Tecnologie VO
potenzialità e problematiche

Marco Molinaro
INAF-OATs
IA2

La missione e le prospettive scientifiche di TNG nell'astrofisica del 2020

Museo Diocesano di Padova
1-3 Marzo 2017

outline

- Open Science & Virtual Observatory
- IA2 & TNG use case
- Status and …

Open Science

Open Science involves transitioning from a system in which it is difficult to access and locate the results of scientific research to one that openly distributes results to all kinds of end users [...]

(Open Science Conference, EU 2016)

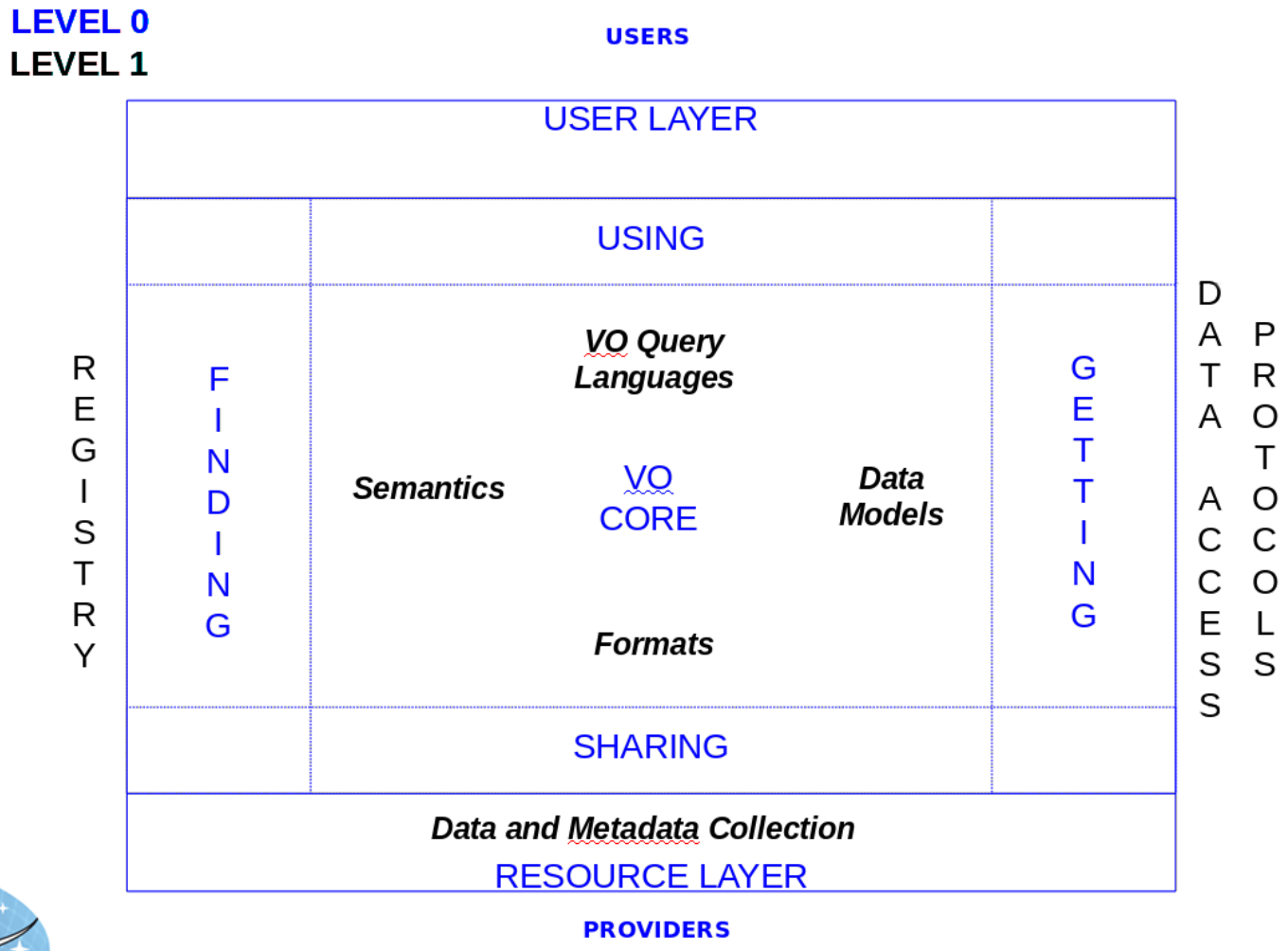
Data should be

- **F indable**
  metadata, persistent identifiers, indexing
- **A ccessible**
  standard communication protocols, accessible metadata
- **I nteroperable**
  common metadata modelling, FAIR vocabularies, metadata referencing
- **R e-usable**
  provenance, licensing, domain-relevant standards

(force11.org/fairprinciples)

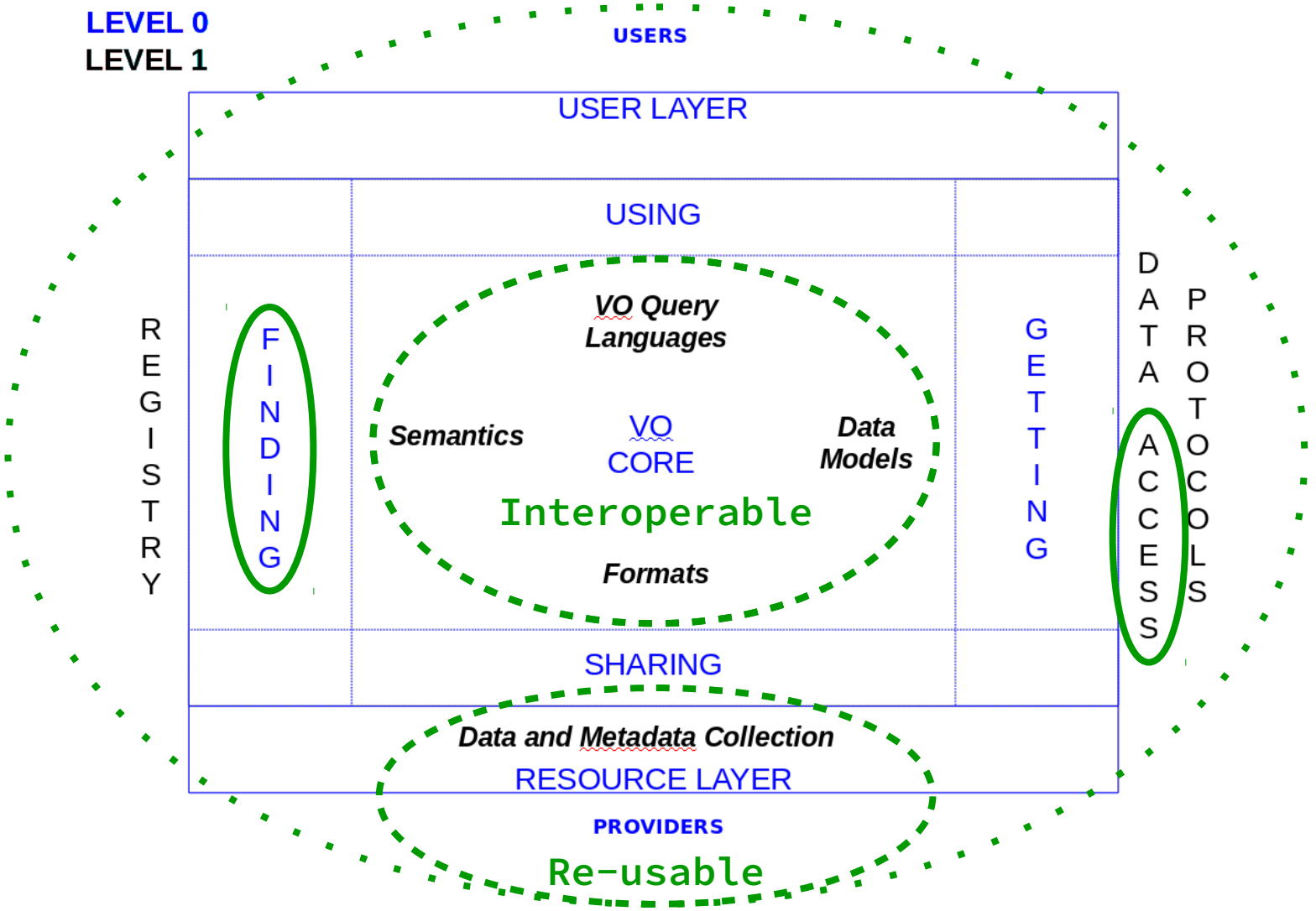
METADATA

Virtual Observatory - Architecture

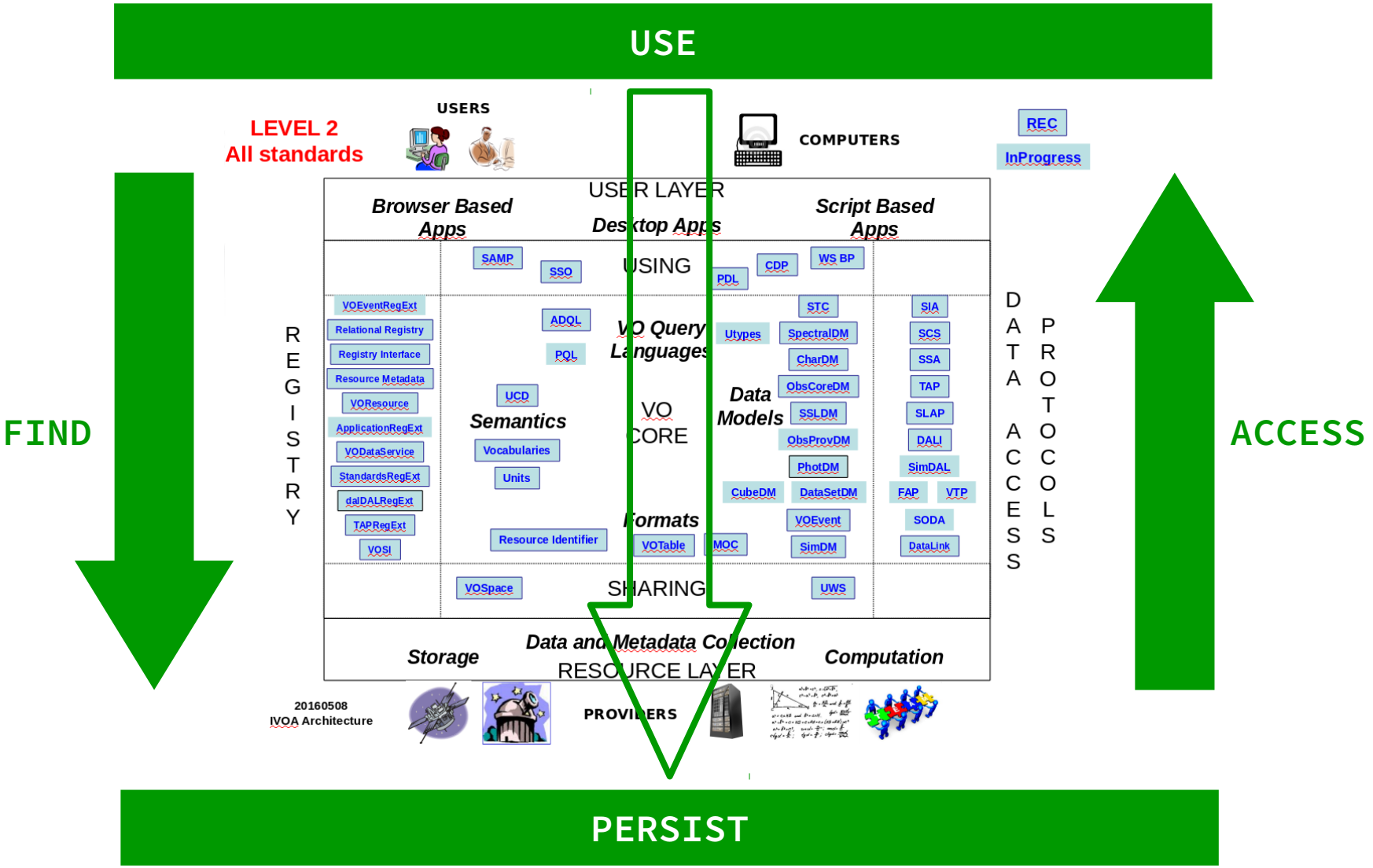


(IVOA Architecture, http://www.ivoa.net/documents/Notes/IVOAArchitecture/)

VO – FAIR analogy



Virtual Observatory framework



Virtual Observatory at work

- **Observations are stored as Resources**
  - Data Collections (e.g. obs made using the same instrument)
  - Data Sets: atomic searchable data (e.g. a FITS image)
  - Table Sets: e.g. for catalogue purposes
- **Services are also stored as Resources**
  - Attached to Data Collections and/or Table Sets
- **Collections and Services are linked and findable**
- **Data Sets and Table Sets are metadata annotated to allow usability and interoperability**
- **Metadata annotation is intended both for**
  - Discovery (findable) purposes
  - Modelling of content (interoperability), including semantics

DataSet
Metadata
Annotation

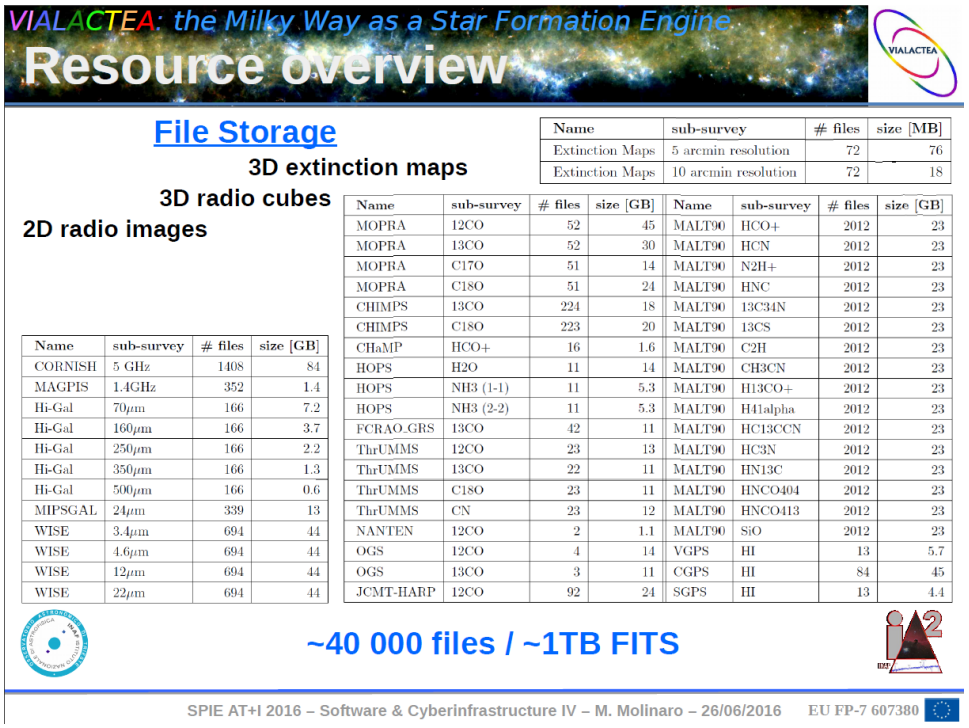
IA2 – Data Provider

- **Italian center for Astronomical Archives (IA2)**
  - Serves data for: TNG, LBT, Asiago, …
    - Attend Friday's talk from C. Knapic to have better details on IA2 archive facilities
  - Provides archival and services for ICT
- **One of its missions is provide VO support to the community**
  - Hosting VO oriented data services
  - Developing tools to facilitate this goal
  - Offering support in VO engagement
    - All of these under the VObs.it umbrella (national project member of the IVOA)
- **Since TNG main archives sits at IA2, this should be the place to provide TNG data as a set of VO resources**

IA2 VO tools & services

- **IA2 developed tool to serve “simple” VO protocols**
  - Software
    - VO-Dance, IA2TAP, TAP_SCHEMA Manager, VAPE
  - Services
    - cone searches, image access, spectra access for
      - WINGS, VIPERS, TNG, TIRGO, Planck ERCSC
- **The knowledge gained has been re-used not only for internal usage, but also for external or hosted projects**
  - VIALACTEA, SpaceInn

The VIALACTEA experience



VIALACTEA: the Milky Way as a Star Formation Engine.

Resource overview

File Storage

3D extinction maps

Name	sub-survey	# files	size [MB]
Extinction Maps	5 arcmin resolution	72	76
Extinction Maps	10 arcmin resolution	72	18

3D radio cubes

Name	sub-survey	# files	size [GB]
MOPRA	12CO	52	45
MOPRA	13CO	52	30
MOPRA	C17O	51	14
MOPRA	C18O	51	24
CHIMPS	13CO	224	18
CHIMPS	C18O	223	20
CHaMP	HCO+	16	1.6
HOPS	H2O	11	14
HOPS	NH3 (1-1)	11	5.3
HOPS	NH3 (2-2)	11	5.3
FCRAO_GRS	13CO	42	11
ThrUMMS	12CO	23	13
ThrUMMS	13CO	22	11
ThrUMMS	C18O	23	11
ThrUMMS	CN	23	12
NANTEN	12CO	2	1.1
OGS	12CO	4	14
OGS	13CO	3	11
JCMT-HARP	12CO	92	24

Name	sub-survey	# files	size [GB]
MALT90	HCO+	2012	23
MALT90	HCN	2012	23
MALT90	N2H+	2012	23
MALT90	HNC	2012	23
MALT90	13C34N	2012	23
MALT90	13CS	2012	23
MALT90	C2H	2012	23
MALT90	CH3CN	2012	23
MALT90	H13CO+	2012	23
MALT90	H41alpha	2012	23
MALT90	HC13CCN	2012	23
MALT90	HC3N	2012	23
MALT90	HN13C	2012	23
MALT90	HNCO404	2012	23
MALT90	HNCO413	2012	23
MALT90	SiO	2012	23
VGPS	HI	13	5.7
CGPS	HI	84	45
SGPS	HI	13	4.4

2D radio images

Name	sub-survey	# files	size [GB]
CORNISH	5 GHz	1408	84
MAGPIS	1.4GHz	352	1.4
Hi-Gal	70μm	166	7.2
Hi-Gal	160μm	166	3.7
Hi-Gal	250μm	166	2.2
Hi-Gal	350μm	166	1.3
Hi-Gal	500μm	166	0.6
MIPSGAL	24μm	339	13
WISE	3.4μm	694	44
WISE	4.6μm	694	44
WISE	12μm	694	44
WISE	22μm	694	44

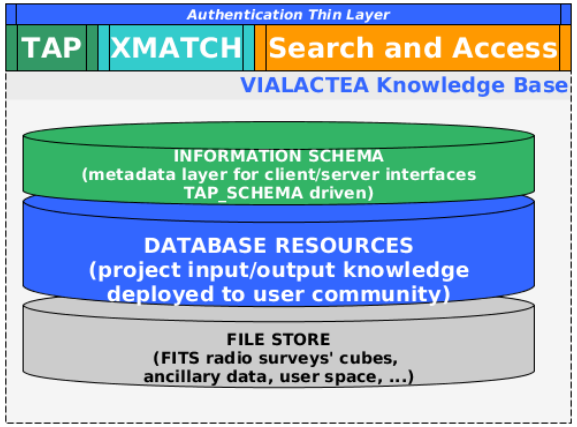
~40 000 files / ~1TB FITS

SPIE AT+I 2016 – Software & Cyberinfrastructure IV – M. Molinaro – 26/06/2016 EU FP-7 607380

- Heterogeneous datasets put together in a common “find, access, cut” scenario
- Work made side by side with scientific community driving the project
- VO kept as a guideline
  - No resources enough to chase compliance

The VIALACTEA experience

- VO middle layer based upon TAP protocol
- Re-use of MOC/HEALPix experience
- Direct use of WCS to automate search algorithm and cut solutions



	SEARCH	CUTOUT	MERGE	VALUES	DEFAULT
surveyname	√		√	<surveys table>	NONE
species	√		√	<surveys table>	NONE
transition	√		√	<surveys table>	NONE
pubdid		√		<search result provided>	NONE
skysystem	√	√	√	GALACTIC, ICRS	GALACTIC
l,b	√	√	√	0/360, -90/+90 [deg]	**MANDATORY**
r	√	√	√	0/2 [deg]	0 [deg]
dl,db	√	√	√	0/2, 0/2 [deg]	0, 0 [deg]
vl,vu	√	√	√	<dataset depending>	<full available range>
nullvals		√		flag key	not present

But…

- **Main products available at IA2 archive for TNG are raw data**
  - Makes it difficult to go beyond basic services
    - Only 3 currently exist
  - Metadata included in FITS headers is not always optimal to allow consistent servicing
    - data, being raw and missing some metadata becomes unused or non-findable on the long term
    - Persistence, made out of simple storage preservation, is hardly useful

What can be done

- **Improve data policy on machine readable services**
  - Will allow resource generation also for proprietary data
- **Connect better to science-ready data brewers**
  - Will improve metadata annotations
    - Mandatory for real FAIR science publishing
- **Don’t disregard bringing use cases to the VO**
  - Multi-messenger astronomy is not done on 1 instrument
  - Better standards and research framework comes only out of collaborative efforts

A little example

- Time Domain is currently a hot topic in IVOA
- GAPS produces Times Series out of HARPS-N observations
- Starting out of already public GAPS time series we are going to contribute to the standardization effort of discovery and access of time series data

Conclusions

- Open Science is a FAIR guideline to collaborative science fostering; VO fairly follows it
- Consider TNG observational results into a broader scenario where archival data is as important as freshly minted observations
- And …

Plea: DON'T!

The Archive

There is no driver for an archive because funding comes from people who write proposals. For those, the archive is only a way to provide data to the PI.

(freely taken from BoF-1, ADASS XXVI Proceedings, Grange & al.)