

X Reunión Científica de la SEA
Valencia, 9 de Julio de 2012

How the V0 helped building the ALMA Science Archive

Juan de Dios Santander Vela (IAA-CSIC)

Felix Stoehr (ESO, ALMA Archive Scientist)

Talk Outline

- * Who am I
- * The problem: Archiving & accessing ALMA data
- * The solution: ALMA Science Archive
- * The enabler: VO Technologies
- * Implementing the ASA
- * Conclusions & Future work

Who am I

EXPERIENCE ON (RADIO) ASTRONOMICAL DATA,
METADATA, VIRTUAL OBSERVATORY,
ASTRONOMICAL OBSERVATIONS, ARCHIVES, AND
THE ALMA SCIENCE DATA MODEL

- * Member of the **AMIGA** international collaboration, based at IAA-CSIC **GROUP INTEREST IN TECH DEVELOPMENTS FOR BETTER SCIENCE**
- * Ph.D. on bringing Radio Astronomical data archives and tools into the VO
- * Applied Scientist at ESO VLT archive, Software Engineer/Astronomy Specialist at ALMA archive (May 2009-Dec 2011)
- * Back to IAA-CSIC as VIA-SKA Project Manager

THE PROBLEM

Archiving & Accessing ALMA Data

Archiving & Accessing ALMA Data



Archiving & Accessing ALMA Data

- * Right now (ALMA Cycle 1)
 - * More than 32 antennas (12m) in Chajnantor
 - * More than 9 antennas (7m) in ACA
 - * Baselines 160m to 1km (more than 500 baselines)
 - * Receivers for 4-bands (3,4,6,9), 8-10-12 GHz bandwidths

Archiving & Accessing ALMA Data

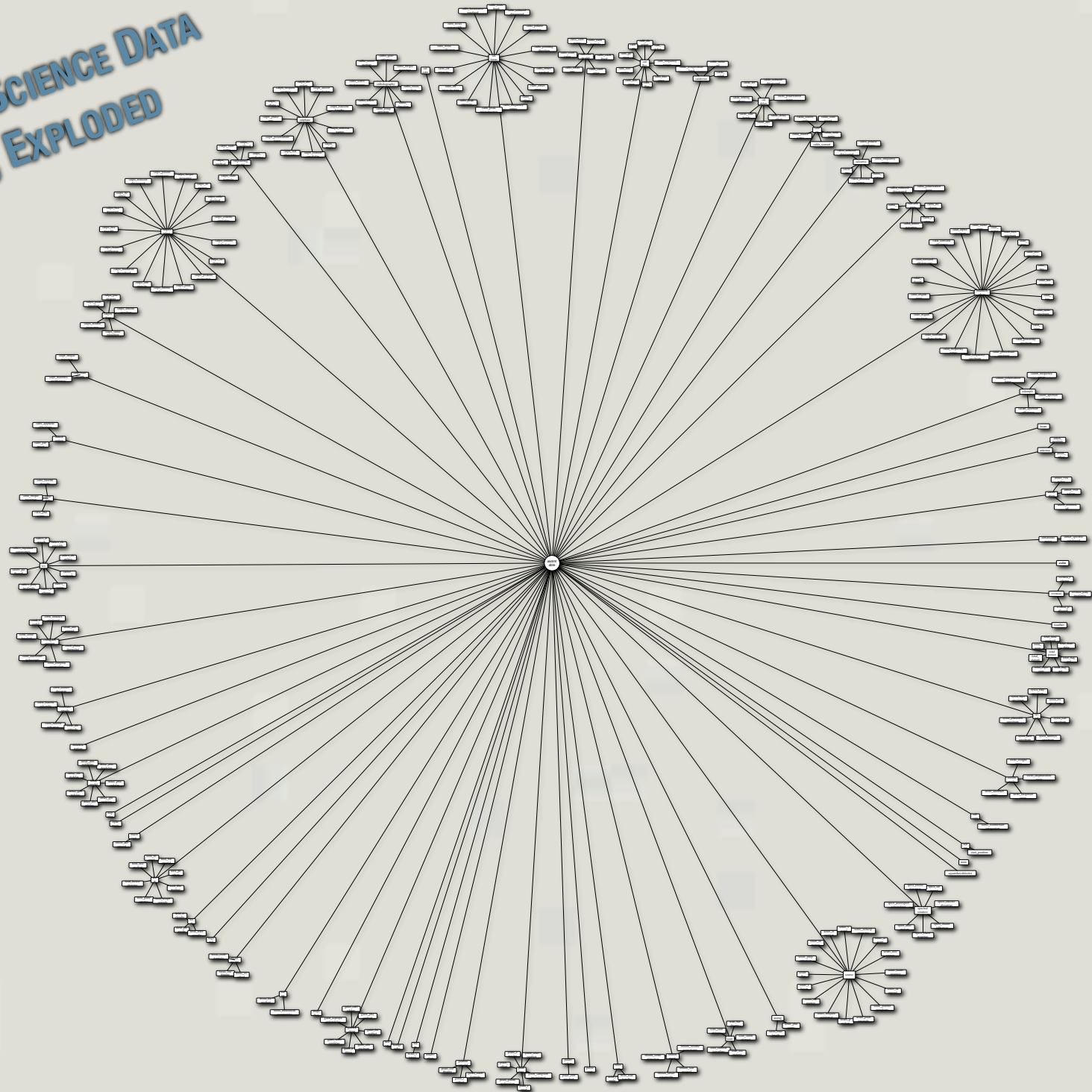
**LARGE AMOUNT OF DATA AND
METADATA TO BE STORED IN ORDER
TO BE LATER PROCESSED**

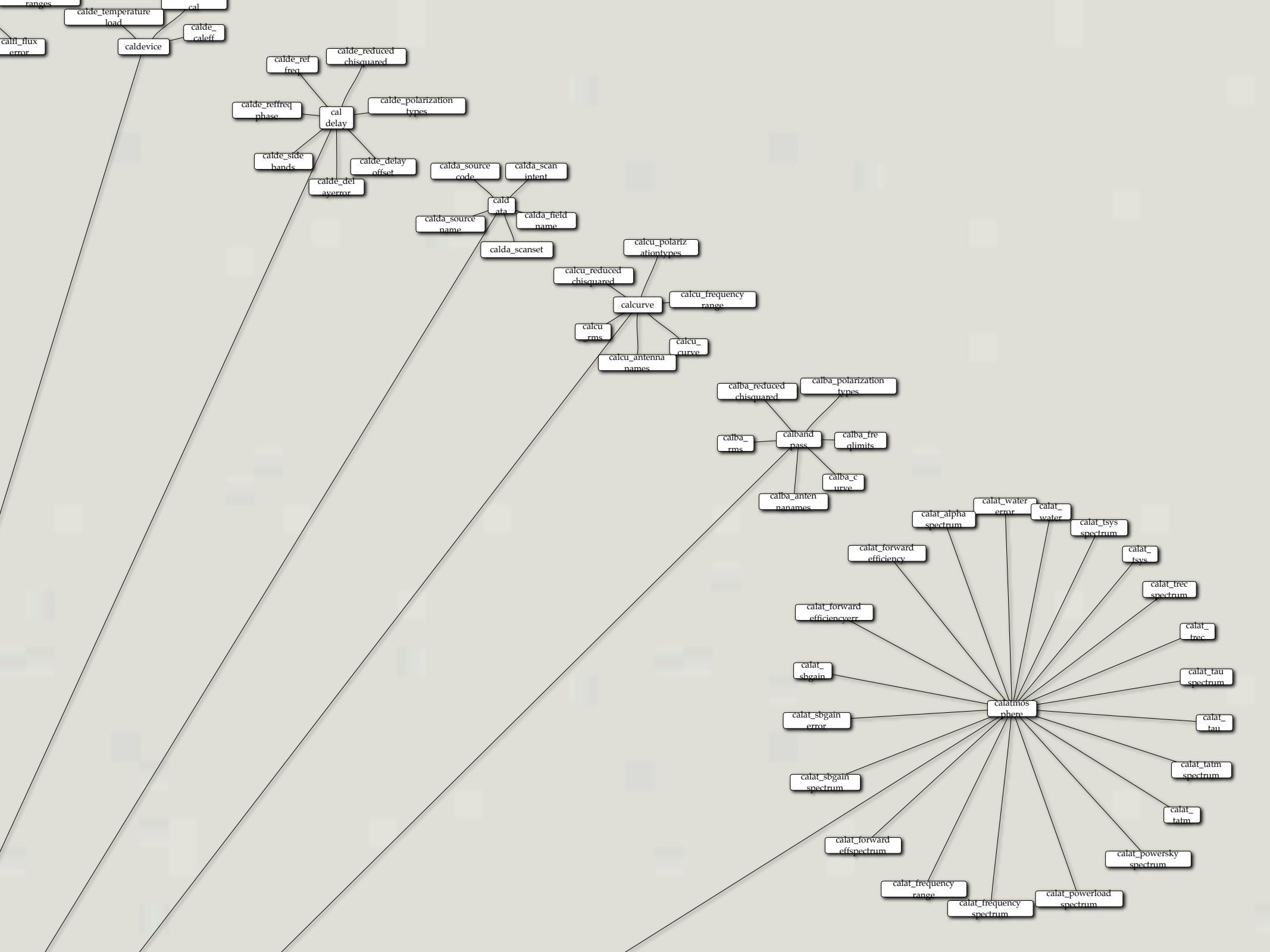
- * When finished...
 - * More than 50 antennas (12m) in Chajnantor
 - * 12 antennas (7m) in ACA
 - * Around 1300 baselines, up to 18 km
 - * Receivers for 10-bands (30-950 GHz),
8-12 GHz bandwidth, SSB/DSB

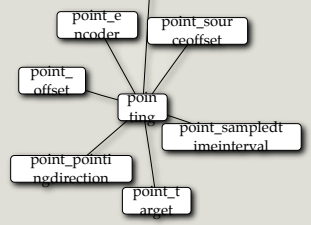
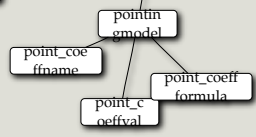
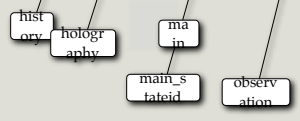
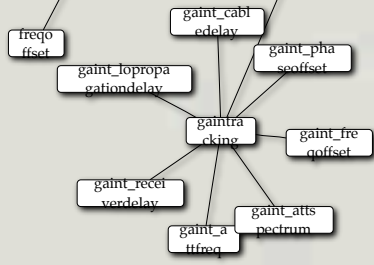
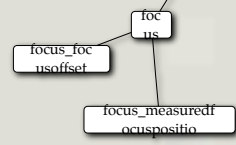
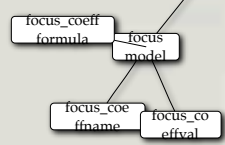
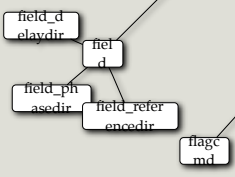
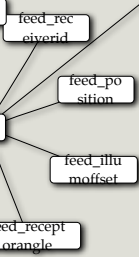
Archiving & Accessing ALMA Data

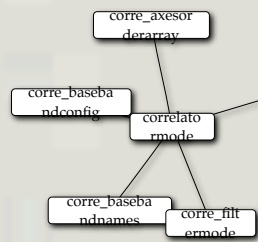
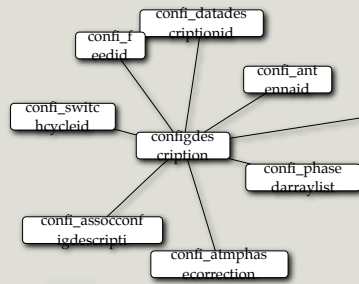
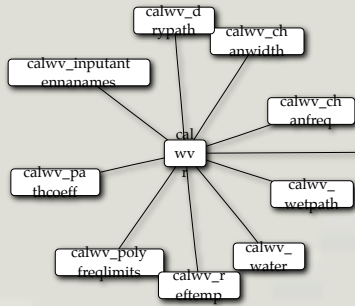
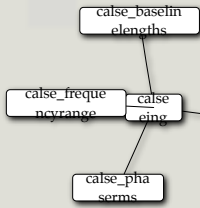
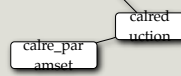
- * Large amounts of entities, data, timestamps to be stored, with high data rate
- * ALMA Science Data Model (Viallefond, Caillat, others) ➔ Science Data Model (ALMA, EVLA...)
 - * XML Database implementation of CASA MS
- * ALMA Project Data Model
 - * Projects, PIs, Science Goals...

THE ALMA SCIENCE DATA MODEL, EXPLODED

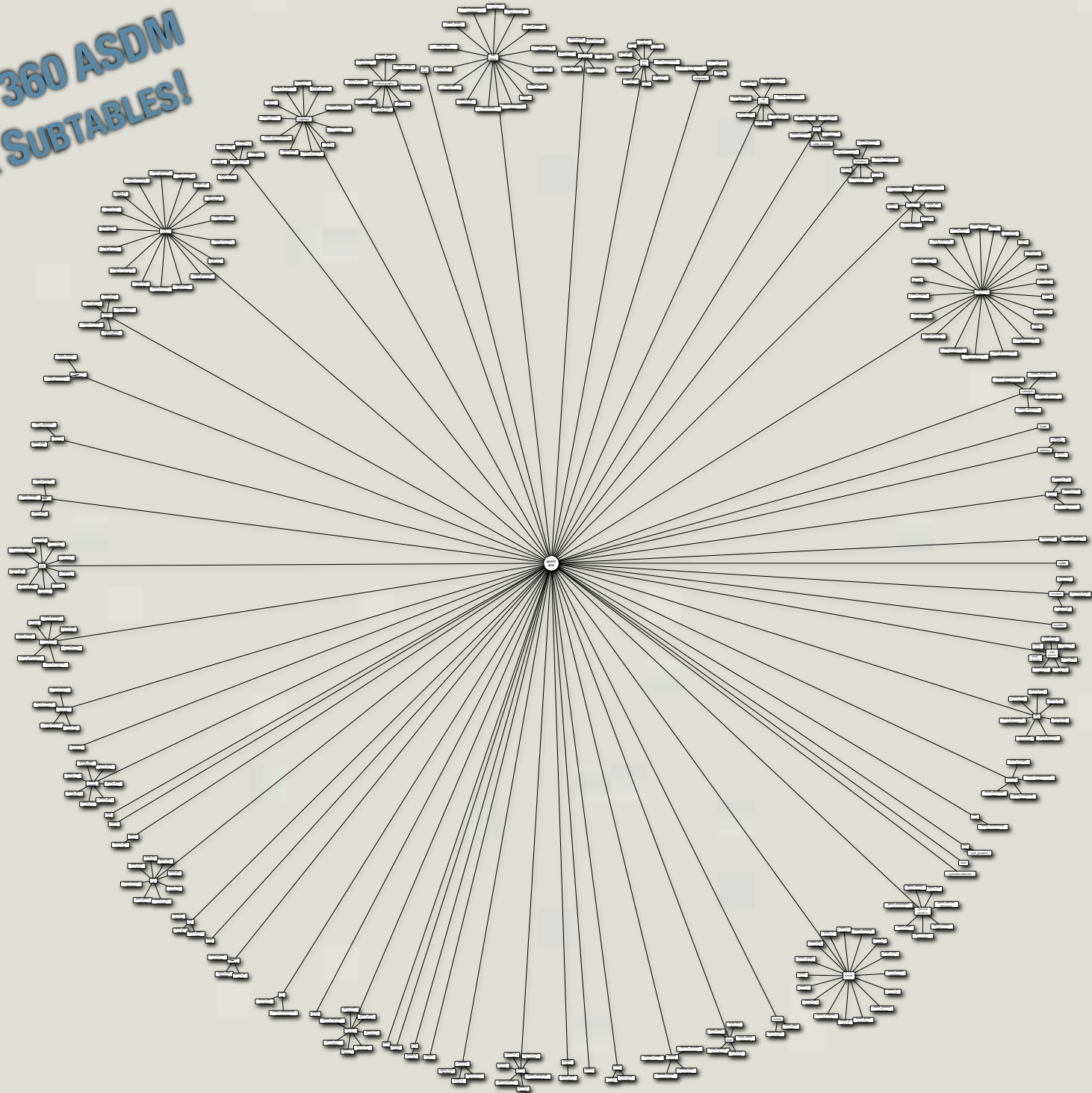








**MORE THAN 360 ASDM
TABLES & SUBTABLES!**



Archiving & Accessing ALMA Data

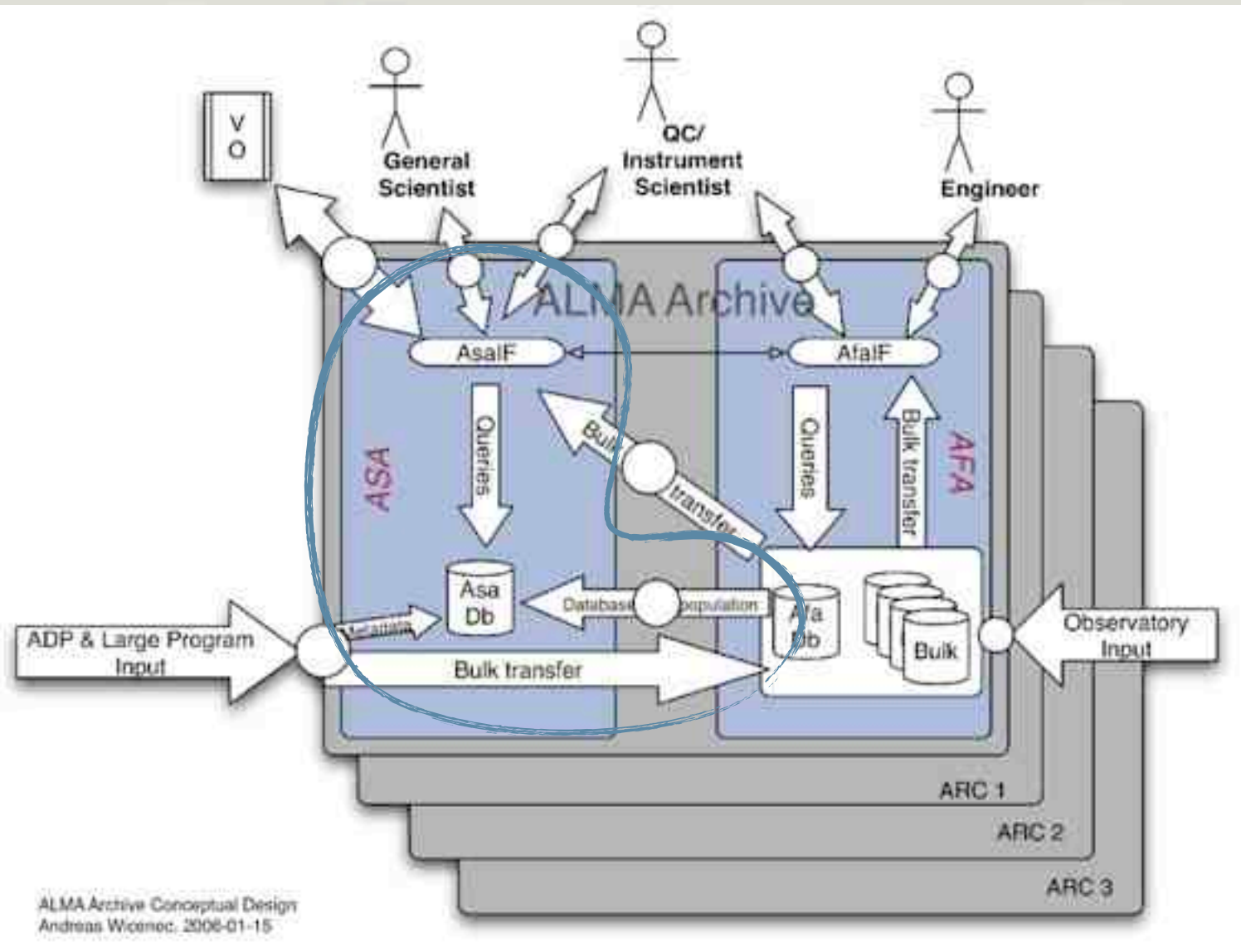
- * ASDM, APDM are CASA-friendly, not astronomer/science friendly
- * ALMA Frontend Archive, **optimized for storage & preservation**, not for data query/retrieval
- * We need an ALMA Science Archive, at least for when there are science data to be released
- * We cannot build it on top of the ASDM

directly

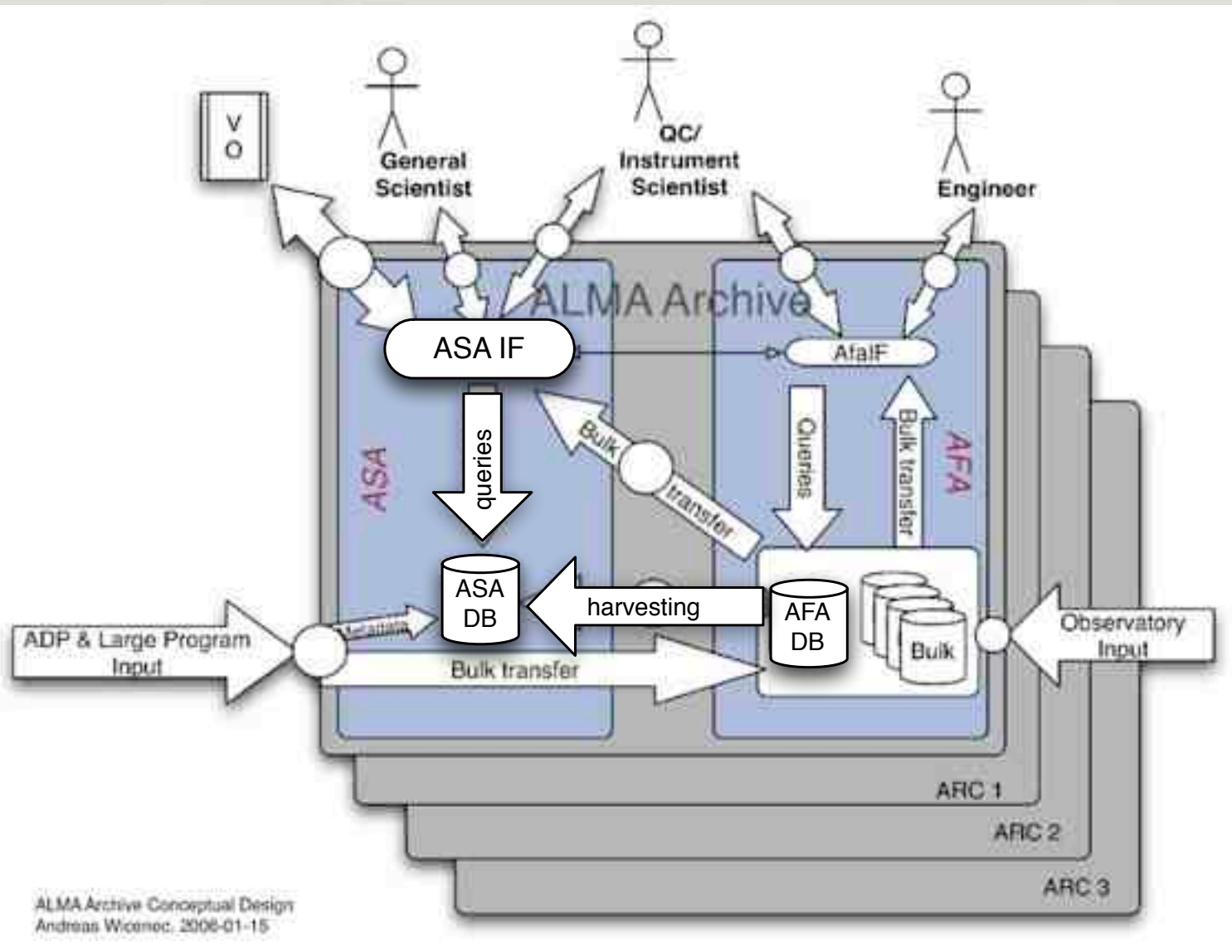
THE SOLUTION

ALMA Science Archive

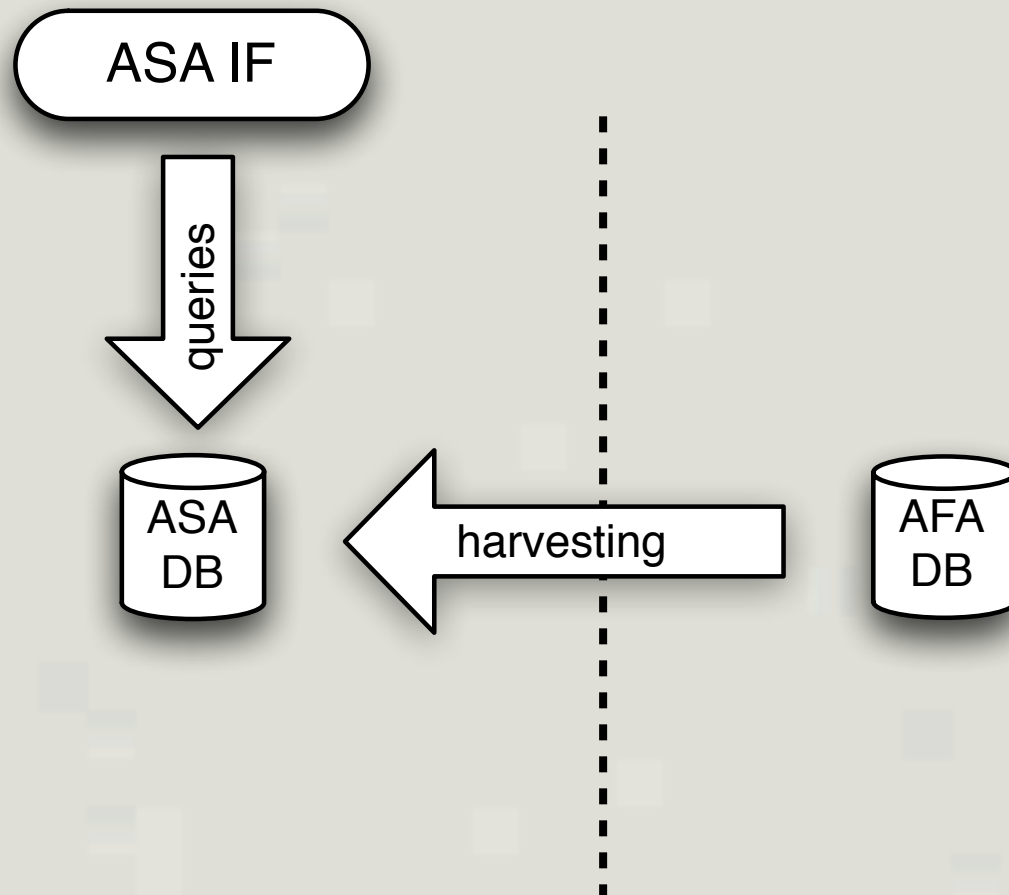
ALMA Science Archive



ALMA Science Archive



ALMA Science Archive



- * No changes to AFA DB
- * Harvesting takes **selected** datasets from AFA DB into ASA DB
- * ASA DB consists of **scientific relevant** parameters
- * ASA IF supports the different clients: engineers, scientists, **and the VO**

ASA Elements: Harvester



- * **Lightweight process.** Current ALMA holdings can be reprocessed in less than 4 hours
- * Ensures **only correct data** enters the ASA
- * Performs tasks such as:
 - * ASDM parsing & consolidation
 - * ASDM consistency checks
 - * coordinate conversion
 - * solar system body identification
 - * baseline reprojection

ASA Elements: ASA DB



- * Plain relational SQL DB
- * Based on Hubble Legacy Archive experience
- * Denormalized Structure
 - * Science
 - * Project
 - * SpectralWindows
 - * Provenance
- * Supports pipeline reduced and externally **reduced visibilities & data products**
- * Allows **joint querying of all product types**

*INSPIRED BY NOA
OBSCORE DATA MODEL,
ASDM*

ASA Elements: Interfaces

ASA IF

- * ASA Interfaces built as web application
- * **VO metadata-driven**
- * Clients supported:
 - * QA/CSV: technical parameters (scheduling blocks, scan intents)
 - * Scientists: science parameters
 - * Resolution (spatial, frequency & velocity)
 - * Positional and target searches
 - * **Programmatic/VO interface**

THE ENABLER

VO Technologies for the ALMA Archive

VO Technologies for the ALMA Archive

- * Virtual Observatory: **federation of archives** sharing a set of **common practices** and **data models** that allow for easy **discoverability** of interoperable data-sets, with **unified description** by means of a **common data model**, within the realm of astronomy.
- * Data model: Description of the **set of entities** needed for information storage in a particular field, specifying both the data being stored, and the relationships between them.

**NEED OF SOFTWARE TO
IMPLEMENT IT**

VO Technologies for the ALMA Archive

- * VO Data Models



- * ASA DB structure inspired from ObsCore, RADAMS, Hubble Legacy Archive

- * VO Software

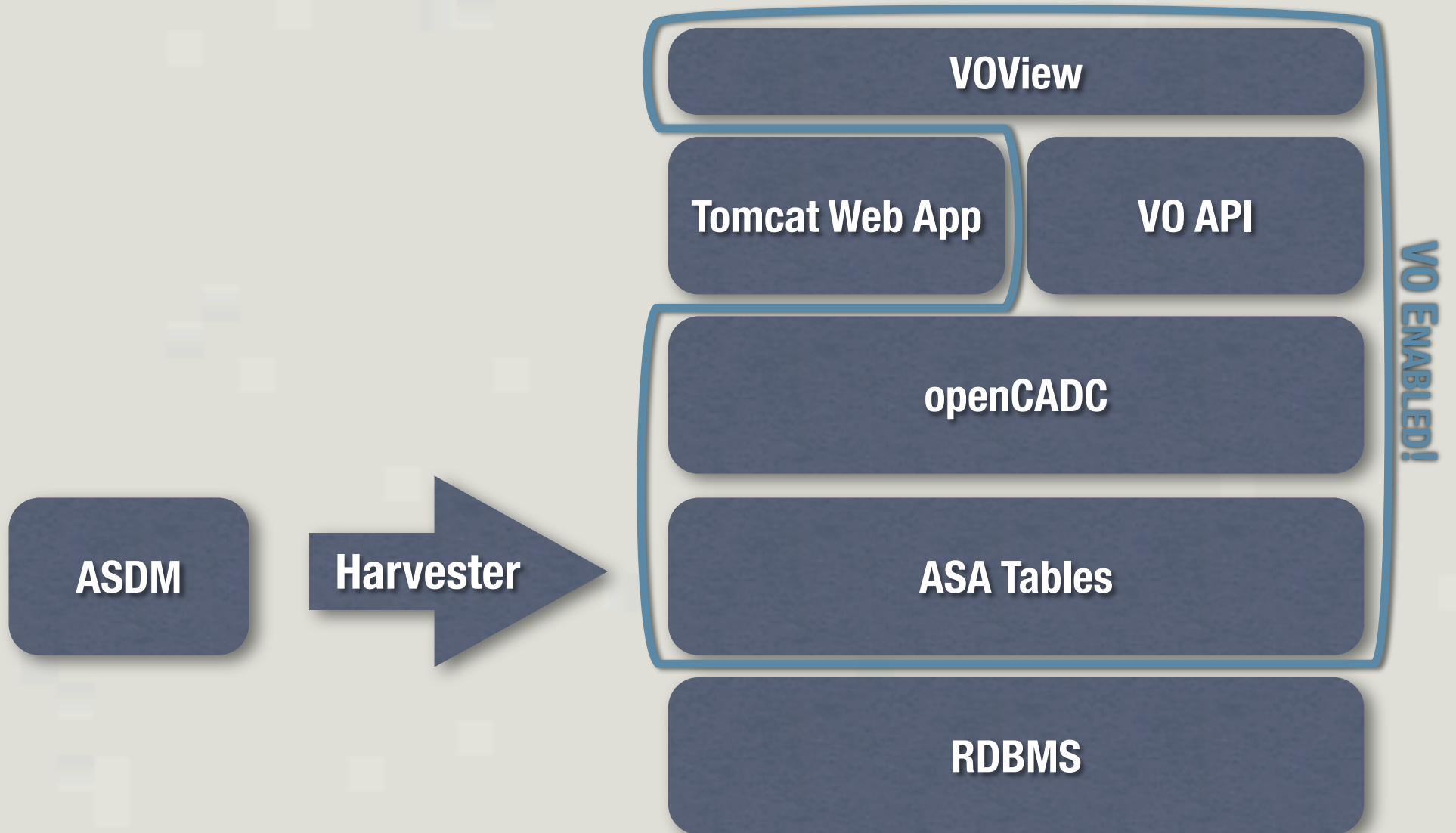
**MAXIMISATION OF SOFTWARE REUSE,
MAXIMISATION OF OUTPUT PER DEVELOPER**

- * openCADC (DB access, VO access protocols, metadata management)
- * VOView (VOTable rendering web component)

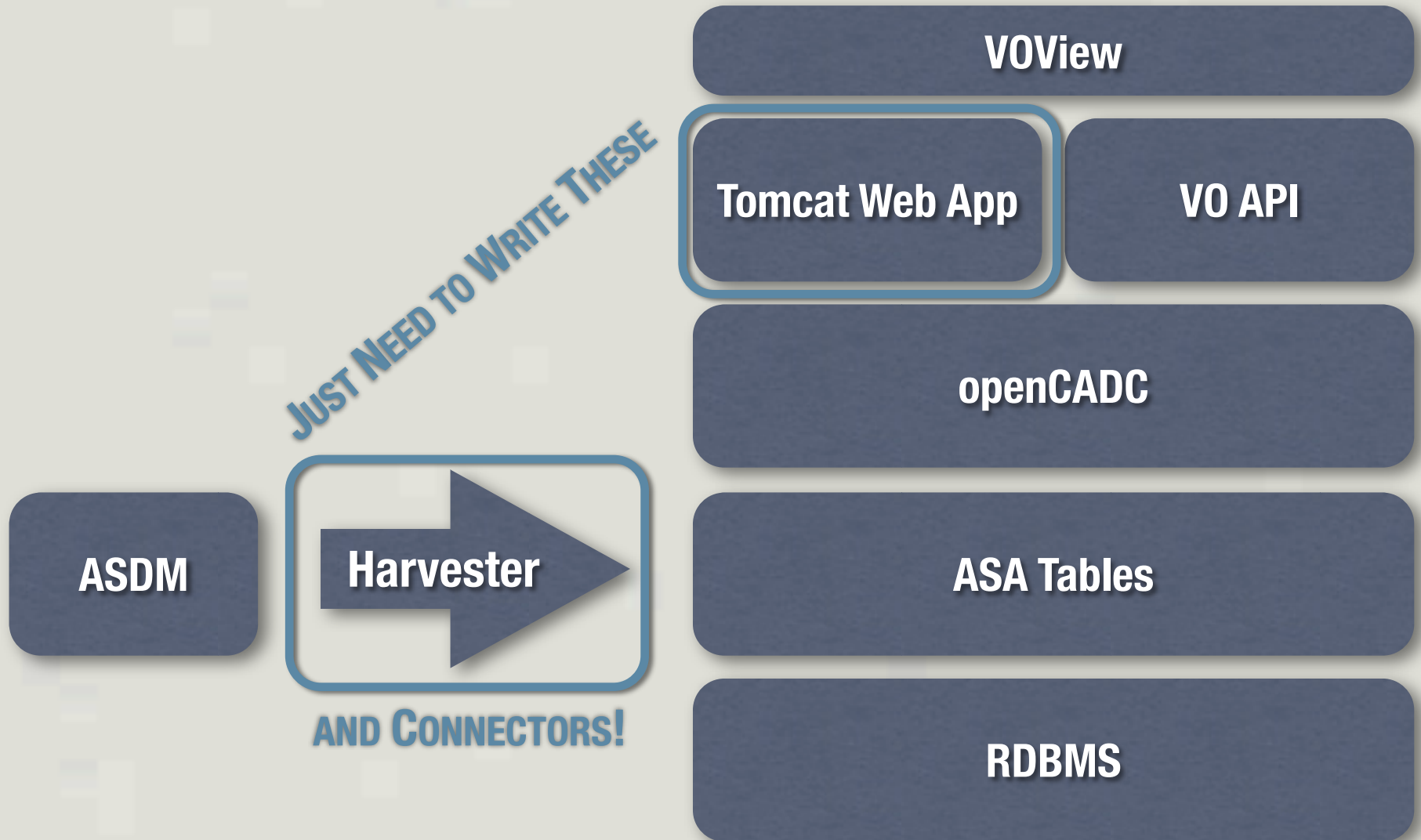
THE RESULT

ASA Implementation

ASA Implementation



ASA Implementation



Demo

Conclusions & Future Work

Conclusions & Future Work

- * The separation of the ASA & AFA allows for **different parameter spaces to be queried** (instrumental vs. scientific) on the ALMA archive
- * Using **IVOA standards** for metadata allows us to use **VO solutions to implement a science archive**
 - * And **reduce the manpower** to implement it!
- * **VO software is mature enough** to implement a

AND YOU GET VO COMPATIBILITY FOR FREE!

Conclusions & Future Work

- ✱ Interface being simplified, even more VO metadata driven
- ✱ Expected **to go public by the end of 2012**, when the first reduced project data end their proprietary periods

Conclusions & Future Work

FUTURE INTERFACE
MOCKUP



Atacama Large Millimeter/Submillimeter Array

In search of our Cosmic Origins

Home -> Archive query -> Science query

Science Query

Query Form

Result Table

Get Data

Search

Reset

Position

Source name (Sesame)
Source name (ALMA)
RA Dec
Search radius

00:10:00

Energy

Band
Frequency
Bandwidth
Spectral resolution
Channels

Time

Observation date
Integration time

Polarization

Polarization type

Observation

Project code
Water vapor
Scan intent
Scheduling Block name
Pad/Antenna name

¡Gracias!

Gràcies!

Thanks!

Questions?