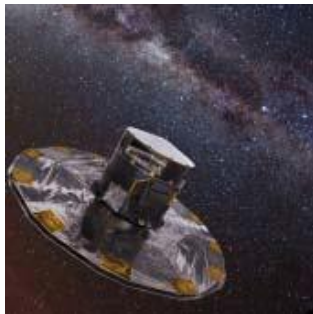


# Data sharing in astronomy – The role of Research Infrastructures in quality and trust

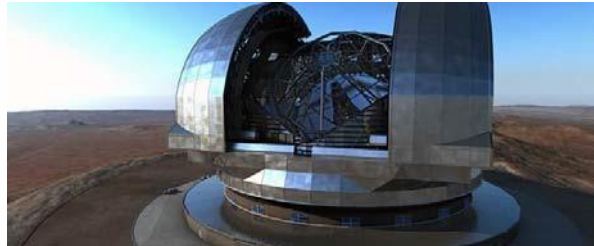
Françoise Genova, CDS, Strasbourg  
astronomical Observatory



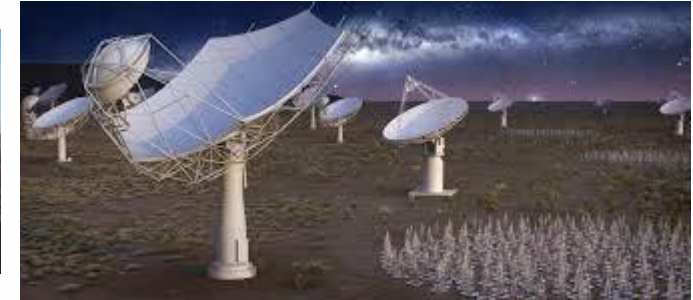
# □ Research Infrastructures in astronomy



Gaia



ELT



SKA

CTA



Planck



Herschel

**+ DATA**

# □ Data as a Research Infrastructure

- Astronomers routinely use data they retrieve on line in their daily research work
- The astronomical data RI has many components
  - Observatory archives
  - Very Large surveys
  - Value-added databases, e.g. CDS (Strasbourg astronomical Data Center)
  - Journals
  - Modeling data
- Astronomers trust the data providers, which have an established role in the community context
- Trust is not only linked to data quality, but also to the « quality » of the different elements of the data sharing system, including the fact that the system is relevant to users' needs

# □ Data sharing in astronomy: accessibility AND reusability

- Early pioneers – IUE 1978-1996, CDS 1972
- International collaboration on standards
  - Format (FITS) 1979
  - Bibliographic id 1989
  - Interoperability of data and tools
    - Standards defined by the IVOA (since 2002)
    - Open and inclusive framework – anyone can « publish » a data resource in the VO, anyone can develop a VO-enabled tool to access data
    - More than 100 « authorities » provide a resource in the VO, including all the large data providers
- Astronomy data is FAIR thanks to the data providers and the VO developers





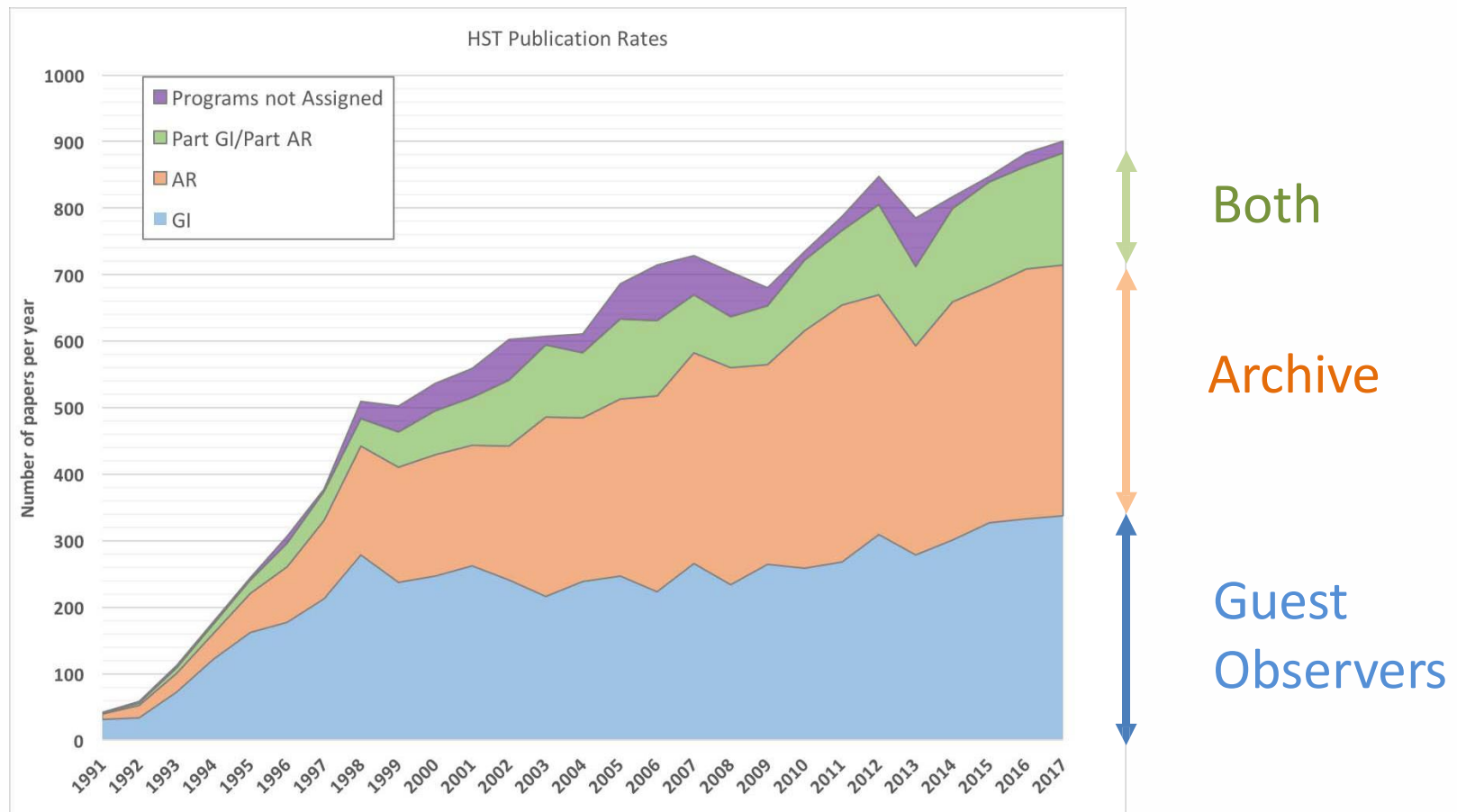
# □ RI data policy in astronomy Observatories

- Observation time is obtained through often tough competitive process
- Observatories make their data available after a proprietary period (in general 1 year)
- Proprietary period an important factor
  - To make the open data policy acceptable by the community
  - To continue to have the best possible observation proposals – ie to build trust in the archive content!

# □ RI data management and quality Observatories

- Data management is included in the mission and budget of the RIs or of the agencies which manage them
- They provide data to observers and make them public in their archives
- Data is Reusable and for most observatories available in the VO (FAI)
- Significant effect on RI impact –  
« good »/useful/ « trusted » data is reused

# □ Publications using HST data



# □ Data management and quality

## Value-added data service - CDS

- CDS is a RI in the French National RI Roadmap
- Fully trusted by the community
  - ~1 000 000 queries/day on the services
  - Services used by observatories, research agencies & journals for their own needs
- Data curation & services to access data
- CDS DSA & WDS certified (now applying to CTS)
  - Already trusted by its community but important wrt. CDS evaluation by the rest of the world including the funders
- Data from published papers, large surveys and selected data from observatories
- Quality ensured by an integrated team of astronomers, specialized librarians and IT engineers
- Expertise built on 46 years: quality of the content, also quality of the services (functionalities, operations) wrt. user needs and expectations





# □ Conclusions

- Data sharing does change the way science is done and boosts the RI impact when well done (ie in a trustable and trusted way)
- Lots of work behind the scene on data management & stewardship, standards and tools
- Quality/relevance rely on expertise built on the long term including disciplinary knowledge and a deep knowledge of their instruments for the observatories
- All disciplines are different but lessons learnt can be shared

# □ Implications

- Very long term endeavour – sustainable support a must
- Data Sharing frameworks should be built taking community requirements and feedback into account, including from RIs
- Enable collaboration at the European & international levels
  - Cluster projects are a good vehicle when well targeted
  - Create/find an appropriate international forum for disciplinary discussions (specific, generic such as RDA)
- *Trust is not only linked to data quality, but also to the « quality » of the different elements of the data sharing system, including the fact that the system is relevant to users' needs*
- Quality/relevance driven by science needs, neither technology nor policy demand nor data conservation – although the three play a role

