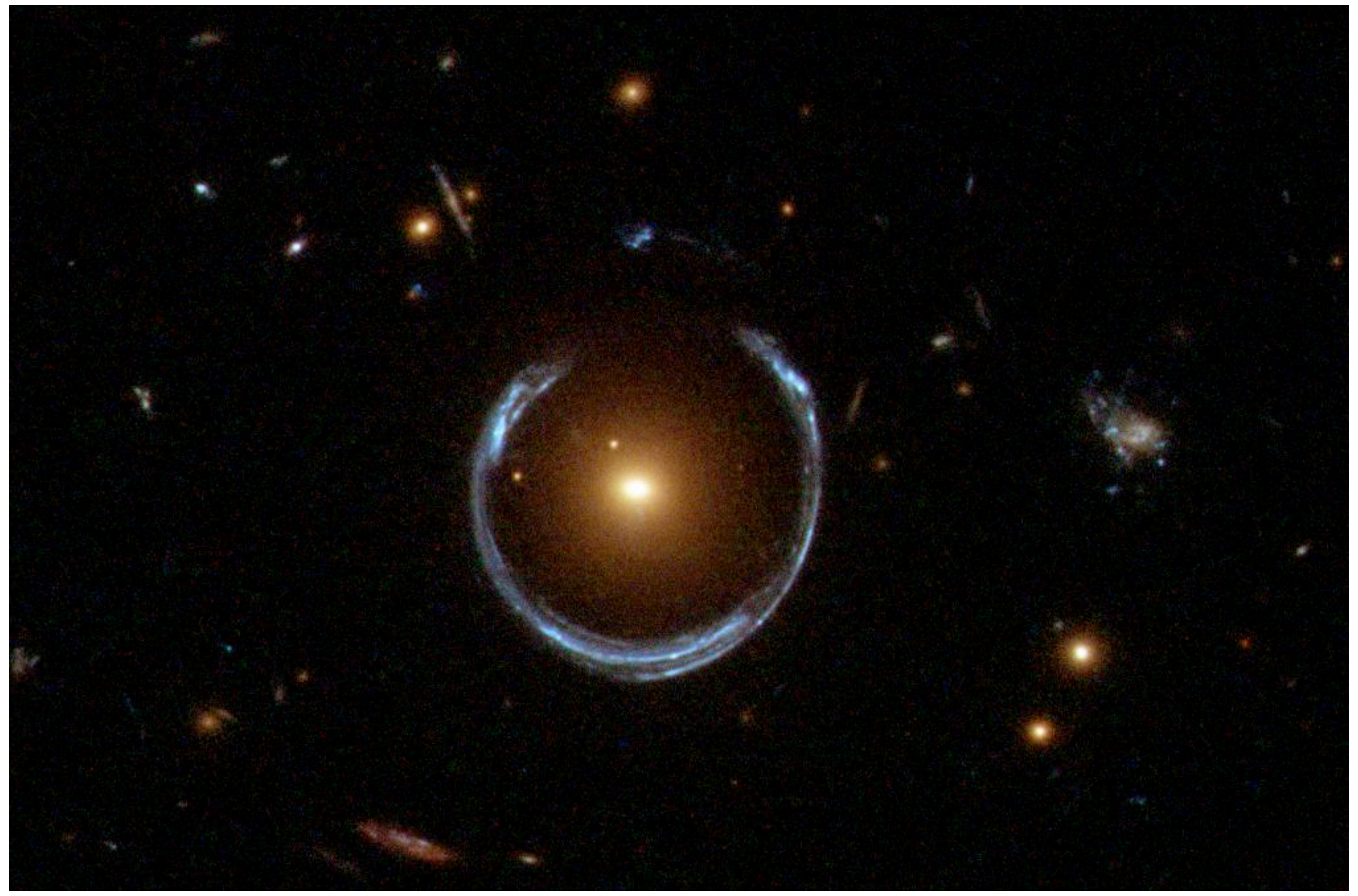
Data reduction in space sciences

Gareth Murphy

What do other scientists do with their data?

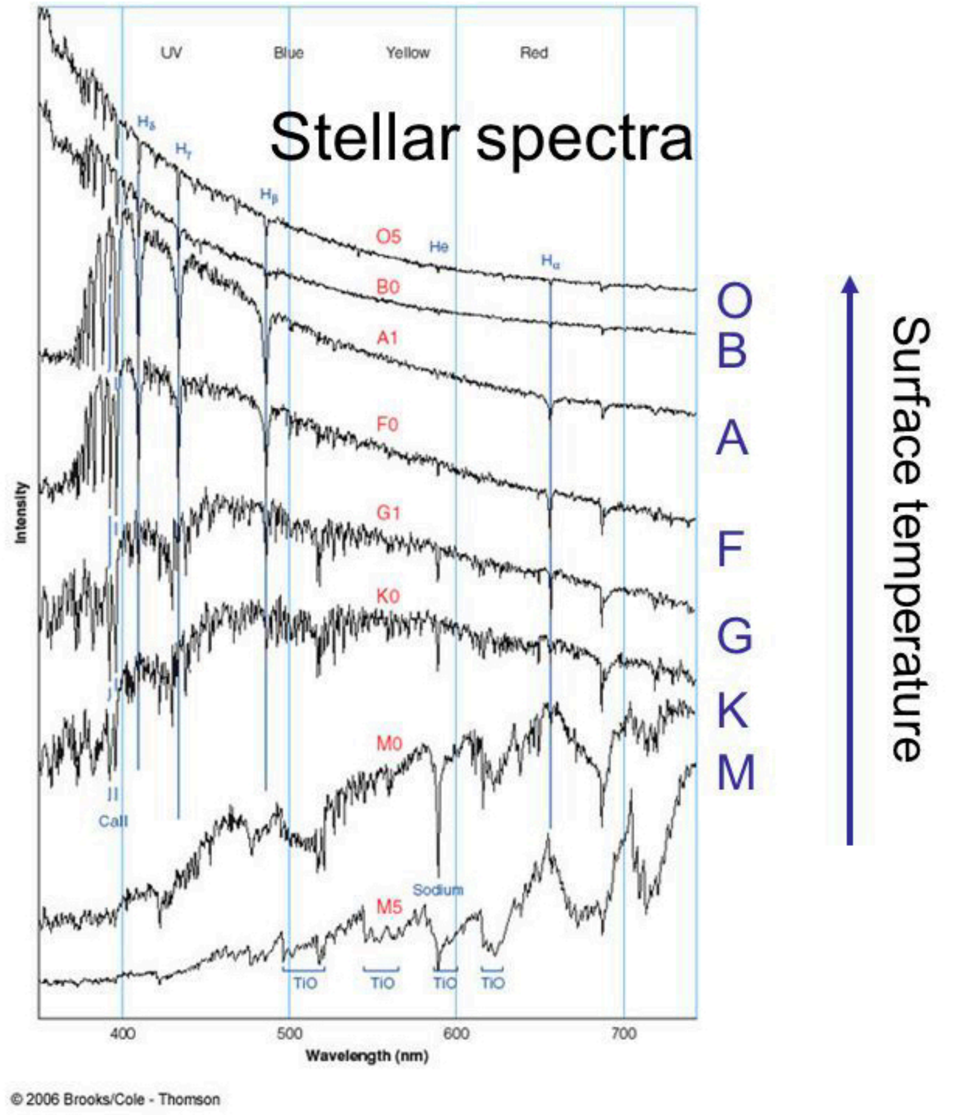
- Can we learn from their efforts?
- Are the use cases the same?
- Astronomers have lots of instruments in one large uncontrolled environment
- Not easy to control instruments in space
- Observations can be decades long
- May not be relevant to neutron sciences

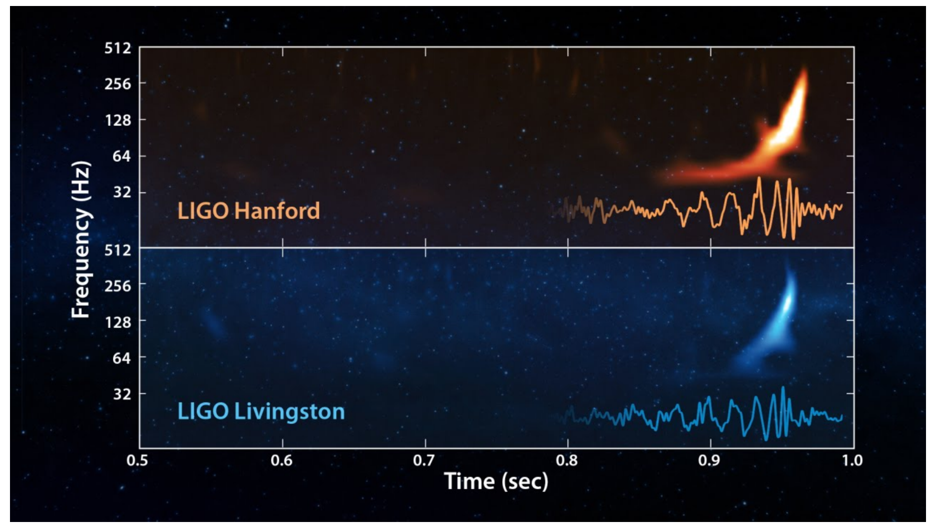


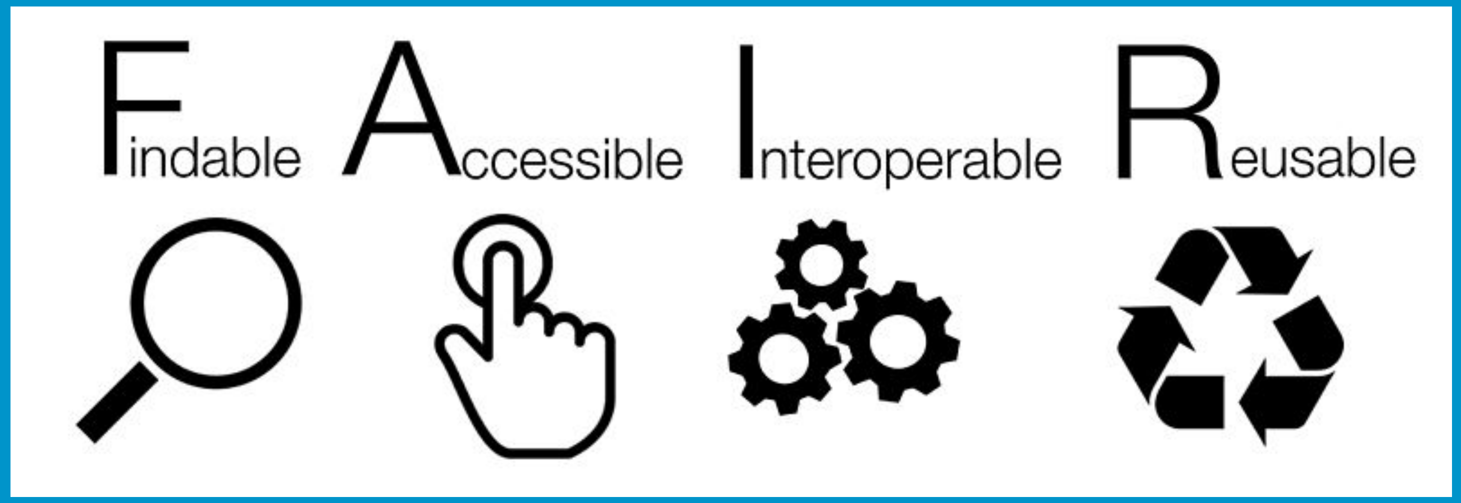
Astronomy data

EUROPEAN SPALLATION SOURCE

- Types of data: Images & spectra of stars, exotic objects and events: gamma ray bursts, gravitational waves
- Metadata: location (in spherical coordinates), time, start and stop time
- File format: FITS (Flexible Image Transport System (70s format)), moving to HDF5 for big data
- Use cases: multi-wavelength, multi-instrument studies, data from 10+ years ago is useful
- Is there common ground with neutron science?







EUROPEAN SPALLATION SOURCE

To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

TO BE ACCESSIBLE:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

TO BE INTEROPERABLE:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards

Getting the data

- ftp, rsync, http download
- Observatory website, e.g. Hubble space telescope
- SIMBAD (Set of Identifications, Measurements, and Bibliography for Astronomical Data) data catalog, CDS , Strasbourg

Astronomy Database: SIMBAD

EUROPEAN SPALLATION SOURCE

Not Secure | simbad.u-strasbg.fr/simbad/sim-fid

Portal | Simbad | VizieR | Aladin | X-Match | Other | Help

SIMBAD: Query by identifiers

other query modes : Identifier query | Coordinate query | Criteria query | Reference query | Basic query | Script submission | TAP | Output options | Help

Query an identifier

Identifier :

Examples
sirius, M31, MCG+02-60-010
How to write an identifier can be found in the dictionary of nomenclature
IAU format can also be used, with the following format:
iau [J|B]1230+08 [* *enlarging-factor*] [= *Object-type*]

you can choose to query : only this object

around the object, define a radius : 2 arc min

submit id | clear

Query a list of identifiers

Enter the name of an ASCII file produced by a text editor containing one identifier per line:

Choose File No file chosen

list display | full display

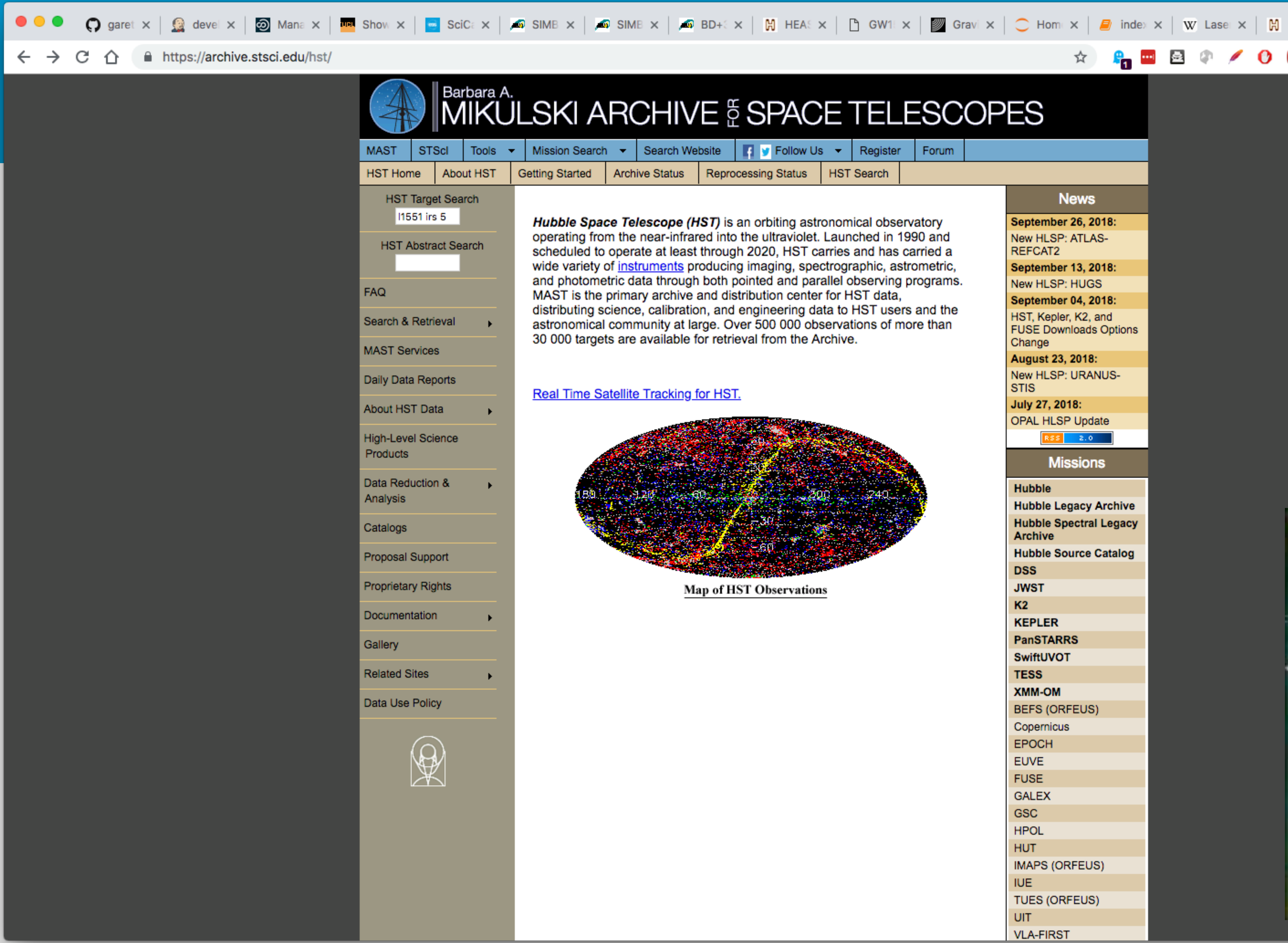
submit file | clear

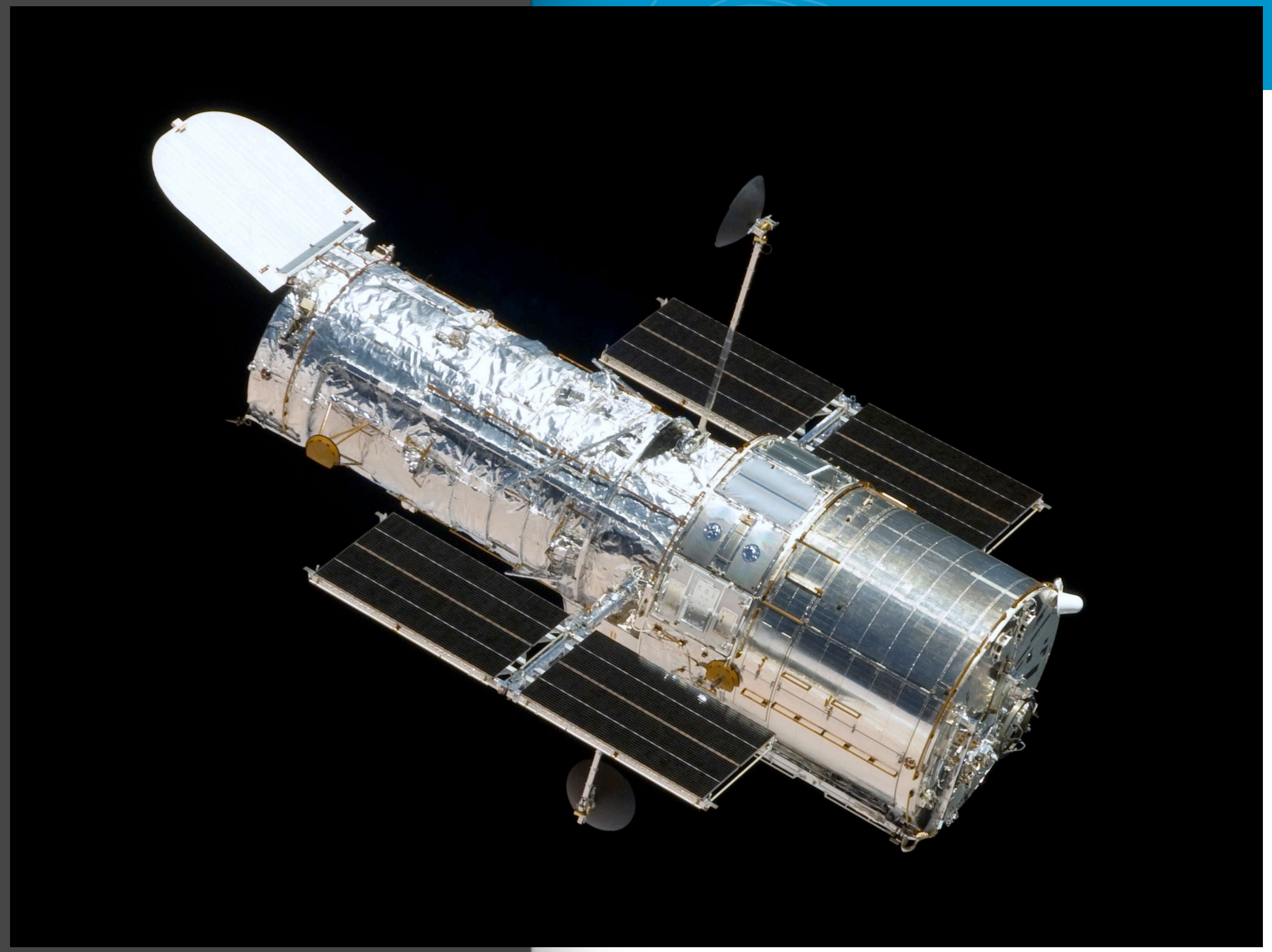
query around the objects with radius : 2 arc min

Only the list display applies here

Query by identifiers can be done by

- full identifiers
- partial identifiers using wildcards ('?' = one char, '*' = any string, including an empty one (no char), '[xyz]' = one char among the list.
  Examples:


https://archive.stsci.edu/hst/
Barbara A.
MIKULSKI ARCHIVE FOR SPACE TELESCOPES
MAST
STScI
Tools
Mission Search
Search Website
Follow Us
Register
Forum
HST Home
About HST
Getting Started
Archive Status
Reprocessing Status
HST Search
HST Target Search
l1551 irs 5
HST Abstract Search
FAQ
Search & Retrieval
MAST Services
Daily Data Reports
About HST Data
High-Level Science Products
Data Reduction & Analysis
Catalogs
Proposal Support
Proprietary Rights
Documentation
Gallery
Related Sites
Data Use Policy
Hubble Space Telescope (HST) is an orbiting astronomical observatory operating from the near-infrared into the ultraviolet. Launched in 1990 and scheduled to operate at least through 2020, HST carries and has carried a wide variety of instruments producing imaging, spectrographic, astrometric, and photometric data through both pointed and parallel observing programs. MAST is the primary archive and distribution center for HST data, distributing science, calibration, and engineering data to HST users and the astronomical community at large. Over 500 000 observations of more than 30 000 targets are available for retrieval from the Archive.
Real Time Satellite Tracking for HST.
Map of HST Observations
News
September 26, 2018:
New HLSP: ATLAS-REFCAT2
September 13, 2018:
New HLSP: HUGS
September 04, 2018:
HST, Kepler, K2, and FUSE Downloads Options Change
August 23, 2018:
New HLSP: URANUS-STIS
July 27, 2018:
OPAL HLSP Update
Missions
Hubble
Hubble Legacy Archive
Hubble Spectral Legacy Archive
Hubble Source Catalog
DSS
JWST
K2
KEPLER
PanSTARRS
SwiftUVOT
TESS
XMM-OM
BEFS (ORFEUS)
Copernicus
EPOCH
EUVE
FUSE
GALEX
GSC
HPOL
HUT
IMAPS (ORFEUS)
IUE
TUES (ORFEUS)
UIT
VLA-FIRST



Query results with metadata

EUROPEAN SPALLATION SOURCE

https://archive.stsci.edu/hst/search.php?target=l1551+irs+5&action=Search&resolver=SIMBAD&radius=3.0&outputformat=HTML_Table&max_records...

Mission Search / Missions / Contacts / STScI / MAST

Columns Help / Archive Status

HST Search Results

Edit Query

Object name l1551 irs 5 resolved by SIMBADCFA (via SANTA cache) to **HBC 393** (FUOr)
RA: 4 31 34.08 **Dec:** 18 8 4.90 **(J2000)**

100 rows displayed, but 105 are available.

Table Info

Plot marked spectra | Submit marked data for retrieval from STDADS
Mark all | Unmark all | Mark public | Unmark public | Mark proprietary | Unmark proprietary

◀ Previous Next ▶ Page 1 of 3

Mark	Dataset	Target Name	RA (J2000)	Dec (J2000)	Ref	Start Time	Stop Time	Exp Time	Instrument	Apertures	Filters/Gratings	Central Wavelength	Proposal ID	Release Date	Preview Name
☐	J92A21031	L1551IRS5	04 31 34.130	+18 08 05.04	3	2004-09-02 18:10:23	2004-09-02 18:22:26	536.000	ACS	WFC1-POL120V	F606W;POL120V	5934.307	10178	2005-09-03 01:31:58	J92A21031
☐	J92A11031	L1551IRS5	04 31 34.130	+18 08 05.04	3	2004-09-02 16:34:24	2004-09-02 16:46:27	536.000	ACS	WFC1-POL120V	F606W;POL120V	5934.307	10178	2005-09-03 01:30:11	J92A11031
☐	J92A11021	L1551IRS5	04 31 34.136	+18 08 05.02	3	2004-09-02 16:18:58	2004-09-02 16:31:01	536.000	ACS	WFC1-POL60V	F606W;POL60V	5934.307	10178	2005-09-03 01:29:40	J92A11021
☐	J92A21021	L1551IRS5	04 31 34.136	+18 08 05.02	3	2004-09-02 17:54:57	2004-09-02 18:07:00	536.000	ACS	WFC1-POL60V	F606W;POL60V	5934.307	10178	2005-09-03 01:31:25	J92A21021
☐	J92A21011	L1551IRS5	04 31 34.140	+18 08 04.92	3	2004-09-02 17:39:31	2004-09-02 17:51:34	536.000	ACS	WFC1-POL0V	F606W;POL0V	5934.307	10178	2005-09-03 01:31:04	J92A21011
☐	J92A11011	L1551IRS5	04 31 34.140	+18 08 04.92	3	2004-09-02 16:03:32	2004-09-02 16:15:35	536.000	ACS	WFC1-POL0V	F606W;POL0V	5934.307	10178	2005-09-03 01:21:33	J92A11011
☐	N92A01010	L1551IRS5	04 31 34.141	+18 08 04.97	3	2004-10-07 20:58:41	2004-10-07 21:11:38	639.639	NICMOS	NIC2-FIX	POL0L	19998.000	10178	2005-10-08 20:36:15	N92A01010
☐	N92A01020	L1551IRS5	04 31 34.141	+18 08 04.97	3	2004-10-07 21:12:25	2004-10-07 21:25:22	639.639	NICMOS	NIC2-FIX	POL120L	19998.000	10178	2005-10-08 20:36:52	N92A01020
☐	N92A01030	L1551IRS5	04 31 34.141	+18 08 04.97	3	2004-10-07 22:39:38	2004-10-07 22:52:35	639.639	NICMOS	NIC2-FIX	POL240L	19998.000	10178	2005-10-08 20:18:56	N92A01030
☐	N92A01040	L1551IRS5	04 31 34.141	+18 08 04.97	3	2004-10-07 22:53:22	2004-10-07 22:58:51	191.833	NICMOS	NIC2-FIX	F160W	16030.400	10178	2005-10-08 20:19:14	N92A01040
☐	W0K20601T	L1551	04 31 34.104	+18 08 07.96	2	1991-08-17 19:10:17	1991-08-17 19:10:47	30.000	WFPC	ALL	F675W	6731.000	1138	1992-08-17 16:07:00	W0K20601T

https://archive.stsci.edu/cgi-bin/mastpreview?mission=hst&dataid=J92A21031

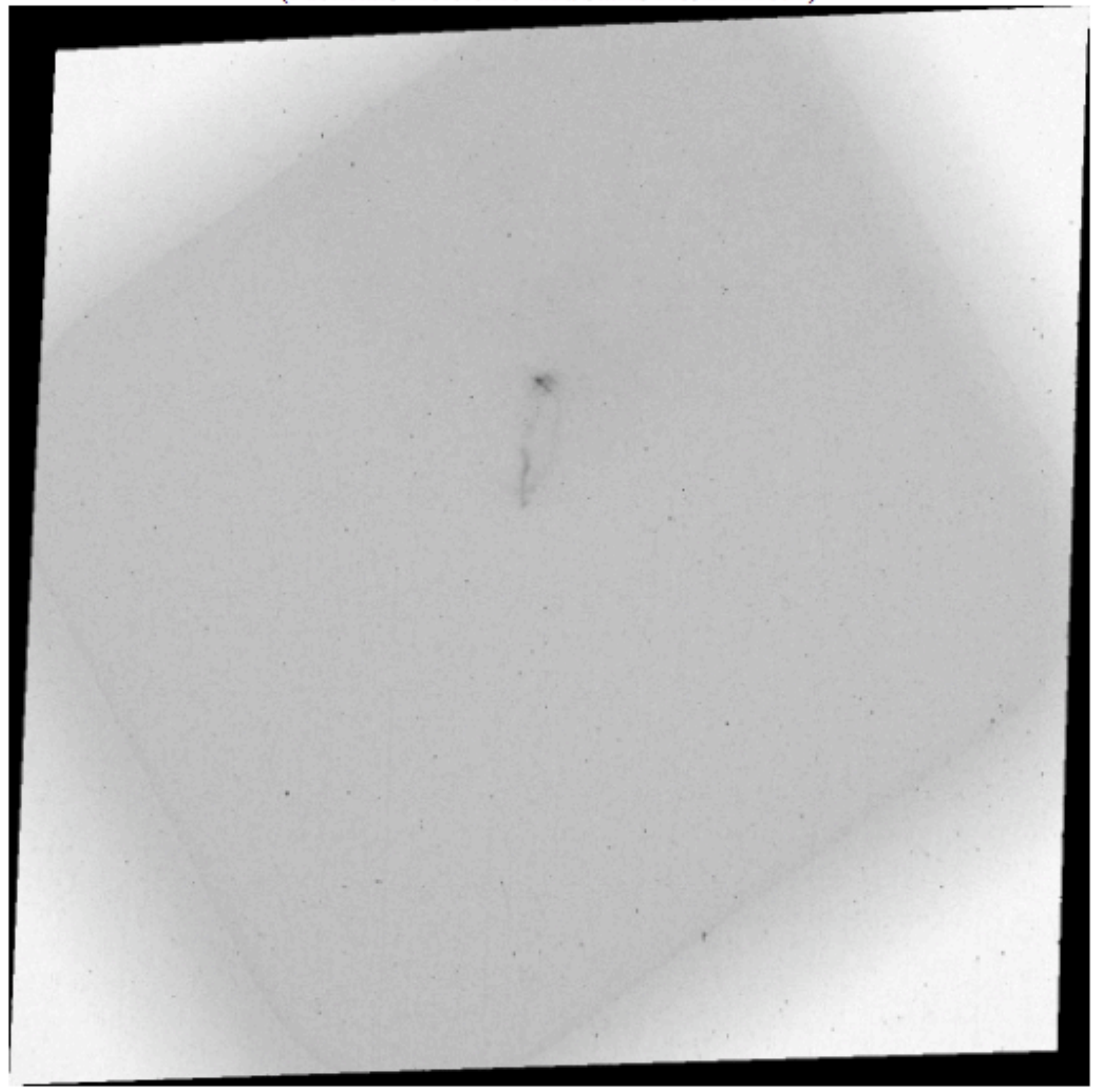
Barbara A. MIKULSKI ARCHIVE FOR SPACE TELESCOPES

MAST | STScI | Tools | Mission_Search | Tutorial | Site Search

About MAST | Getting Started

Preview for J92A21031

(Publication reference: ads/Sa.HST#J92A21031)



Preview calibrations are uncertain so preview data should be used for diagnostic/quick-look purposes only.

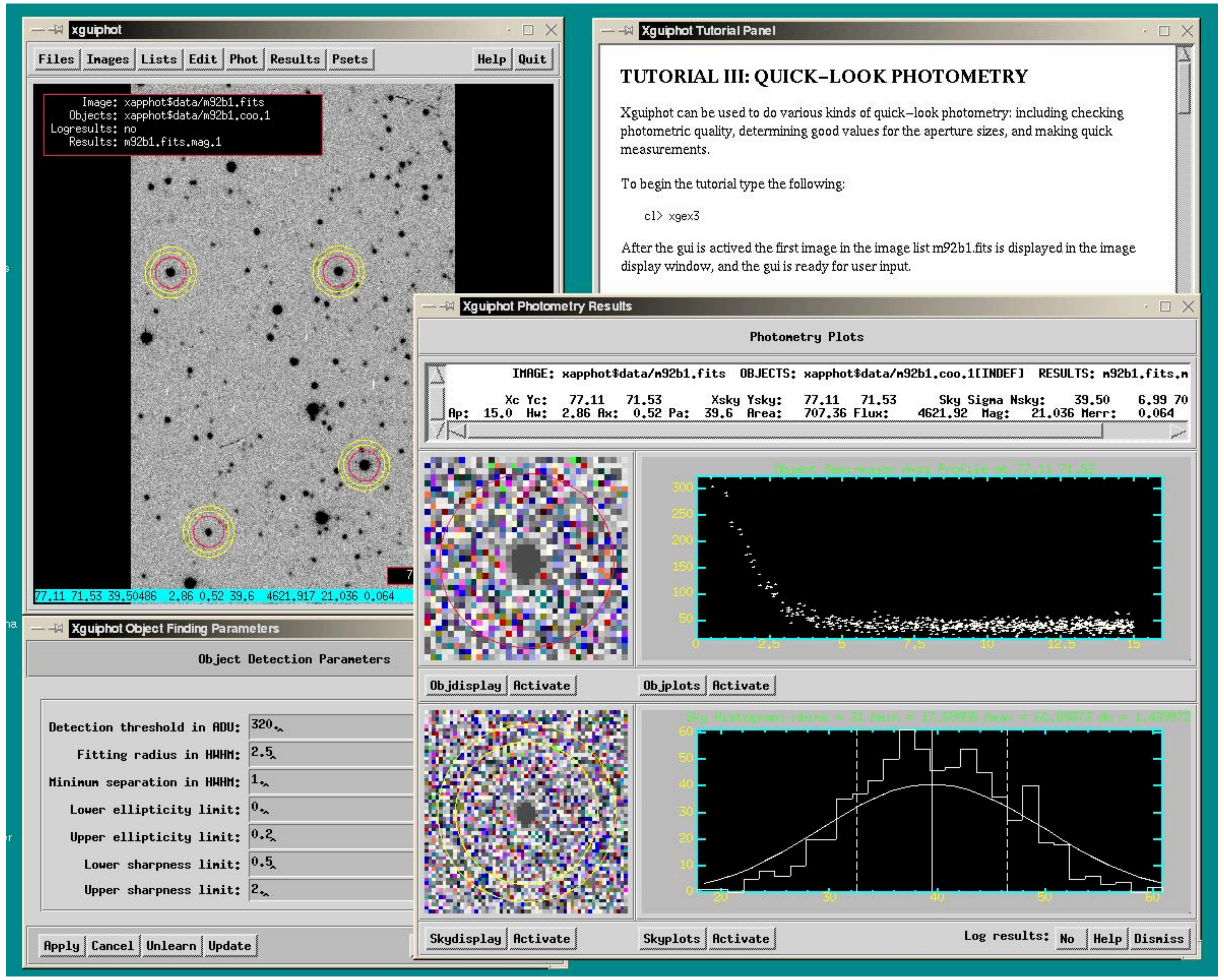
interactive display

Preview in FITS format More preview format options

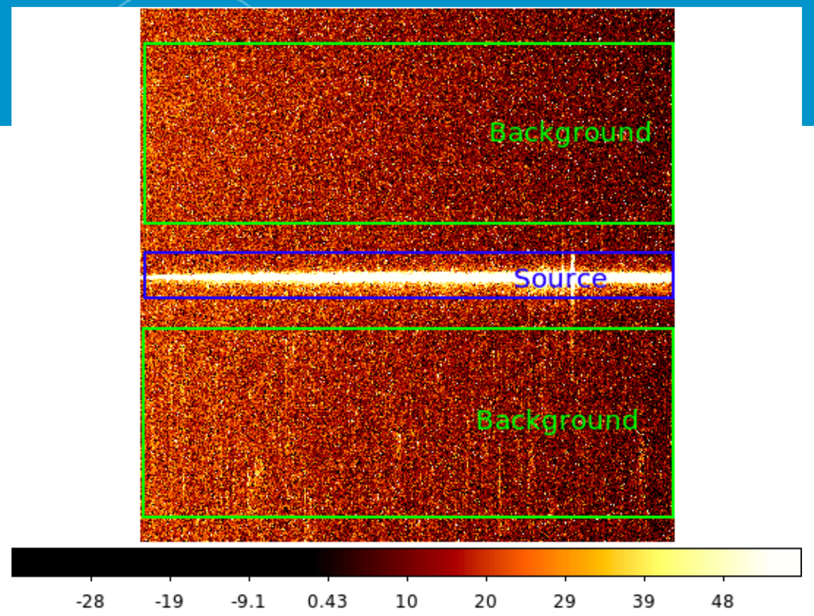
Exposure Information

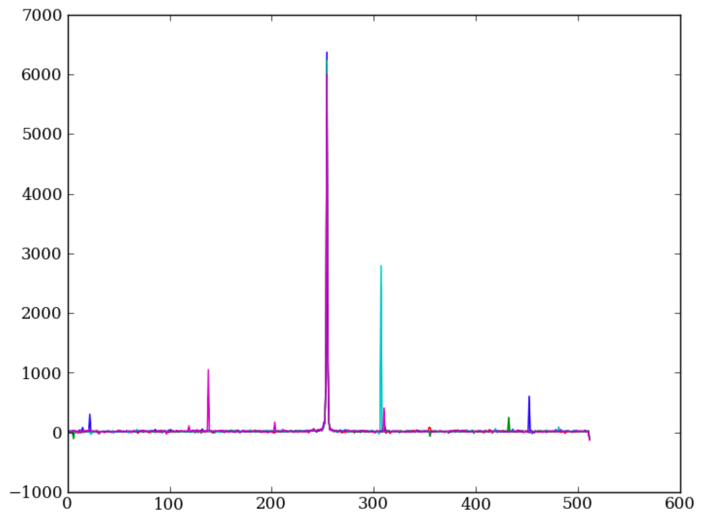
Target Name: L1551IRS5	**Observation Date:** Sep 2 2004 6:10PM	**Instrument:** ACS
RA: 04 31 34.13	**Exp Time:** 536	**Filter/Grating:** F606W;POL120V
Dec: +18 08 05.04	**Release Date:** Sep 3 2005 1:31AM	**Aperture:** WFC1-POL120V

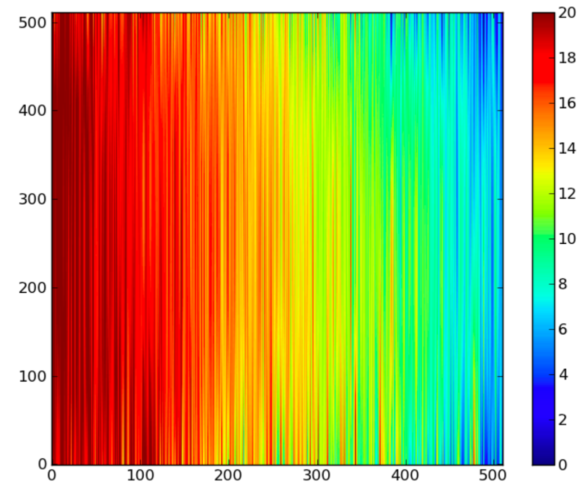
Traditional data reduction methods has problems

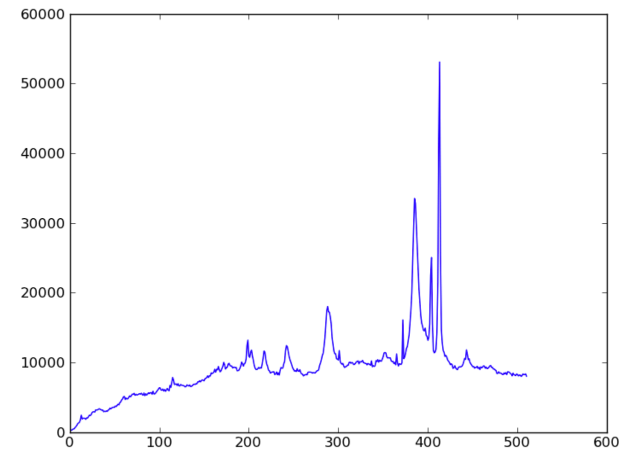


- Image Reduction and Analysis Facility
- Written in obsolete SPP and CL languages (1982)
- Hard to change
- Typical operations:
  - Look for sources
  - Fit curves with images
  - Subtract cosmic rays (median filter)
  - Fit for background and subtract



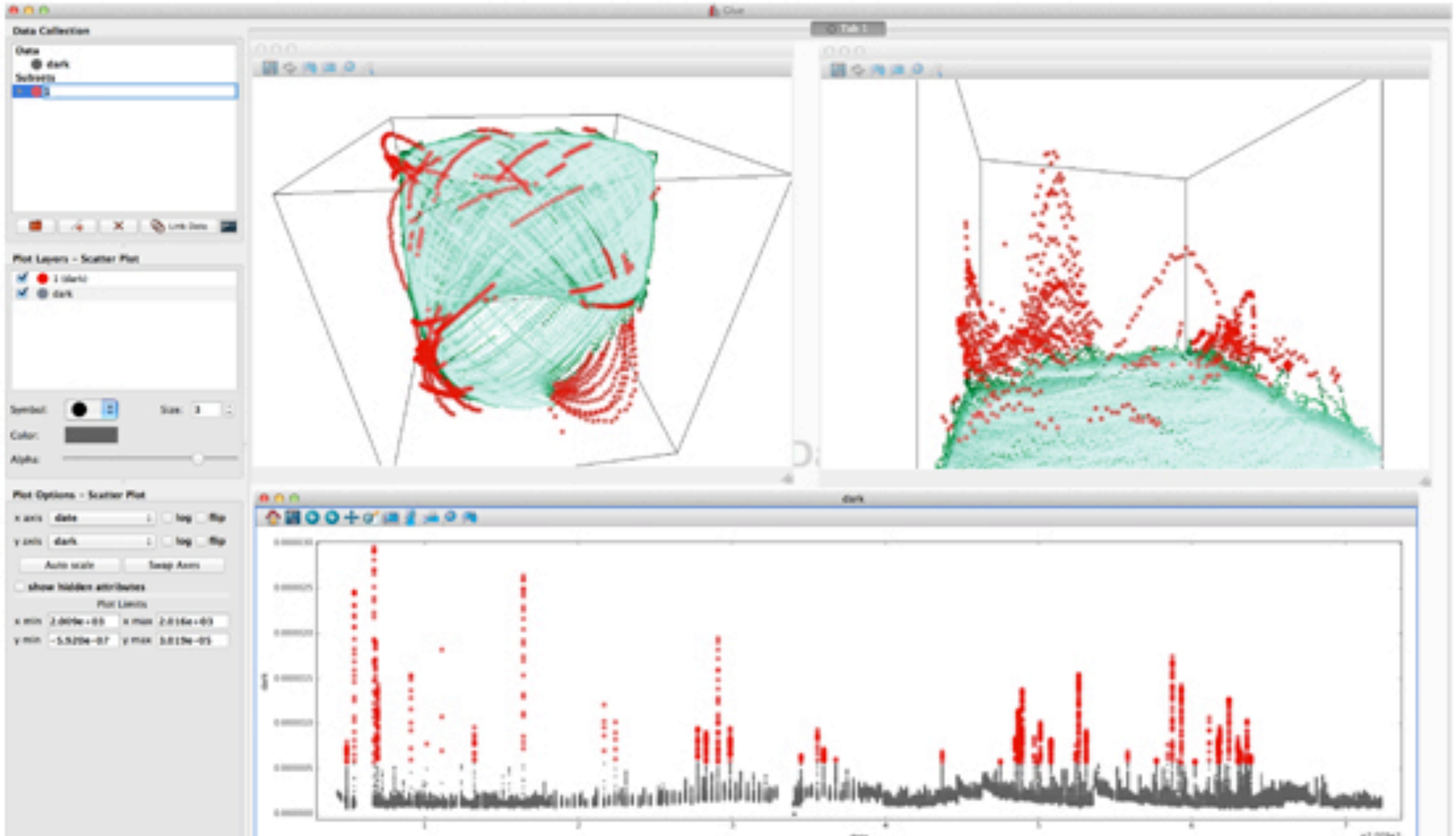


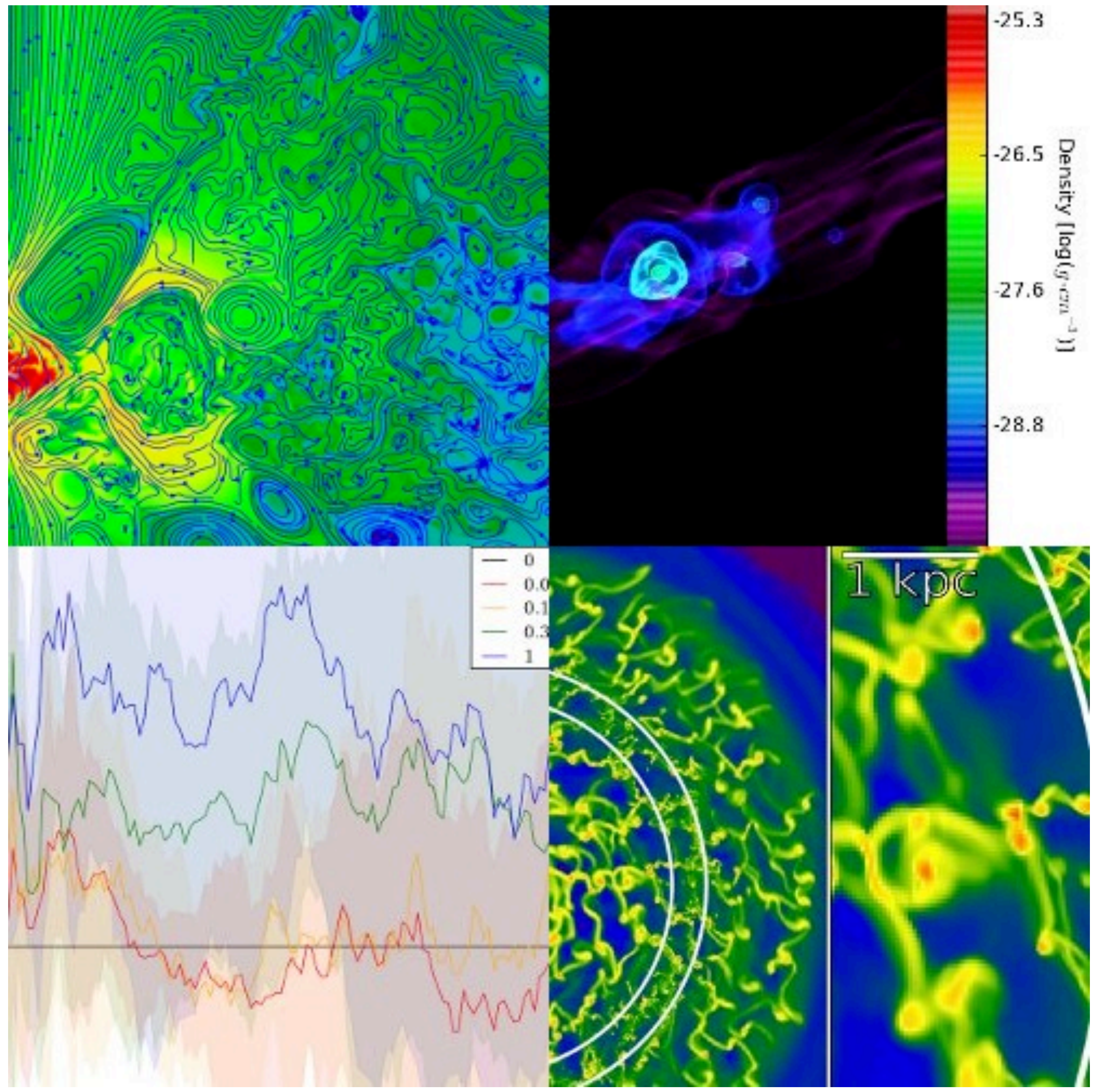




The new way: PyRAF, AstroConda

- Old IRAF wrapped in python
- No need to reimplement legacy software
- "Painless" transition to modern software



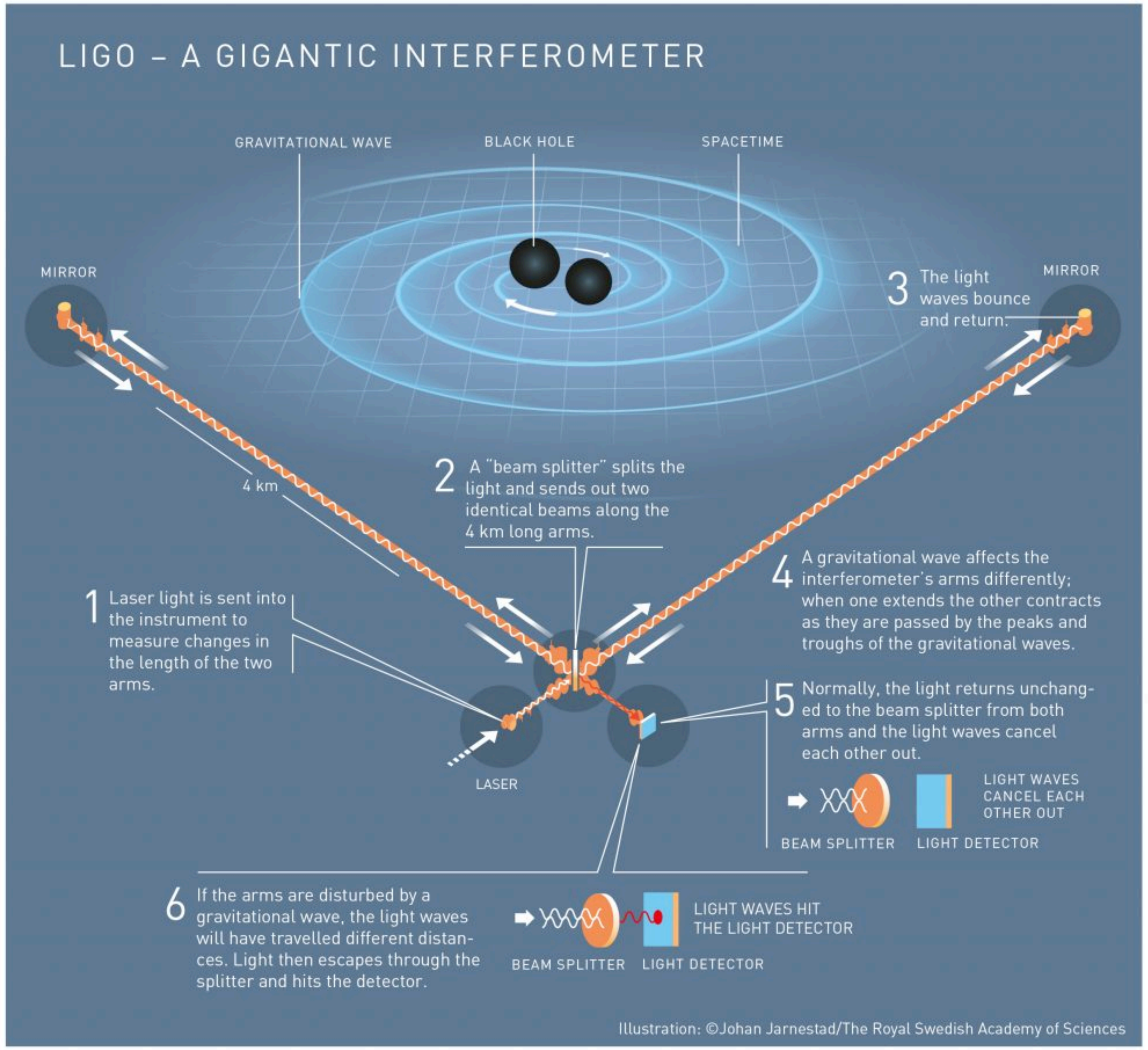


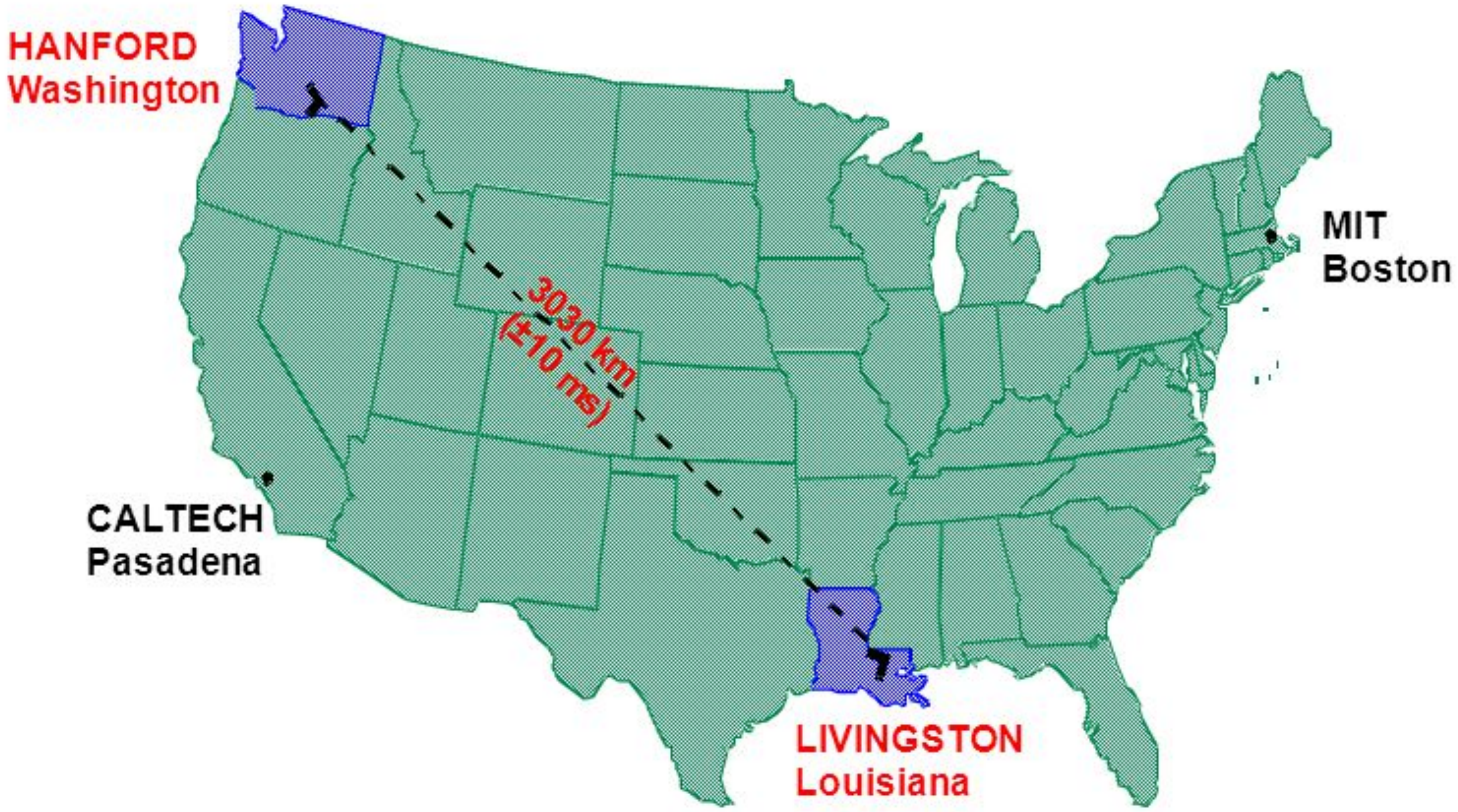
Problems

EUROPEAN SPALLATION SOURCE

- Reliance on old software
- Hard to transition to new data format
- New detectors can break free of constraints
- E.g. gravitational waves had less pre-existing workflows
- LIGO observatory

Laser Interferometer Gravitational Wave Observatory





All data free and online as observed - no embargo

- (Small print: Not all analysis tools are online)
- Data 3 formats, HDF5, gzipped ASCII and custom
- Can be mounted as local directory over CernVM-FS
- iPython/Jupyter notebooks for analysis
- Lots of data but few events - 1 event expected every 6 months! (6 since 2015)
- 1st neutron star collision in 2017

https://hub.mybinder.org/user/losc-tutorial-l-_event_tutorial-v299i1ce/notebooks/index.ipynb

jupyter index (unsaved changes)

File Edit View Insert Cell Kernel Widgets Help

Download the data on a computer with a python installation

If you are using a pre-configured setup (eg, in binder), great! You don't have to download or set up anything.

Otherwise, to begin, get the necessary files, by downloading the zip file and unpacking it into single directory:

- LOSC_Event_tutorial.zip

This zip file contains:

- this IPython notebook LOSC_Event_tutorial.ipynb, and LOSC_Event_tutorial.py code.
- python code for reading LOSC data files: readligo.py.
- the event data files (32s sampled at 4096 Hz, in hdf5 format, for both LIGO detectors).
- waveform templates (32s sampled at 4096 Hz, in hdf5 format, for both plus and cross polarizations).
- a parameter file in json format

You will also need a python installation with a few packages (numpy, matplotlib, scipy, h5py, json).

- For hints on software installation, see https://losc.ligo.org/tutorial00/
- The tutorial should work on python 2.6 and above, including python 3, as well as in recent versions of Ipython.
- You *might* see Warning or FutureWarning messages, which tend to be associated with different versions of python, Ipython, numpy, etc. Hopefully they can be ignored!
- the filetype "hdf5" means the data are in hdf5 format: https://www.hdfgroup.org/HDF5/
- NOTE: GPS time is number of seconds since Jan 6, 1980 GMT. See http://www.oc.nps.edu/oc2902w/gps/timsys.html or https://losc.ligo.org/gps/

Set the event name to choose event and the plot type

In [1]:
```
#-- SET ME   Tutorial should work with most binary black hole events
#-- Default is no event selection; you MUST select one to proceed.
eventname = ''
eventname = 'GW150914'
#eventname = 'GW151226'
#eventname = 'LVT151012'
#eventname = 'GW170104'

# want plots?
make_plots = 1
plottype = "png"
#plottype = "pdf"
```

In [2]:
```
# Standard python numerical analysis imports:
import numpy as np
```

https://hub.mybinder.org/user/losc-tutorial-l-_event_tutorial-v299i1ce/notebooks/index.ipynb

jupyter index (autosaved)

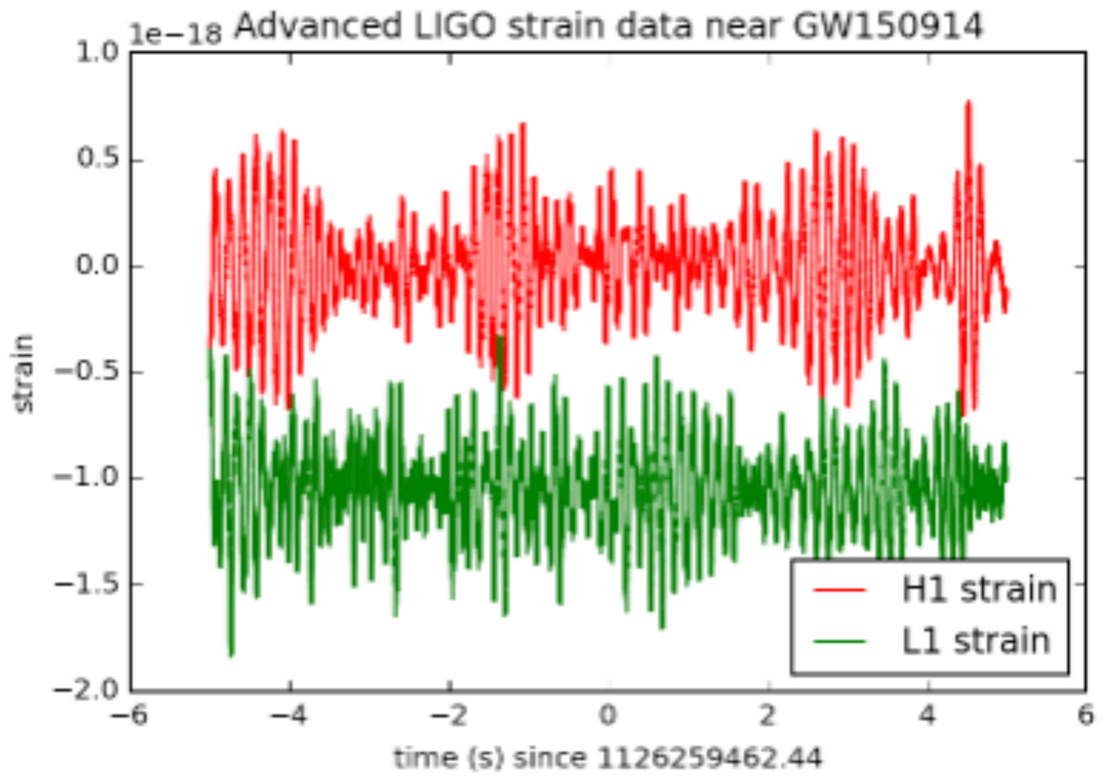
File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 2

```
For L1, 32 out of 32 seconds contain usable DATA
```

In [7]:

```
# plot +- deltat seconds around the event:
# index into the strain time series for this time interval:
deltat = 5
indxt = np.where((time >= tevent-deltat) & (time < tevent+deltat))
print(tevent)

if make_plots:
    plt.figure()
    plt.plot(time[indxt]-tevent,strain_H1[indxt],'r',label='H1 strain')
    plt.plot(time[indxt]-tevent,strain_L1[indxt],'g',label='L1 strain')
    plt.xlabel('time (s) since '+str(tevent))
    plt.ylabel('strain')
    plt.legend(loc='lower right')
    plt.title('Advanced LIGO strain data near '+eventname)
    plt.savefig(eventname+'_strain.'+plottype)
```

```
1126259462.44
```



The data are dominated by **low frequency noise**; there is no way to see a signal here, without some signal processing.

Plot the Amplitude Spectral Density (ASD)

Plotting these data in the Fourier domain gives us an idea of the frequency content of the data. A way to visualize the frequency content of the data is to plot the amplitude spectral density, ASD.

The ASDs are the square root of the power spectral densities (PSDs), which are averages of the square of the fast fourier transforms (FFTs) of the data.

They are an estimate of the "strain-equivalent noise" of the detectors versus frequency, which limit the ability of the detectors to identify GW signals.

EUROPEAN SPALLATION SOURCE

https://hub.mybinder.org/user/losc-tutorial-l-_event_tutorial-v299i1ce/notebooks/index.ipynb

jupyter index (autosaved)

Hz and harmonics, calibration lines, control dither lines, etc) and some (60 Hz and harmonics) are unwanted. We'll return to these, later.

You can't see the signal in this plot, since it is relatively weak and less than a second long, while this plot averages over 32 seconds of data. So this plot is entirely dominated by instrumental noise.

The smooth model is hard-coded and tuned by eye; it won't be right for arbitrary times. We will only use it below for things that don't require much accuracy.

Binary Neutron Star (BNS) detection range

A standard metric that LIGO uses to evaluate the sensitivity of our detectors, based on the detector noise ASD, is the BNS range.

This is defined as the distance to which a LIGO detector can register a BNS signal with a single detector signal-to-noise ratio (SNR) of 8, averaged over source direction and orientation. Here, SNR 8 is used as a nominal detection threshold, similar to typical CBC detection thresholds of SNR 6-8.

We take each neutron star in the BNS system to have a mass of 1.4 times the mass of the sun, and negligible spin.

GWs from BNS mergers are like "standard sirens"; we know their amplitude at the source from theoretical calculations. The amplitude falls off like 1/r, so their amplitude at the detectors on Earth tells us how far away they are. This is great, because it is hard, in general, to know the distance to astronomical sources.

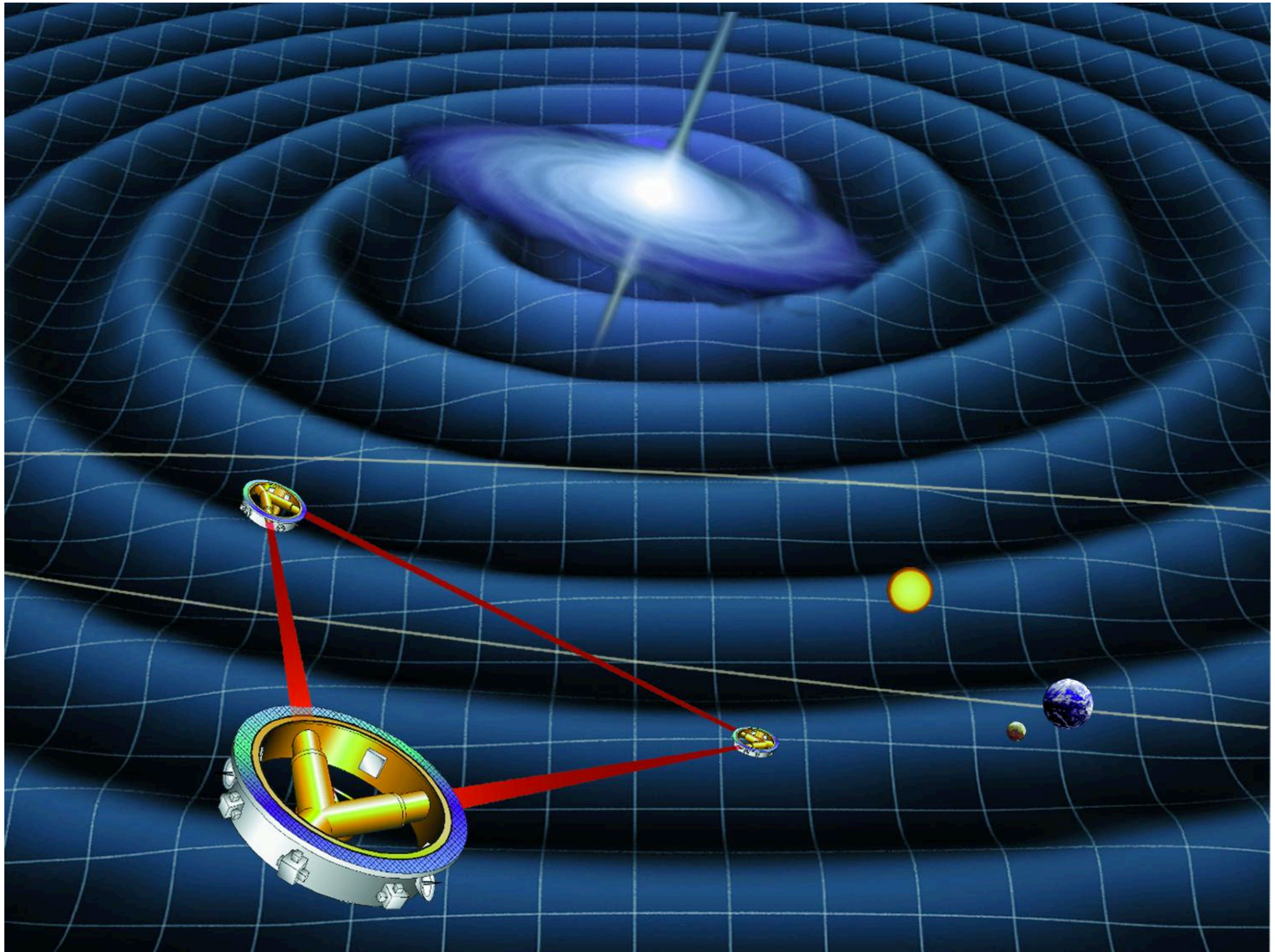
The amplitude at the source is computed in the post-Newtonian "quadrupole approximation". This is valid for the inspiral phase only, and is approximate at best; there is no simple expression for the post-inspiral (merger and ringdown) phase. So this won't work for high-mass binary black holes like GW150914, which have a lot of signal strength in the post-inspiral phase.

But, in order to use them as standard sirens, we need to know the source direction and orientation relative to the detector and its "quadrupole antenna pattern" response to such signals. It is a standard (if non-trivial) computation to average over all source directions and orientations; the average amplitude is 1./2.2648 times the maximum value.

This calculation is described in Appendix D of: FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries B. Allen et al., PHYSICAL REVIEW D 85, 122006 (2012) ; http://arxiv.org/abs/gr-qc/0509116

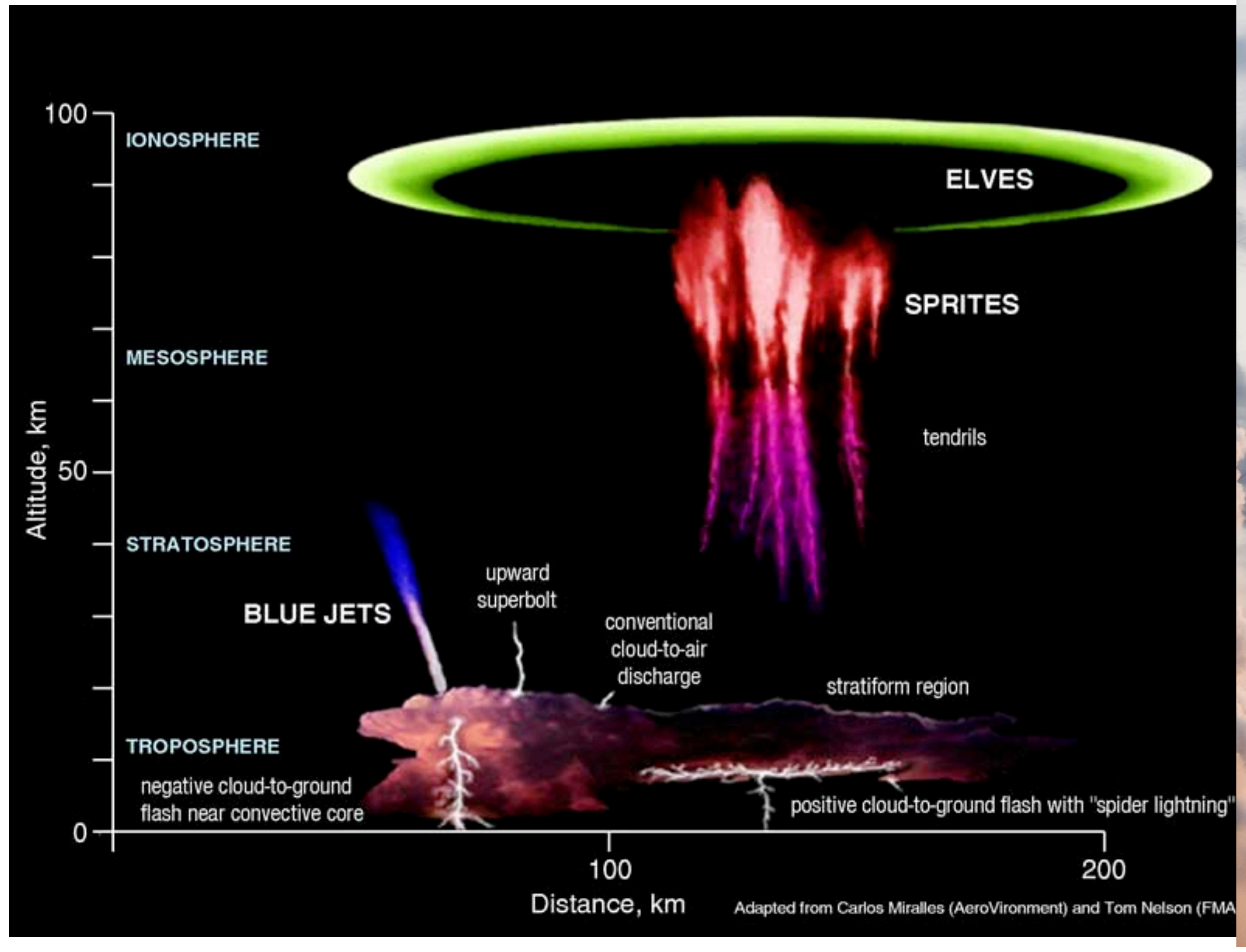
In [9]:
```
BNS_range = 1
if BNS_range:
    #-- compute the binary neutron star (BNS) detectability range

    #-- choose a detector noise power spectrum:
    f = freqs.copy()
    # get frequency step size
    df = f[2]-f[1]

    #-- constants
    # speed of light:
    clight = 2.99792458e8                # m/s
    # Newton's gravitational constant
    G = 6.67259e-11                      # m^3/kg/s^2
    # one parsec, popular unit of astronomical distance (around 3.26 light years)
    parsec = 3.08568025e16               # m
```

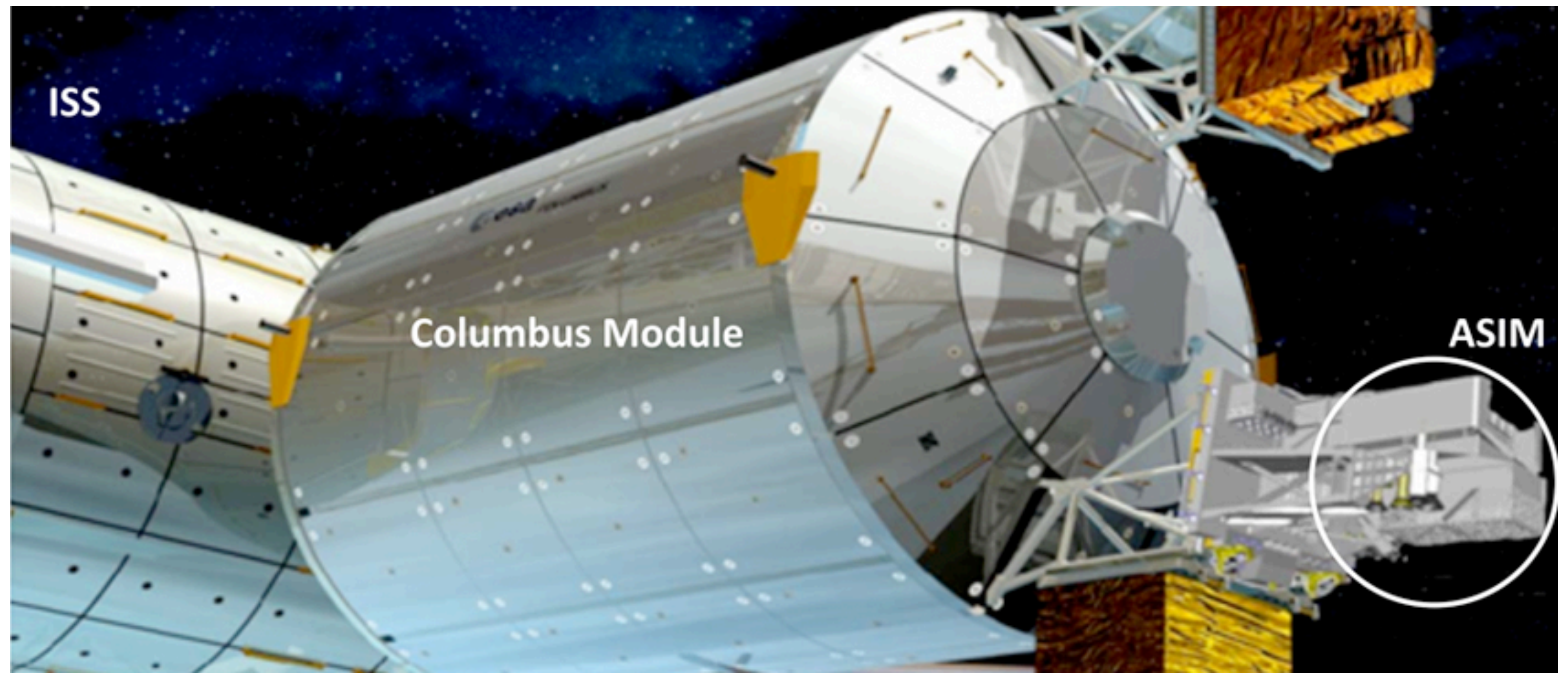
Result of LIGO success: will be LISA



Atmosphere-Space Interaction Monitor

- Datafile are split and compressed and downloaded in small packets ~ 1 MB over a slow connection (~56 k) from satellite, e.g. from space station to White Sands, NM, to Brussels, to Copenhagen
- Reassembled into original files - Level 0
- Calibration files downloaded and used to calibrate observations
- Data combined with other sources (e.g. cloud observations from METEOSAT)

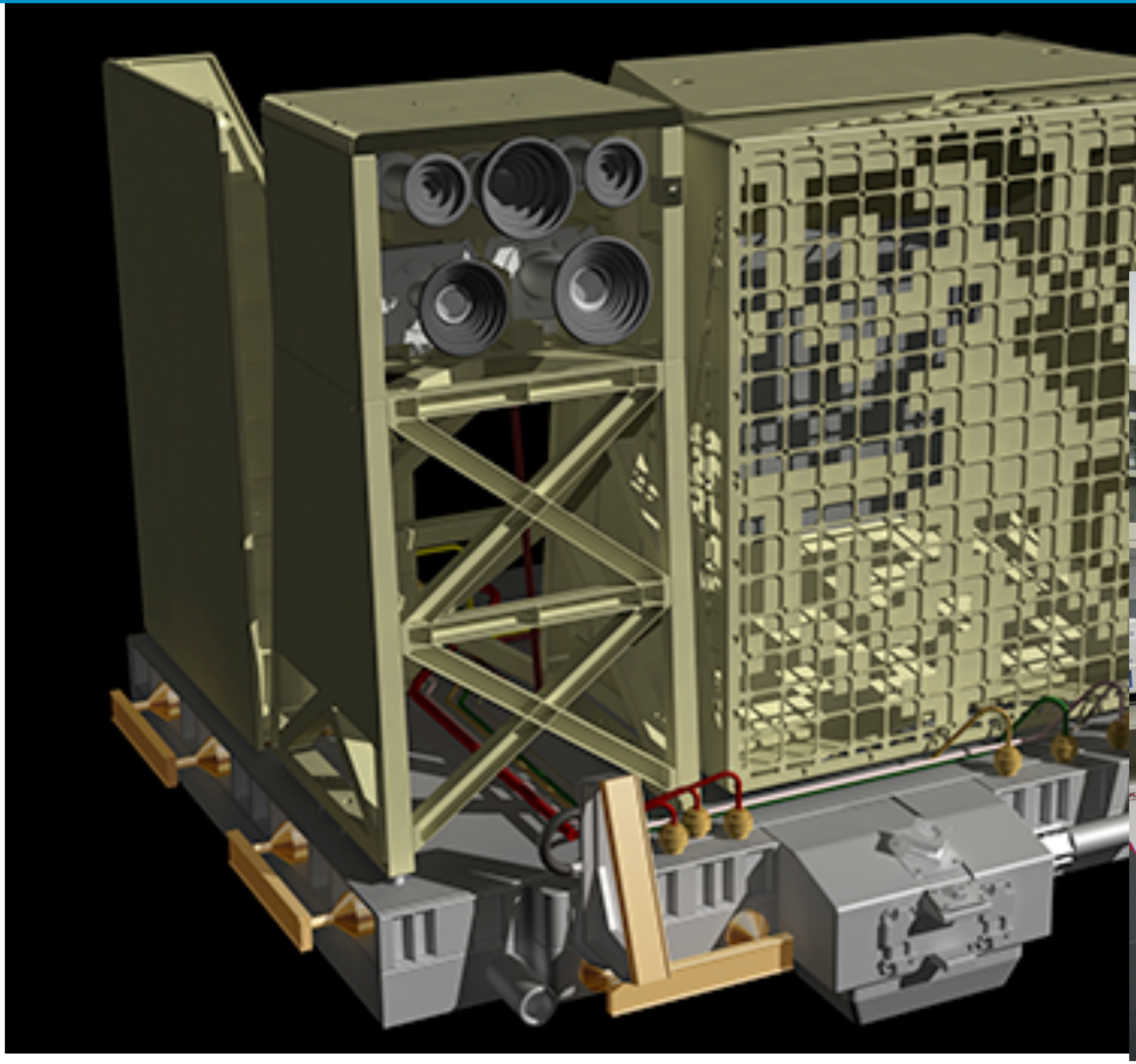


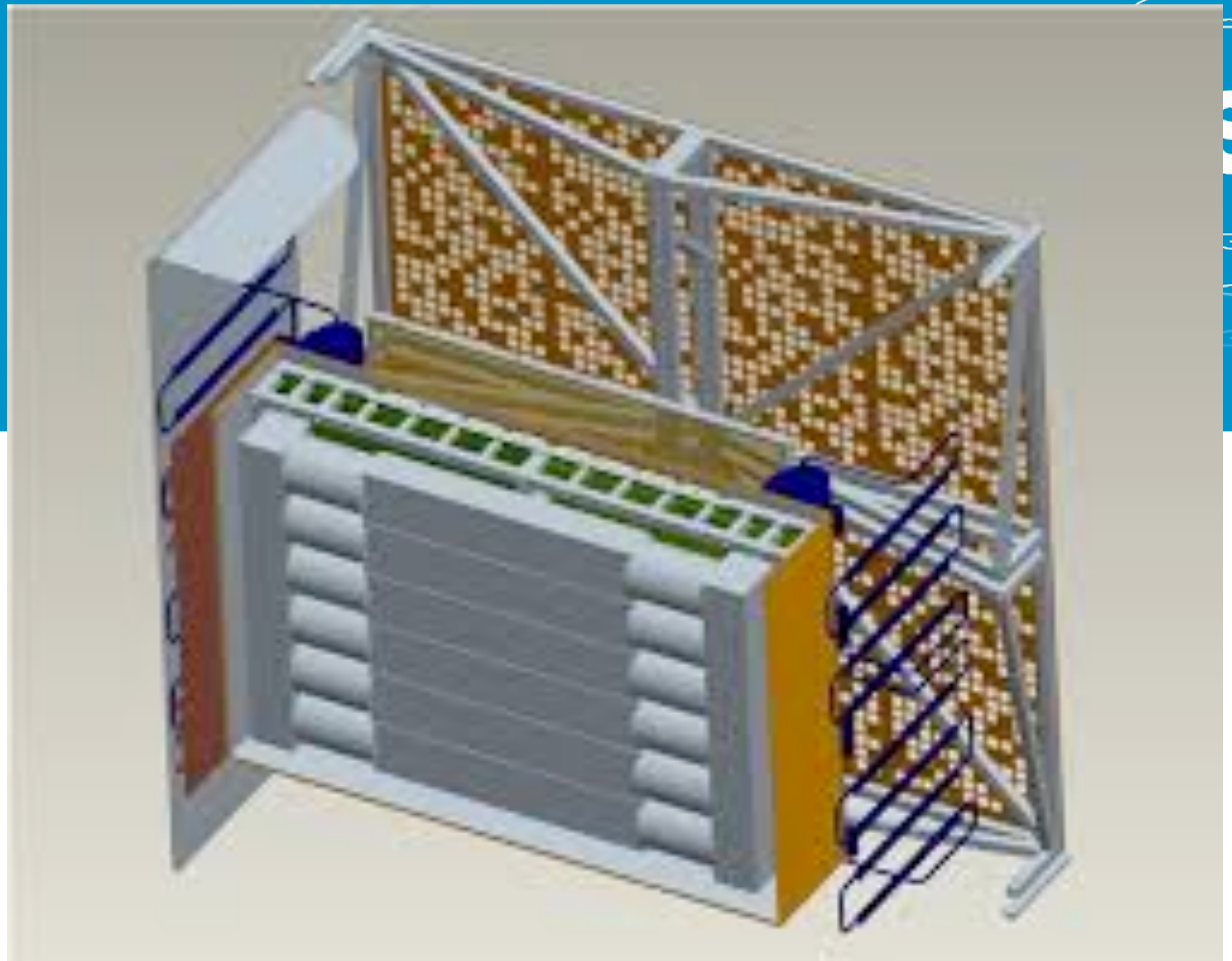


MXGS and MMIA

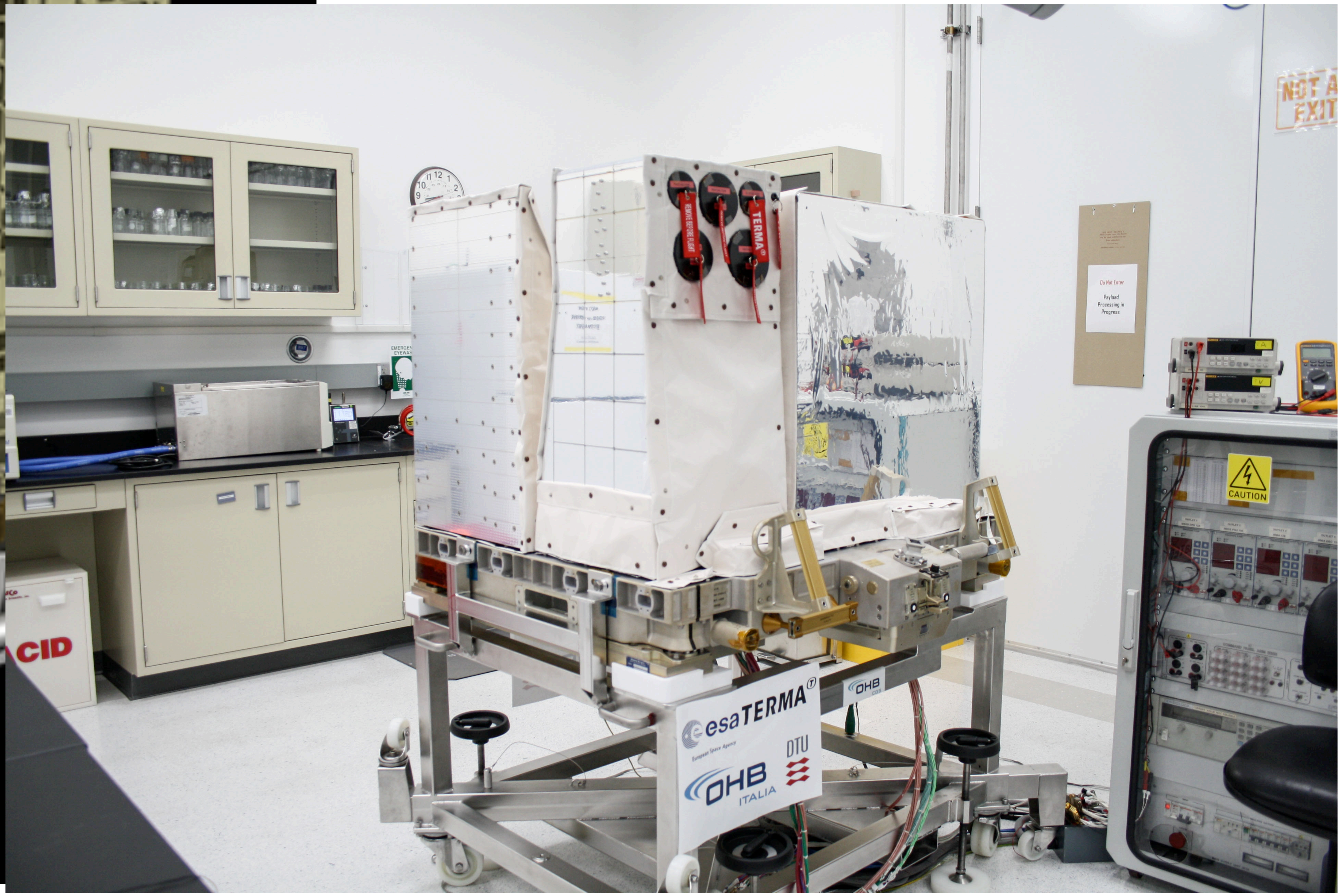
Modular X-ray and Gamma-ray Sensor &
Modular Multi-Imaging Assembly

EUROPEAN
SPALLATION
SOURCE





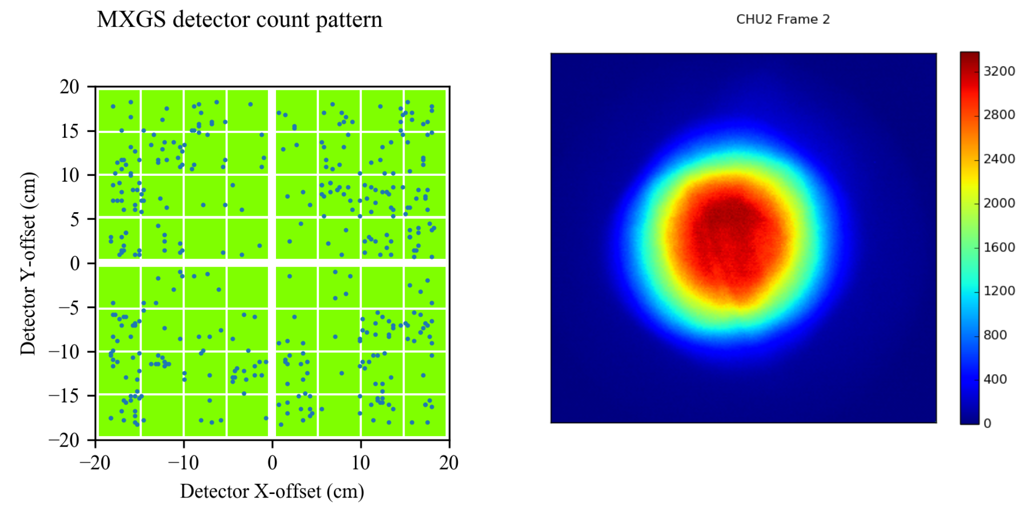
BGO, CZT



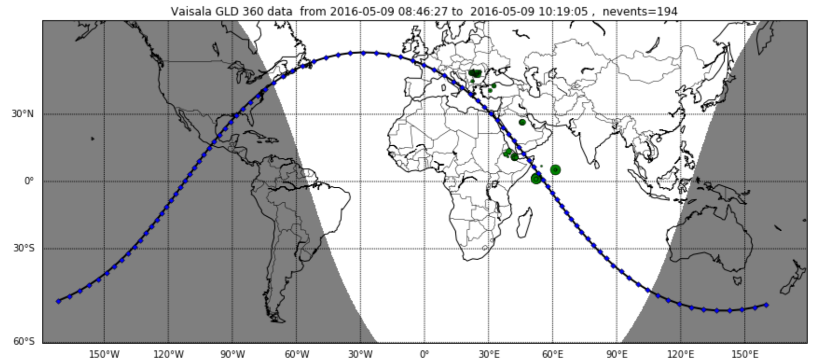
DATA PROCESSING LEVELS

```
001100000011000
000111010001100
0001100000010
```

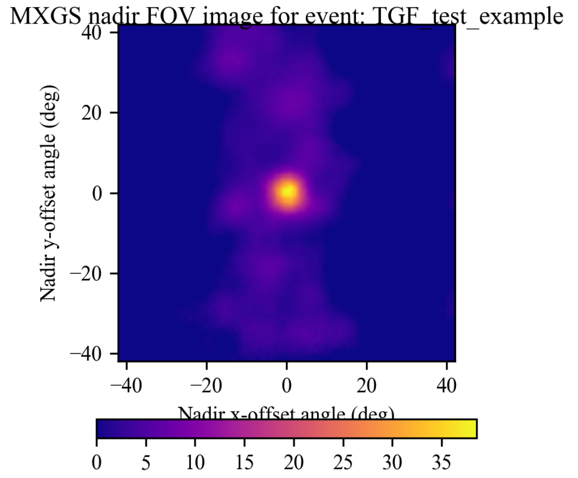
Level 0 - Reconstructed **unprocessed** instrument data at full resolution

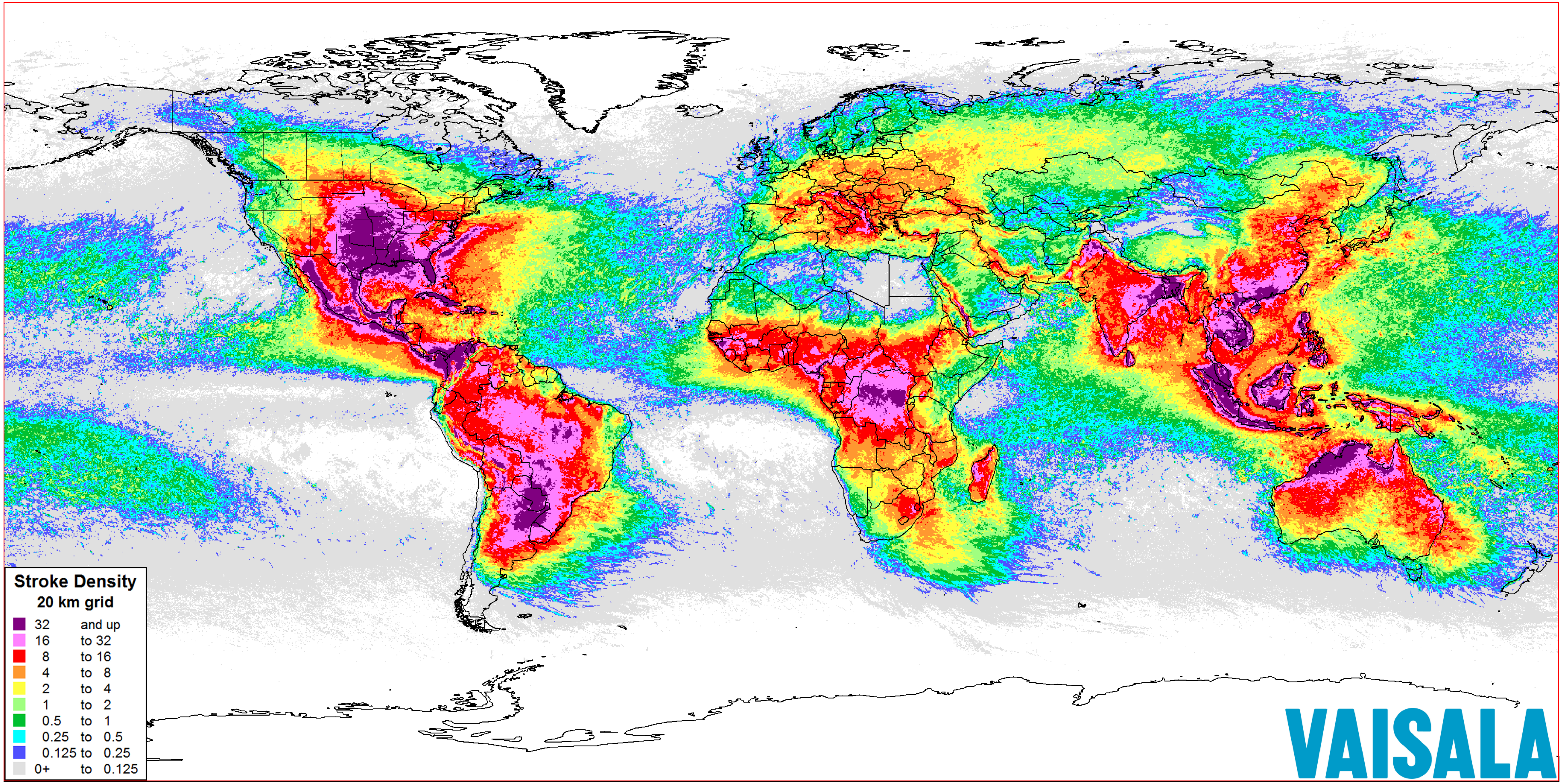


- Level 1 - Reconstructed instrument data **time referenced**, and annotated with ancillary information, including **calibration** coefficients and **georeferencing**

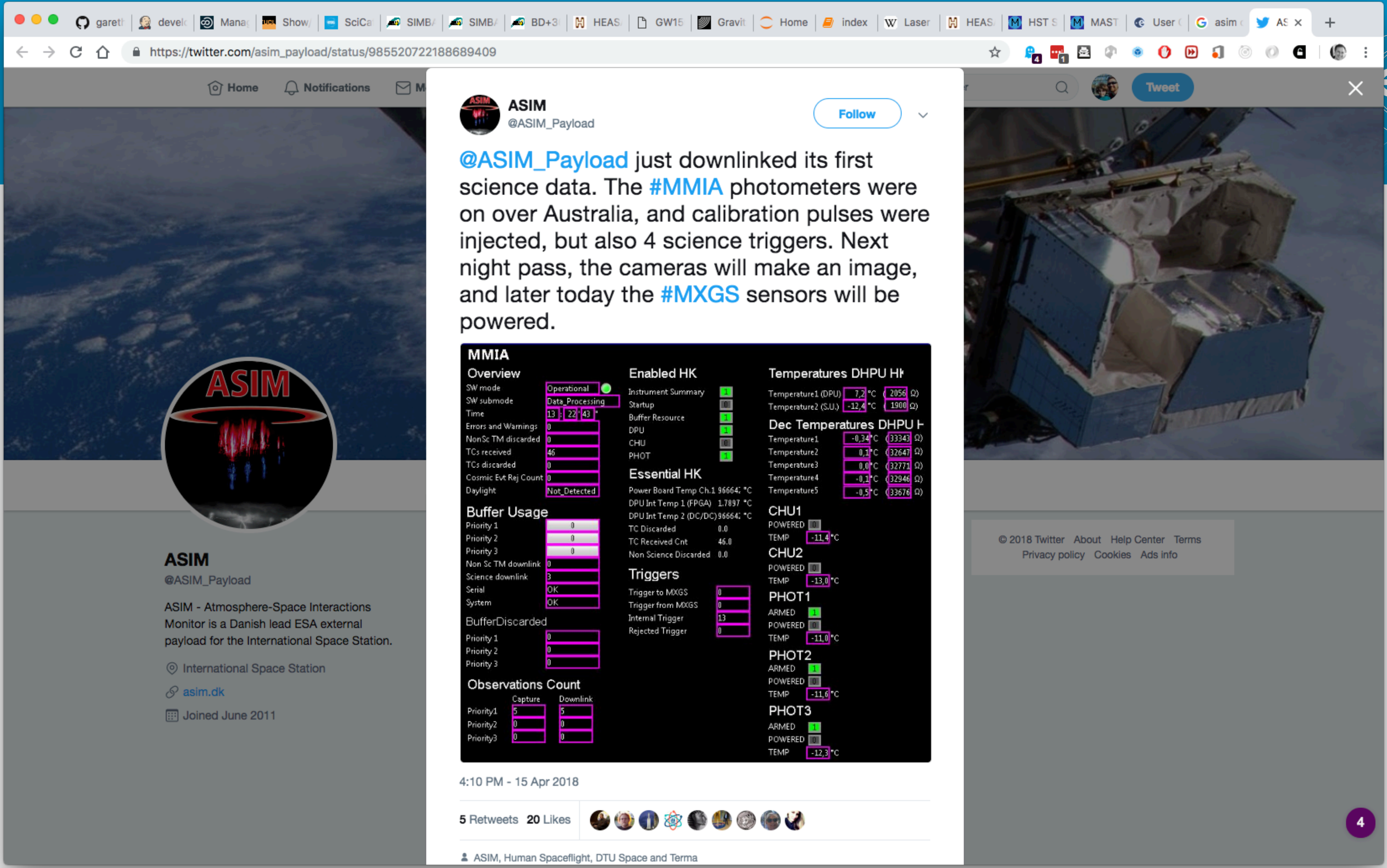


Level 2 - **Derived** environmental variables at the same resolution and location as the Level 1 source data.

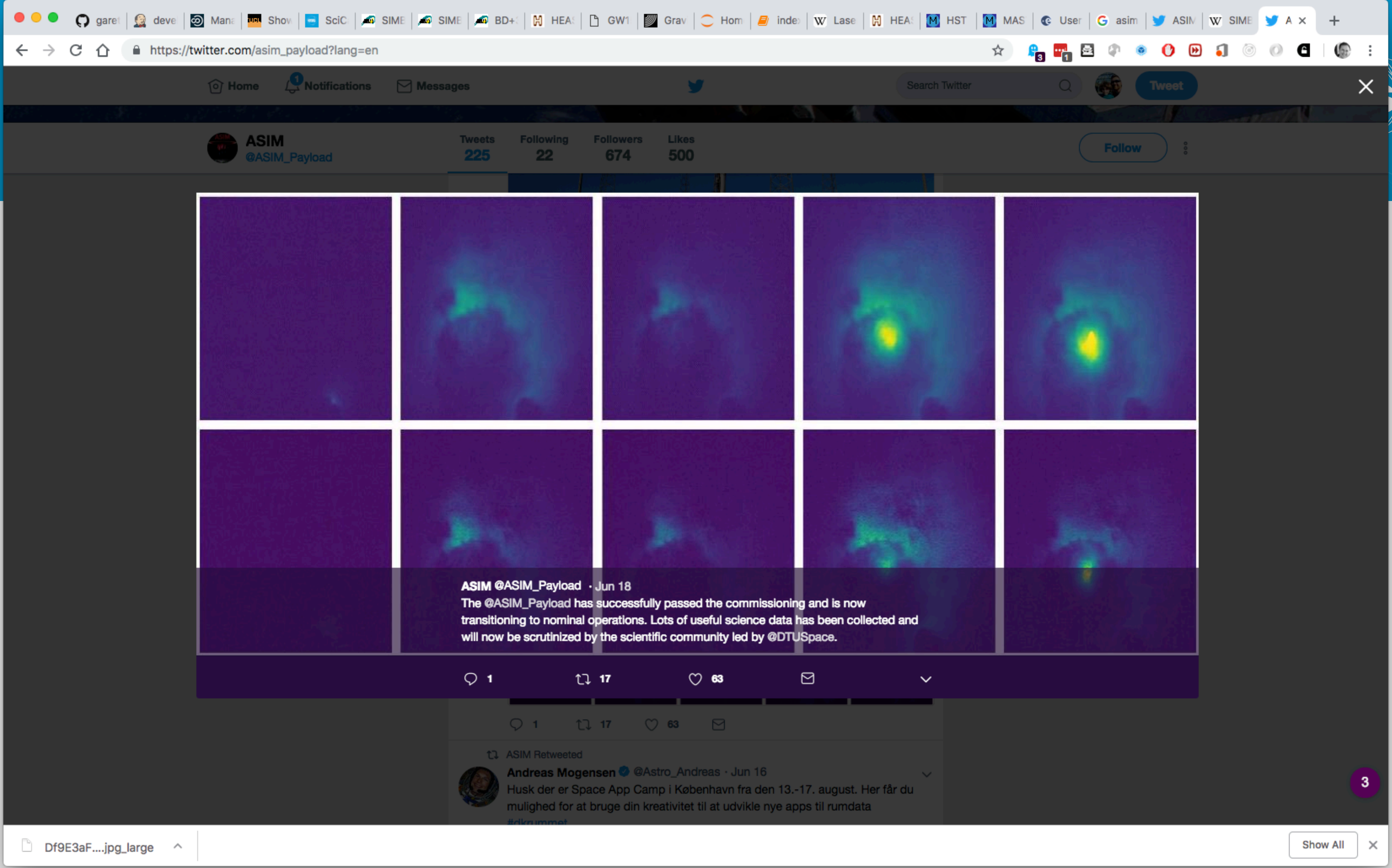




Stroke Density Map - 20 km grid Average: 2012-2016 GLD360 v2.0 data: 7,828,464,140 strokes


https://twitter.com/asim_payload/status/985520722188689409
Home
Notifications
Tweet
ASIM
@ASIM_Payload
Follow
@ASIM_Payload just downlinked its first science data. The #MMIA photometers were on over Australia, and calibration pulses were injected, but also 4 science triggers. Next night pass, the cameras will make an image, and later today the #MXGS sensors will be powered.
MMIA
Overview
SW mode
Operational
SW submode
Data_Processing
Time
Errors and Warnings
NonSc TM discarded
TCs received
TCs discarded
Cosmic Evt Rej Count
Daylight
Not_Detected
Buffer Usage
Priority 1
Priority 2
Priority 3
Non Sc TM downlink
Science downlink
Serial
System
BufferDiscarded
Observations Count
Capture
Downlink
Enabled HK
Instrument Summary
Startup
Buffer Resource
DPU
CHU
PHOT
Essential HK
Power Board Temp Ch.1
DPU Int Temp 1 (FPGA)
DPU Int Temp 2 (DC/DC)
TC Discarded
TC Received Cnt
Non Science Discarded
Triggers
Trigger to MXGS
Trigger from MXGS
Internal Trigger
Rejected Trigger
Temperatures DPHU HK
Temperature1 (DPU)
Temperature2 (S.U.)
Dec Temperatures DPHU H
CHU1
POWERED
TEMP
CHU2
PHOT1
ARMED
PHOT2
PHOT3
4:10 PM - 15 Apr 2018
5 Retweets 20 Likes
ASIM, Human Spaceflight, DTU Space and Terma
ASIM
@ASIM_Payload
ASIM - Atmosphere-Space Interactions Monitor is a Danish lead ESA external payload for the International Space Station.
International Space Station
asim.dk
Joined June 2011
© 2018 Twitter About Help Center Terms
Privacy policy Cookies Ads info

EUROPEAN SPALLATION SOURCE



Conclusions

EUROPEAN SPALLATION SOURCE

- Astronomers are getting larger amounts of data as detector size increases
- Transitioning gradually to new formats e.g. HDF5
- As much info/metadata online and available in a single source as possible
- Opportunities for multi-wavelength/multi-instrument studies (e.g. gravitational waves and X-ray)