO standards.

VO protocols Are At The Heart of ESASky

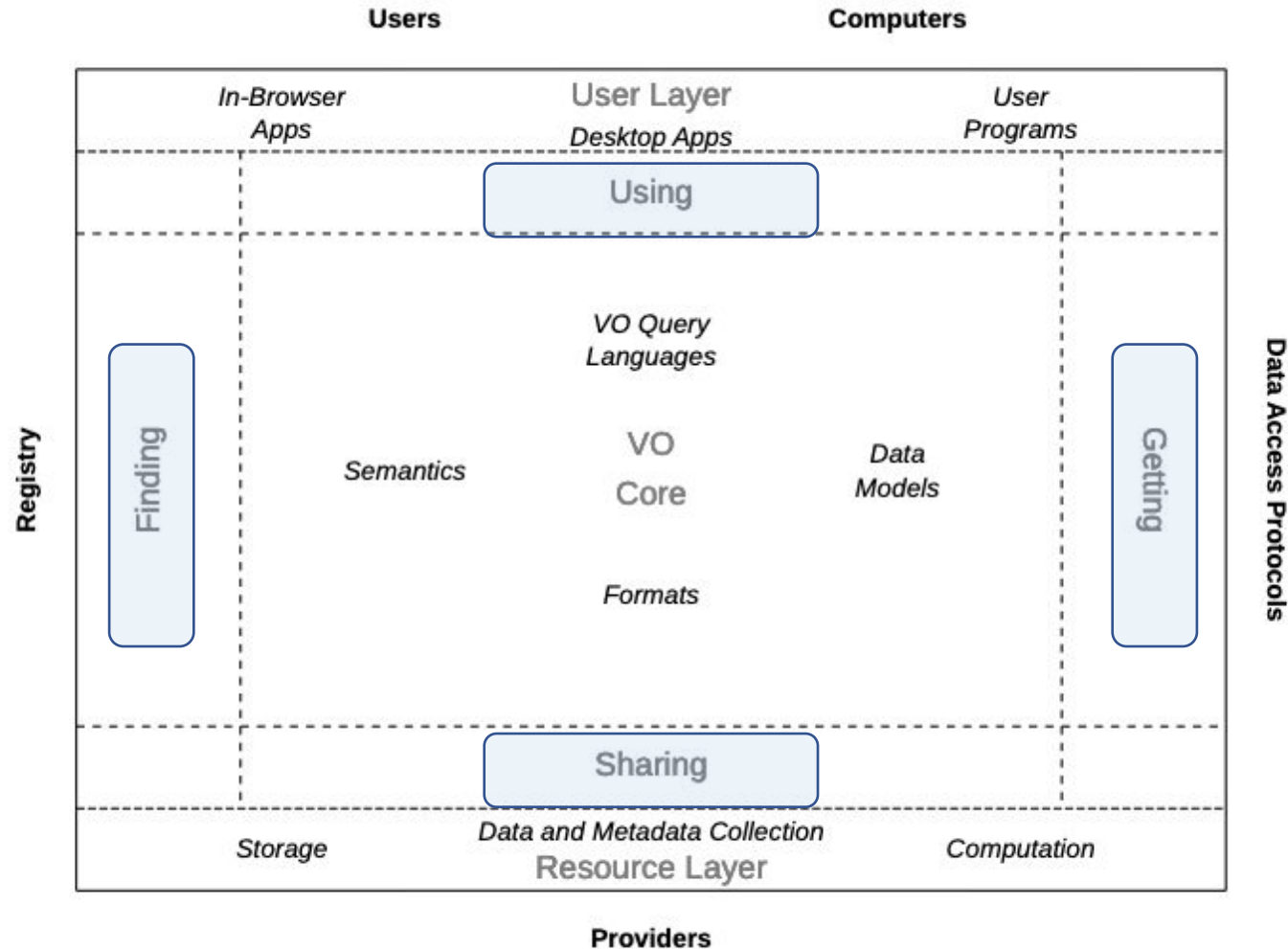


IVOA in the Community – ESASky

FAIR Principles and The IVOA

- The IVOA Architecture has strong parallels with the FAIR framework.
 - Focus on processes to move metadata and data through architecture rather than properties of the service or data
- To a large degree, implementing IVOA standards goes a long way to implementing FAIR principles.
- There are a few specifics that IVOA standards do not provide (out of scope or we are just implementing)

IVOA Architecture and FAIR Principles



IVOA Standards are (Mainly) Findable

- F1. (meta)data are assigned a globally unique and persistent identifier
 - F2. data are described with rich metadata (defined by R1 below)
 - F3. metadata clearly and explicitly include the identifier of the data it describes
 - F4. (meta)data are registered or indexed in a searchable resource
-
- F1. IVOIDS are not citable
 - F4. IVOA standards do not require that the data identifier be returned in all cases
 - Left up to service providers

IVOA Standards are (Mainly) Accessible

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available
- A2. IVOA standards have no rules about reliability and longevity.
- Again, left up to service provider
- IVOA monitors services and provides regular weather reports on availability

IVOA Standards are (Mainly) Interoperable

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
 - I2. (meta)data use vocabularies that follow FAIR principles
 - I3. (meta)data include qualified references to other (meta)data
-
- IVOA standards implement vocabularies and in particular Universal Content Descriptors that mean I1, I2, I3 are all met but ...
 - ... **Strictly: Need a cross-referencing framework between data that describe the same object or phenomena - in development.**

IVOA Standards are (Mainly) Re-usable

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards.
-
- R1.1 IVOA has no specifications about licenses - this is up to the provider
 - R1.2 There is an IVOA Provenance Data Model was recently published and not all providers implement it.

FAIR, Machines & IVOA

- FAIR principles emphasize **machine-actionability**, a prerequisite to develop smart data discovery & computational agents based on ML/AI techniques
- Data need to be FAIR, but it's not enough for ML/AI: FAIRness addresses intrinsic features of datasets, but what about quality? Ethics? Privacy? And in particular, what about their **utility for a given problem?**
- FAIR principles allow zero-order ML: find usually elusive patterns in datasets. But what about connecting patterns to higher order abstractions, like humans do? (even Deep Learning is only *faking* comprehension).
- Utility of data is context- and field-dependent, and its decoding by machine requires that **semantics** is added to the data.
- In Astronomy, **IVOA is the natural venue to augment the FAIR framework to allow ML/AI agents to extract semantic information from data and perform decision based on such information.**

Bedtime Entertainment

- Simon O'Toole. Invited presentation at ADASS XXX (Nov 2021) “Fair Standards for Astronomical Data.” <https://youtu.be/IBzGBEWF7Rs>
- IVOA Architecture Document.
<https://ivoa.net/documents/IVOAArchitecture/20211101/EN-IVOAArchitecture-2.0-20211029.pdf>