



IVA MOMCHEVA

DATA SCIENCE MISSION OFFICE

CONSOLIDATING THE LEADERSHIP AND OVERSIGHT OF [ARCHIVE] ACTIVITIES WILL LEAD TO GREATER COORDINATION, ELIMINATE DUPLICATION, ENABLE MORE EFFICIENT USE OF RESOURCES AND ALLOW THE INSTITUTE TO TAKE FULL ADVANTAGE OF SYNERGIES BETWEEN THE DATA SCIENCE WORK ON MULTIPLE MISSIONS.

WE RECOMMEND THAT STScI FORM A DATA SCIENCE MISSION OFFICE AS SOON AS POSSIBLE.

Big Data @ STScI



Arfon Smith
Mission Office Head



Joshua Peek
Project Scientist



Iva Momcheva
Mission Scientist



Michael Fox
Infrastructure Engineer

WHO ARE WE?

AREAS OF RESPONSIBILITY

- ▶ Data management
 - ▶ Data systems
 - ▶ Data processing
 - ▶ Data analysis tools
 - ▶ Archival work

Not in DSMO purview: planning, ETC, pipeline development

SCHEDULED HAND-OFFS

DMS area	April '17	May – June '17	July – Sept '17	FY18 onwards
ASB	Planning	Dual mission	Full responsibility	Full responsibility
DPAS (TBD)	Planning	Planning?	Dual mission	Full responsibility
DSB (TBD)	Planning	Planning?	Dual mission	Full responsibility
DAT (SSB)	Planning	Dual mission	Full responsibility	Full responsibility



WHAT WILL DSMO DO FOR HUBBLE?

IMPROVE INFRASTRUCTURE

- ▶ Recommend improvements to current infrastructure, process and management
- ▶ Update and consolidate toolchain across branches
- ▶ Explore cloud infrastructure for reprocessing, access and delivery of data



IMPROVE ARCHIVAL SERVICES

- ▶ Better support for large datasets (GO & archival large programs)
- ▶ Recommendations for delivering data and software
- ▶ Tools and access for HLSPs
- ▶ Consolidate services around archive
- ▶ Connect HST with other facilities

Credit: Katie Peek; Source: MAST, STSCI (data); Hubble Heritage Team (STScI/AURA)

How to Read the Chart Below

Each dot represents a Hubble observation referenced in a published paper.

Blue dots (nearly 14,000) relate to Hubble's archive: instead of asking for new observations, the researcher answered a question using existing images.

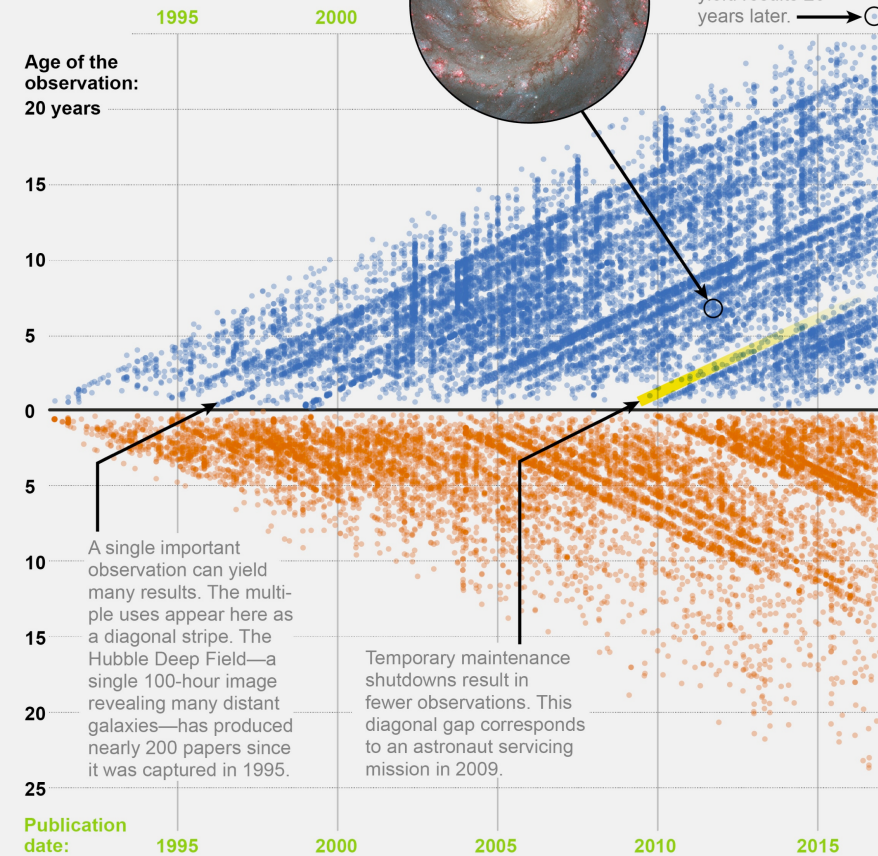
Orange dots (nearly 9,000) relate to new data: a researcher makes a specific observation request to answer a question.

An additional 10,000 references to Hubble observations appeared in papers using a mix of the telescope's new and archival images.

Data are current as of early December 2016.

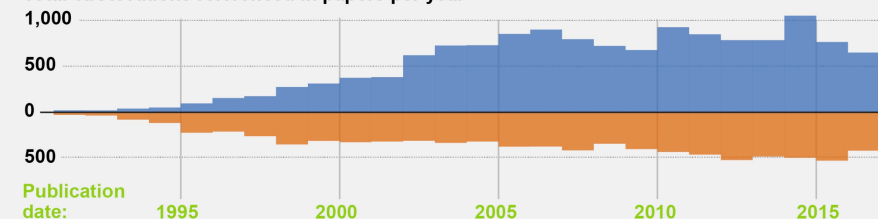
Observations—when Hubble points at an object and collects data—can be beautiful and useful. This image of the Whirlpool galaxy from 2005 helped astronomers identify which star exploded as a supernova in 2011.

Hubble's earliest observations still yield results 26 years later. → ○



After Hubble's eventual demise, results based on new data (*orange*) will disappear, but those based on the archives (*blue*) will continue.

Total observations referenced in papers per year



IMPROVE DATA ANALYSIS TOOLS

- ▶ Dedicated data analysis tools team
- ▶ Update toolchain to Python
- ▶ Community and cross-AURA collaborations



THANK YOU!