The ALMA Archive and the VO

Felix Stoehr
Subsystem Scientist of the ALMA Science Archive

Outline

- ALMA
- Science Archive
- VO and ALMA
- Data-centre perspective

- Who? Joe Schwarz, Andreas Wicenec, Gary Fuller, Sandra Etoka, Alisdair Manning, Brian Glendenning, Mark Lacy, Stéphane Leon, Erik Muller, Masao Saito, Holger Meuss, Alessio Checucchi, Justo Gonzales, Matthias Bauhofer, Juande Santander, Christophe Moins, Robert Kurowski, Karla Parussel, Paolo Nunes, Stefano Zampieri, Norm Hill, Viola Wang, Stewart Williams and many more

ALMA

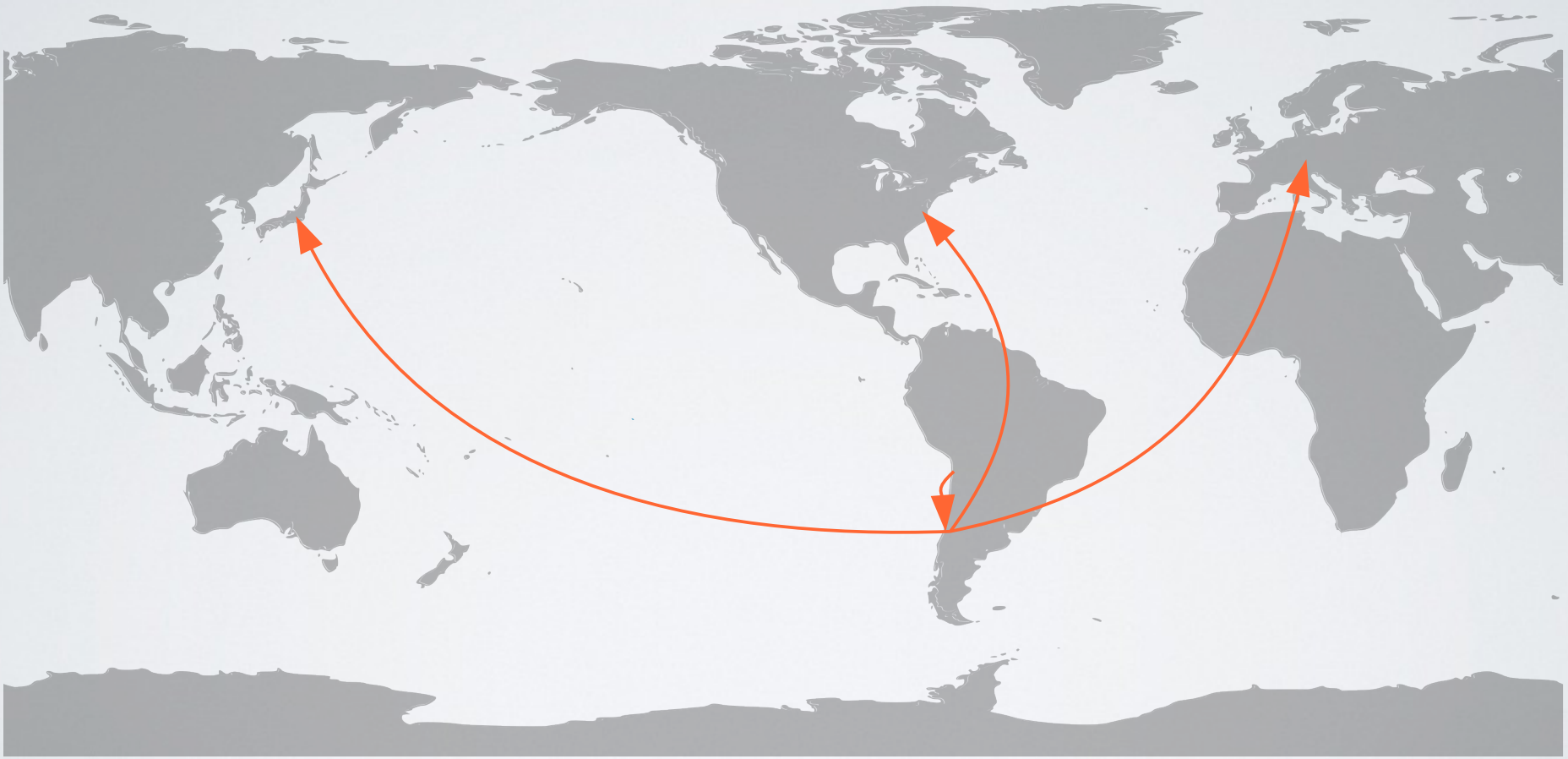
ALMA First Light

ALMA IN A NUTSHELL

- 66 antennas at 5000 m elevation in the Atacama Desert
- Interferometry with baselines up to 16 km
- Wavelength range from 3 mm to 400 μm (84 to 702 GHz)
- Build by ESO, NRAO, NAOJ in cooperation with Chile
- Data-rate in full operations: 6.6 Mbytes/s = 200 TB/year

- Timeline:
  - Early Science Cycle 0 data taking is under way
  - Early Science Cycle 1 Call for Proposals 31st of May
  - Cycle 1 data taking around beginning of 2013
  - Completion of the full array: second half of 2013

Data flow



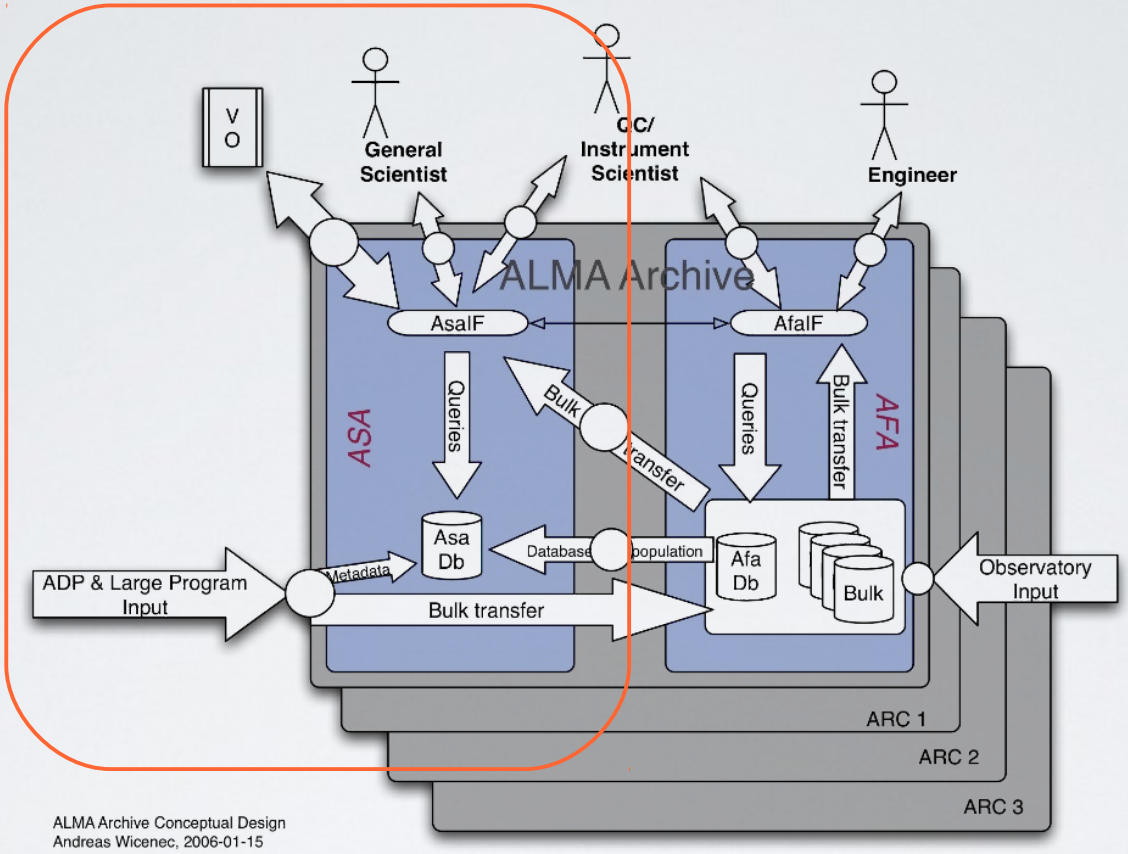
Science archive rationale

- As the success of ALMA is measured by the **science output of the community**, the goal is to render the ALMA user-experience for PIs and archival researchers as perfect as possible

- Creating a Science Archive that allows archival researchers to discover and retrieve the data they can use helps **maximise** the **scientific return** of ALMA

- Great return-for-investment ratio

ALMA Science Archive



Etoka, Fuller & Wicenec (2007), Etoka et al. (2008)

http://almasw.hq.eso.org/almasw/pub/Archive/ASADocuments/ASA_Requirements_2007-04-30ALMA-70.50.00.00-005-A-SPE.pdf

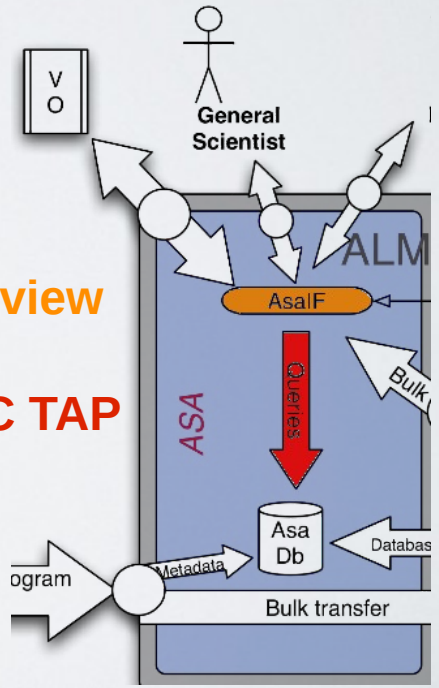
http://wiki.astrogrid.ac.uk/pub/Astrogrid/RadioAgenda/ALMA_Science_Archive_Oxford_081206.pdf

VO PLANS FOR ALMA

- **ASA 2.1-R6 The Archive shall be VO compliant**

- ALMA will provide
  - ObsTAP/TAP/SIAPv2 including STC-S footprints for the **science-grade products (data cubes)** once the pipeline creates these products routinely (2013/2014)
  - a WebSAMP enabled search interface

- ALMA (re)uses
  - Voview
  - OpenCADC TAP



VO plans for ALMA

- ALMA has the aim of providing in the future
  - A cutout service for data cubes
  - VOSpace (read-only for requests, read/write for user-contributed data)
  - User processing using UWS

- ALMA has no plans for providing
  - SSAP
  - IVOA-Single-sign-on
  - Simulations through the VO
  - Products other than cubes through the VO (e.g. moment maps, raw visibility/single-dish data)

How datacentres tick

- **Slowly**
- Large amounts of **inhomogeneous** of metadata and products: **very large effort** to bring even a small service
- Resources are **very limited**
- Strong **prioritization** and **long-term planning** in place: managers must be convinced of **added value** through VO.
  - "What can be done with VO that could not or not well be done before?"
  - "How much effort does it save us to use VO technologies?"
  - "Do we need TAP or can we do with keyword/value (Solr)?"
- The best moment of introducing VO services or technologies is when observatories/archives are **newly constructed** or when the infrastructure is **overhauled** completely

Can not "Just use PostgreSQL"

For a data centre, this might actually mean:

- Conduct a RDBMS cost study
- Get approval and prioritization from management
- Plan for the hardware including standby and backup servers
- Purchase request of the hardware with maintenance contract
- Purchase request for external PostgreSQL on-site support
- Secure DBA support (maybe changing outsourcing contracts)
- Additional PostgreSQL DBA training for current DBAs
- Secure a budget and issue a call for tender
- Write software requirements
- Wait for the availability of software development resources, development, testing and roll-out

Wishlist

- **Standard** tools that many others use (do not abandon voview)
- Drop-in **generic** tools (Java is great here) that do not require a complete system in place (OpenCADC TAP). This makes technolgy take-up cheap.
- Final **metadata definition for SIAPv2**
- **IVOA data identifiers** to link data and publications: unique, persistent and resolving service to multiple sites
- **Homogeneous and slick user-experience** of WebSAMP enabled tools (voview, Aladin, Topcat, voplot)

Completely personal

- Keep it **simple**, go for the 80/20 solution
- No "**evolving standards**": No resources available to keep up.
- **Backwards compatibility** ("Note that all prior clients will not work with this new version of the interface)"
- **Less is more**: Too many standards/protocols. We will not have the energy to implement all of them.
- ObsTAP is great, **push for it**!
  - Simple
  - Reasonably easy to fill: potential uptake is large
  - Describes the **4 physical** axes: can ask **scientific questions**!
  - Mandatory keywords: allow **blind cross-archive searches** other than just position

Even more personal

- The real power is in the discovery of data that goes **not** through the positional search. "Give me all spectra of quasars with z>1.5 and resolving power>1000 ", “Give me places on the sky that have been observed in the IR and in the UV”
- Interoperability also means that the **critical mass** is achieved. Better having less protocols and large uptake
- It would be great to have **solid statistics** to convince managers of the usefulness of the VO, e.g.
  - The number and relevance of **science** papers in refereed journals that only could be done thanks to the VO
  - A **measure** of how much the work of the scientists is actually made easier due to the VO