ASTERICS and EOSCpilot

Connections between astrophysics and the European cloud

Fabio Pasian and Marco Molinaro

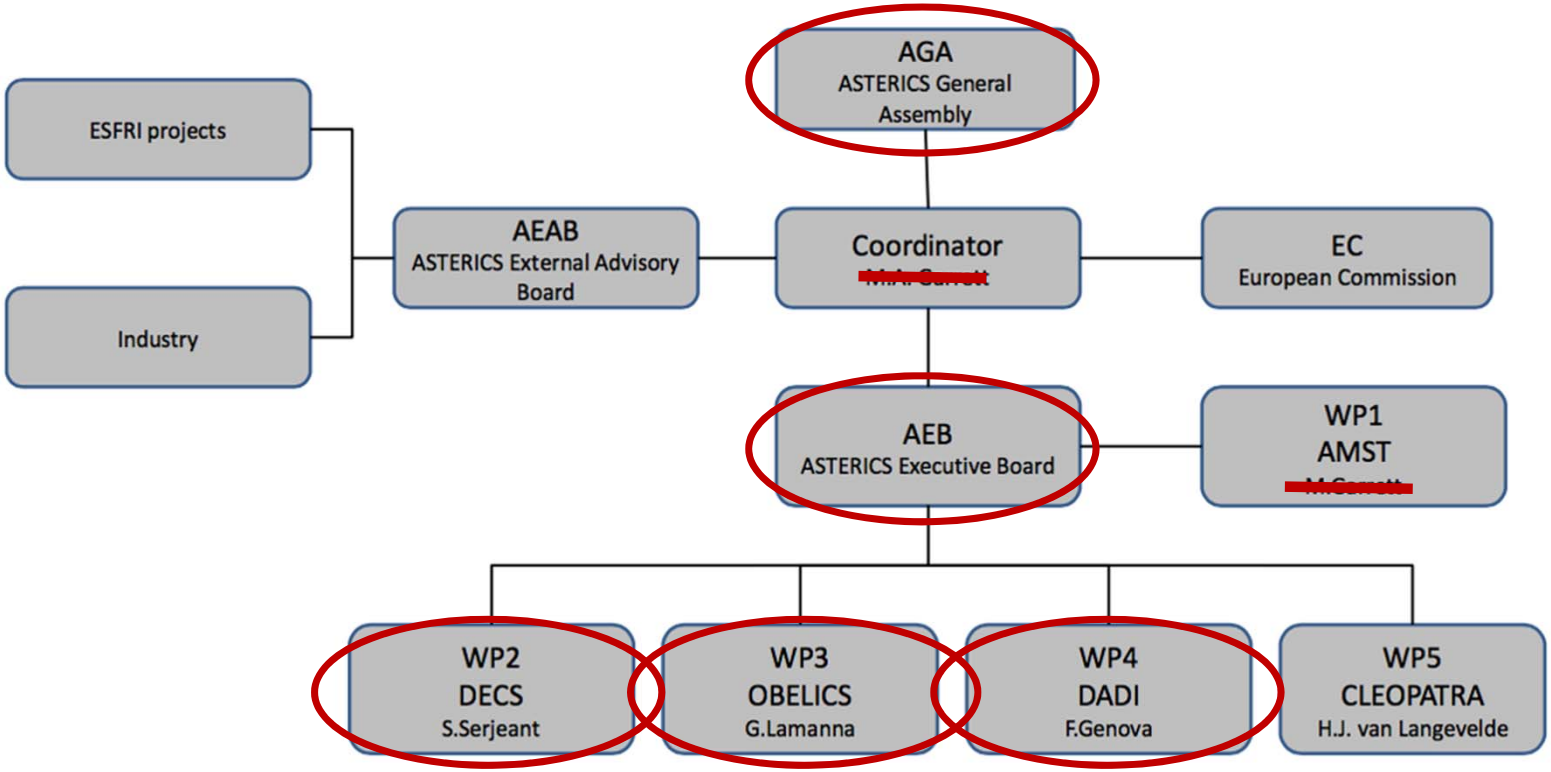
INAF Osservatorio Astronomico di Trieste, Italy

ASTERICS and EOSCpilot projects

ASTERICS

- H2020 project, funded at €15 million for 4 years, started 1 May 2015
- 22 partners, led by ASTRON
- Cluster (CTA, SKA, KM3Net, precusors, world-class facilities + ELT, EST)
- Bringing together the astronomy, astrophysics and astroparticle physics communities on a European scale for the first time
- Programme closely modelled on ASTRONET priorities → aims to tackle common problems, solutions, and cross-facility synergies
- Strong focus on “interoperability” - enabling multi-messenger / multi-wavelength astronomy
- Best practice platforms, interfaces, education, impact monitoring

Governance & Management



WP2 – DECS

- Dissemination & public engagement
- Audiences: scientific & technical communities, academia, private industry, other public research centres, SMEs, policy makers, general public
- Open ESFRI facilities to wider stakeholders through citizen science (Science 2.0)
- Coordinated citizen science experiments to open ESFRIs & pathfinders/precursors to public
- Educational resources & efficacy metrics

WP3 – OBELICS

- Software interoperability; establishing open standards and software libraries
- Training in parallel programming and big data frameworks
- Adapt & optimise extremely large database systems for ESFRIs
- Demonstrate data integration across ESFRI & pathfinder projects using data mining tools & statistical analysis techniques on petabyte data sets

WP4 – DADI

- Train & support ESFRI staff in Virtual Observatory
- Train & support wider astronomical community in scientific use of VO framework in particular for pathfinder data, & gather their requirements & feedback
- Adapt VO framework & tools to ESFRI project needs → extend IVOA standards

WP5 – CLEOPATRA

- Technology development for fibre connectors; relaying alerts; data streaming software; data dissemination (including ALMA & E-ELT); advanced scheduling algorithms
- Builds on WRE (White Rabbit Ethernet) and the EC DG-EXPReS/NEXPRES projects

INAF in ASTERICS (I)

- Italian participation: INAF (third-largest partner 1.46 M€) + INFN (240 K€)
- Chair of General Assembly participates in Executive Board (F. Pasian)
- DECS (ref. M. Ramella) with OATs:
  - T2.4* - co-ordinate translation and adaption of MPE resources, and testing in educational environment
- DADI (ref. M. Molinaro) with OATs:
  - T 4.1* - Support to astronomy ESFRI projects
  - T 4.2 - Support to the astronomical community
  - T 4.3 - Updates of the VO framework

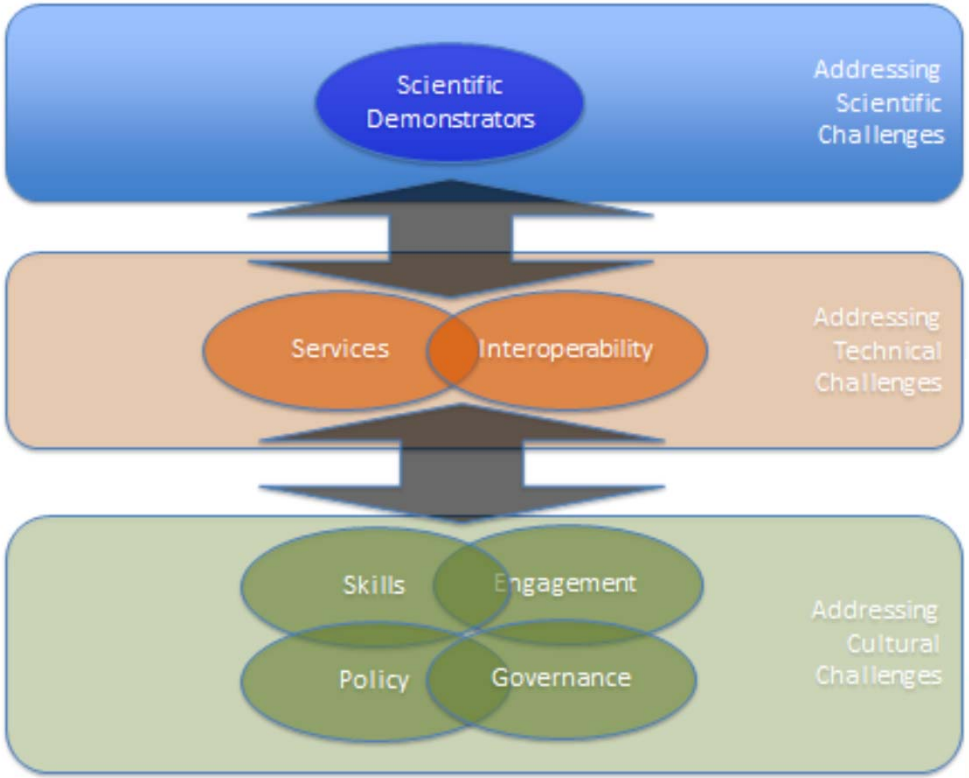
INAF in ASTERICS (II)

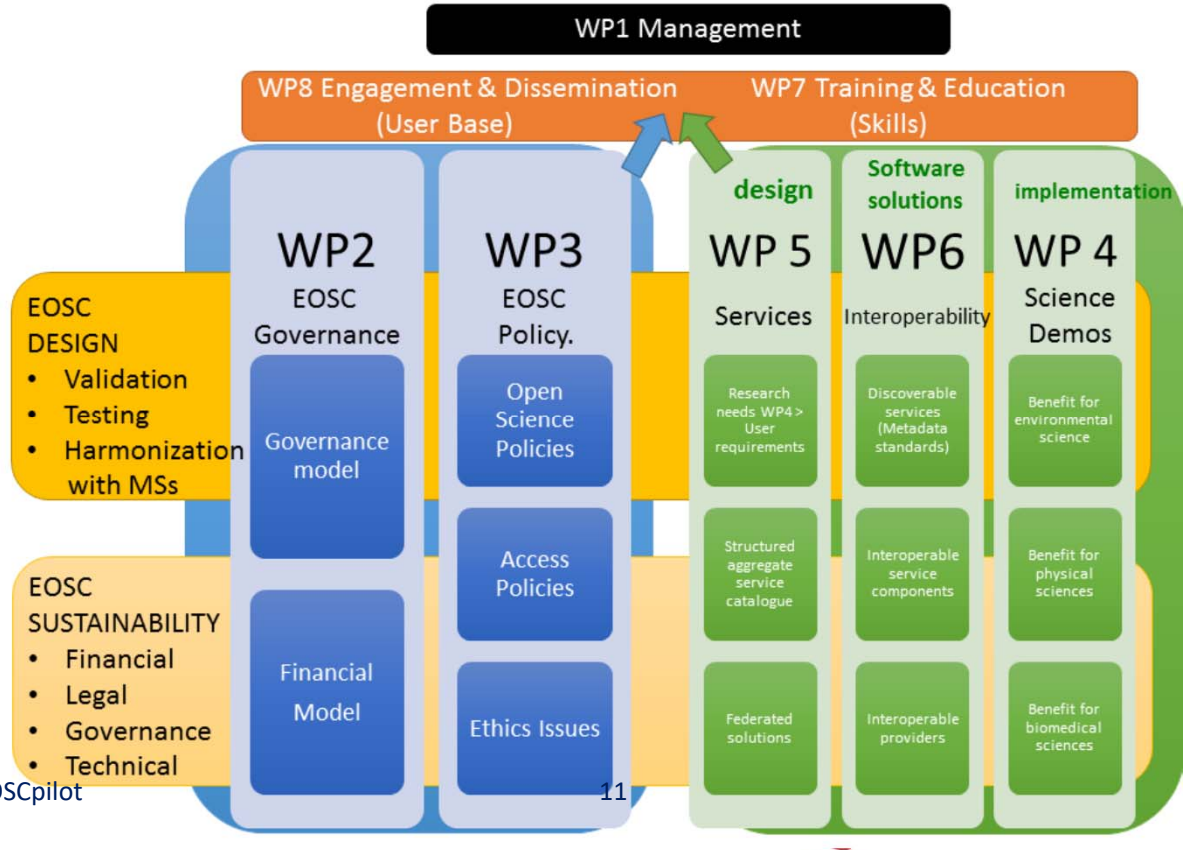
- OBELICS (ref. L.A. Antonelli) with OAR, OATs, OACt:
  - T3.2.3* - optimised handling of secondary data streams and meta-data
  - T3.2.4* - low-power computer platforms for data-driven scalable parallel programming
  - T3.3.2 - prototype benchmarks for testing (e.g. XLDB initiative)
  - T3.3.3 - open services for data integration (including VO-integration)
  - T3.3.4 - software frameworks for data catalogues and query solutions to maximise data integration
  - T3.4.1 - open source software libraries for Peta-scale analysis/mining
  - T3.4.2* - workflow architectures + improving existing authorisation, authentication and accounting protocols
  - interface and complementarity with DADI

EOSCpilot

- H2020 project, funded at €10 million for 2 years, started 1 Jan 2017
- 33 partners, led by UK STFC
- Multi-domain
- Support the first phase in the development of the European Open Science Cloud (EOSC) by:
  - establish the governance framework for the EOSC
  - develop a number of pilots that integrate services and infrastructures
  - engage with a broad range of stakeholders
- Small participation from astronomy, partners are INAF (services and interoperability + pilot: VisIVO) + ASTRON (pilot: LOFAR archive)

EOSCpilot: structure





INAF in EOSCpilot

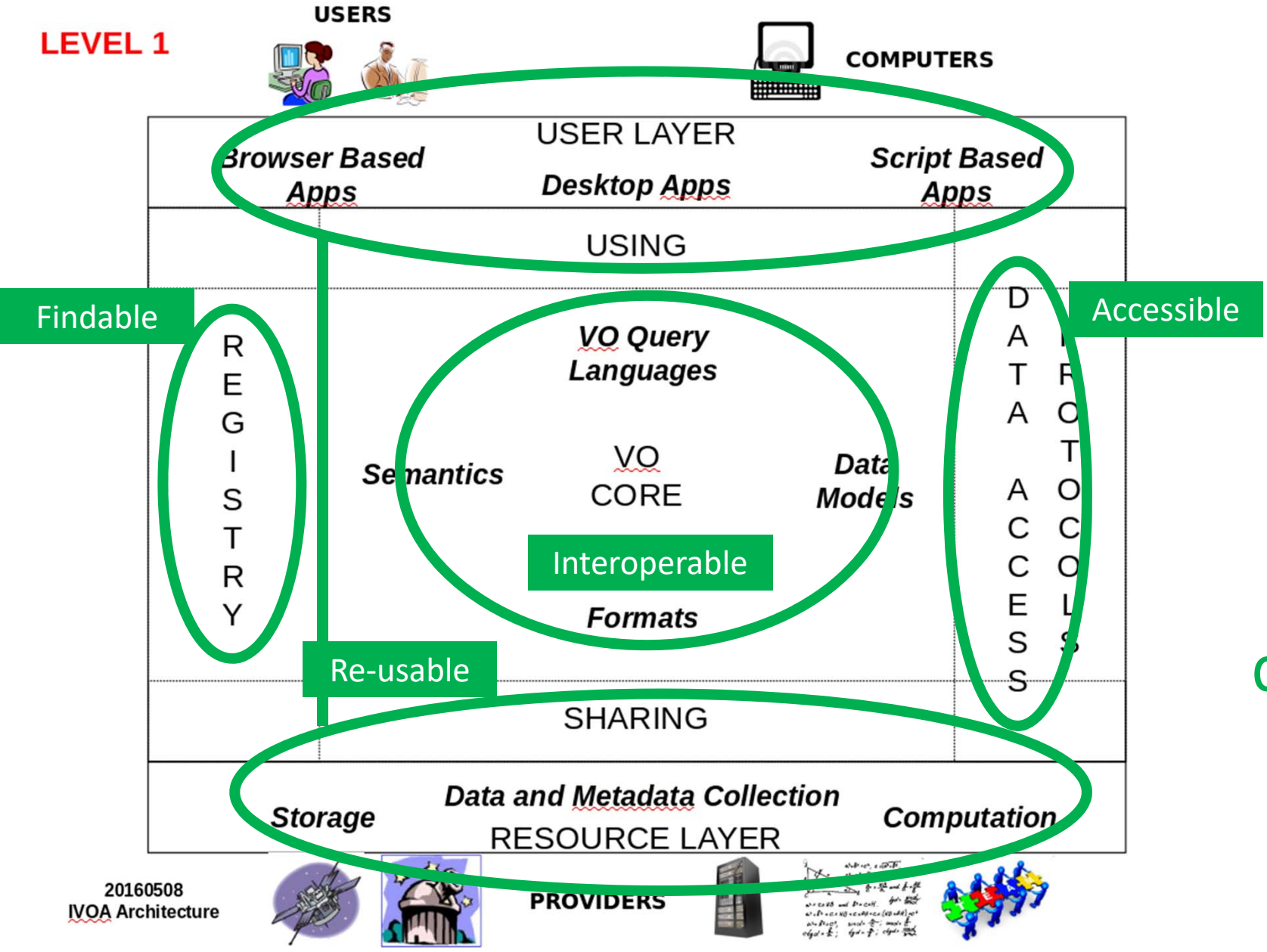
- INAF participation: 110 K€ + support for science demonstrator
- WP5 (Services):

  INAF (6 PM) gives input to

  1. definition of the overall EOSC architecture
  2. definition of the EOSC service portfolio
  3. definition of the EOSC federated service management
- WP6 (Interoperability):

  INAF (6 PM) gives input to

  1. sub-task 6.2.3 Data Interoperability (as part of the FAIR principle): co-responsible of preparation of final report
- Science demonstrator: VisIVO



IO SONO UN PRECARIO DELLA RICERCA

FAIR principles and the VO

VO-compliant data are FAIR !

Asterics
Astronomy ESFRI & Research Infrastructure Cluster

EOSCpilot
The European Open Science Cloud for Research Pilot Project

FAIR compliance is not enough! (I)

- 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)

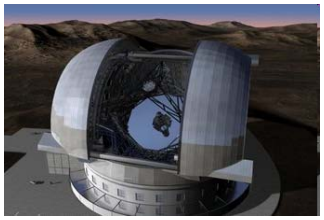
- Data providers shall make efforts to **guarantee preservation** of their data over the **longest possible** period of time (include this in their Data Management Plans)

FAIR compliance is not enough! (II)

- To be effective and enhance current capabilities, 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

- 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)

computing facilities



Archive

Archive

Archive

Archive

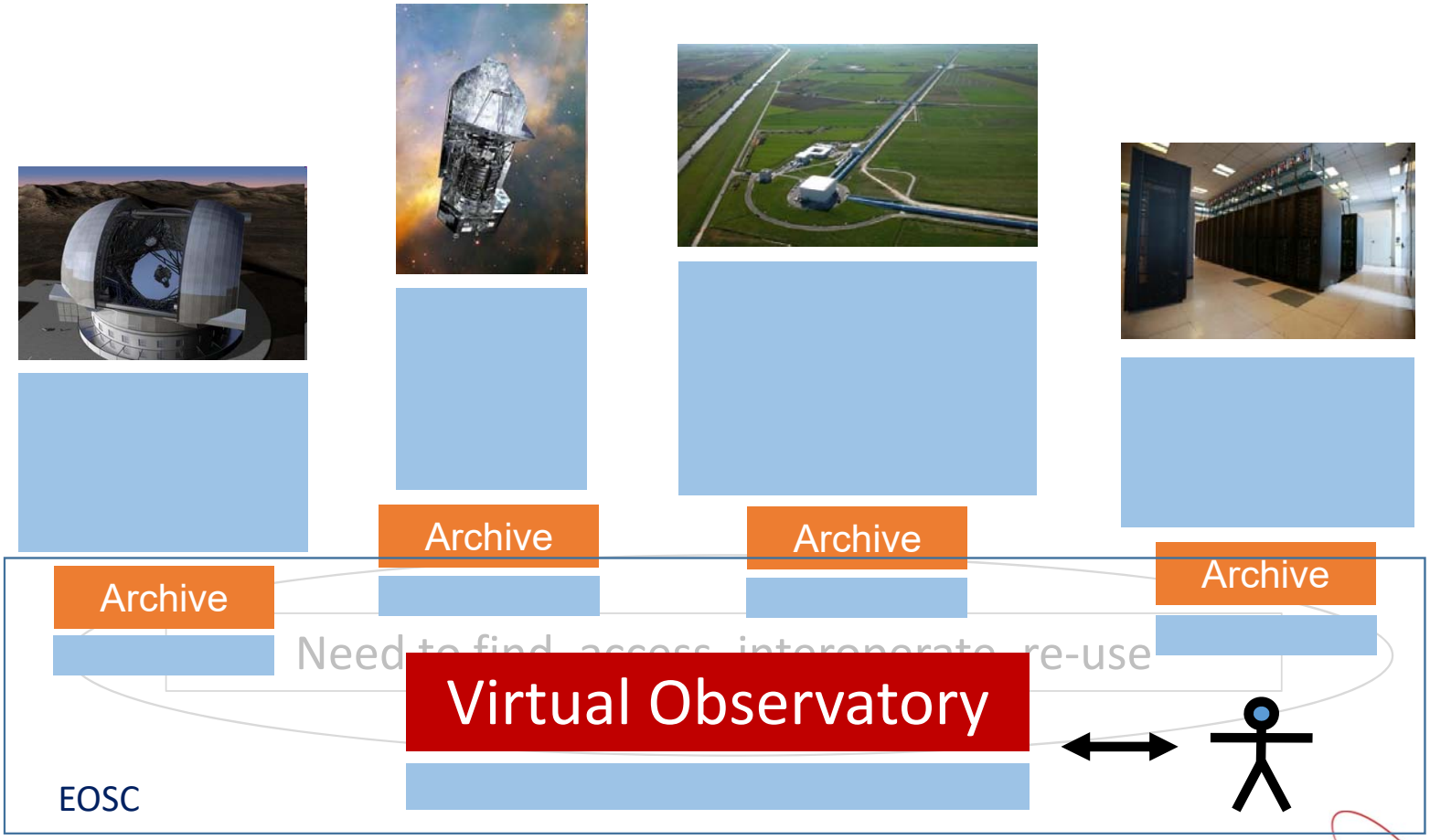
Need to find, access, interoperate, re-use

Virtual Observatory

Astro Research Infrastructures and EOSC

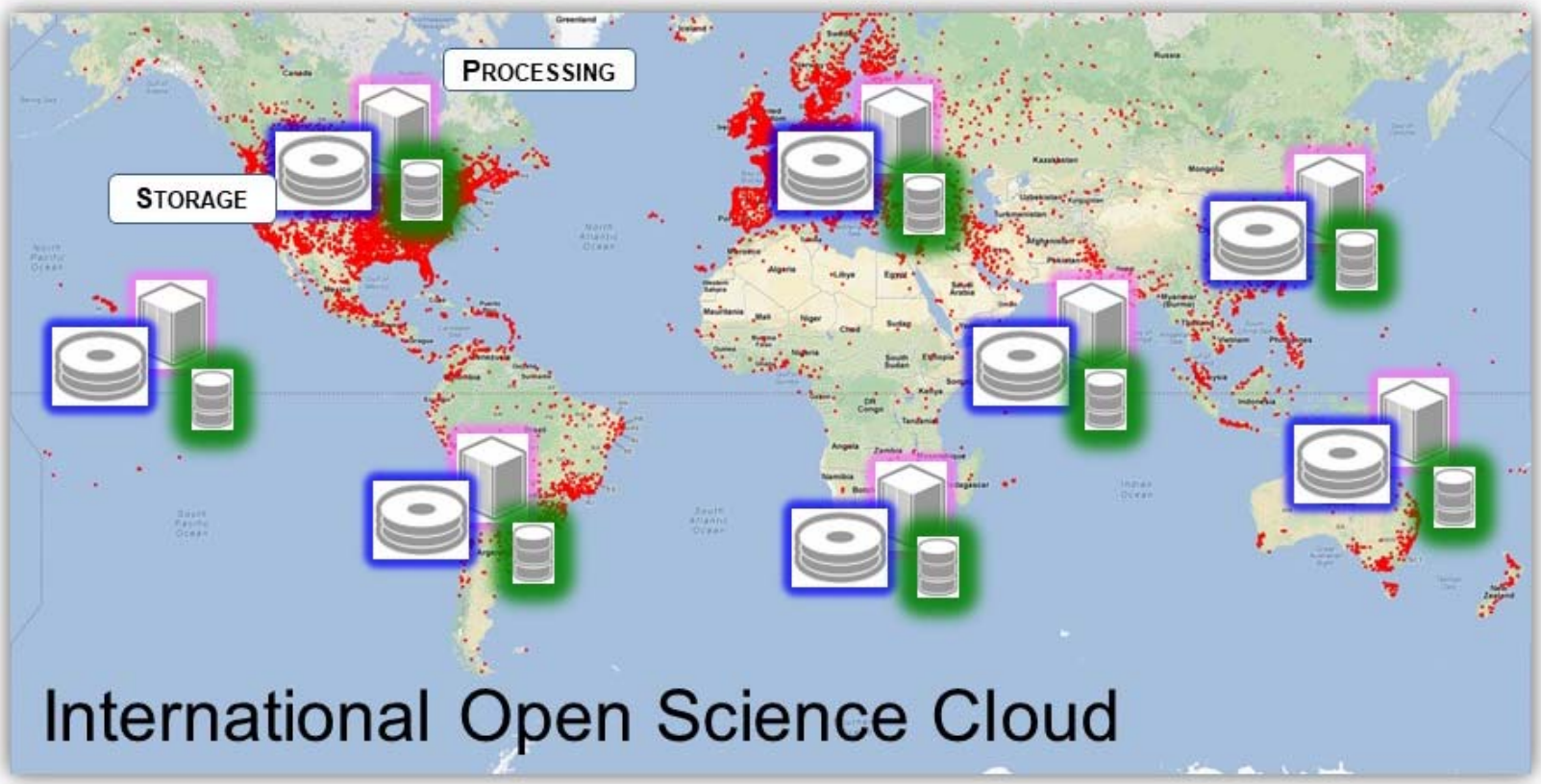
- Each RI produces data at a continuous rate, and performs basic processing (calibration, geometric registration, etc.): for that purpose it is expected to have its own dedicated e-infrastructure
- EOSC should provide data-computing interoperability mechanisms, to allow processing on archived data at the archive location (data centre) within a uniform environment (cut-out service → catalogues → classification → Machine Learning → user-provided code → …)
- From interoperable data collections to an integrated system of services for data-intensive astronomy → science

computing facilities



Shared international platform

IO SONO UN
PRECARIO
DELLA RICERCA



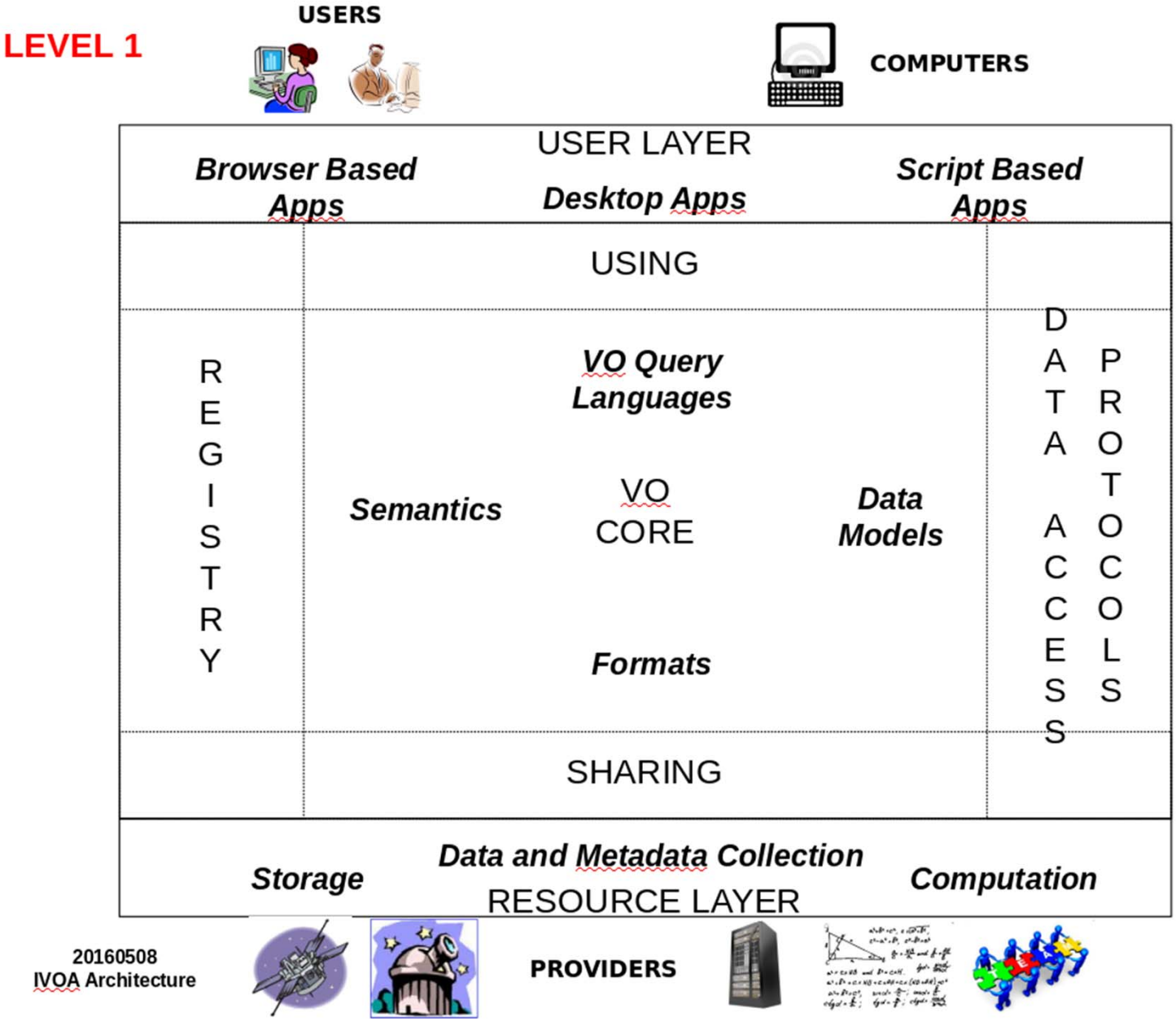
(David Schade, CADC, UN Open Universe Workshop, 22 Nov 2017)

Status and perspectives

- EOSCpilot Stakeholders Forum (28-29 Nov 17, Brussels), w/ASTERICS
  - splinters/panels devoted to policy, governance, architecture, engagement
- ~~EOSC is going to happen!~~ → EOSC is happening!!
  - the EC is investing a lot in this initiative
  - there will be no other significant EC funding for facility computing
- EOSC is **NOT** Astro-driven!
  - but, as a community, we have a chance of providing input (through ASTERICS)
- EOSC-INFRA H2020 call dedicated to integration in EOSC of Clusters of Research Infrastructures → towards an ASTERICS-in-EOSC proposal?!
  - Opportunity for our community to merge the participation in two projects within a coherent framework

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.

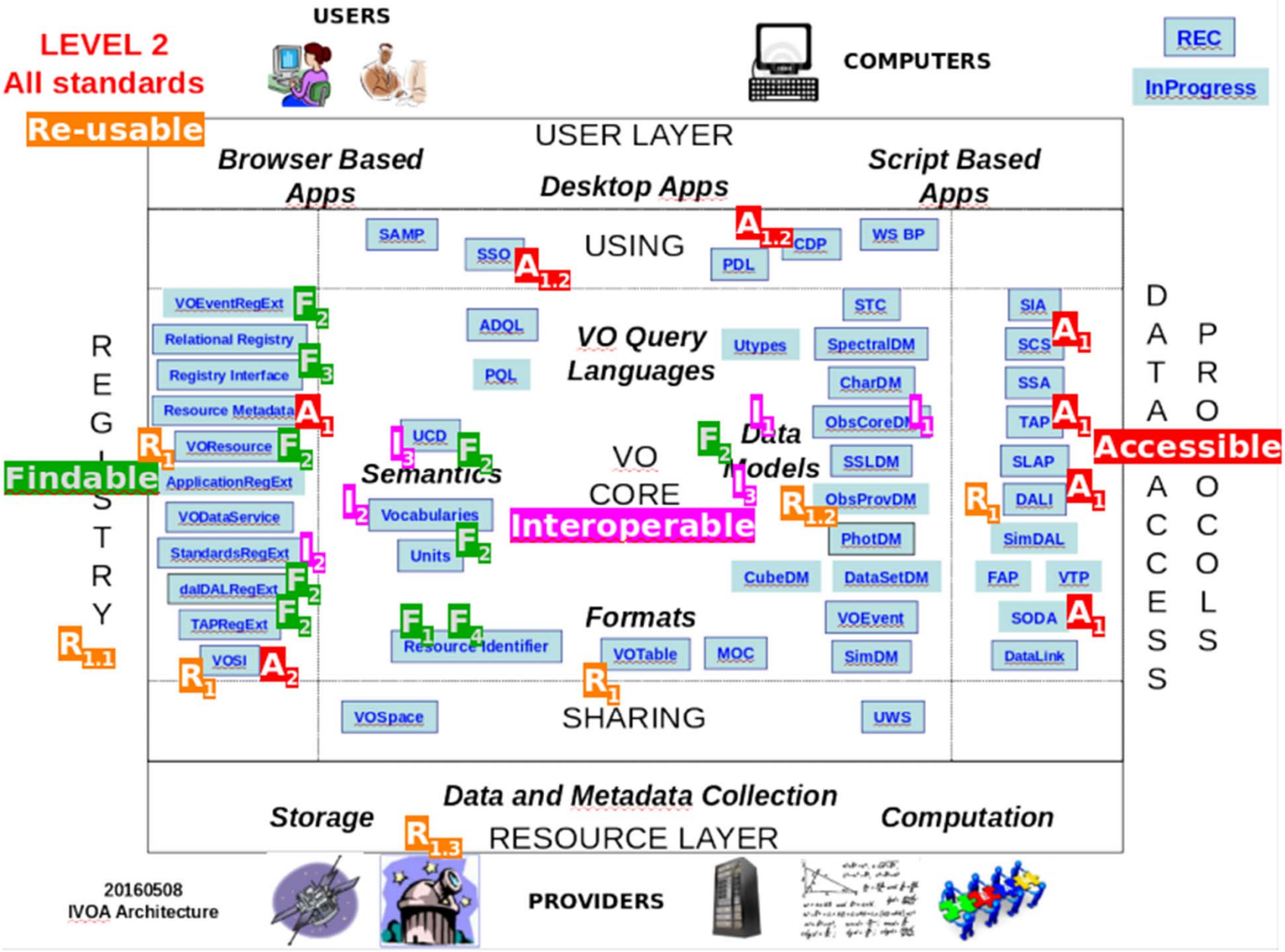


IO SONO UN PRECARIO DELLA RICERCA

FAIR principles and the VO

Asterics
Astronomy ESFRI & Research Infrastructure Cluster

EOSCpilot
The European Open Science Cloud for Research Pilot Project



IO SONO UN PRECARIO DELLA RICERCA

FAIR principles and the VO

FITS Format

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