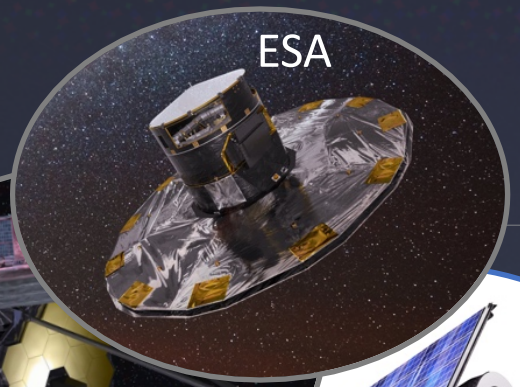
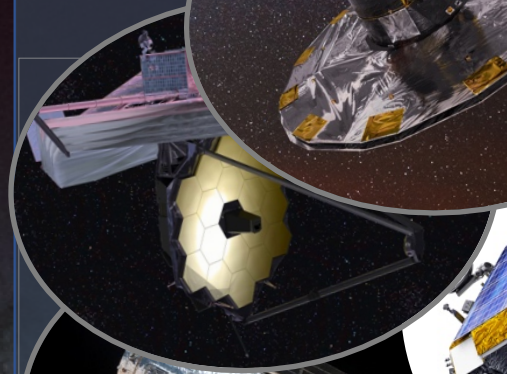
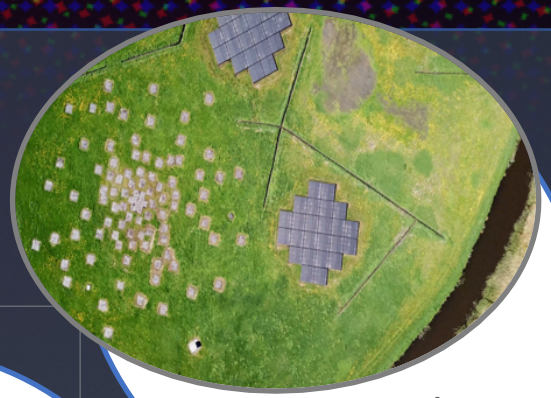
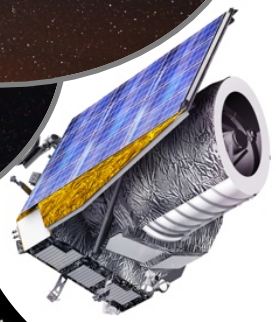
ASTRO@NFDI



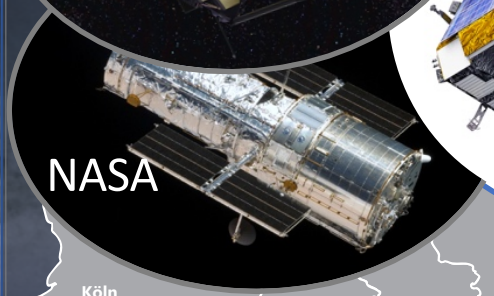




LOFAR + data archive Jülich



DLR



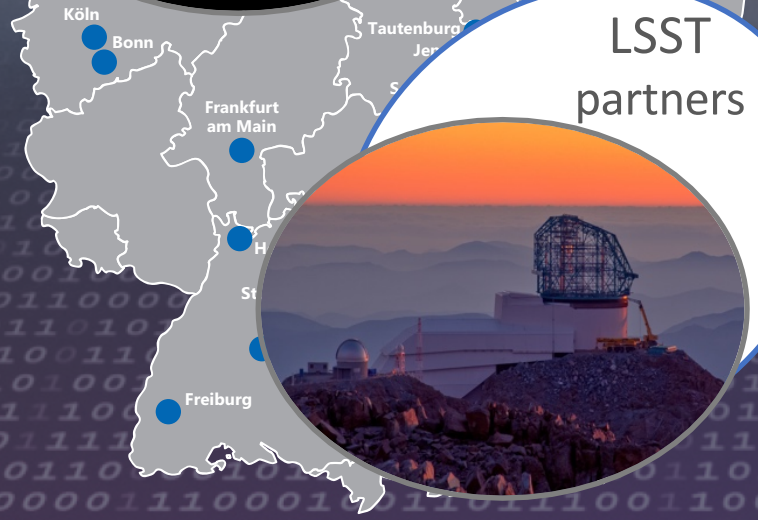
FZJ

MPCDF

LRZ

R D S

KAT. Komitee für Astro. Teilchen. Physik

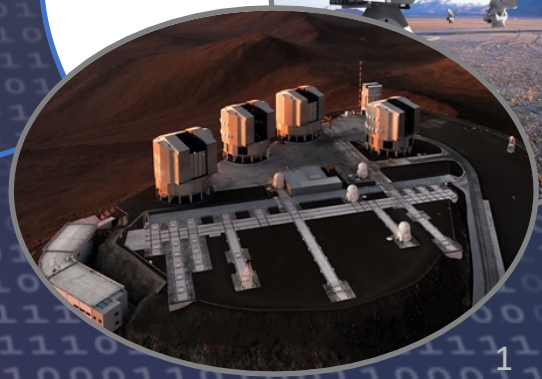


LSST partners

VDR + SKA

CTA data center

ESO



Astro@NFDI

- Rat deutscher Sternwarten
  - 3 Helmholtz Institutes
  - 2 Leibniz Institutes
  - 7 Max-Planck Institutes
  - 26 University institutes
  - 1 state institution
  - 1 private institution
- Komitee für Astroteilchenphysik
- 3 HPC centers + HTW
- 4 ESFRI Projects (ESCAPE)
- 7 international organizations

Who we are:

- About 3000 researchers in the area of astronomy, astrophysics and particle astrophysics (“astrophysics”) in Germany
- About 50 university and non-university institutions (Helmholtz, Leibniz, Max-Planck) + HPC Centers
- Tightly linked to international and intergovernmental organization and facilities (e.g. ESO, SKA, CTA, EST)
- Organized in the Council of German Observatories (RDS) and the Committee for Astro-Particle Physics (KAT)

R D S

KAT. Komitee für Astro. Teilchen. Physik

Astro@NFDI: “Capture data today to answer questions of tomorrow”

- Infrastructure project driven by and for the astrophysics community
- Serves the needs of a diverse and dynamic community
- Addresses fundamental and growing data challenges that astrophysics is encountering now, and other sciences soon
- Unique dissemination channel for training, education and public engagement in data science
- Ready-to-go exploratory development platform for data-driven science
- Professionally managed like a large instrumentation project
- With a long record in data management, accessibility, interoperability and reusability

Why astronomy & astrophysics is arguably different than other sciences

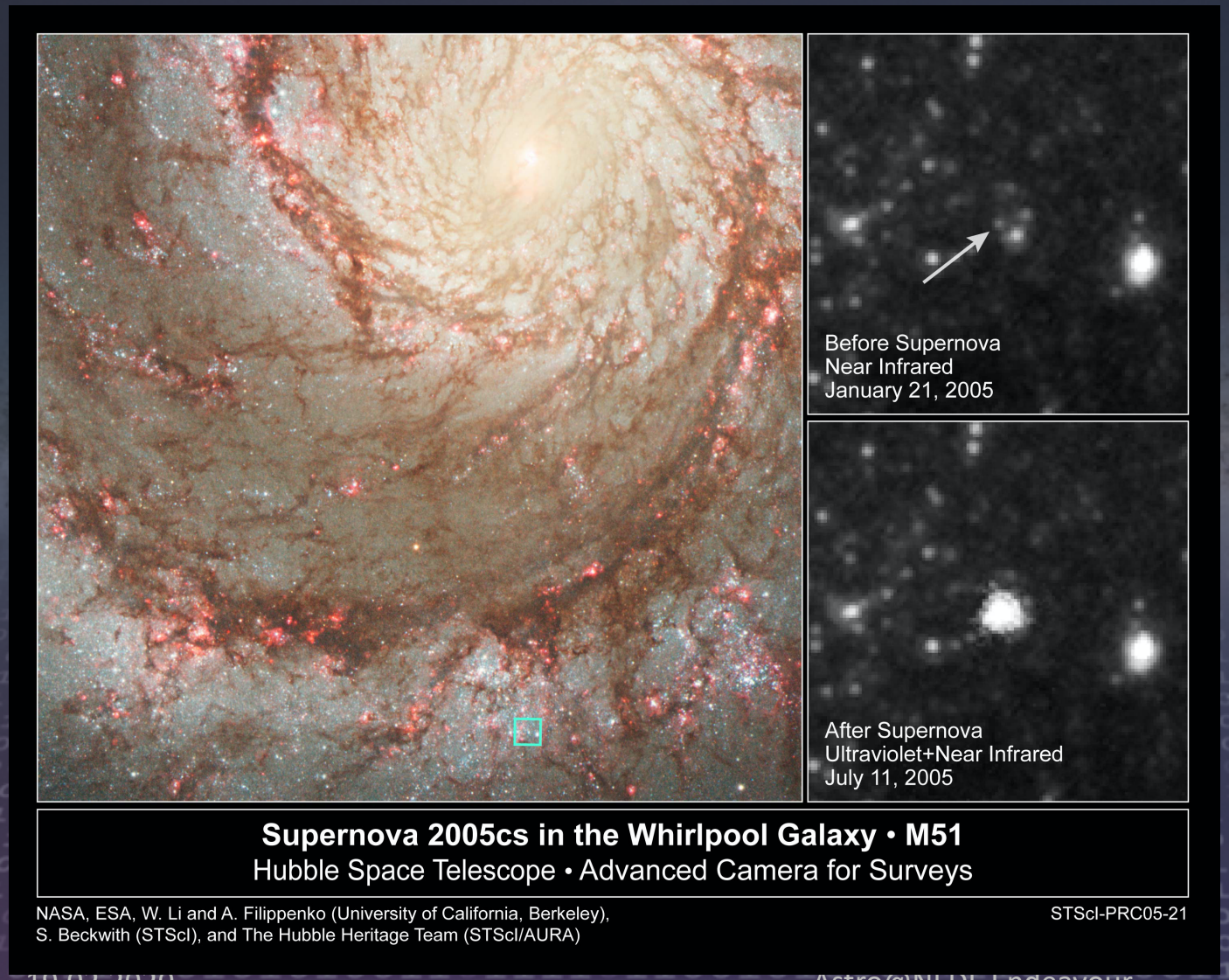
Experiment

Physics: Controlled & focused experiments to approach the fundamental theory of space, time and matter

Observatory

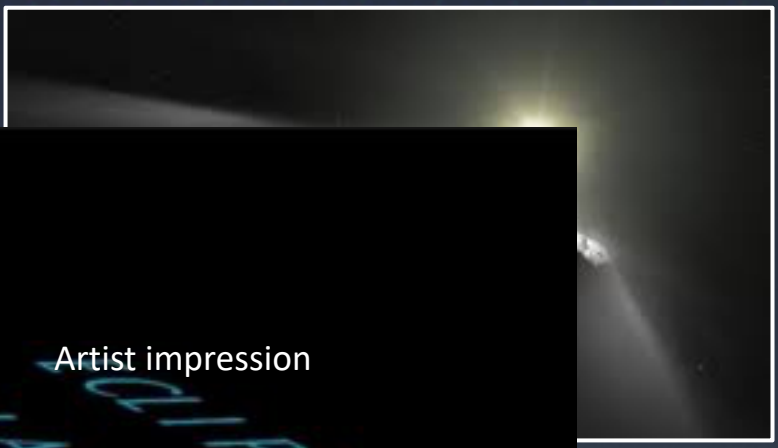
Astronomy: Cannot do experiments with the Universe! Explore the diversity of objects and processes ⇒ Multiple Data Irreversibility challenges!

Reusability



More than 50% of publications using the Hubble Telescope are based on archival data

Reusability



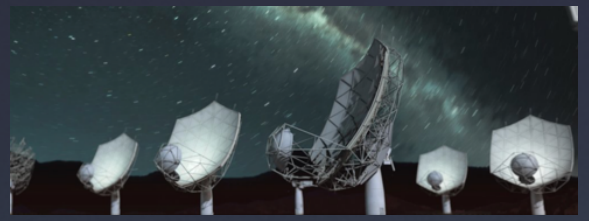
Artist impression

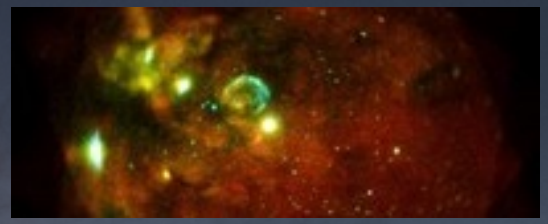
2017

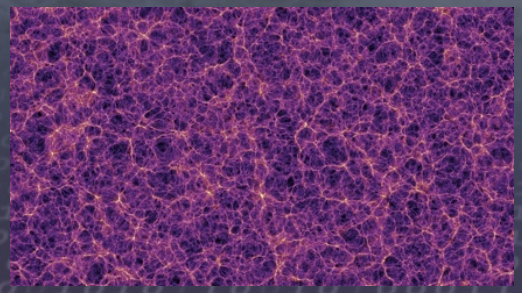
Oumuamua: first extrasolar object detected to fly through solar system

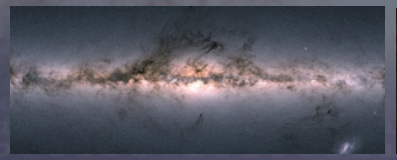
Unfortunately no data of the approach available

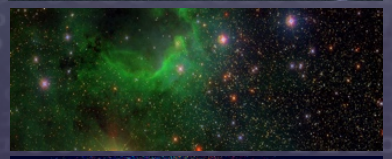
Our Data – now and in the future

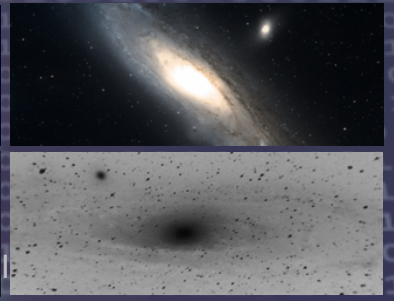


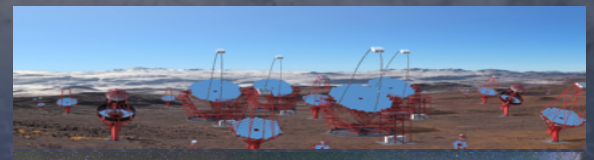


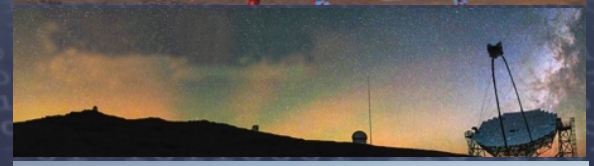


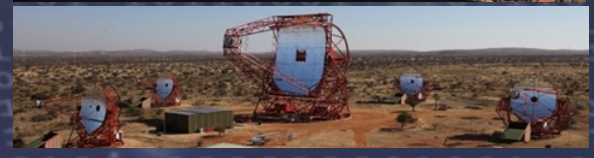




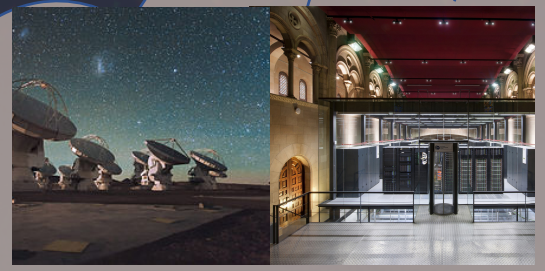








TA2: Data infrastructures



Data sources:
Observations, Simulations
Expected data production:
several Exabyte/year

Metadata
Raw data

Pipelining,
Quality Control
Processing
Curating

Metadata
Science ready data

Archives,
Data Publishing

Standards,
API

Open Data

Metadata
Science Data

Metadata
Science Data

Scientific articles

Pipelining, CRE

Federated Archives

Standards,
API

Metadata
Science Data

Metadata
Science Data

Collaborative Research Environments
Institute IT

Scientific article

Open Access

Data Flows
Scientific Workflows

Astro Science Data Portal

Work in international Bodies
Support Community Efforts
& Open Source

AAI
Resource Allocation
Accounting

Data Archive Building
Archiving Procedures
DOI

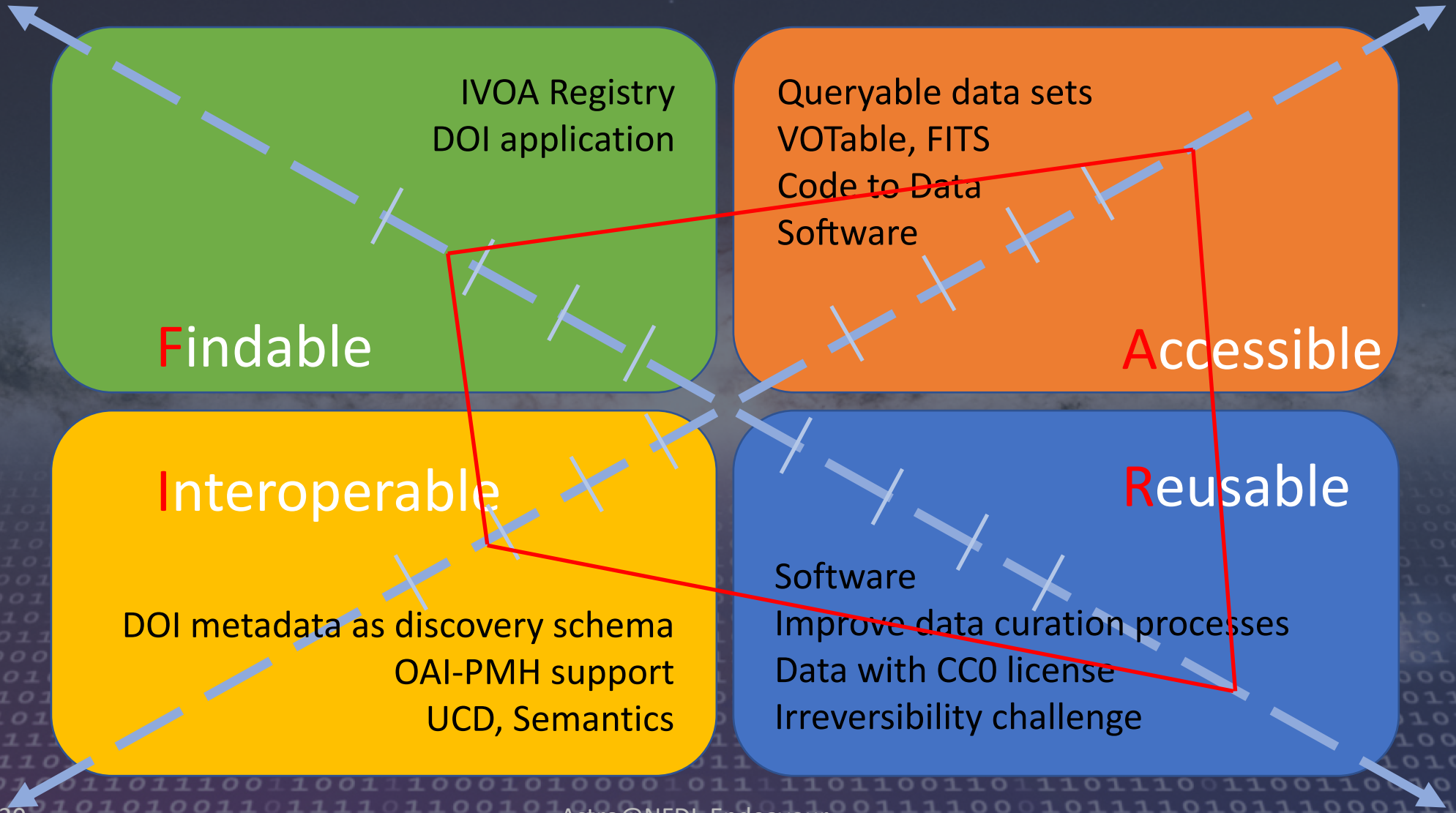
Maintain Data Services
Initiate further development of
Standards & Protocols

Collaborative Research
Environments
Federated Archives

Data Management Support
Software Management Support
Consulting

Astro@NFDI, FAIR Data (now)

www.astro-nfdi.de/how-fair-is-your-data/



Astro@NFDI: IVOA Approach

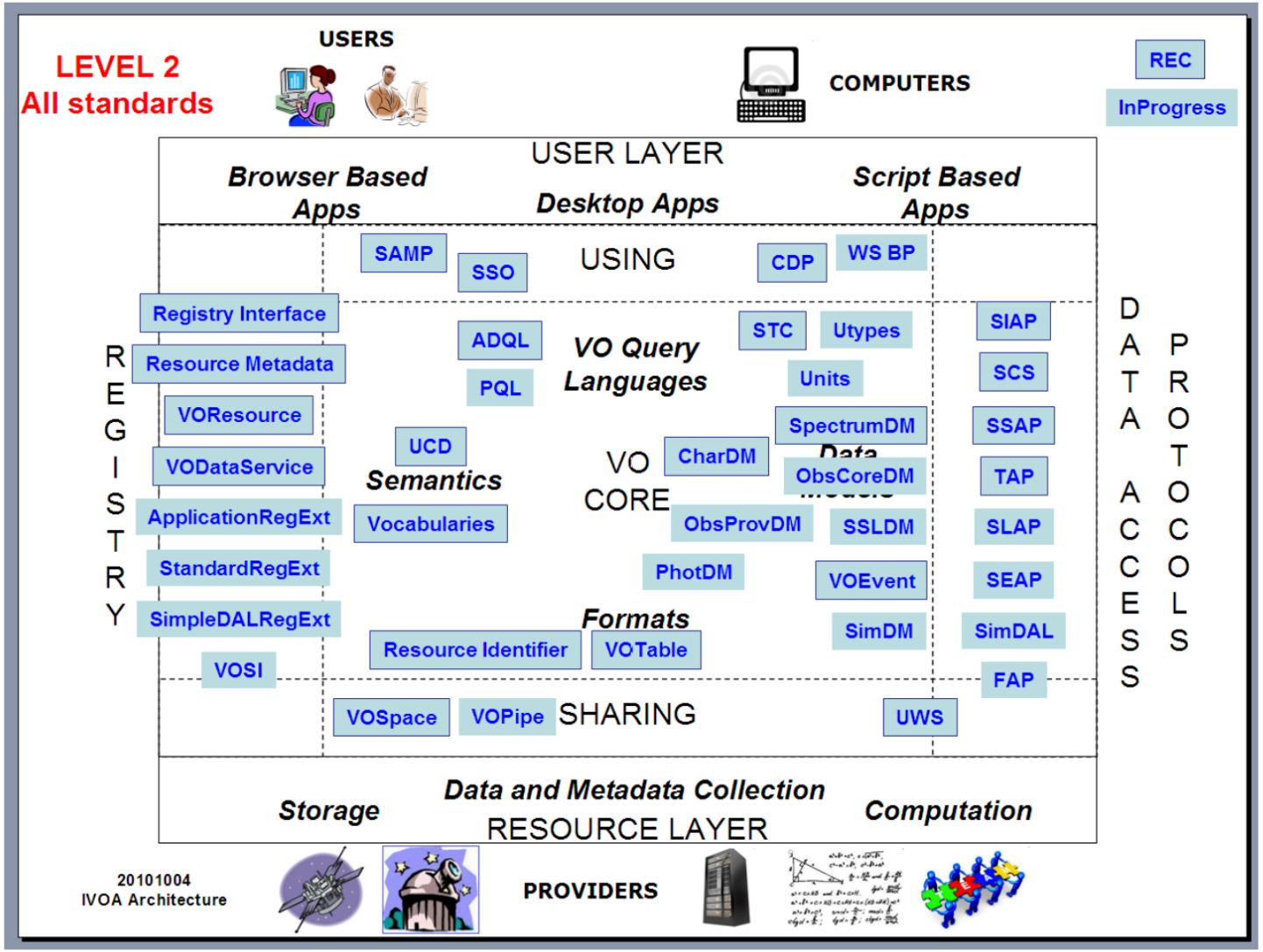
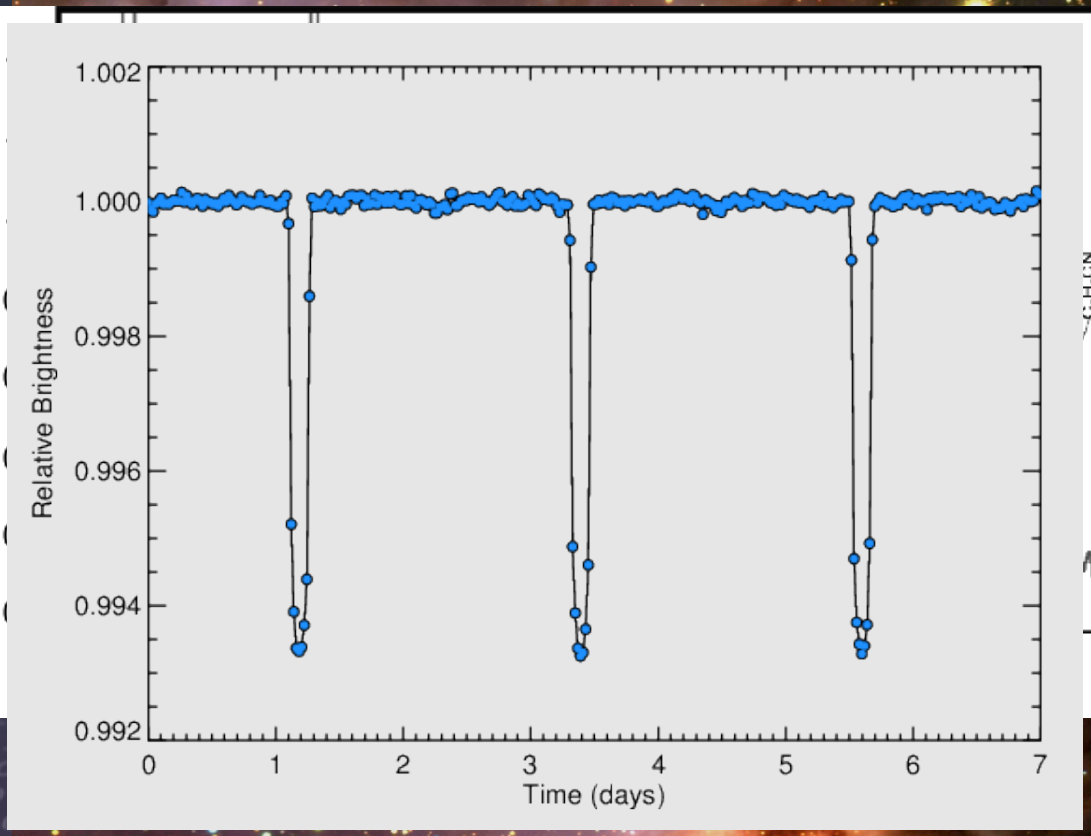


Figure 3: IVOA Architecture Level 2

TA3: Diverse requirements …



Multiple wavelength

Time domain

Spectra

Simulations

Time baseline

Data set size

etc. …

TA3: Diverse requirements ...

Task Area 3

Federate generic services

- serves them all
- unites them
- opens up new interdisciplinary research

Multiple wavelength

Time domain

Spectra

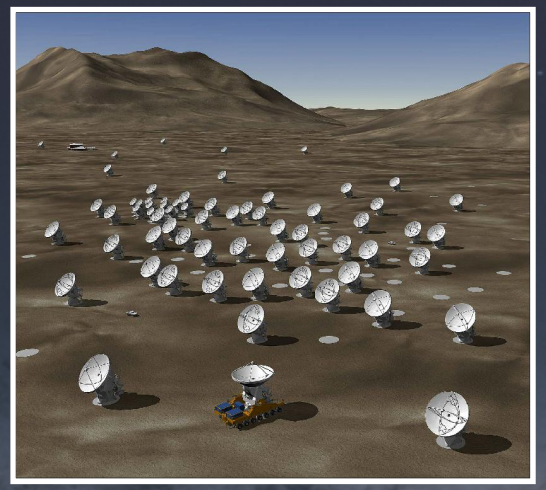
Simulations

Data size

Images

etc. ...

Workflows - Task Area 4



Original data

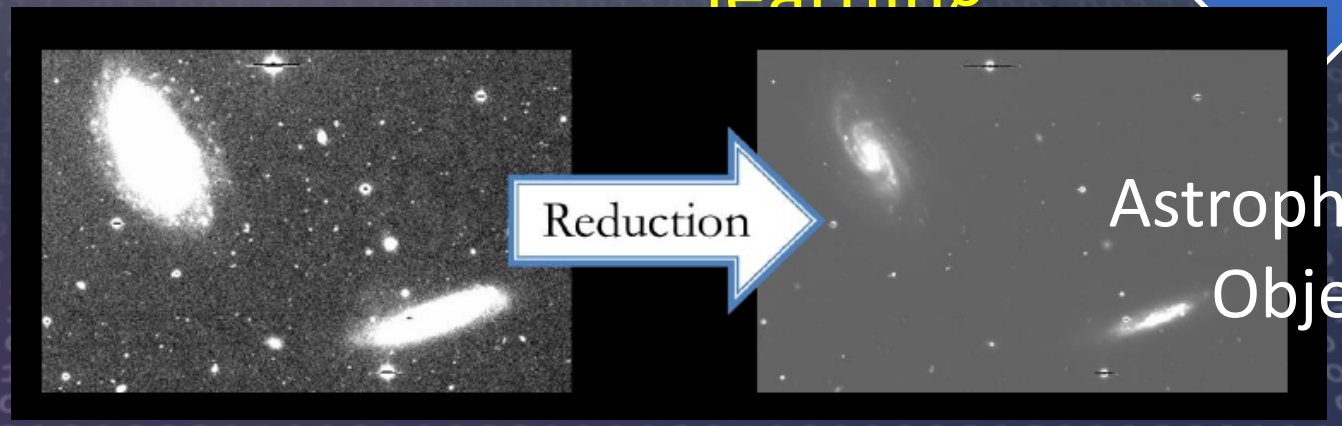
Combining single telescopes

Data reduction

Combining different wavelength

Final image

e.g. Machine learning

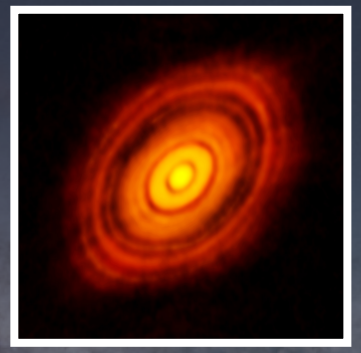


Astrophysical Object

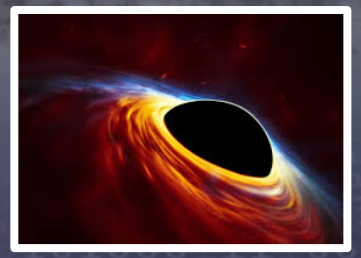
Different Data Workflows for different communities

Defining standards and designing general software modules for

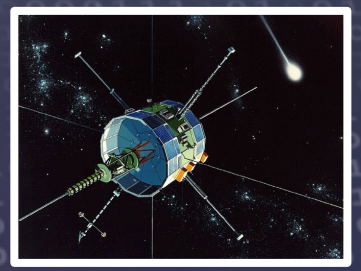
- Data analysis of observational data



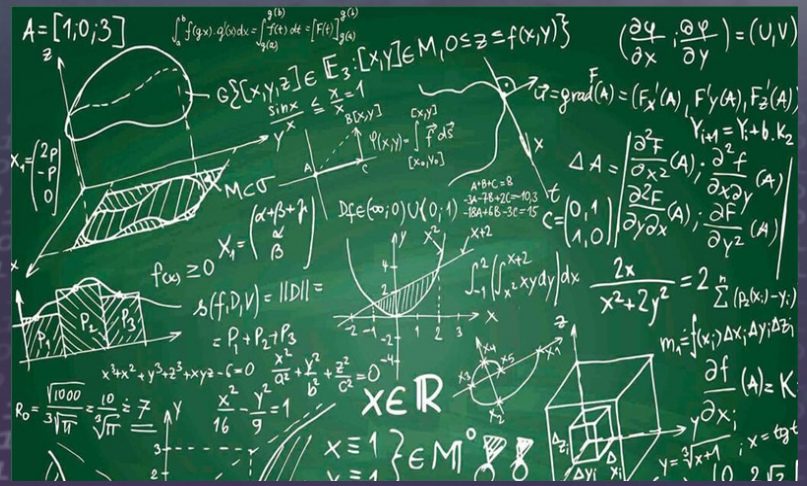
- Astrophysical Simulations



- Instrument Design

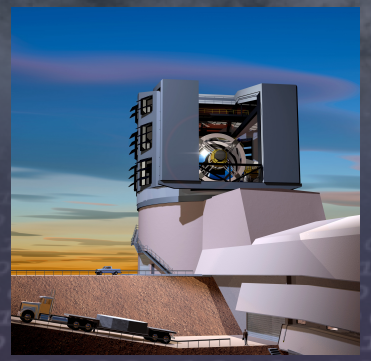
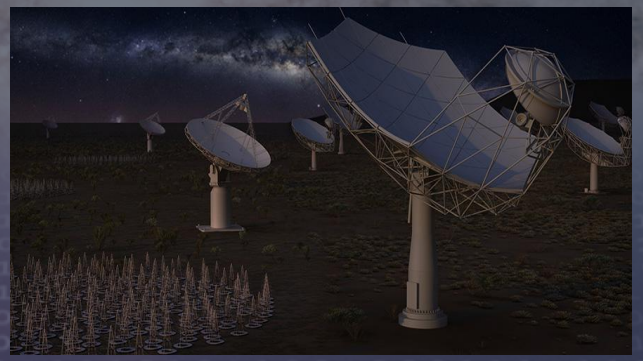
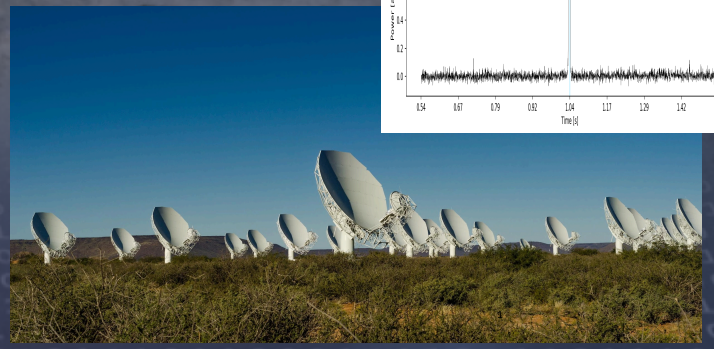
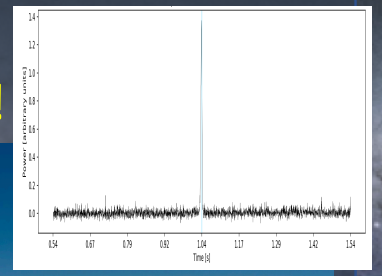


Change of research methodology



TA5: Data Irreversibility Challenges

- Previously: (most) data could be stored and reanalyzed/reused in data life cycle
- Now: data rates increasingly large – even after on-line analysis (FPGA, GPU, HCP, AI etc.)
  only parts of data can be stored
- Soon: only tiny fraction of data can be stored: dramatic loss of information

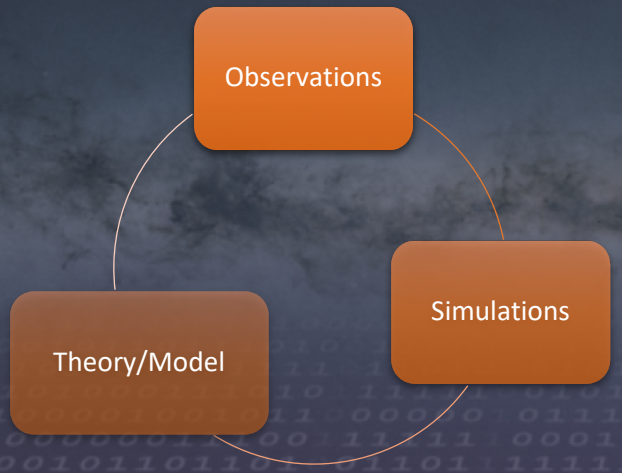
LSST	CTA	SKA	MeerKAT – now!	
			 	
20 TB/night	40 TB/night	5 EB/day	3 PB/day	raw data
1.5 PB/yr	4 PB/yr	300 PB/yr	4 PB/yr	archived

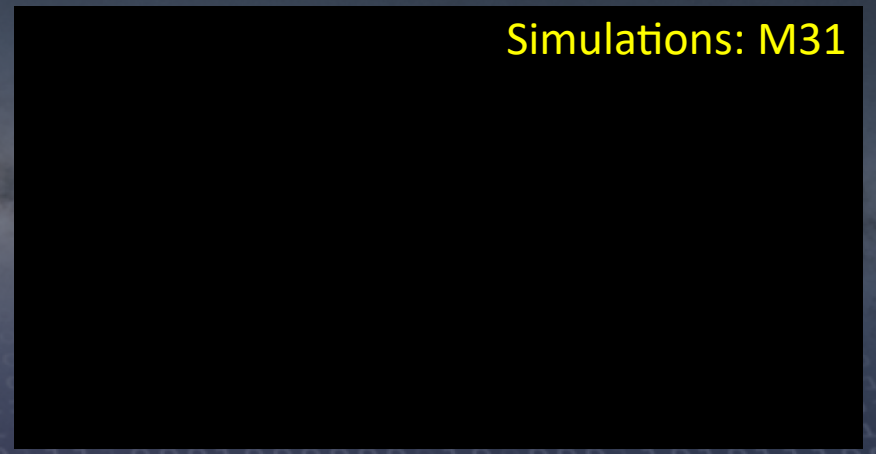
Simulations and data irreversibility

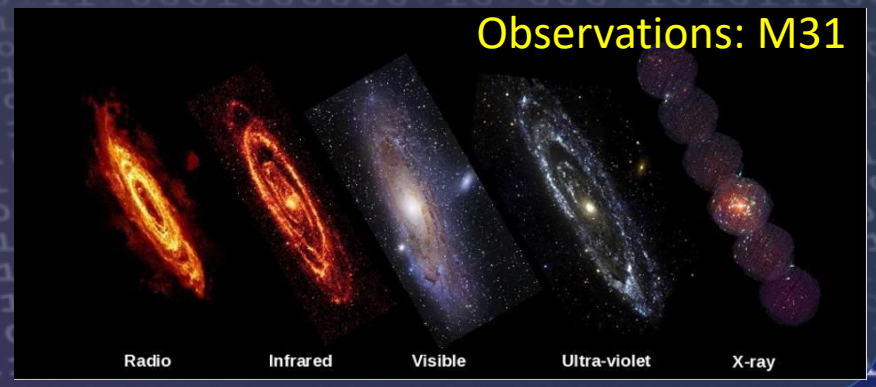
Impact of information loss needs to be described, understood and accounted for

a) Simulations must be used to gauge nature and impact of information loss, while...

b) To infer physics, observations in presence of information loss must be compared with simulations considering the limitations due to information loss:







Today, simulations are huge too.

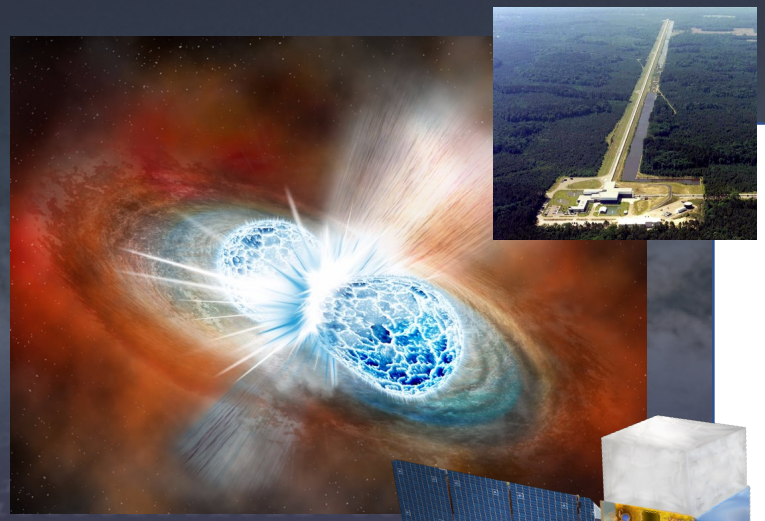
Example “Illustris TNG”: 19 Million CPU hours,

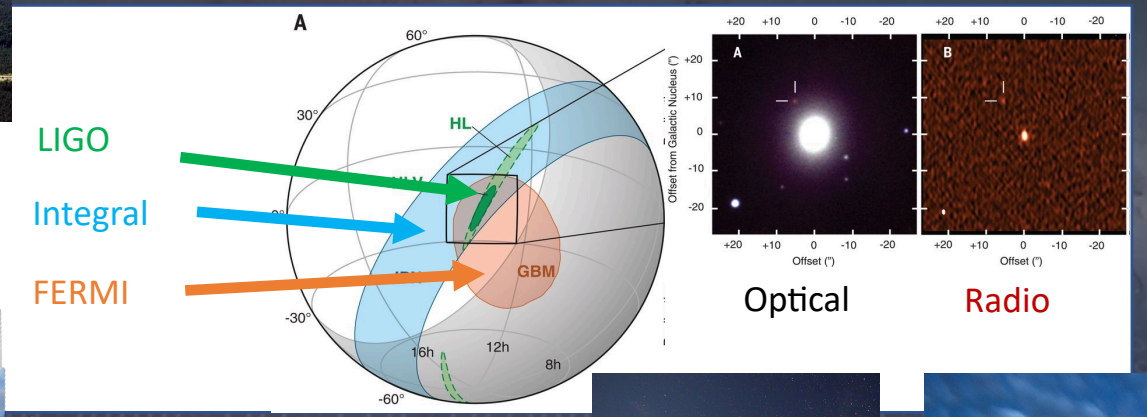
>13 Billion particles, 250 TB total volume

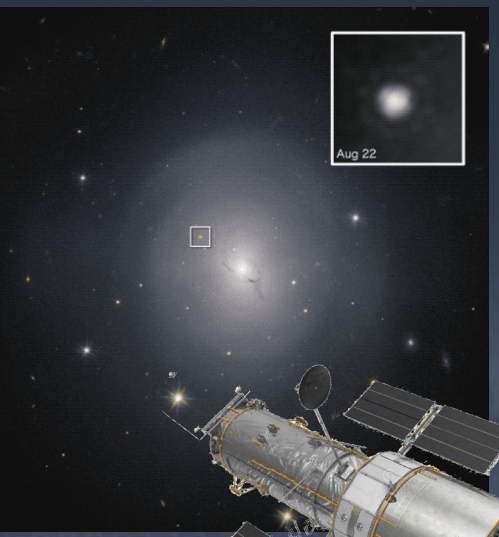
“Universe in time, space and more...”

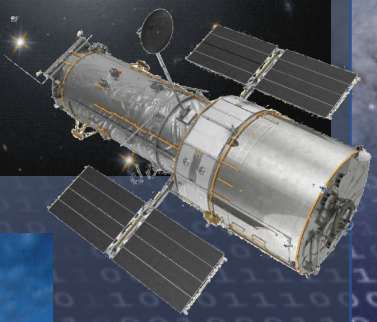
Adapted life-cycle needs to be fast, dynamical and scalable

- Archives must be re-analysed, re-evaluated and insight must be fed back into online analysis
- Online analysis must be fast and reliable with low false-alarm probability





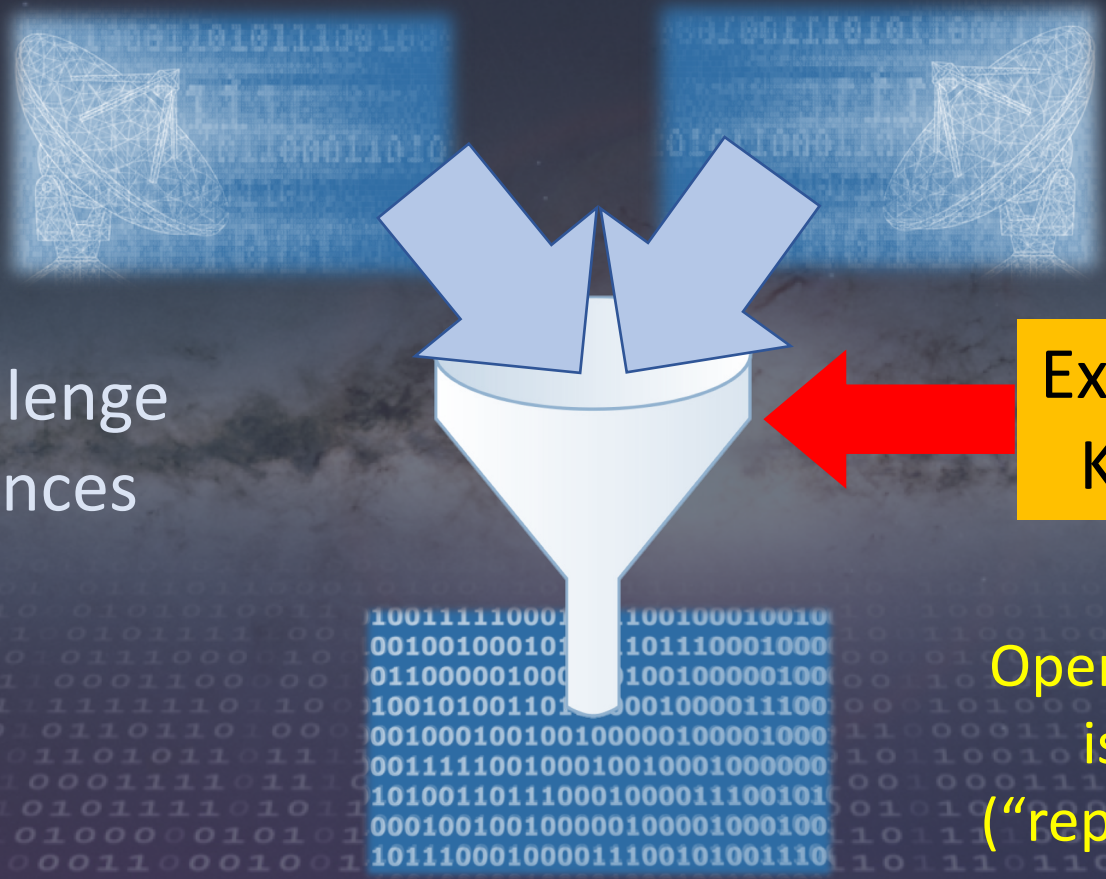




Triggers are *very complex and interwoven* between different instruments and archives

Our laboratories and experiments cannot be controlled

Unusual complexities & diverse external interferences affecting online processing severely



A generic challenge
for most sciences

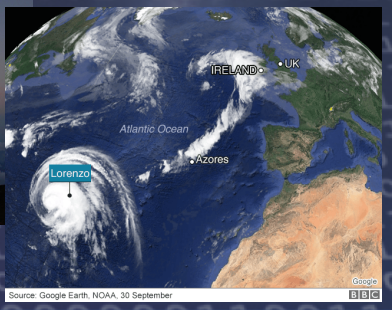
External interference:
Known & unknown

Openness of data
is essential
(“reproducibility“)

“We need to solve it for astronomy, in collaboration with others, for the benefit of all.”
(with other foreseeable positive consequences, e.g. green computing)

TA6 Synergies & Cross-cutting

- Despite different, community-specific needs, many challenges are common.
- Different communities have advanced on individual aspects to different degrees. Sharing expertise and common developments in cross-cutting topics are desirable.
- Challenges are preferences to develop own solutions and reluctance to share services.
- Especially *new challenges* for Astro@NFDI community will emerge in other fields and considerable gains are possible by cooperative approaches.
- NFDI will profit from extensive ASTRO@NFDI's expertise with common services, common access, interchangeable formats, archives developed for diverse communities.

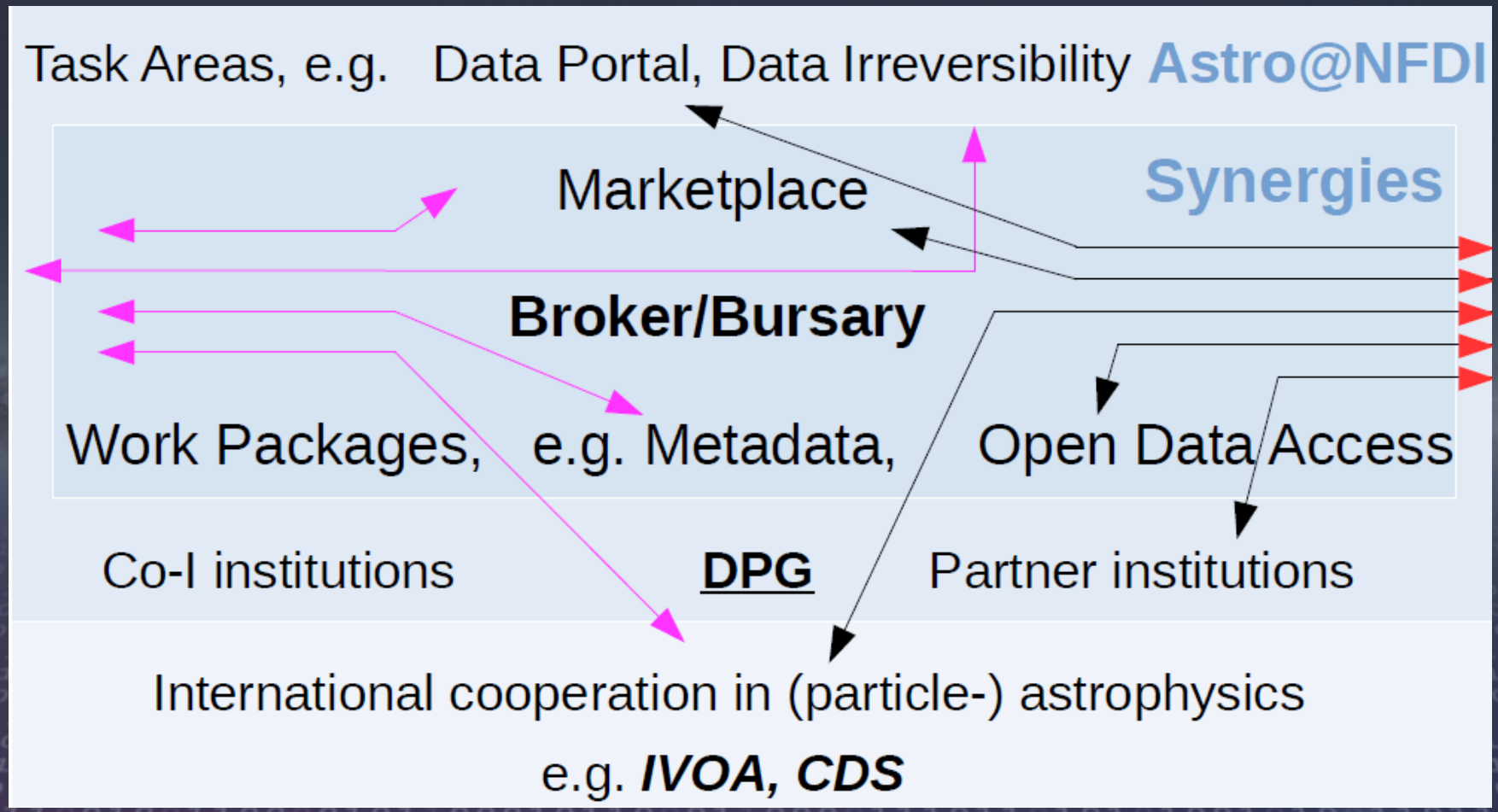


TA6: Cross-cutting activities & synergies

Year 1

PAHN-PaN

DAPHNE
FAIRmat

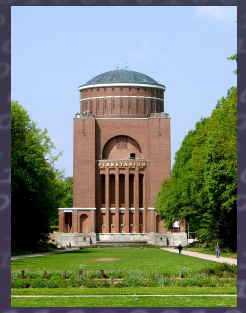


NFDI4Chem
NFDI4Earth
NFDI4Ing
NFDI4HPC
RSE-4-NFDI
NFDI4Life
NFDI-Neuro
GeRDI4NFDI
AI4NFDI
CompeNFDI

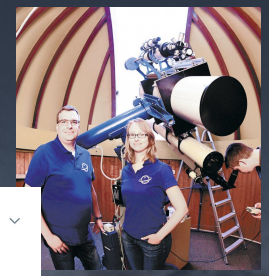
TA7: Training, Education & Outreach: Challenges

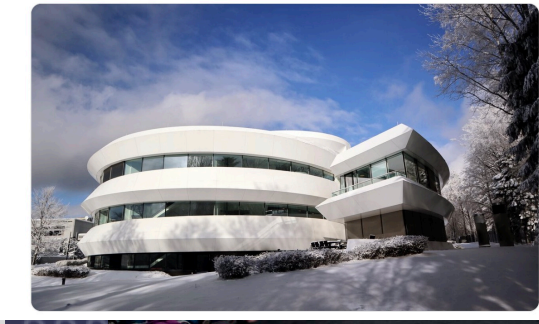
Unlike any other sciences, dense and su ria, local societies & observatories thanks to hug nd data





Training

Expert users

Data & training resources

Analyze, promote & foster

Education

Teachers Students

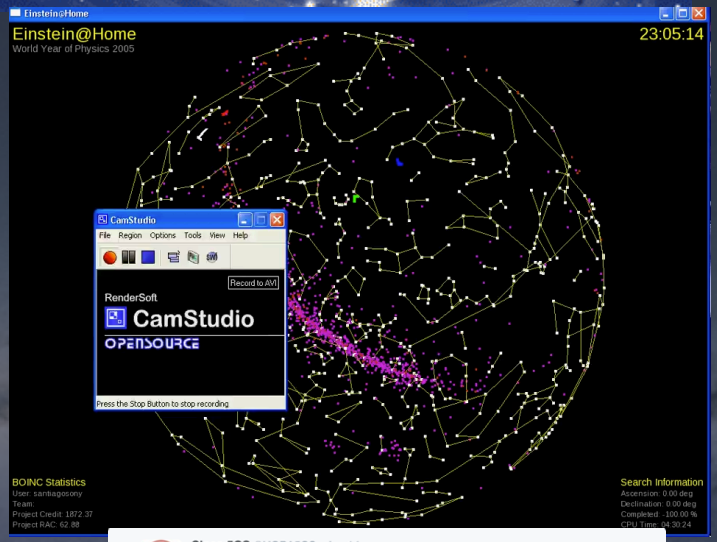
Edu resources
Visiting seminars
Career prospects
Gender gap

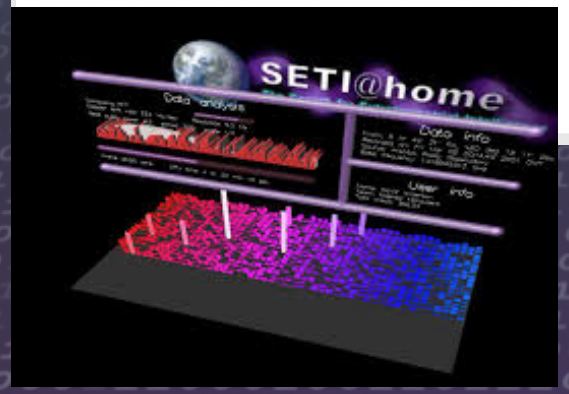
Outreach

Media Amateurs Schools

Communication training & resources
Pedagogical approaches
School curricula
Citizen science
Active engagement

Citizen Science

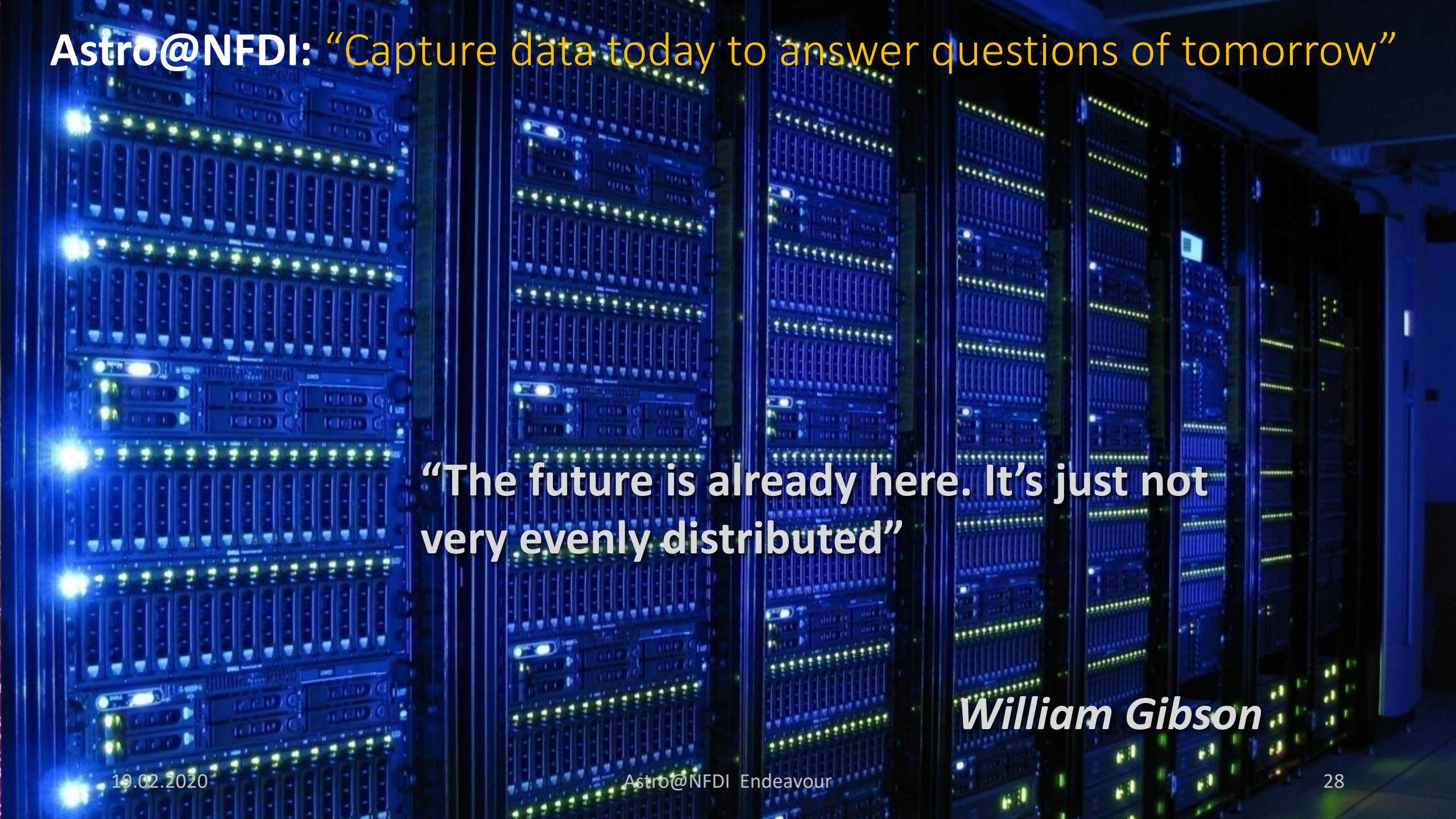




Astro@NFDI: “Capture data today to answer questions of tomorrow”

- Infrastructure project driven by and for the astrophysics community
- serves the needs of a diverse and dynamic community
- addresses fundamental and growing data challenges that astrophysics is encountering now, and other sciences soon
- Unique dissemination channel for training, education and public engagement in data science
- Ready-to-go exploratory development platform for data-driven science
- Professionally managed like a large instrumentation project
- With a long record in data management, accessibility, interoperability and reusability

Astro@NFDI: “Capture data today to answer questions of tomorrow”



“The future is already here. It’s just not very evenly distributed”

William Gibson