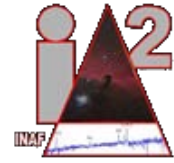


VO Day & Friends  
17-19 December 2013  
INAF-OACT



# Data Publishing in the VO

What is it? Why should I? How to?

**Marco Molinaro**  
INAF-OACTs IA2



# Outline

---

- ▶ Purpose of this talk
- ▶ What means VO publishing
- ▶ Why publish data to the VO
- ▶ How to publish data to the VO

# User or Provider?

---

- ▶ This workshop is for VO tools users
  - ▶ Data is *accessed* from the tools
  - ▶ Data *has been provided* by someone else
    - ▶ May be *YOU*
- ▶ Trusting the data you use is important
  - ▶ VO may obfuscate confidence
    - ▶ WWW works in a similar way
      - do you trust wikipedia?
      - do you trust your online bank accounting?
- ▶ This talk is about getting some grips about how the data you use in the tools is accessible to them
  - ▶ It is about trusting the VO

# ...some information

---

- ▶ What
  - ▶ Details on the architecture behind the VO astrophysical resources
- ▶ Why
  - ▶ Good reasons to let data be accessible through VO
- ▶ How
  - ▶ A minimal set of tips to get yourself the *providers' badge*

# What is VO Publishing

---

- ▶ Publishing in general is
    - ▶ Offering data to other people
  - ▶ Accounts for data
    - ▶ Access
    - ▶ Readability
    - ▶ Search/Seek
  - ▶ VO infrastructure adds also the goal for
    - ▶ Interoperability
- 
- ▶ Level -1: Web, Data Formats
  - ▶ Level 0: Registry
  - ▶ Level 1: Data Models
  - ▶ Level 2: Protocols

# Level -1 / basic access

---

## ▶ Main goals: Access & Readability

### ▶ Access

- ▶ HTTP-based web portals & services
  - Quite-easy way of starting data sharing
  - Quite common solution

### ▶ Readability

- ▶ File format dependent
  - Usually takes place after data retrieval
    - Unless portal-embedded tools are provided
- ▶ VOTables, FITS, VOEvent are the VO preferred
  - Other formats are possible (e.g. CSV, HDF5)
    - Meta-description un-avoidable

# Level 0 / Registry

---

## ▶ Seeking & Access

### ▶ Seek & Exhibit

- ▶ Let community find your services
- ▶ IVOA provides «*registration*»
  - Describe your resources (metadata)
  - Get back an «IVORN»
    - Unique identifier, like: **ivo://authority/custom/local/id**
  - Works for almost everything: web pages, FTP servers, ...
    - Interoperability issue

### ▶ Access

- ▶ Finer solution
  - Resolve IVORN into consumable URL endpoints
    - See *Level 2* for the *consuming* details

# Level 1 / Data Modeling

---

## ▶ Readability & Interoperability

### ▶ Readability

- ▶ File formats (*Level -1*) let you access stored values
  - «I'm a table, I've named columns with descriptions and defined data types.»
- ▶ Data Models improve data processing by
  - Identifying the scope: spectra, images, time series
  - Providing semantics: machine readability to automate data processing
    - Coordinate specifications (space, time, band, ...)

### ▶ Interoperability

- ▶ Same scope -> same model -> compliant description
  - Processing works seamlessly above the actual data format



# Level 2 / Protocols

---

## ▶ Access & Interoperability

### ▶ Access

- ▶ Define properly the way services are accessed
  - Parameters, constraints, ...
- ▶ Scope targeted
  - Catalog-like: SCS
  - Images: SIA
  - Spectra: SSA
  - Generic tabular data: TAP
    - Uses data models to provide semantics
- ▶ Reasonable VO tools support (see tutorials!)

### ▶ Interoperability

- ▶ Common interface access -> easier resource merging

# Validation

---

- ▶ That's nice! Does it work?
  - ▶ VO contribution means enriching a collaborative, distributed information system
    - ▶ No need to be too strict
    - ▶ But bad resources cause bad user experience
  - ▶ Validation of resources mitigates this issue
    - ▶ It works! Is it compliant?
- 
- ▶ More on protocols and validation in the «*How To*» section

# Why Publish to the VO

---

- ▶ IVOA mission
  - ▶ Framework & Architecture
  - ▶ Goal
- ▶ Data Center reality
- ▶ Converge on a solution
- ▶ Why is all this worth?

# IVOA, Mission ...

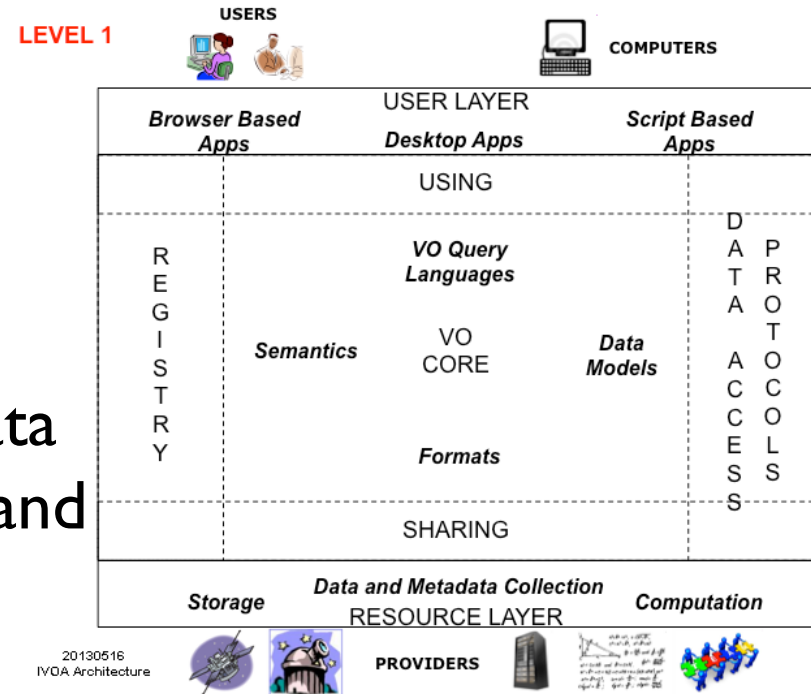
---



- ▶ **Mission**
  - ▶ Allow **astronomers** to interrogate multiple data centers in a seamless and transparent way
  - ▶ Provide new powerful analysis and visualization **tools** within that system
  - ▶ Give **data centers** a standard framework for publishing and delivering services using their data

# ... Architecture & Goal

- ▶ Standardization of
  - ▶ data and metadata
  - ▶ data exchange methods
- ▶ Use of a service registry
- ▶ A framework which enables data centers to provide competing and co-operating data services



## ▶ To enable science!

- ▶ Data discovery, Efficient data access, Interoperable analysis tools, Interoperable data, Scalable visualization and computing, Data Mining, ...

# Data Centers / Data Providers

---

- ▶ **Sketched-up reality**
  - ▶ Providers have constraints on data serving
    - ▶ Primary users' requirements
- ▶ **Converging on a solution with**
  - ▶ Integrate data potential exploitation
    - ▶ Primary user requirement
  - ▶ with full data interoperability
    - ▶ IVOA goal
- ▶ **DC perception of VO**
  - ▶ Pro: increased scope
  - ▶ Con: extra software and maintenance
    - ▶ Plus retro-fitting and data modeling

# Perspective

---

- ▶ (sort of) Goal switch
  - ▶ Move from: publish data **IN** the VO
  - ▶ To: publish data **WITH** the VO
- ▶ Will not avoid efforts
- ▶ Will take advantage on existing intellectual investment of the IVOA

# Why

---

- ▶ Leverage the intellectual investment from a wide community
  - ▶ Scientific
  - ▶ Technical
- ▶ Fundamental components for data centres
  - ▶ No longer an “extra” layer
- ▶ Increase data visibility
- ▶ Exploit data potential
  - ▶ Interoperating different datasets
  - ▶ Capitalize non-primary data content



# How to Publish to the VO

---

- ▶ Putting together
  - ▶ Formats
  - ▶ Protocols
  - ▶ Models
  - ▶ Registry
- ▶ To reach a goal
  - ▶ **interoperable astrophysical data resources**
- ▶ **But in practice?**

# Getting started...

<http://wiki.ivoa.net/twiki/bin/view/IVOA/PublishingInTheVONew>

## Publishing Data into the VO

### ↓ [Publishing Data into the VO](#)

- ↓ [0. Introduction](#)
- ↓ [1. Questions & Answers about what type of data to be published into the VO](#)
- ↓ [2. Registry for VO Data Service Discovery](#)
- ↓ [3. Toolkits to publish data into the VO](#)
- ↓ [4. Developer's corner : other useful software tools and libraries for VO development](#)

### 4. Developer's corner : other useful software tools and libraries for VO development

Tool / Library Name	Contact Point	Date of last Release	
<a href="#">STIL</a>	<a href="#">MarkTaylor</a>	12/2012	Table I/O and processing library in Java, including full VOTable I
<a href="#">STILTS</a>	<a href="#">MarkTaylor</a>	3/2013	Table command-line tools suite. Includes <a href="#">votlint</a> (VOTable valida
<a href="#">SAVOT</a>	<a href="#">AndreSchaaff</a>	4/2010	Simple Access To VOTable, parsing / editing / writing VOTable
<a href="#">CDS UWS Library</a>	CDS cds-question (at) unistra.fr		A UWS (Universal Worker Service) is defined by IVOA ( <a href="http://www.ivoa.net/documents/latest/ADQL.html">http://www.ivoa.net/documents/latest/ADQL.html</a> ). To sum up ADQL adds to the bas corresponds to a cone search around (10°,5°) with a radius of 2
<a href="#">CDS ADQL Library</a>	CDS cds-question (at) unistra.fr	3/2012	ADQL (Astronomical Data Query Language) is a SQL-like langu
<a href="#">UCD assignment tool</a>	sebastien.derriere (at) astro.unistra.fr	10/2008	The interface allows to find the UCD corresponding to a descrip

### 3. Toolkits to publish data into the VO

Toolkit Name	Contact Point	DAL Services Supported
<a href="#">SAADA</a>	Laurent Michel	SIA/SSA/CS/TAP (including an ObsTap view mapper)
<a href="#">DALServer</a>	USVAO	SIA/SSA/SCS/SLAP; TAP implemented but not yet integrated
<a href="#">VO-Dance</a>	<a href="#">VObs.it</a>	SCS, SIAP, SSAP; TAP (using a different application)
<a href="#">DaCHS</a>	<a href="#">MarkusDemleitner</a> , <a href="#">GAVO</a>	SCS, SIAP, SSAP, TAP, OAI-PMH
<a href="#">OpenCADC</a>	Patrick Dowler CADC	
<a href="#">DALToolkit</a>	ESA-VO	SIAP, SSAP
<a href="#">DSA</a>	<a href="#">AstroGrid</a>	

# Way to the goal

---

- ▶ Identify your path to goal, consider:
  - ▶ Technological expertise
    - ▶ Including willingness for coding
  - ▶ Dataset complexity
    - ▶ Single table, set of images, images and catalogues, multiple tablesets, tablestes with image and/or spectra access, ...
  - ▶ Structural starting point
    - ▶ From sparse files to fully structured archive
- ▶ but also...
  - ▶ Dataset content dynamics
    - ▶ One-shot DR / incremental dataset
  - ▶ Release data policy
    - ▶ Public, private, mixed

# Help yourself or find support

---

- ▶ Identify your path helps in
  - ▶ Not losing your mind while diving in the IVOA-Doc ocean
    - ▶ TECHNOLOGICAL EXPERTS must consider themselves good divers!
- ▶ First step in defining your strategy
  - ▶ Clear ideas of what you want to expose
  - ▶ How you want to structure the deployed resource(s)
- ▶ Ask your national VO representatives for
  - ▶ Help in defining your path
  - ▶ Existing support solution
  - ▶ Knowledge and expertise

# VO Resource Registration

- ▶ Resource
  - ▶ Data collections
  - ▶ Services
    - ▶ Registry
- ▶ Curation and Content
- ▶ Capabilities and Interfaces
- ▶ Where?
  - ▶ RofR
  - ▶ full harvestable
  - ▶ publishing

<b>Insert Resources</b>
Resource
Organisation
Authority
Data Collection
Service
Registry
Table Service
Data Service
Catalog Service
Cone Search (CS)
Open Sky Node (OSN)
Simple Image Access (SIAP)
Proto Spectral Access (PSAP)
Simple Spectral Access (SSAP)
Simple Line Access (SLAP)
Theoretical Spectral Access (TSAP)
Table Access (TAP)

<http://registry.euro-vo.org>



# VO Resource Validation

**AIDA** Astronomical Infrastructure for Data Access

---

The Euro-VO projects:
VOTECH
EuroVO-DCA
EuroVO-AIDA

- EURO-VO Registry
- Search Resources
- Insert Resources
- Update Resources
- Validate Resources
- EURO-VO Registry Resource Details

Member of

Powered by

## EURO-VO Full Harvestable VO Resource Registry

Welcome to the EURO-VO Full Harvestable VO Resource Registry.

Validate Resources

CS Validator

SIA Validator

SSA Validator

SLA Validator

is, you can have a look at the [IVOA Resource Registry](#) specifications.

registry data:

es in the Registry. For example, clicking on the "Simple Image Access" will "Image Access" in the registries around the world.

Due to the fact that the Catalog Service Resource type allows for one entry per table, and that CDS contains thousands of tables, and in order to not clobber the access, we have separated the Catalog Service in CDS and non-CDS searches for commodity.

- Insert Resources
 

Allows the insertion of a new Resource in the EURO-VO Registry.
- Update Resources
 

Allows the edition of a Resource that resides in the EURO-VO Registry. Resources can only be updated in the Registry where they have been introduced, and not in registry that harvest them from other places. Consult the registry specification for more details (see above URL).
- Validate Resources
 

Allows the validation of various types of DAL Services Resource.

If you have any question regarding the EURO-VO Registry, please send a note to our [Registry manager](#).

Version: 1.1
 co-funded project



# VO Resources

---

- ▶ 14366 VO Resources
  - ▶ As of 13.12.2013
  - ▶ Taken from GAVO registry, RegTAP interface

```
SELECT
      COUNT(*)
FROM rr.resource
WHERE
      res_type != 'vstd:standard' OR
      res_type != 'vg:registry' OR
      res_type != 'vr:organization'
```