Facilities Overviews: Themes

Tom McGlynn

Presentations

- May 2015: Sesto
 - CDS: Landais: Integrating VOServices into VizieR
 - CADC: Gaudet: An Integrated VO-Enabled Framework
- Oct 2015: Sydney
 - ChinaVO: Li: The Architecture and Maintenance of the China-VO System
- May 2016: Cape Town
 - GAVO: Demleitner: VO Protocols Implementation at GAVO
 - IRSA: Landry: VO Protocols Implementation at IRSA
- Oct 2016: Trieste
 - ESA/GAIA: Gonzalez:Administrating a heavily used TAP instance: Gaia Archive operations for DR1
 - SVO:Alacid:VO Operations at the Spainish Virtual Observatory
- May 2017: Shanghai
 - China-VO: Li: China-VO Operations Report
 - WFAU: Voutsinas: WFAU Operations Report

- Oct 2017: Santiago
 - ChiVO: Camilo Nunez and Mauricio Araya: ChiVO
 - NASA/HEASARC: McGlynn: Usage of the VO at the HEASARC
- May 2018: Victoria
 - CADC: Gaudet: A Second Look at VO Operations
 - ESA: Merin: ESASky Updates and Operations
 - ESA/GAIA: Salgado: GAIA Archive for Release 2
 - IRSA: Groom: IRSA Operations
- Nov 2018: College Park
- May 2019: Paris
 - PADC: Erard: VO Operations at the Observatoire de Paris
 - MAST: Dower: VO Operations at MAST

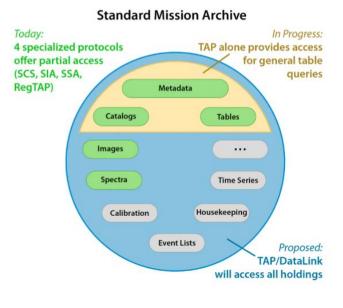
Institutions Presenting

- CDS
- CADC x 2
- China VO x 2
- GAVO
- IRSA x 2
- ESA/GAIA x 2
- SVO

- WFAU
- ChiVO
- HEASARC
- ESA/ESA Sky
- PADC
- MAST

Enabling a common interface

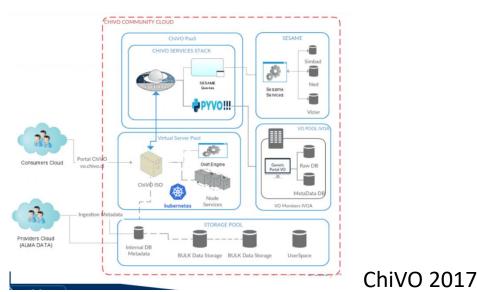
- Handling a legacy of diverse missions with different missions
- Providing a common interface across diverse archives

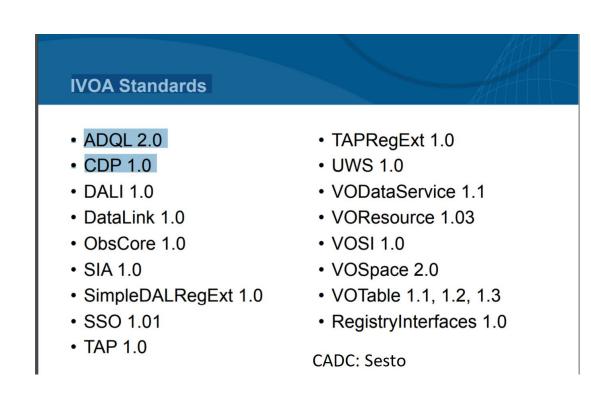


NAVO VO interface

Entering the VO requires implementing a wide variety of interconnected protocols

- Requires care in understanding dependencies and interrelationships of standards
- Versions of standards matter
- Dependencies on some protocols (e.g., registry) may be non-obvious
- Substantial barrier to entry that can be alleviated by reuse of existing code

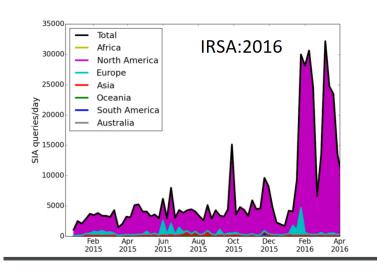




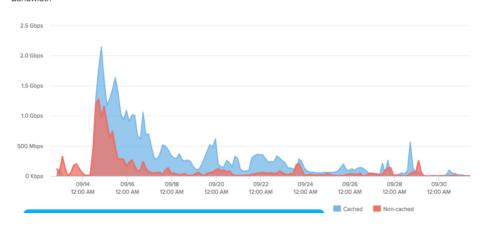
Usage is highly variable

- Need to size to anticipated peaks
- Data releases
- External events which drive interest
- Mirrors to support distribution of broadly popular data

Simple Image Access Traffic

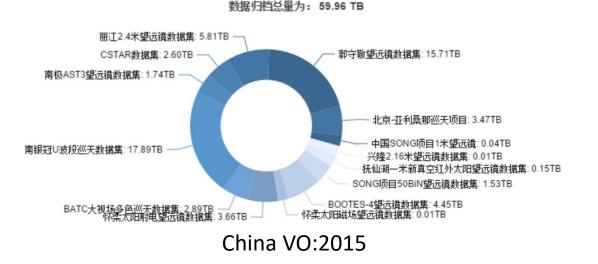


Bandwidth



Lots of Data

- ...and growing TB -> PB and not too long PB -> EB
- Frequent use of cloud and cloud architectures but so far these are not in commercial cloud





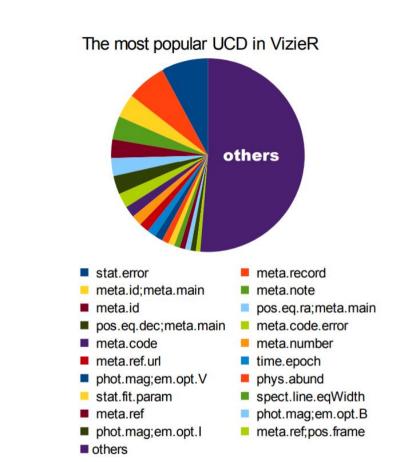
Archives are dynamic and need to facilitate ingest of data

- Standardized approaches for externally provided data
- Need to understand how to best preserve metadata while transforming to VO standards.
- Can derive VO from broader standards.
 - The VO models are <u>views</u> on CAOM:
 - ObsCore: observation.intent = "science" and plane.calibrationLevel is not null
 - SIAv2: ObsCore and plane.dataProductType in ('image', 'cube')
 - SIAv1: observation.intent='science' and plane.calibrationLevel > 1 and plane.dataProductType = 'image' and artifact.productType = 'science'
 CADC 2018

4. Publication workflow

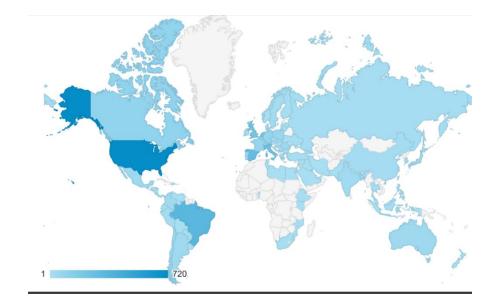
- 1. Provider upload (mail, rsync, http)
- 2. Subset on developer machine, RD in SVN
- 3. SVN checkout on server, non-linked service
- 4. Provider clarifications/feedback
- 5. Repeat until consensus
- 6. Publish to VO/web page

GAVO Worflow: 2016



Our services are broadly used

- Services are used worldwide
- Archives use data at other archives
- Many services are used not only by professional astronomers but also by public.
- Archives support one another (e.g., by sharing HiPS)

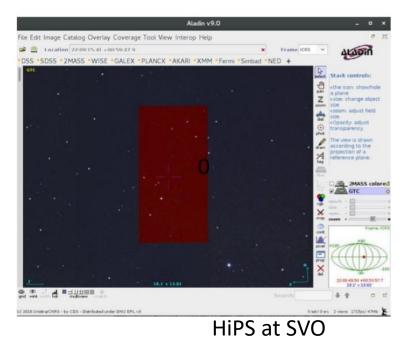


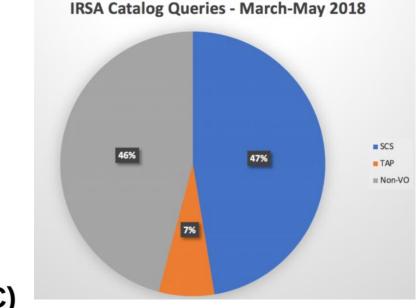
ESASky Usage

Archives evolve

- Implementations of protocols are prioritized to meet perceived needs.
- VO support needs to be balanced against legacy interfaces
- New standards are implemented as their utility becomes apparent.

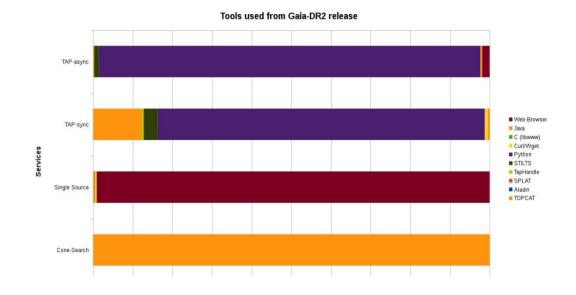
<u>HIPS</u> (GTC)





Lots to learn about how users access archives

- What clients do they use?
- Are they using PyVO or other such libraries?
- VO vs non-VO interfaces
- Simple protocols (e.g., conesearch) most commonly used but may be used inefficiently.



Closing thoughts

- Lots of different approaches taken in these talks addressing different aspects of how VO can support usage of our archives: what standards are used, how do we maintain it, usage, volumes,... These different approaches can be very enlightening and have helped me to reconsider how I use the VO.
- Variability and size of VO data usage can be intimidating, but we need to plan for it.
- Some archives build on VO protocols as fundamental, others add VO interfaces to existing frameworks. This is not 1-1 with whether they had legacy data.
- Institutional adoption of VO is complex requiring understanding of myriad standards:
 - Can probably do better in making it easier for institutions to join VO. Libraries like DaCHS, CDS HIPS and MOC libraries and CADC codes should be promoted.
 - Do we need secondary documentation beyond standards for implementors given that on entry a user needs to understand 10 or more that interrelate in non-obvious ways (VOTable, UCDs, DMs, Registry,).