Open Science, FAIR data & gli standard VO

implementazione della pubblicazione dati presso IA2

Marco Molinaro
INAF – Osservatorio Astronomico di Trieste

Outline

- Open Science & FAIR principles
- Virtual Observatory FAIRness
- INAF IA2 VO publishing architecture
- Conclusions

Open Science & FAIR principles

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)

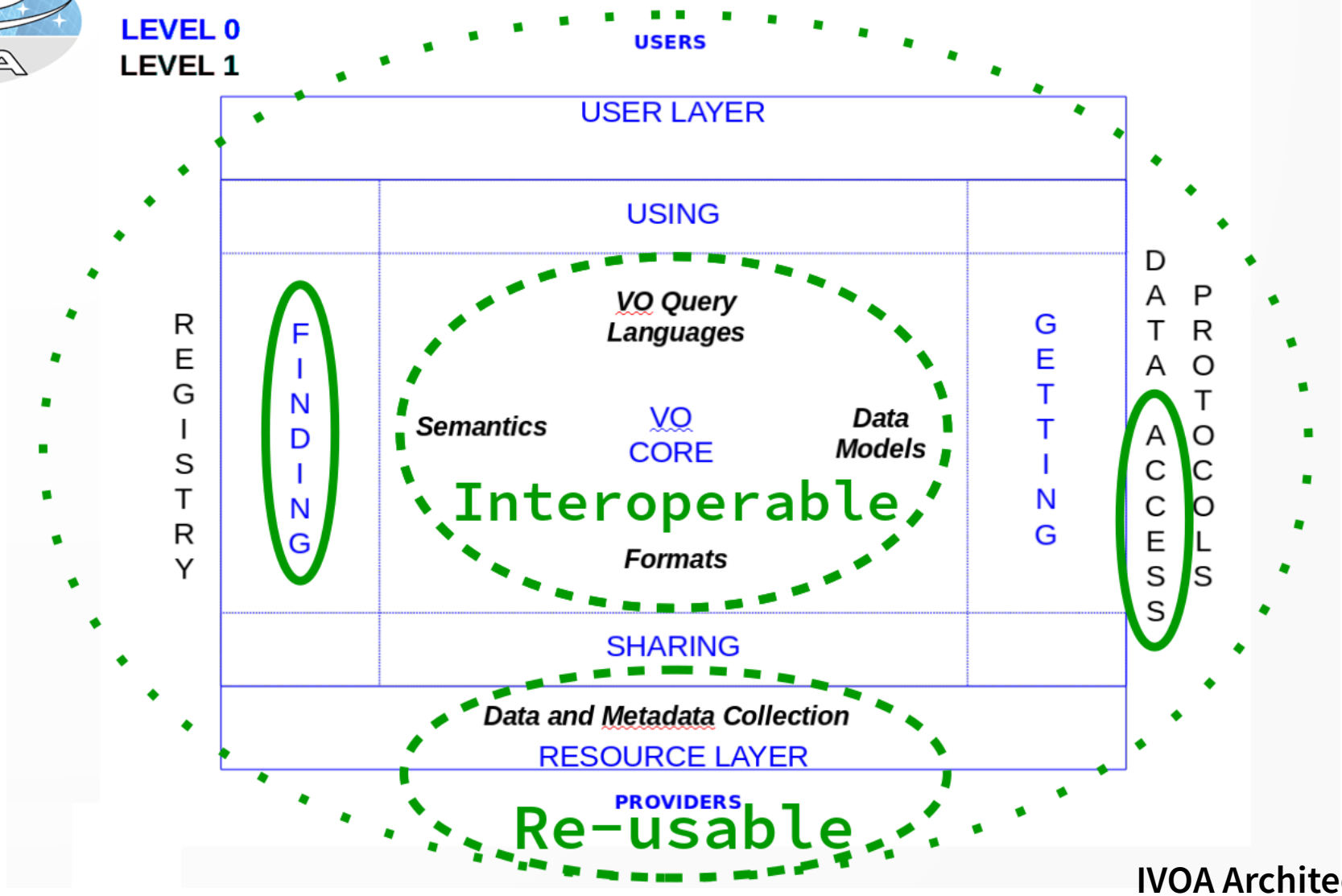
Data Driven Astrophysics

Current scenario

- Scientific publications driven by archival data are comparable or larger in volume then first investigation results;
- Scientific research is multi-instrument, multi-band, compares observations and computational data, multi-messenger;
- Current data amounts exploitation requires large collaborations or subsequent re-analysis;
- Scientific research funding is commonly public money, thus its results must be public as well in the end.

An effective exploitation and proper re-use of current astrophysical resources requires an effort in homogenising the methods by which resources are searched and accessed.

IVOA “FAIR” Architecture

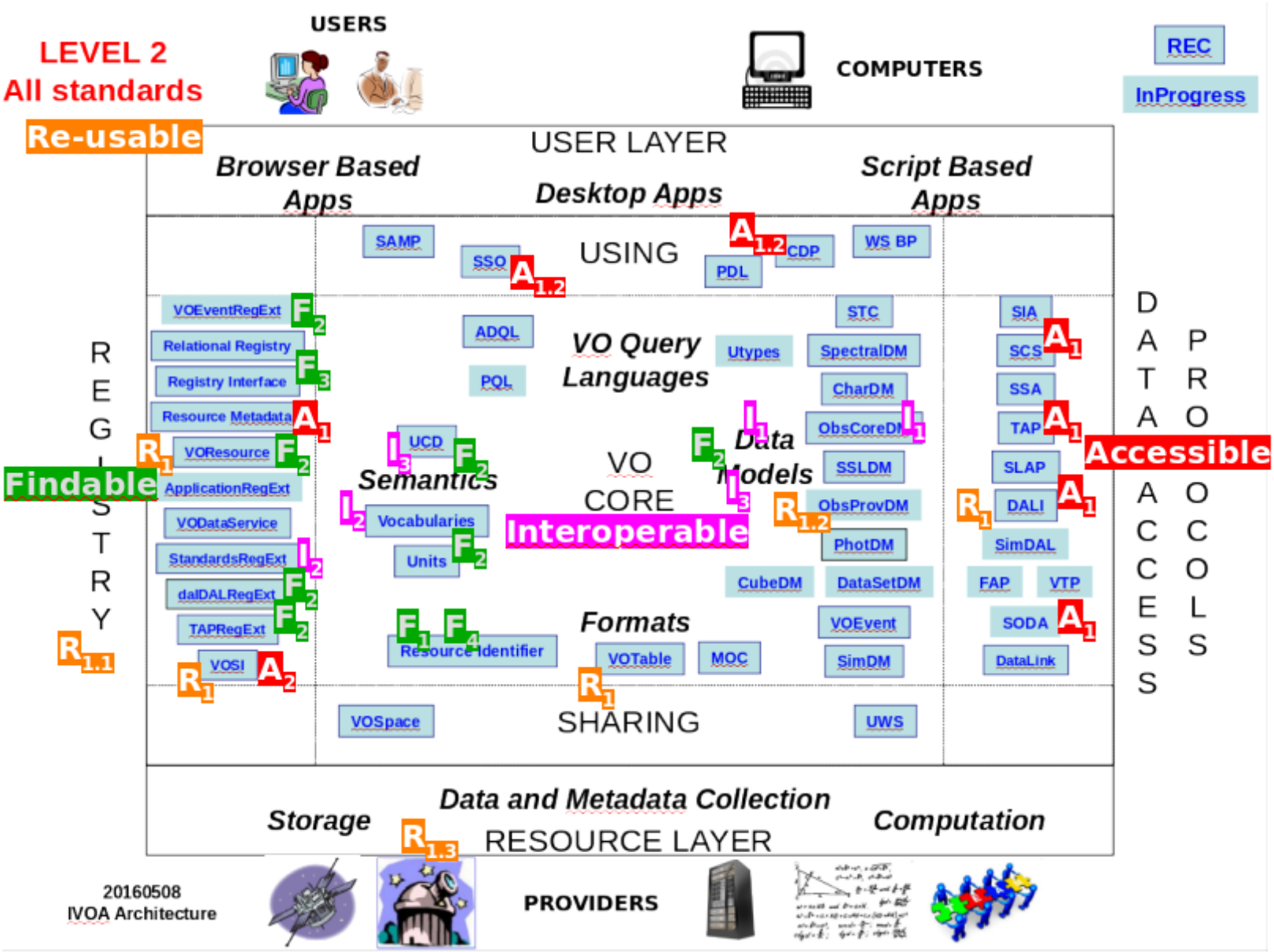


IVOA Architecture

IVOA “FAIR” Standards

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

IVOA “FAIR” Standards - map

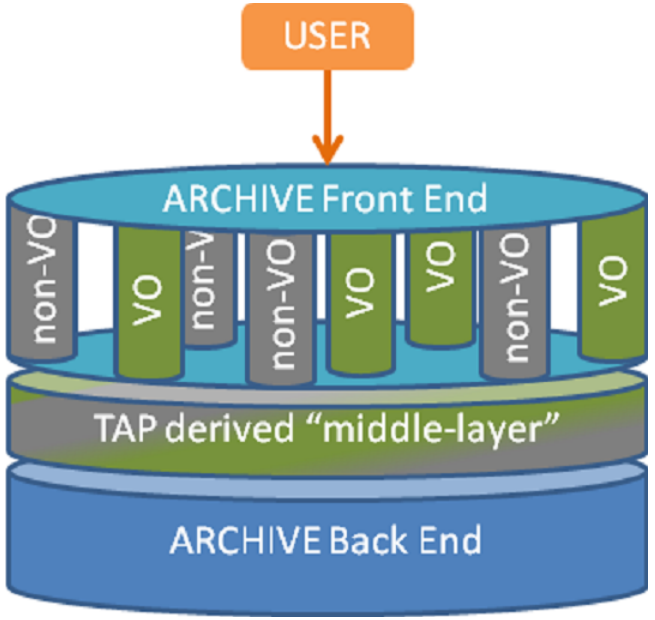


Data FAIRness & the Data Provider

- **Astrophysical data resources deployed through VO standards are FAIR**
  - Not all astrophysical resources are VO available → **NEXT SLIDES**
    - Efforts are needed
    - Deeper interaction among data providers and the IVOA
  - FAIR data principles are not enough
    - Data access services and computational resources are needed on top of data resources
    - IVOA and data providers should work together on standardising efforts

IA2 VO architecture - overview

- **Metadata annotations should be uniquely attached to resources**
  - As far as standards allow it
  - TAP_SCHEMA based solution
    - Specific services consume the metadata
      - Standardised / Custom protocols
  - Observational Core metadata / CAOM metadata easy to support
- **TASMAN admin tool to manage TAP_SCHEMA(ta) instances**
- **Modular solution**
  - Message driven to help scaling
  - Configurable/Replaceable modules as much as possible



TAP_SCHEMA & Observational Models

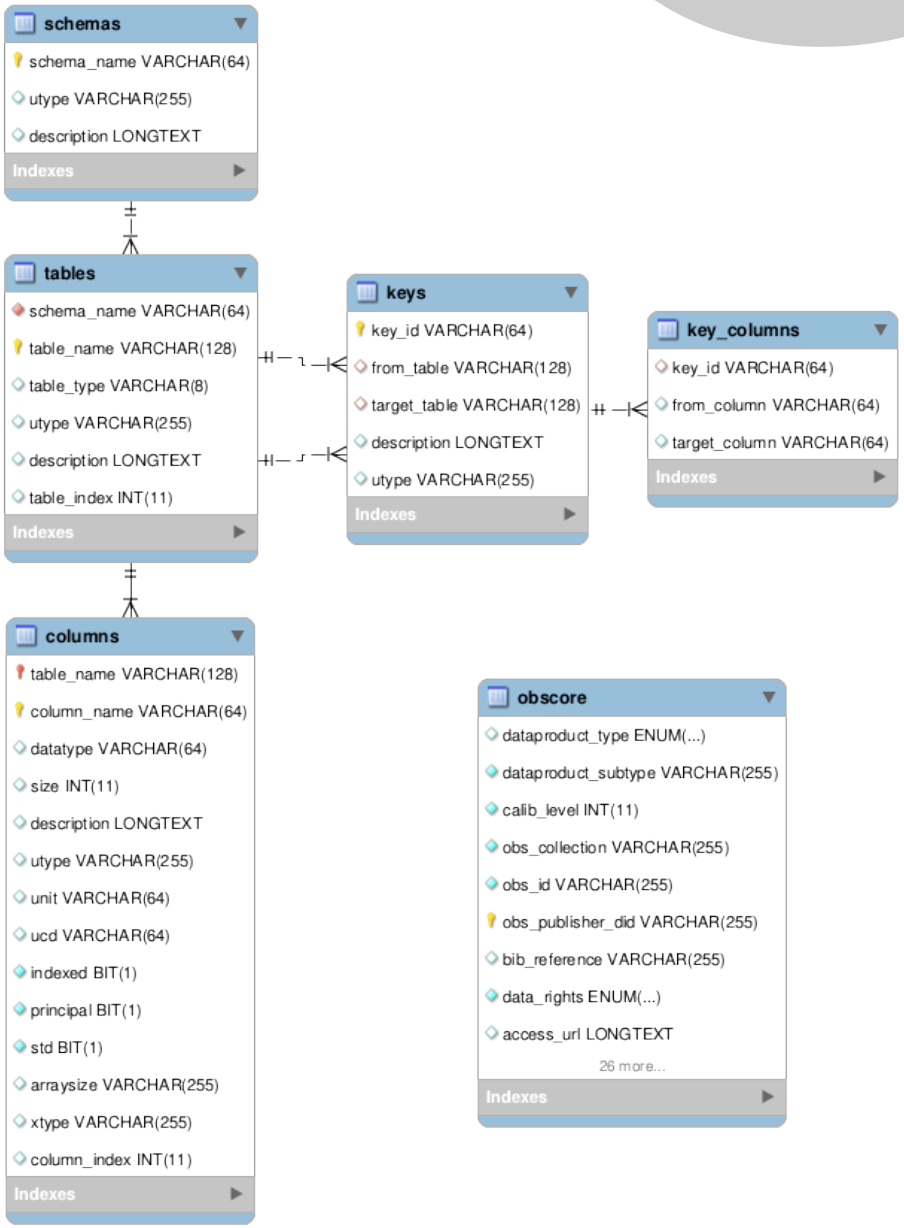
- **VO aware annotation of (tablesets)**
  - DB schemata
  - Tables
  - Columns
  - References

TAP schema

- **VO driven deployment of**
  - Datasets
  - Bundles

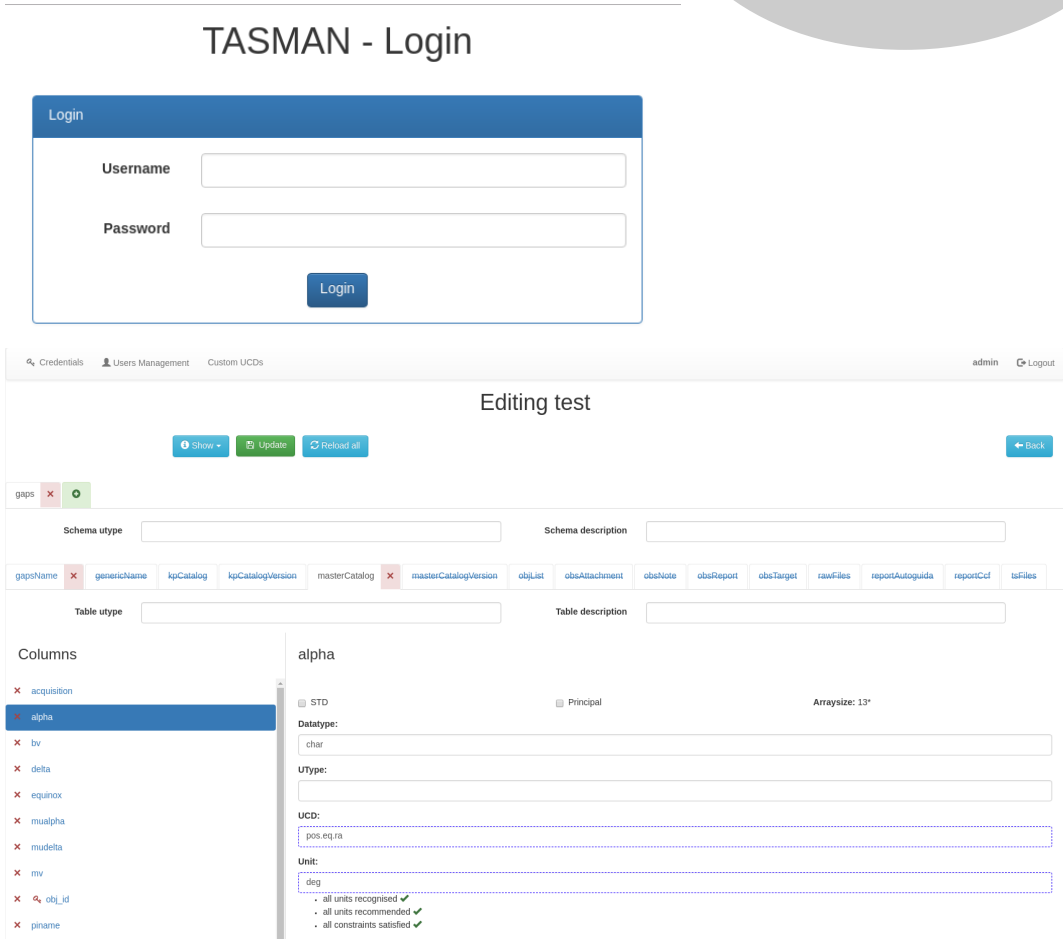
ObsTAP table

- **Common Archive Observation Model**
  - Deploy dataset complexity/specifics
- **TAP_SCHEMA MANager**
  - To help the annotation process
  - To maintain TAP_SCHEMA and VO related schemata/tables



TASMAN

- **Helps managing**
  - TAP_SCHEMA information
    - Datatype mapping
    - UCD suggestion & validation
    - VOUnits validation
    - Deployed/Hidden tables/columns
  - ObsCore basic validation
    - Field specifics and completeness
  - Multiple REC support
    - TAP-1.0 / TAP-1.1 schemata
    - ObsCore-1.1
    - Custom extensions



Browser based

Java application

Embeds container

MySQL & PostgreSQL ready

TAP services toolkit(s)

- **Recently started developing a toolkit from a new library set**
  - Configuration based
  - Spherical geometry & ADQL aware
  - Tesselletion ~ready

TAPlib
(G. Mantelet - GAVO)

- **Despite youth nearly ready for final testing**
  - PostgreSQL + PgSphere for easier ADQL mapping
  - (nevertheless) MySQL capable
- **Working on adding DataLink responses**
  - Allow custom services on top of TAP/ObsCore (but not only) results
- **Currently named TIPTAP**
  - **T**IPTAP: **I**NAF **I**A2 **P**ublishing **T**ool for **T**able **A**ccess **P**rotocol
- **Existing IA2 TAP services still provided by “IA2TAP” deployments**

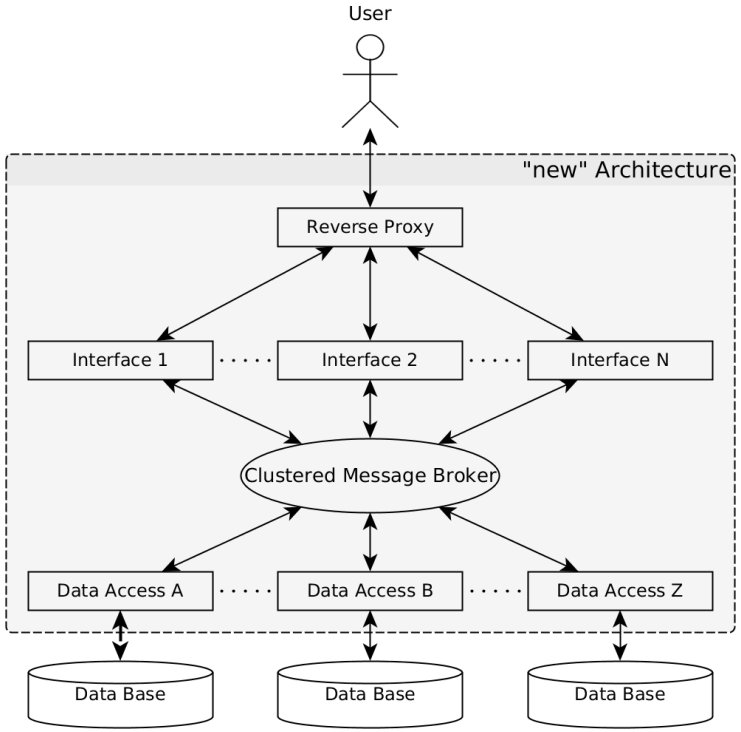
OpenCADC fork

ConeSearch modular toolkit

- **Message driven modular solution**
  - Allow for multiple languages (if needed)
  - Replace current monolithic service deployer
- **Developed modules**
  - HTTP I/F (to accept queries)
  - Data Access (to query the tables)
    - RDB backend
      - PostgreSQL+PgSphere
      - MySQL
  - Configuration repository
    - 1 module – N services
  - Logging module
- **Advanced testing phase/ready for production**
  - Forerunner to other “simple” protocols re-coding

Modules, proxies and broker can be distributed

Help in scaling the system



Other building blocks

- **VIALACTEA Knowledge Base & services**
  - Repository of catalogues
  - Discover, cutout & merge custom services
    - Serving ~20 different surveys at once
    - WCS geometry handled using AST
    - Using AMQP
- **GAPS TimeSeries**
  - Starting up, pushing feedback to VO for exoplanets
- **SSAP**
  - VIPERS DR1, and soon DR2 – based on VO-Dance
- **SIAP-1.0**
  - WINGS, based on VO-Dance
- **SpaceInn**
  - TAP solution based on openCADC IA2TAP tool

What's missing...

- **Most of configuration admin**
  - Service & resource description included
- **Re-deploy of resource layer**
  - PostgreSQL based
  - VO-meta porting of VLKB
- **Missing protocol modules**
  - SIA/SSA, [Datalink, SODA]
  - Other? What about Theoretical/Simulated data?
- **Homogeneous logging**
- **No-browser A&A mechanism**
- **An "Italian" publishing registry?**

...many things

Conclusions

- Work in progress re-shaping the VO publishing solutions at IA2
- Fits into the current revived interest in data driven open science
- Resources are needed, but scientific community input is always a high level requirement

If you can find a path with no obstacles, it probably doesn't lead anywhere.

(Frank A. Clark)

Ackowledgements

- **IA2 tools, present and past developers**
  - Sonia Zorba
    - TASMAN
  - Nicola Fulvio Calabria
    - SCS
  - Robert Butora
    - TIPTAP
    - VLKB services
  - Francesco Cepparo
    - message driven architecture (VOBall)
  - Omar Laurino
    - initial VO-Dance efforts
  - Pietro Apollo
    - Initial IA2TAP efforts
- **External libs/code/services**
  - CDS UCD tools
    - http://cds.u-strasbg.fr/UCD/tools.htx
  - UCIDY (UCD validator)
    - https://github.com/gmantele/ucidy
  - UNITY (VOUnits validtor)
    - https://bitbucket.org/nxg/unity
  - TAPlib (TAP/UWS/ADQL libraries)
    - https://github.com/gmantele/taplib
  - AST library (WCS manipulation)
    - http://starlink.eao.hawaii.edu/starlink/AST
  - PgSphere (PostgreSQL spherical geometry add-on)
    - http://pgsphere.github.io/index.html
  - OpenCADC source code
    - https://github.com/opencadc

Thank you!