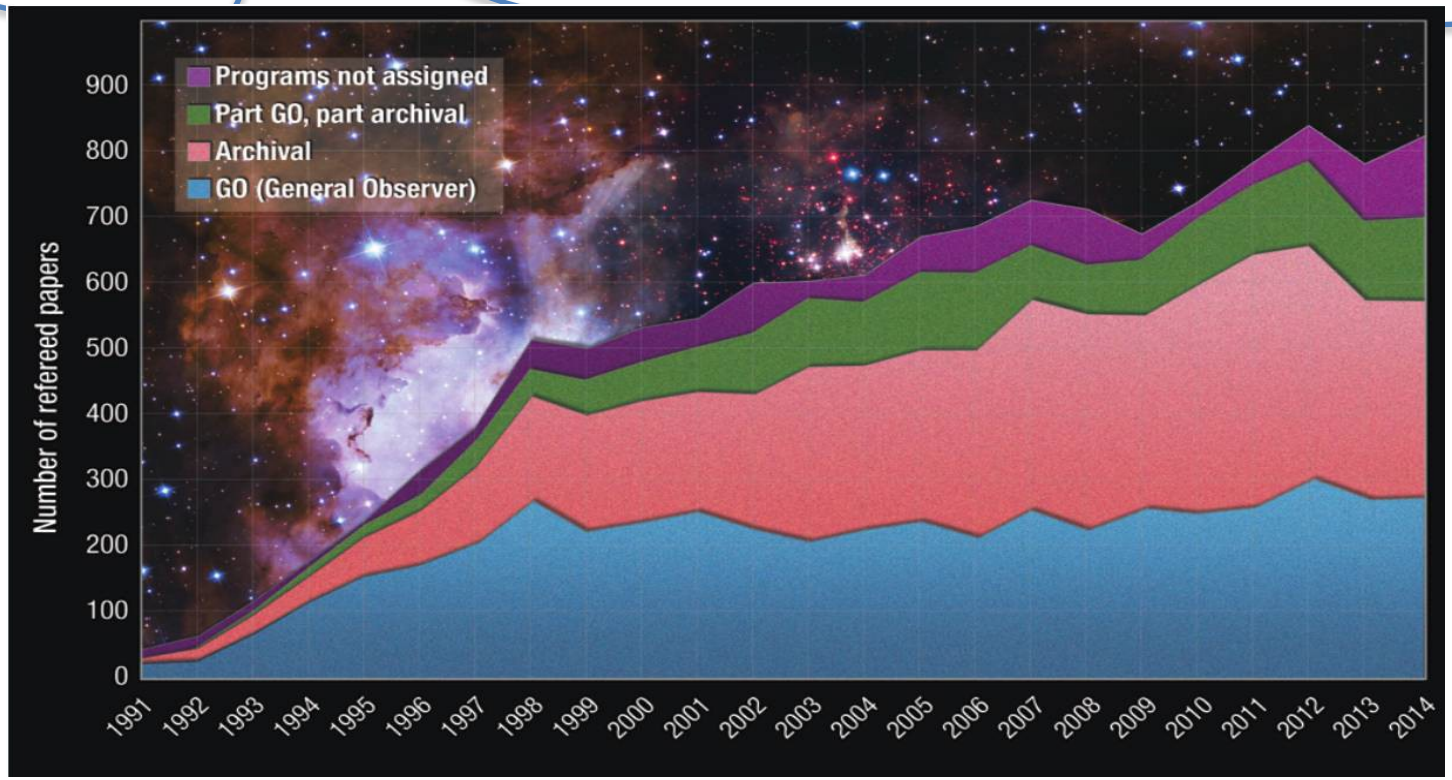
INAF Strategy in the Big Data Era

R. Smareglia
Head of ICT and Data Management office of INAF

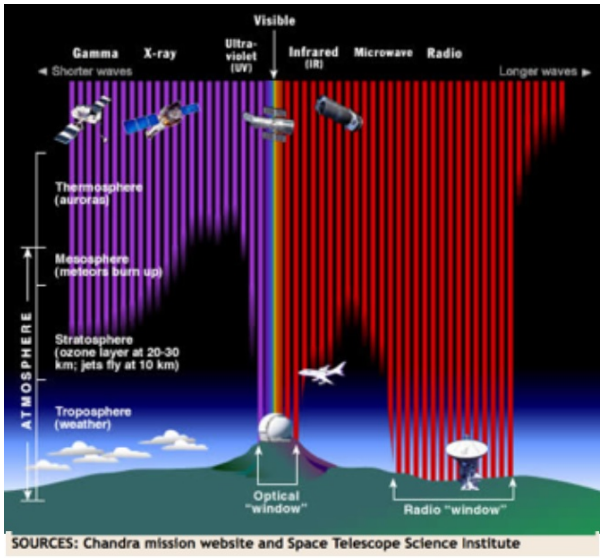
GARR Workshop, Venezia 15-17 Novembre 2017

L'importanza dell'Archivio



HST Newsletter: "*At the present time, approximately **half of the refereed publications** based on Hubble observations are derived purely **from archival data**, and, every year, this number is slightly higher than the number of publications based on new observations. .... the Hubble Archive has become a goldmine for the astronomical community....*"

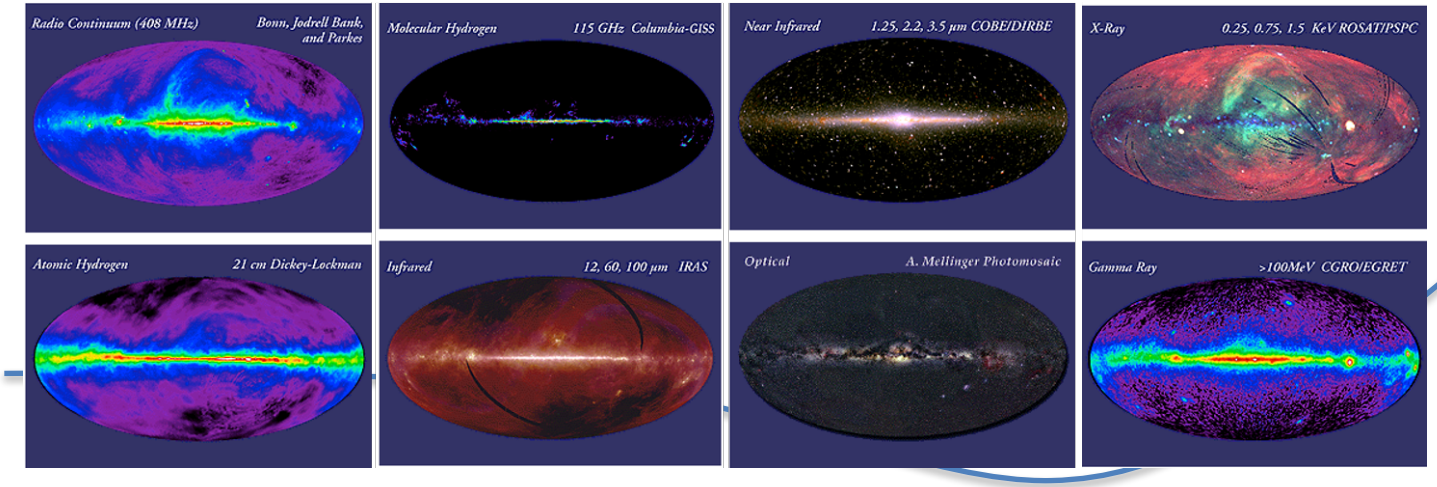
Cosa guarda l'Astronomia



Una visione parziale delle osservazioni astronomiche nei prossimi 10 anni

Alla lista poi mancano:

- Diversi osservatori
- Le missioni planetarie
- Esopianeti
- Osservazione del Sole
- Cosmic ray exp.



Richiesta Principale:
Interoperability ➔ VO ➔ EOSC

Distributed resources

- International team members can bring regional resources
- Big data: moving code to data
- Resources are not simple

Science teams

- Science teams are international virtual organisations
  - Forming around a given multi-year project
  - Handling large datasets
  - Faced with acquiring and building project infrastructure
- Require infrastructure
  - Larger datasets
  - Data management, data distribution, data processing
  - Challenging a team's ability to produce and maintain infrastructure
- May have access to national and regional infrastructure

The VO is a paradigm for Supporting interdisciplinary and collaborative research in astronomy and exploiting the full power of growing and emerging data sets

The VO is a framework

- For data centers to provide co-operating data services,
- For software providers to offer a variety of compatible analysis and visualization tools and user interfaces

Resource ➔Intellectual Propriety

- INAF had more then 90 software package developed, some public, many “locally” to be engineered
- Raw data is public, but “science ready” data is not yet.

➔ Using DOI to suggest share experience and work (software, date, gray articles, ..)

Quindi…

- Sistemi Monolitici non sono la soluzione,
  ➔ ma neanche il troppo distribuito
- Nessuno ti regala niente
  ➔ le partnership hanno pro/contro

- Domanda e': cosa vuole INAF:
  - Creare una cultura del "Big data analyst" per essere pronti tra 5-10 anni
  - Sviluppare la cultura della Proprieta' Intelettuale su SW e Dati
  - Non Demandare completamente ad "altri"


Grazie per l’attenzione