# ALMA Archive development

Felix Stoehr

Subsystem Scientist of the ALMA Science Archive

# Outline

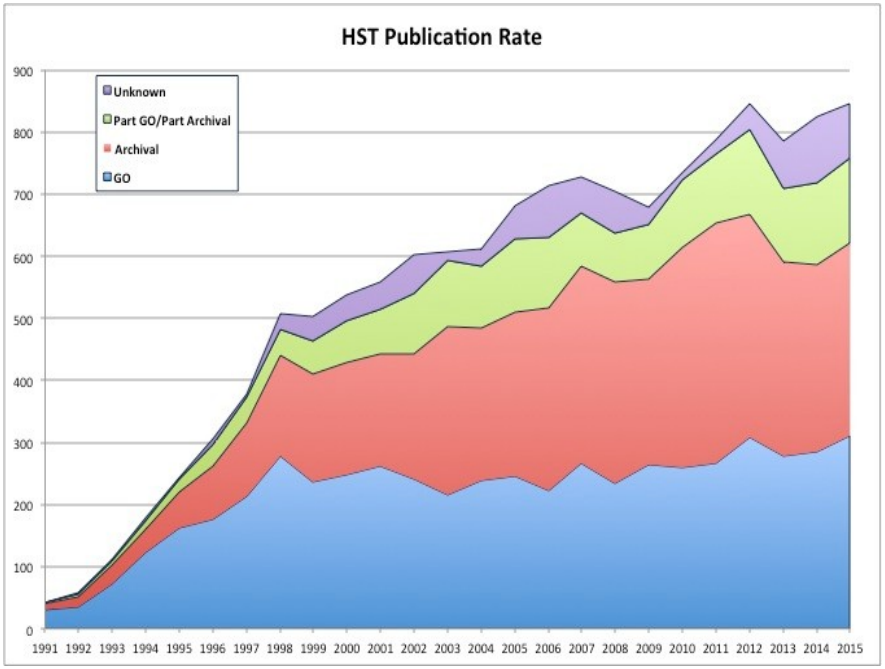
- Science Archive Rationale
- Astronomy of the future
- Archive development plan
- Potential Development Projects

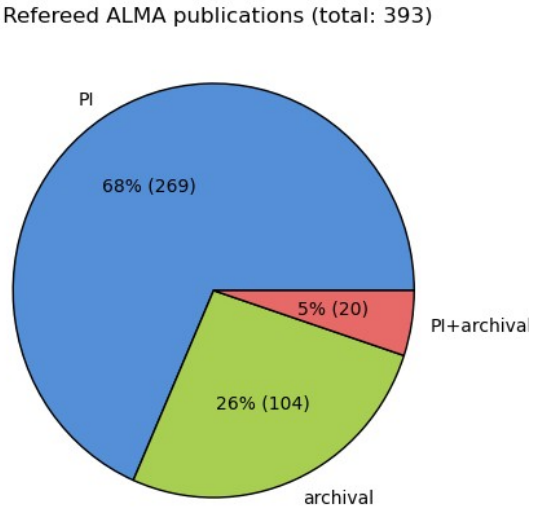
# Rationale

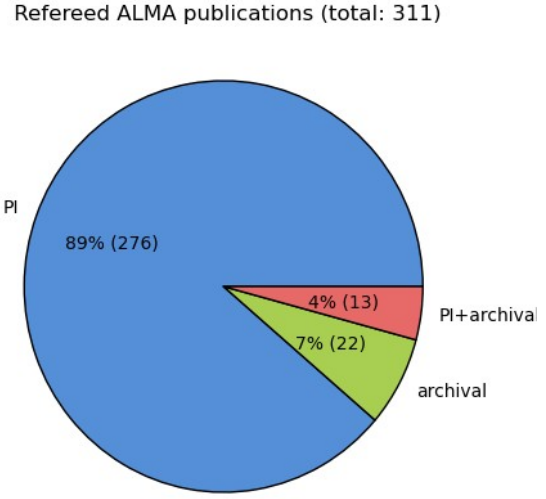
# Science Archive Rationale

- The **success** of an observatory is measured by **the science output of the community**
- **We can not make our own success**: **someone else** is responsible for our success
- The goal for us therefore must be to render the **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

# Science Archive Rationale



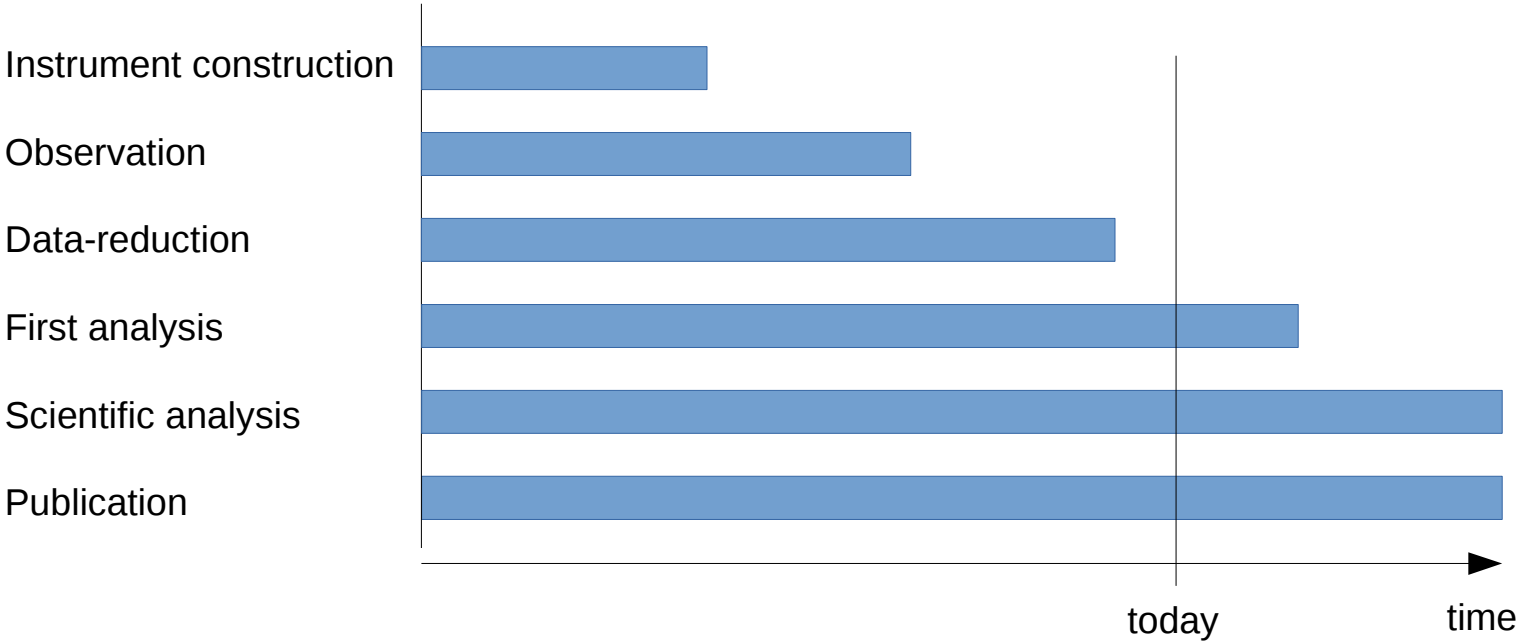



# Future of astronomy

# Evolution of the workload

- What do PIs do?




- Workload is shifting **from PIs to observatories**
- More complex telescopes and more data: **pipelines** and **science-grade** data-products will become integral part of the telescope/instrument design (e.g. ALMA pipeline)
- Next step: **first science analysis** done at the observatories (e.g. ALMA development program ADMIT)

# Future of Astronomy

- ALMA will produce about the same amount of data in one year as ESO has produced in its first 50 years. (And ESO will, too!)
- LSST, LOFAR, SKA, PanStars, Euclid, Gaia, ELTs

- T. Tyson: “Astronomy is **transformed** from being a data-starved science to one where data is overabundant”
- Multi-wavelength science: **less time** per wavelength regime
- **Astronomers do not scale**: bytes/astronomer grow exponentially

- My prediction:

  Now: astronomers **compete for observing time**
- Future: - observatories will **compete for astronomers** to work with their data

  - astronomers will **still compete for observing time**

# Evolution of Science

- This will have a **deep impact** on how science is done
- Astronomers will become **consumers of data** instead of being co-producers. They can **concentrate** on the science
  - Good: astronomy makes best use of astronomers, its future rare resource. → The science output will increase.
  - Risk: astronomers may not understand data limitations
- Observatories will have a much **larger responsibility**
- In some areas: **data-mining** will gain importance
- It will be easier to have an **archival career**

# Science Archive

# Science Archive Design

- speak the **language of the scientists.** Query by **physical concepts**

> **DANIEL DURAND TEST** of science archives:
> You know have one, when you can find your own data by only giving physical parameter constraints

- provide only **relevant** information, keep hurdles **small**
- Strive for great user-interface design
- help users to **quickly decide** whether or not the data returned in a given search is relevant for them
- All public data and metadata is accessible **anonymously**
- **PIs are normal archive users with more access rights**
- Authors are **required** to acknowledge the use of ALMA data using a standard data tag

# ALMA development

- Full generic products for all raw data (pipeline!)

- Collapsed rows (multiple ASDMs, mosaics)
- More and better metadata (rms, redshifts/rest frequencies?)
- Access to individual products
- VO (ObsTAP, SIAPv2, DataLink, cutout using DataAccess)
- Previews
- CARTA integration
- ADMIT integration
- PPI
- Faster search
- visual search and selection (AladinLite)
- ...
- currently 205 tickets are waiting for implementation for Archive

# Development projects

# Development projects

**Reimaging project**

- Once pipeline imaging is accepted by the project:
  - Run pipeline imaging on all calibrated MS from Cycles 0, 1, 2 and 3 (manual and pipeline calibration)
  - prototype is working (290 MS reimaged so far)
  - do QA2 on the products (?)

- probably should run at (EU) ARC for better data access. JAO & Italian ARC node showed already interest

- Depends on Pipeline's timescales

# Development projects

**Pipeline reprocessing as a service**

- NA has developed the software for the VLA
- prototype is working
- Should use VO standards (i.e. UWS)

- Could also be a normal ALMA Archive development but resources are extremely tight

# Development projects

**ADMIT2**

- ADMIT was a NA development program
- Software has been delivered to the project (final: November 2016)
- Improvements are certainly possible and desireable

- Would probably make most sense if the same NA team would continue but it could make sense to have other teams join

# Development projects

**ALMA Source list**

- Provide a curated catalogue of all ALMA sources using the best data-reduction and combining all observations of a source
- Longer-term project (e.g. HST delivered their first list after 25 years of operations)

- Very strong science component

# Development projects

**Recommender system**

- Allow archival researchers to see “similar” data or similar publications or similar proposals
- Idea and proof-of-concept study using machine learning by Alejandro Barrientos (JAO)
- Is planned as a standard ALMA Archive development but probably could be singled out as a development program

- Probably a single FTE would be more than enough

# Development projects

**Non-archival development program:**

**Scheduler**

- global optimization solution for the entire cycle running e.g. every 30 minutes using
- Taking into account: weather, weather prediction, historical weather, bands in antennas, source positions/LST ranges, calibrators, ranks, antenna move lead time, antenna move cost, antenna move, constraints (staff, transporter), start and end antenna configuration for the cycle as boundary conditions, observatory shutdown, scheduled EOC time, expected QA0 failure rates for the different modes (taken from the last cycles), time estimates for the SBs and fractional execution counts, ToOs, Timeconstraint observations, newly added DDTs, number of antennas including the scheduled maintenance of antennas, subarrays, ACA standalone, ...
- Can provide full statistics and schedule from day 1 of a Cycle which could be accessible also to PIs

- This is an ALMA development but could maybe singled out, maybe in cooperation with SAP