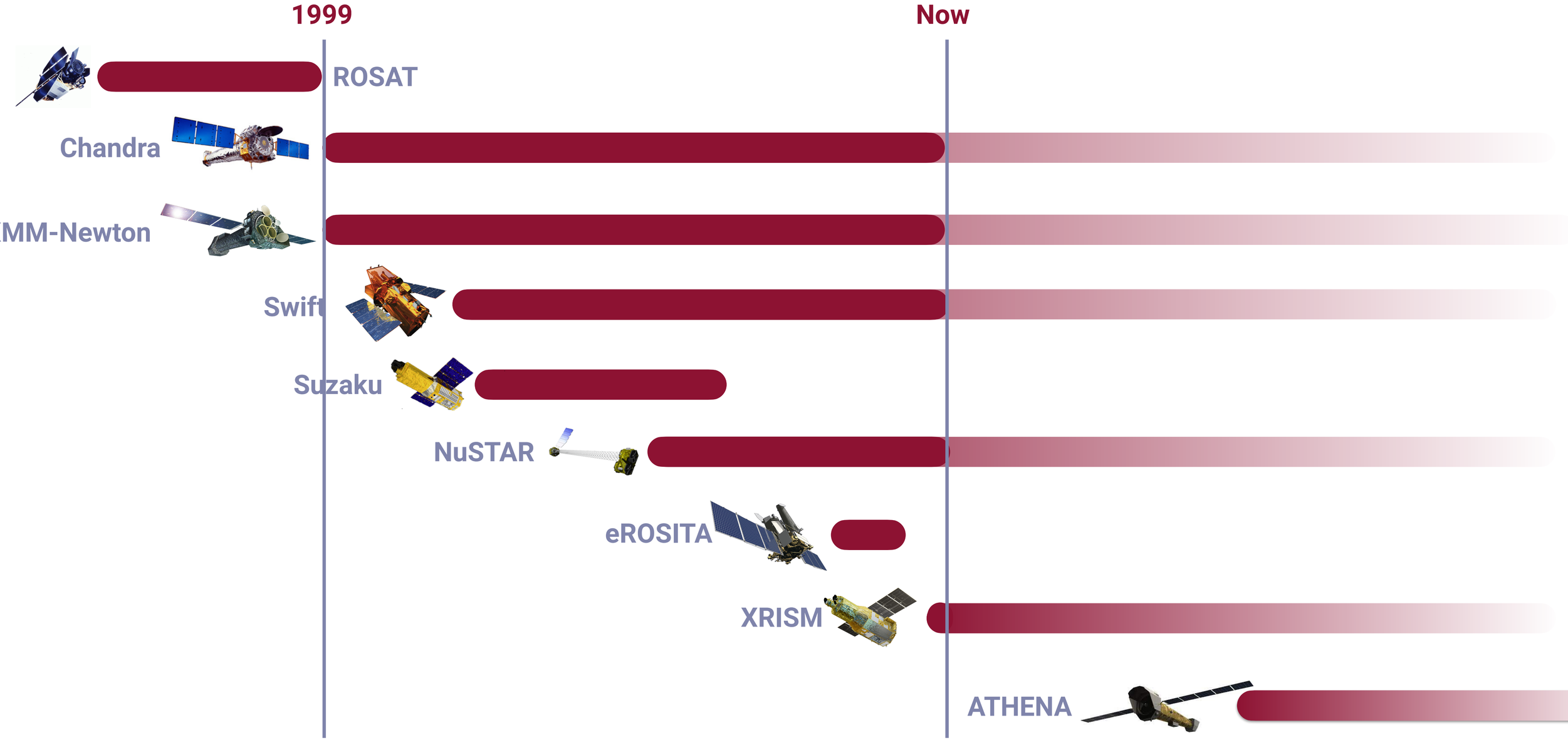
CENTER FOR ASTROPHYSICS
HARVARD & SMITHSONIAN

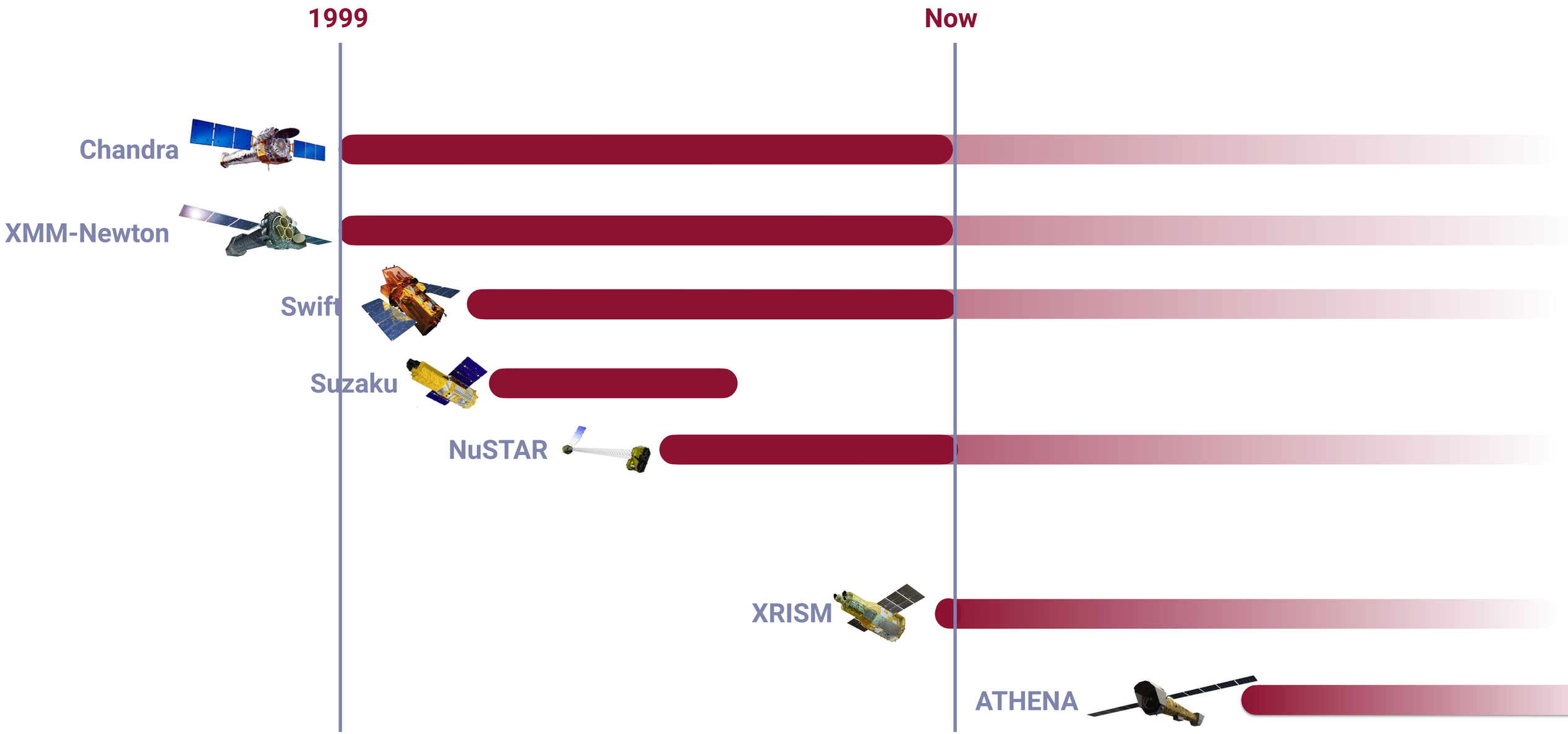
Paving the way to a bright X-ray future: the role and legacy of the Chandra X-ray Observatory

R. D'Abrusco
and CDA & CSC teams

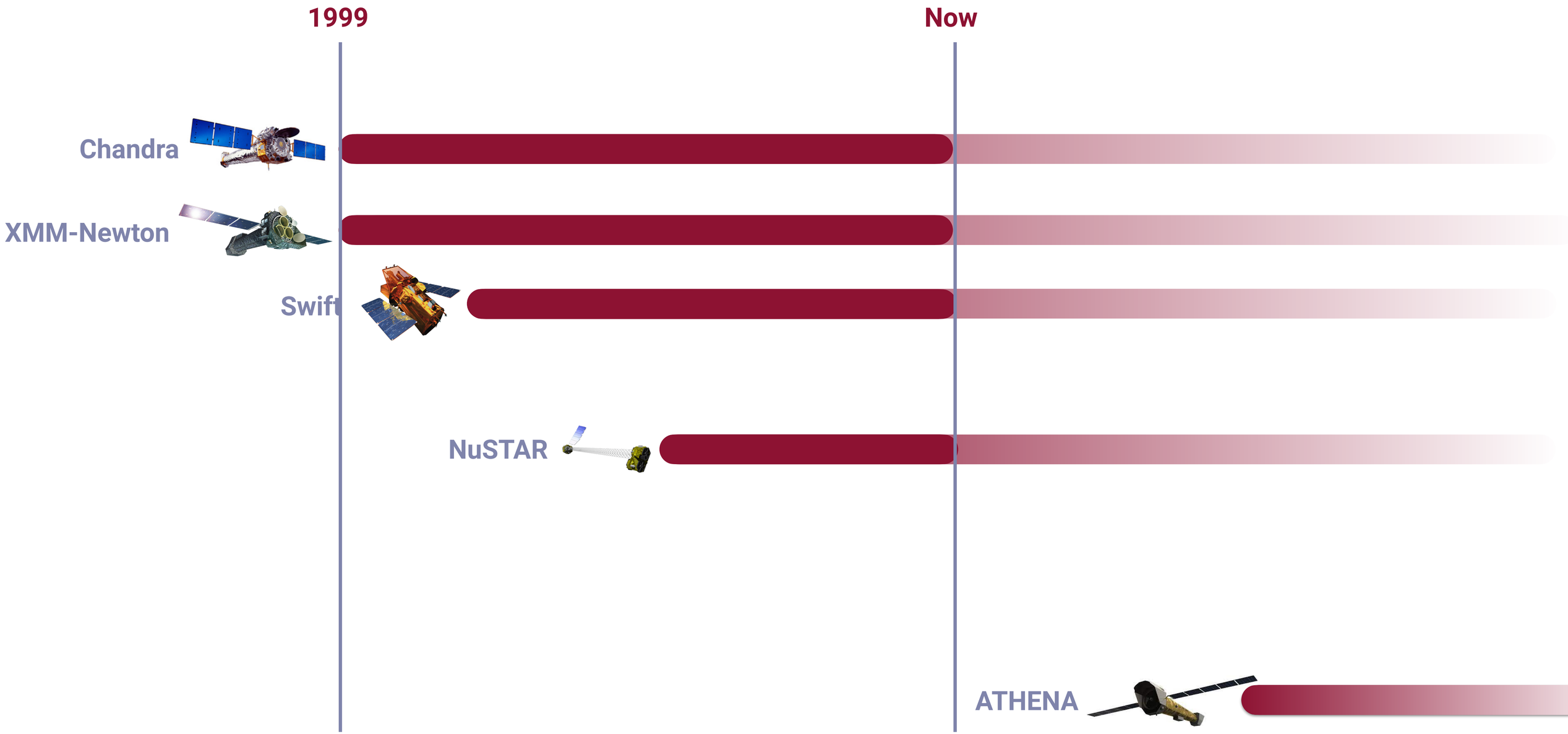
Golden age of X-ray astronomy



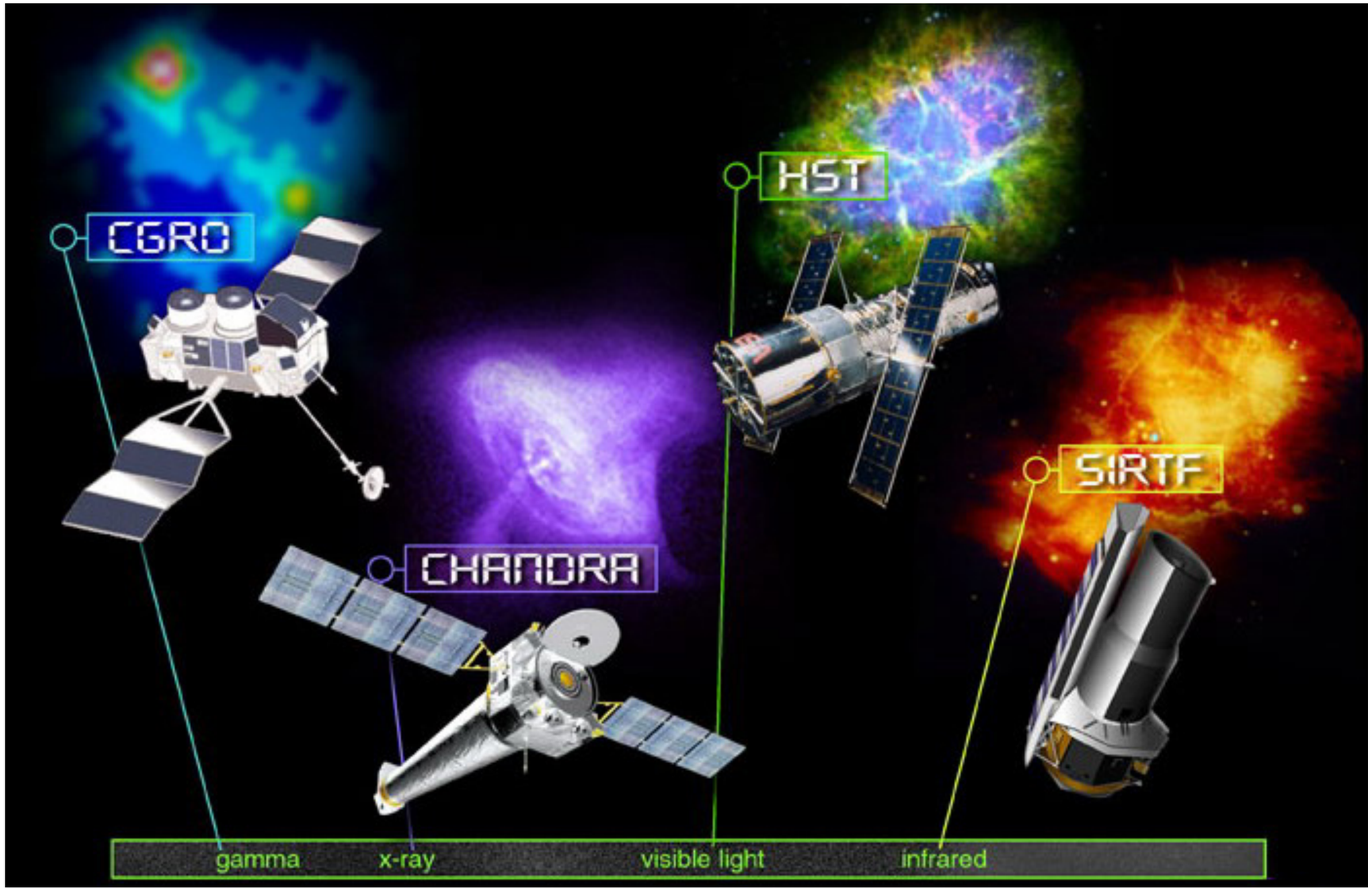
Golden age of X-ray astronomy



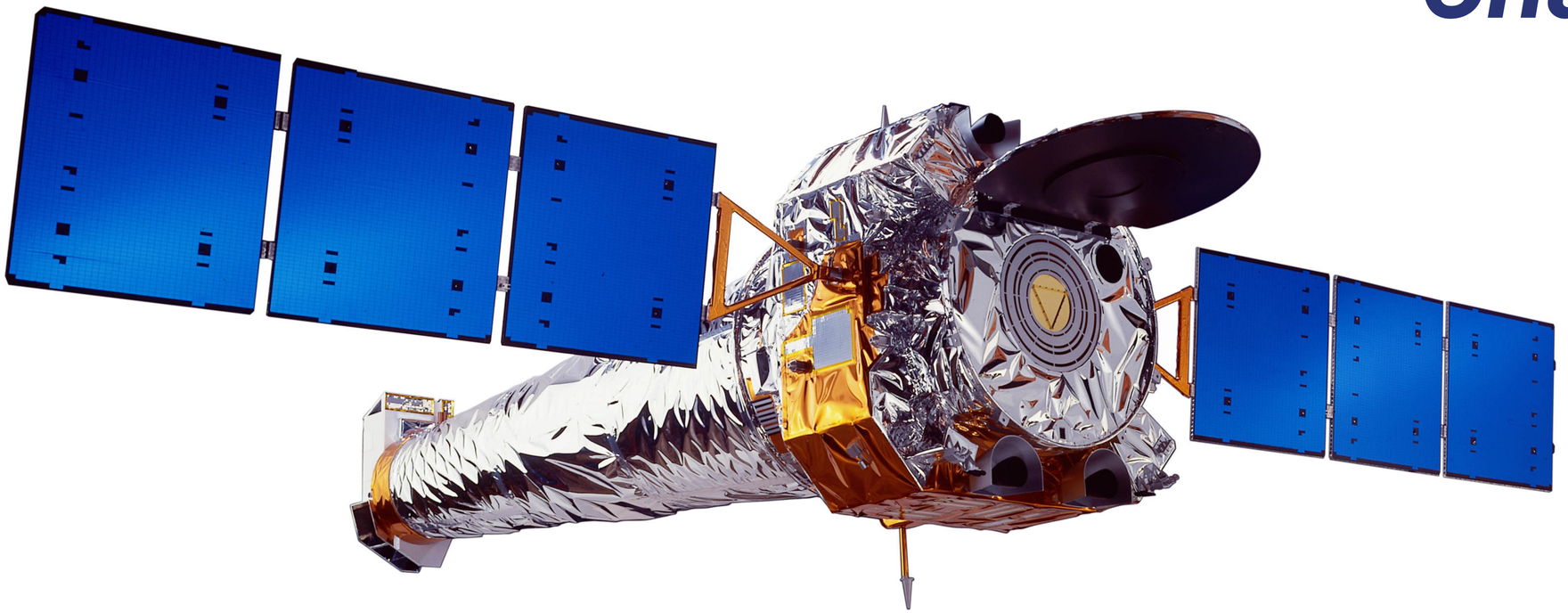
Golden age of X-ray astronomy



NASA Great Observatories



Chandra 101



23 July 1999

- High spatial resolution imaging+moderate spectral resolution+timing
- Each detected photon is labeled in a 4-dimensional space:
  - Spatial information (x and y)
  - Energy
  - Time of arrival
- Sub-arcsecond astrometric accuracy in most cases

A legacy of numbers

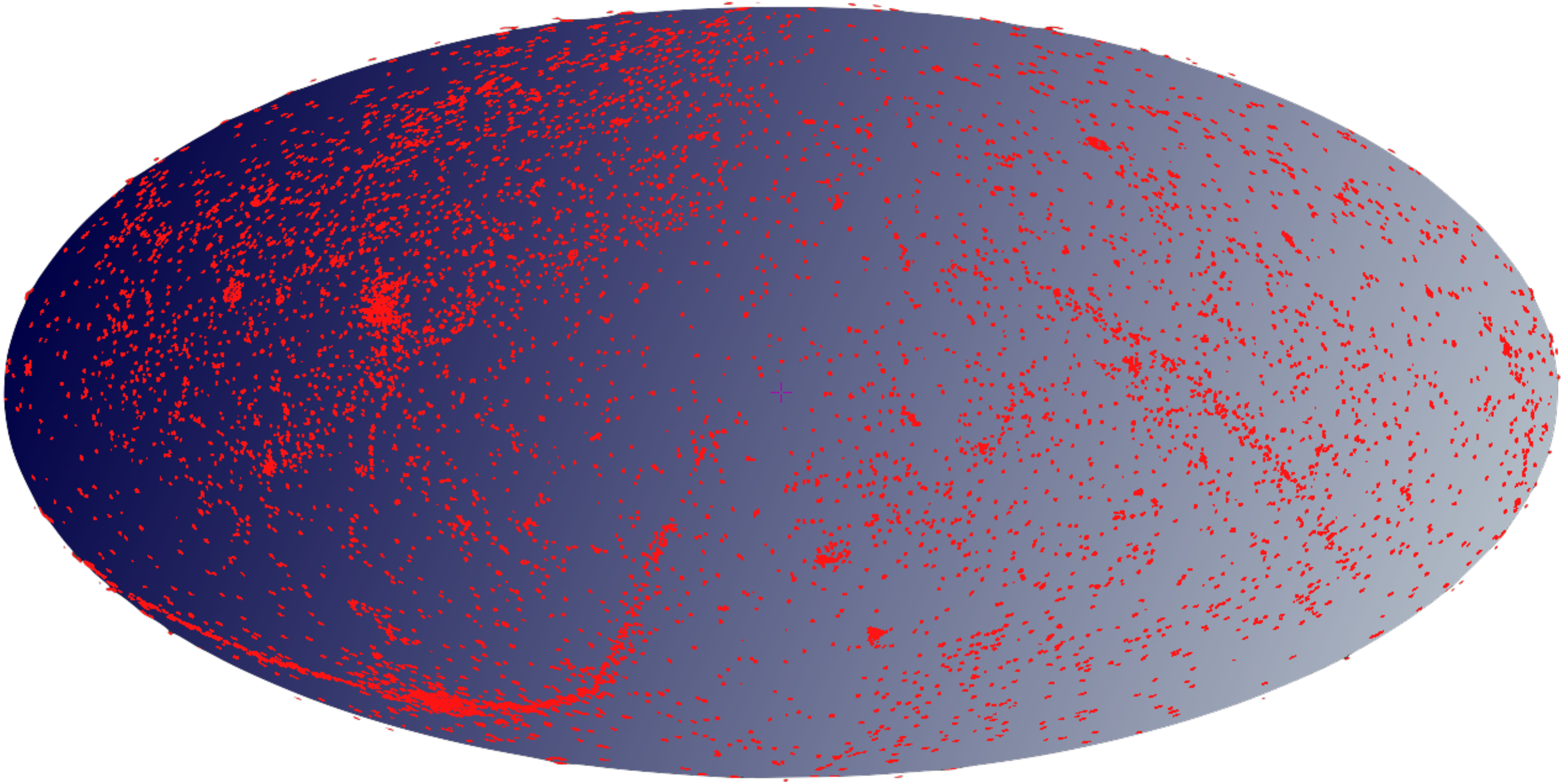
- Balanced mixture of programs of different types (~70% for GOs, ~10% TOOs+DDTs, ~10% GTOs, ~5% CALs) since Cycle 1
- New types of programs introduced to simplify scheduling and maximize the mission scientific legacy
  - *Chandra* Cool Targets (CCTs) (since Cycle 20)
  - *Chandra* Legacy programs (Cycles 26-27)

A legacy of numbers

- Balanced mixture of programs of different types (~70% for GOs, ~10% TOOs+DDTs, ~10% GTOs, ~5% CALs) since Cycle 1
- New types of programs introduced to simplify scheduling and maximize the mission scientific legacy
  - *Chandra* Cool Targets (CCTs) (since Cycle 20)
  - *Chandra* Legacy programs (Cycles 26-27)

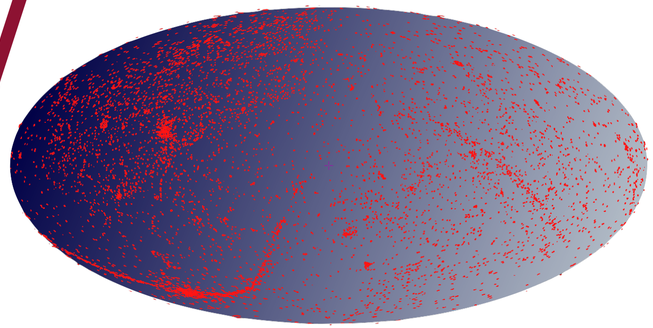
- *Chandra* has collected so far ~24.1k single scientific observations
- Observed for ~16.4 years over the ~24.5 years since launch (~67% efficiency)
- Collected ~8.8 billions photons (average of ~17 counts/s)
- Covered ~2.9% of the sky (it would be ~6.5% of the sky with no overlap among observations)

Highly correlated



A growing archive

Chandra observations

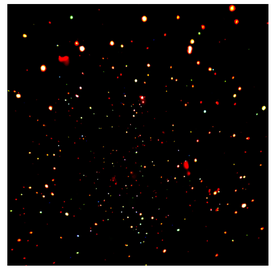


Scientific explosion

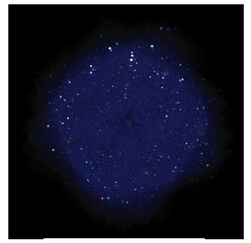
Sgr A*

Intentional & Unintentional aggregates

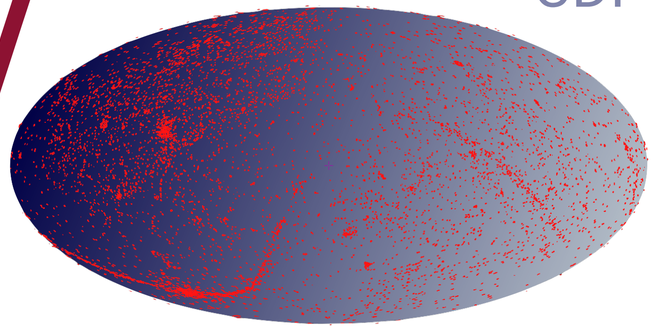
Chandra **observations**



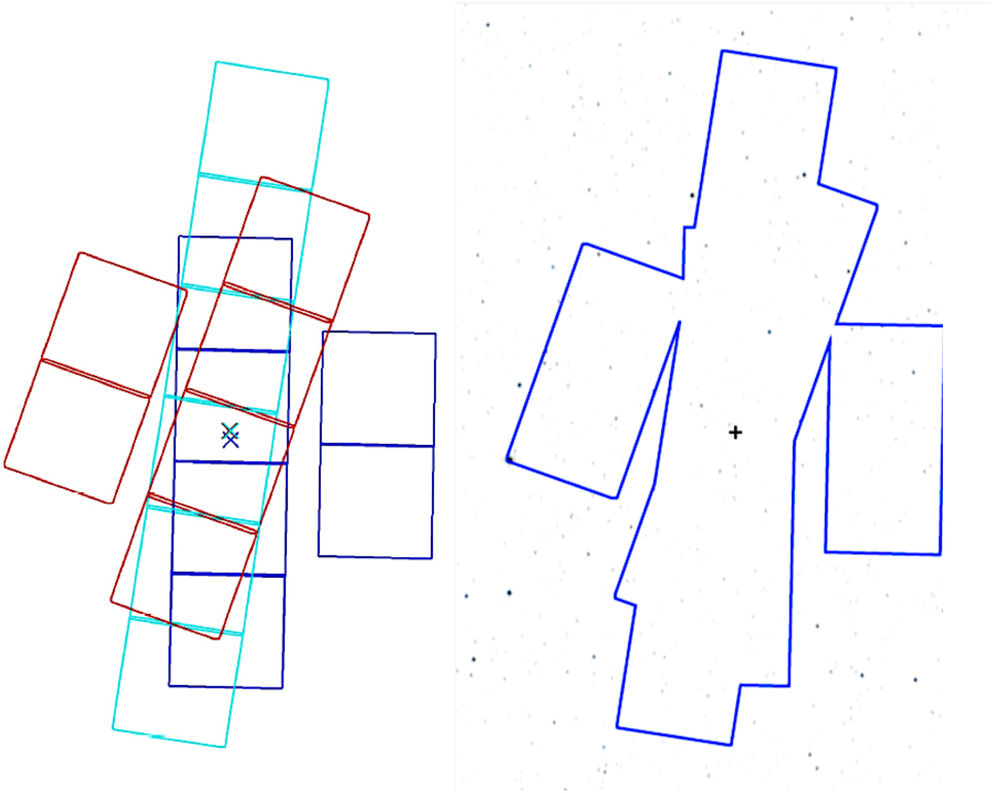
CDF-N

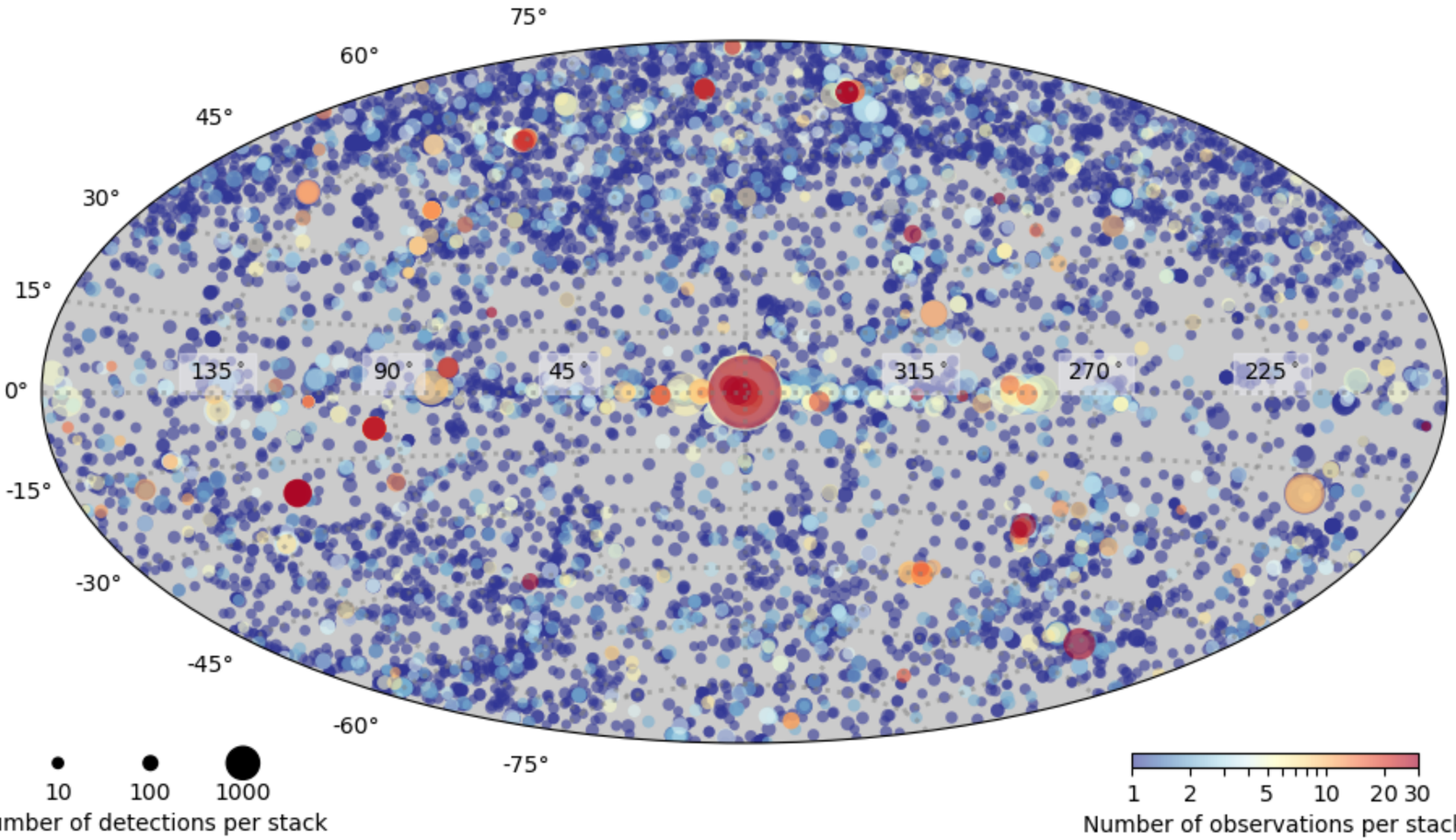


CDF-S



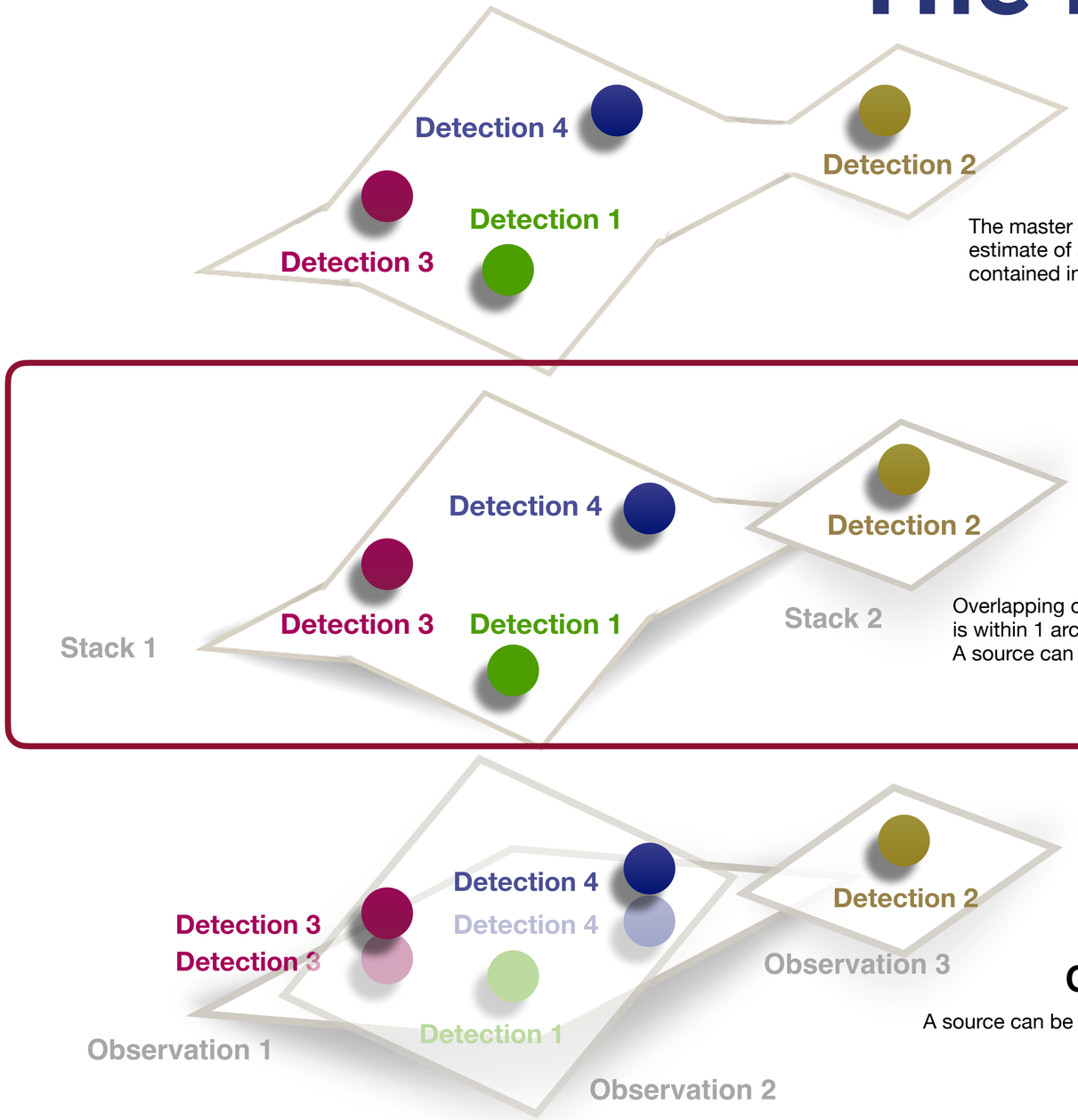
Chandra Source Catalog 2.1





Since v2.0, the CSC building blocks are **stacks**, obtained by merging observations by the same instrument and with aimpoints within 1’.

The full CSC hierarchy



Master Level

The master level properties are the best possible estimate of a source properties given the information contained in all contributing stacks.

Stack Level

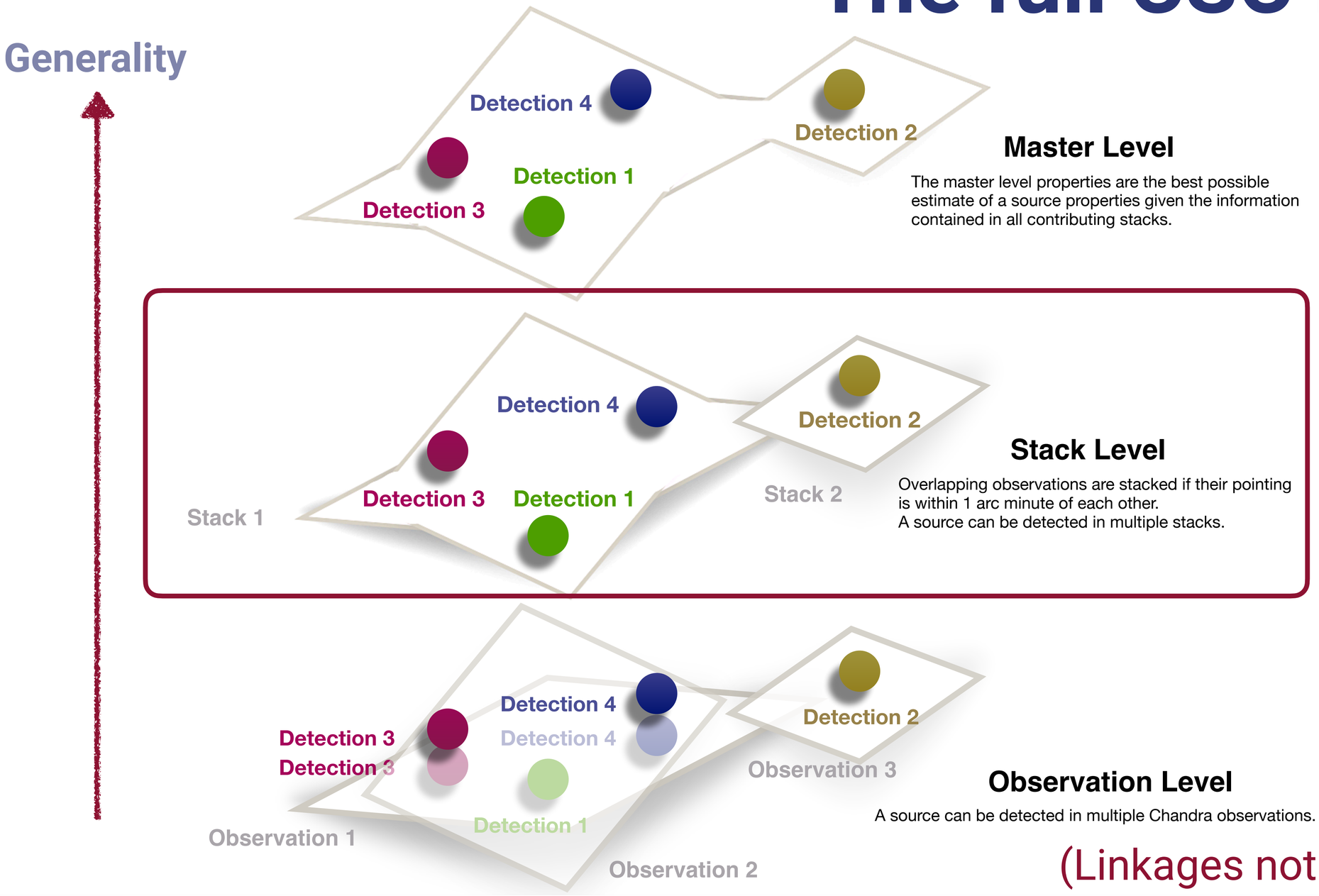
Overlapping observations are stacked if their pointing is within 1 arc minute of each other.
A source can be detected in multiple stacks.

Observation Level

A source can be detected in multiple Chandra observations.

(Linkages not shown for clarity)

The full CSC hierarchy

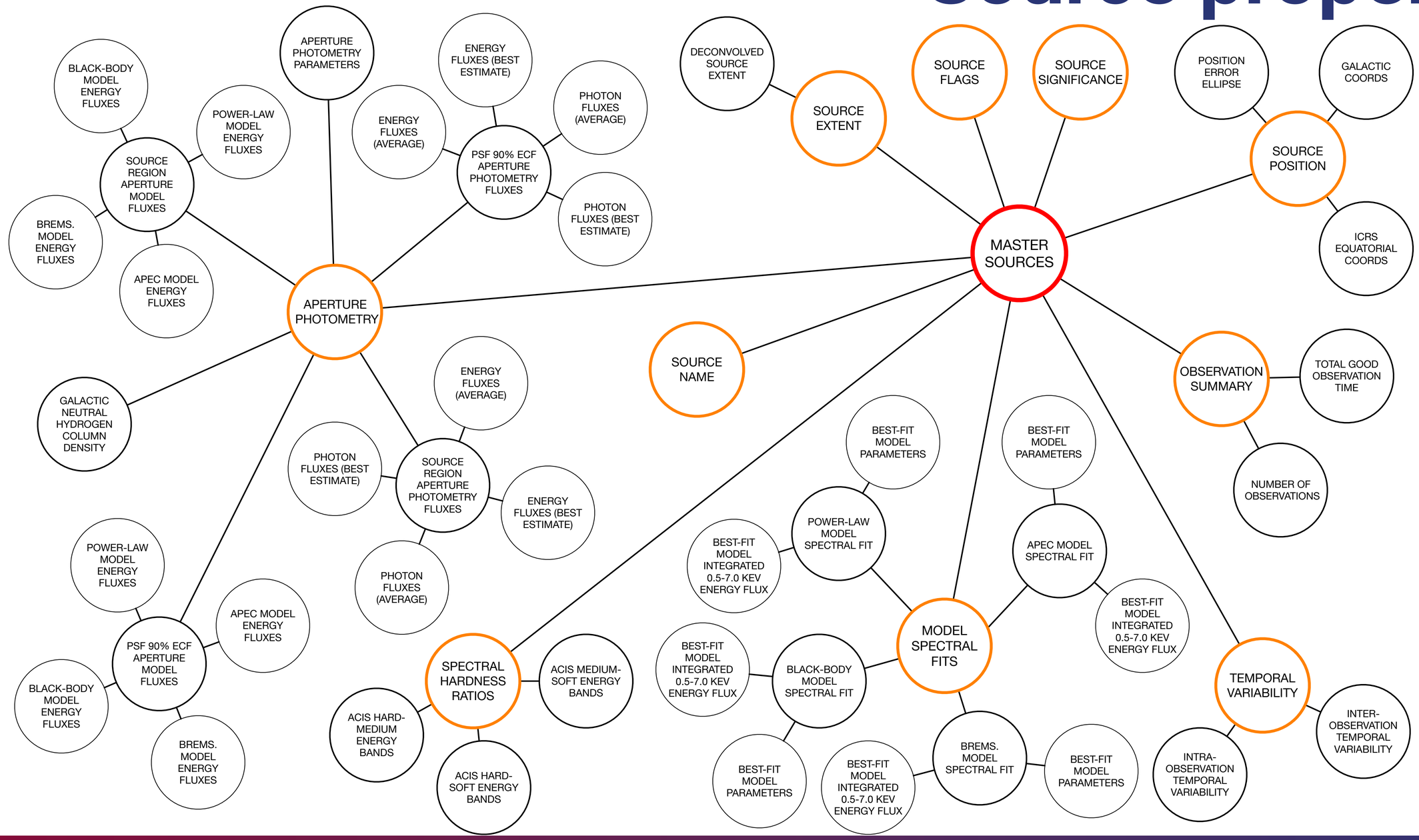


(Linkages not shown for clarity)

CSC 2.1

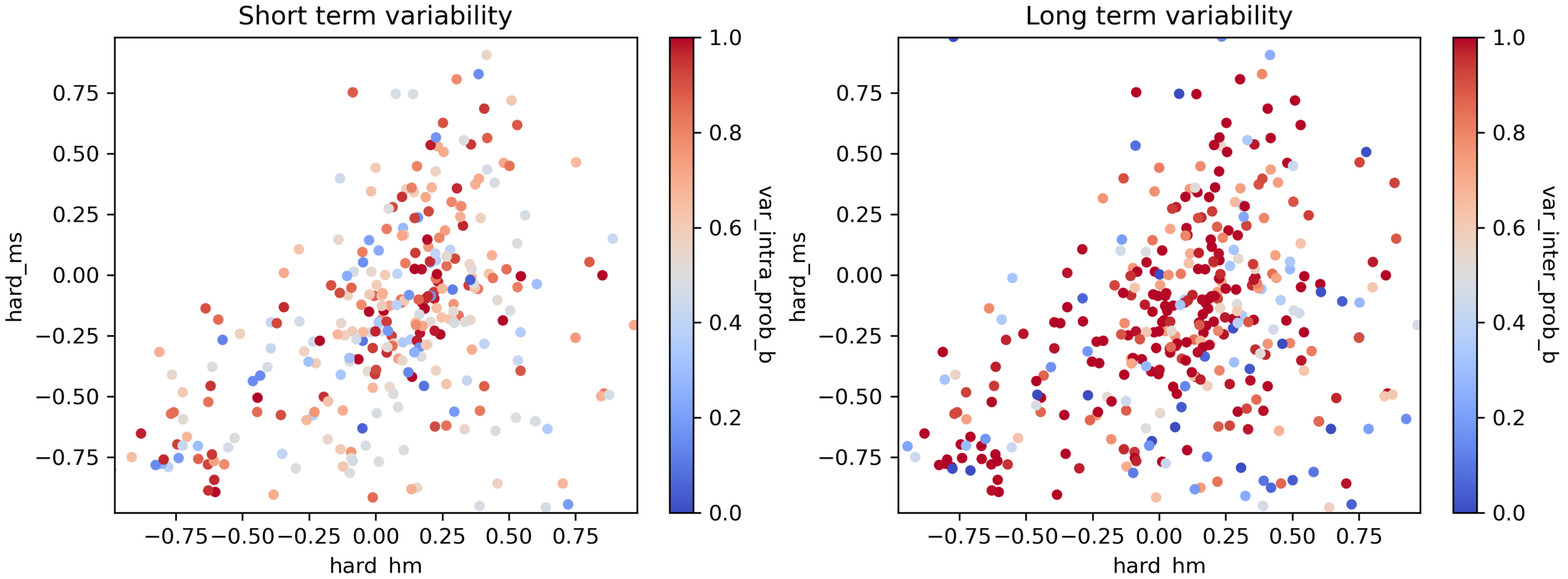
- Includes all observations that were publicly available on 12/31/2021 (early observations, HRC-S and gratings not included).
- ~414,000 individual sources on the sky, which is ~100k more than previous version CSC 2.0.
- >40% increase in sky coverage over CSC 2.0.
- Astrometry tied to the Gaia reference frame.
- Improved aperture photometry algorithm.
- Improved source position fitting algorithm.

Source properties



Why being granular?

Uniform computation of properties for all sources/detections enables data-driven exploration/discovery.



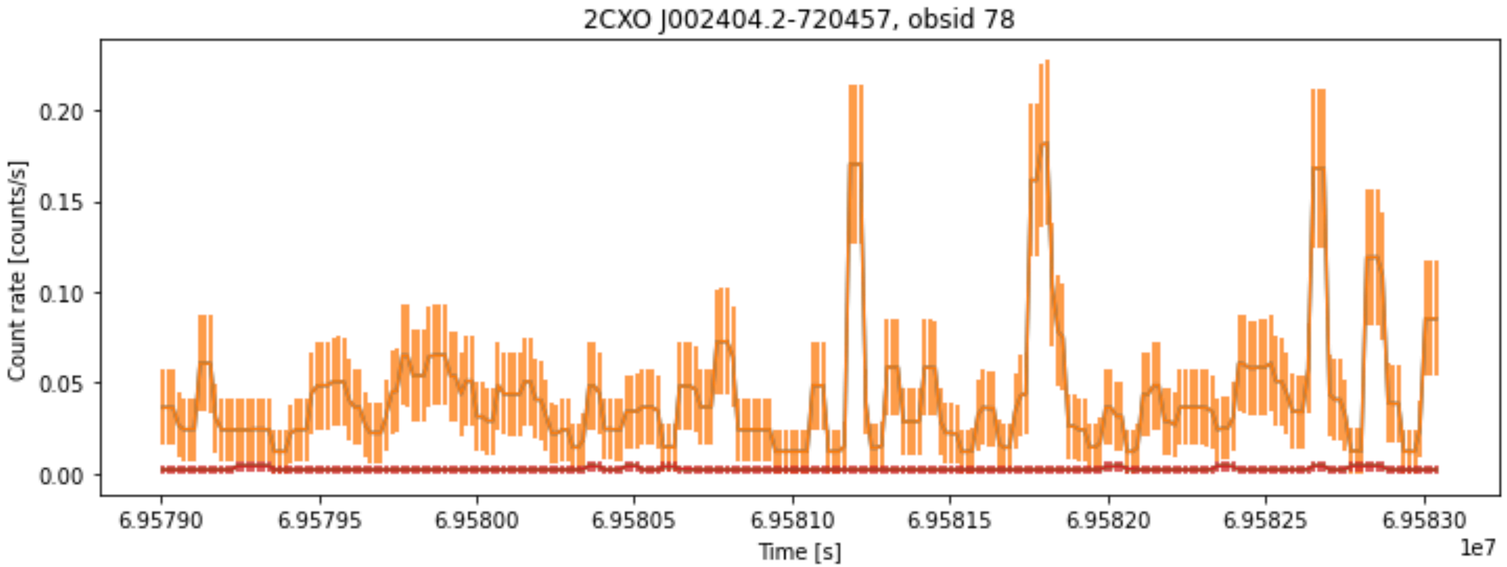
M51 has been observed 27 times between January 2000 to August 2021, with single observation exposures spanning the ~1ks to ~182ks range.

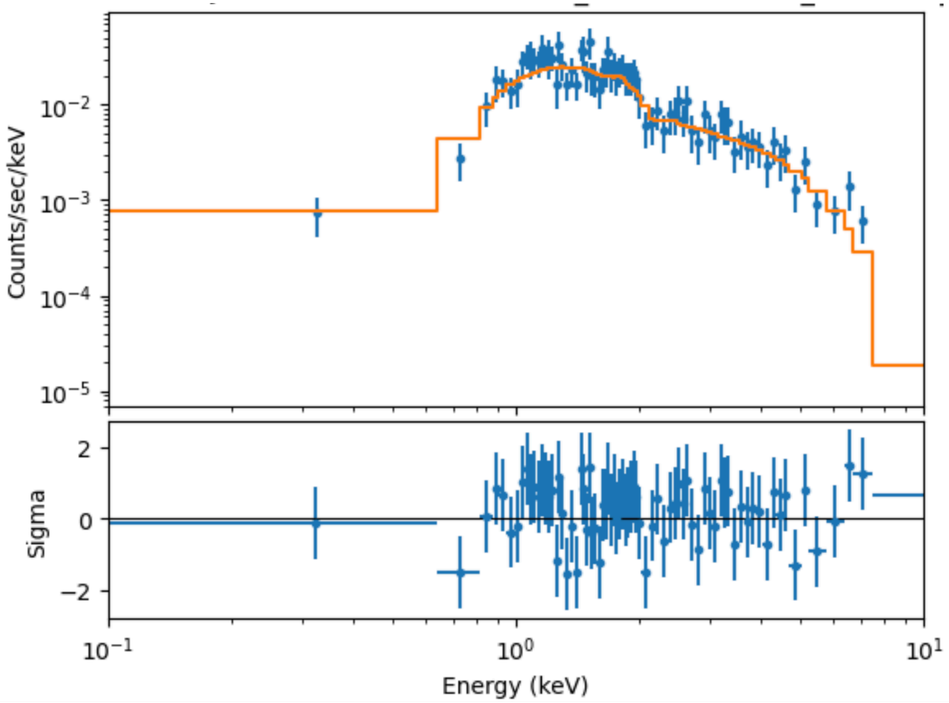
CSC2.1 data products

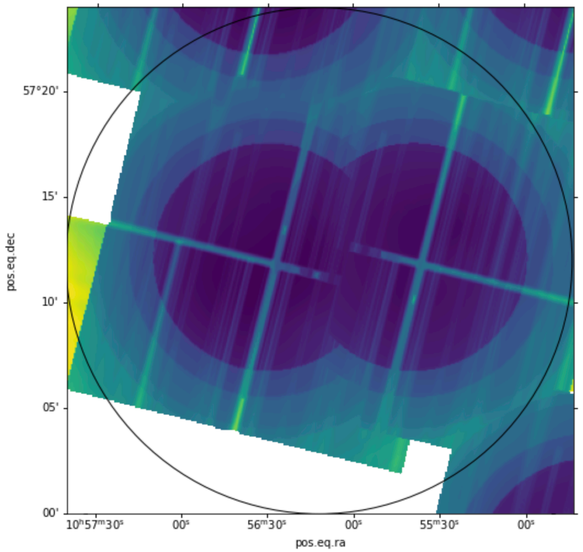
- **Region:**
  - **Master:**
    - Bayesian Blocks source properties
    - Per-Master source region aperture photometry PDF
  - **Stack:**
    - Stack Source Region Event List
    - Stack Source Region Image
    - Stack Source Region Exposure Map
    - Stack Source Region
    - Stack Source Region Draws
    - Valid Stack Source Region Aperture Photometry PDF
  - **Observation:**
    - Event List
    - Image
    - Point Spread Function
    - Exposure Map
    - Spectrum
    - ARF
    - RMF
    - Light Curve
    - Source Region
    - Valid Per-Obsid MLE source fit draws
    - Per-Obsid Source Region Aperture Photometry PDF

- **Full Field:**
  - **Stack:**
    - Stack Event List
    - Stack Image
    - Stack Background Image
    - Stack Exposure Map
    - Stack Field Of View
    - Stack Limiting Sensitivity
    - Stack Merged Source Detection List
  - **Observation:**
    - Event List
    - Image
    - Background Image
    - Exposure Map
    - Aspect Solution
    - Aspect Histogram
    - Bad Pixel Regions
    - Field Of View
    - Pixel Mask
    - Extended Source Region

Source/detections







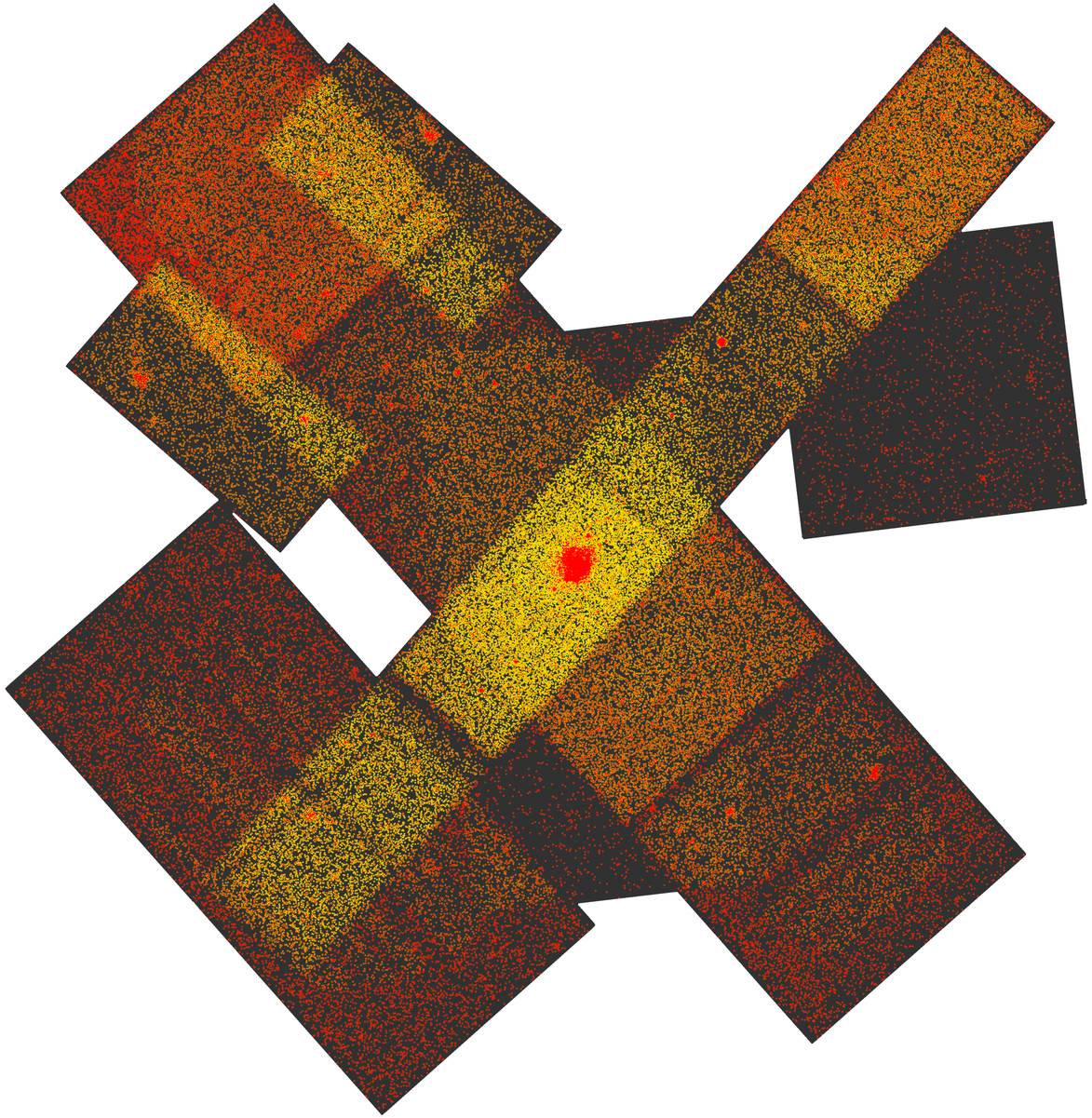
- Bayesian blocks for master sources
- All auxiliary files needed for standard processing of *Chandra* data at each level

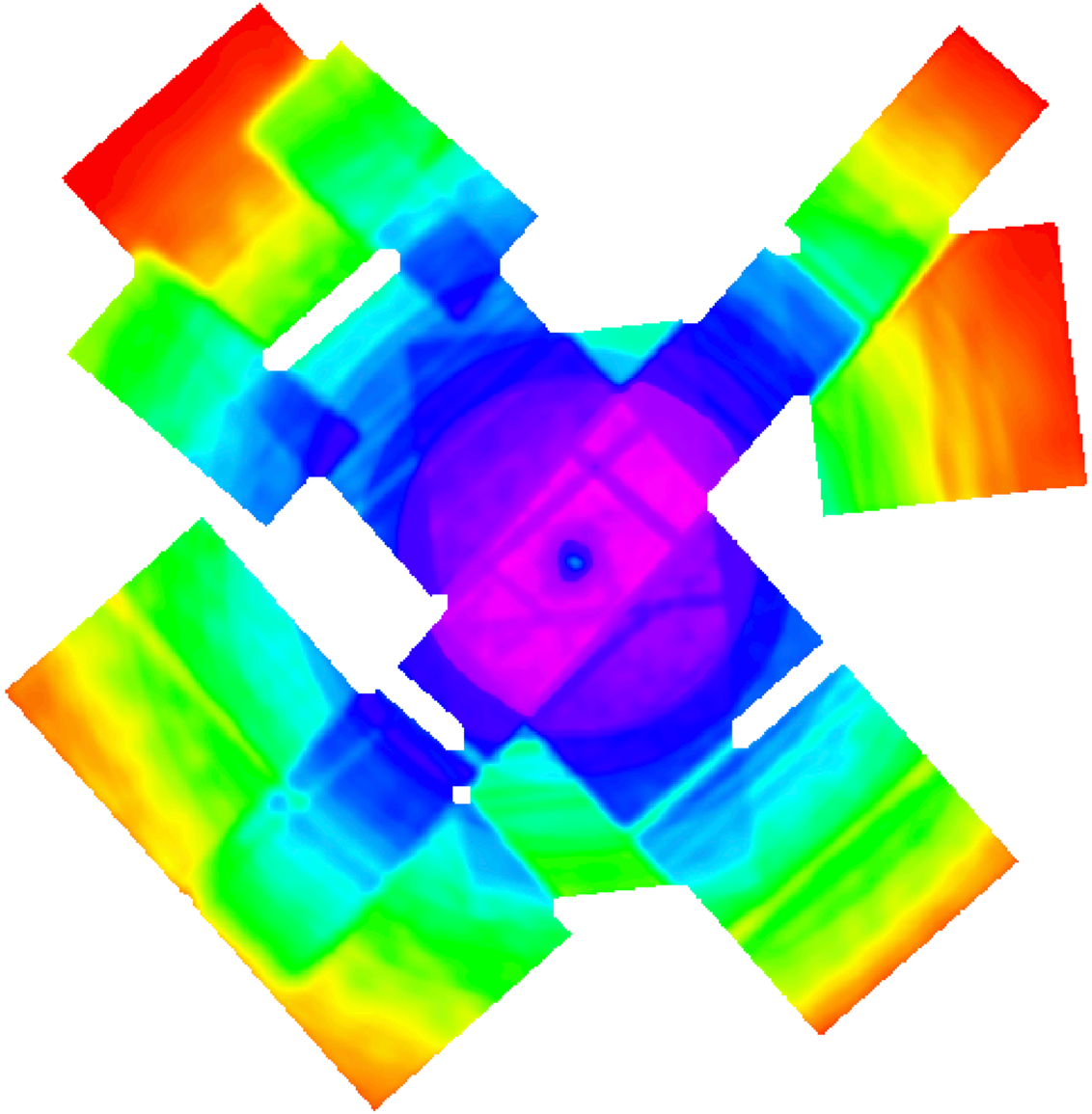
CSC2.1 data products

- **Region:**
  - **Master:**
    - Bayesian Blocks source properties
    - Per-Master source region aperture photometry PDF
  - **Stack:**
    - Stack Source Region Event List
    - Stack Source Region Image
    - Stack Source Region Exposure Map
    - Stack Source Region
    - Stack Source Region Draws
    - Valid Stack Source Region Aperture Photometry PDF
  - **Observation:**
    - Event List
    - Image
    - Point Spread Function
    - Exposure Map
    - Spectrum
    - ARF
    - RMF
    - Light Curve
    - Source Region
    - Valid Per-Obsid MLE source fit draws
    - Per-Obsid Source Region Aperture Photometry PDF

- **Full Field:**
  - **Stack:**
    - Stack Event List
    - Stack Image
    - Stack Background Image
    - Stack Exposure Map
    - Stack Field Of View
    - Stack Limiting Sensitivity
    - Stack Merged Source Detection List
  - **Observation:**
    - Event List
    - Image
    - Background Image
    - Exposure Map
    - Aspect Solution
    - Aspect Histogram
    - Bad Pixel Regions
    - Field Of View
    - Pixel Mask
    - Extended Source Region

Full fields



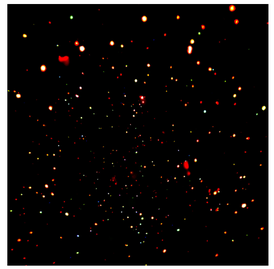


Scientific explosion

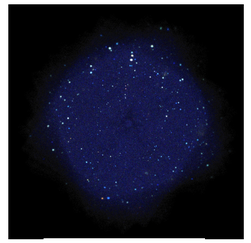
Sgr A*

Intentional & Unintentional aggregates

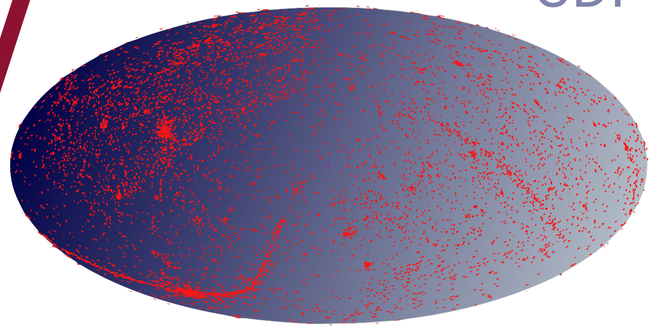
Chandra **observations**



CDF-N



CDF-S



A growing archive

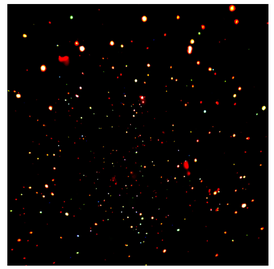
Chandra Source Catalog

(Complete tessellation of the *Chandra* sky)

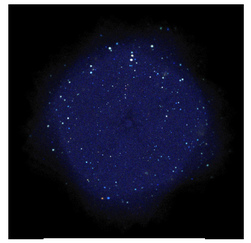
Intentional & Unintentional aggregates

Chandra observations

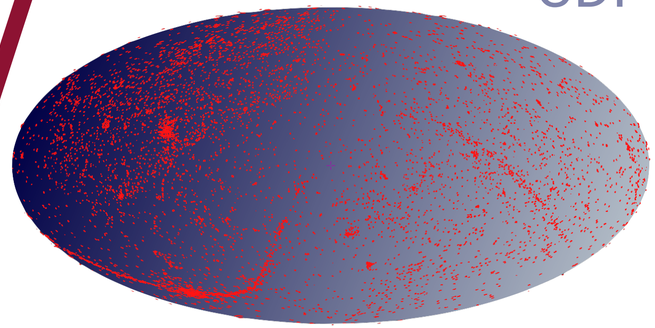
Sgr A*



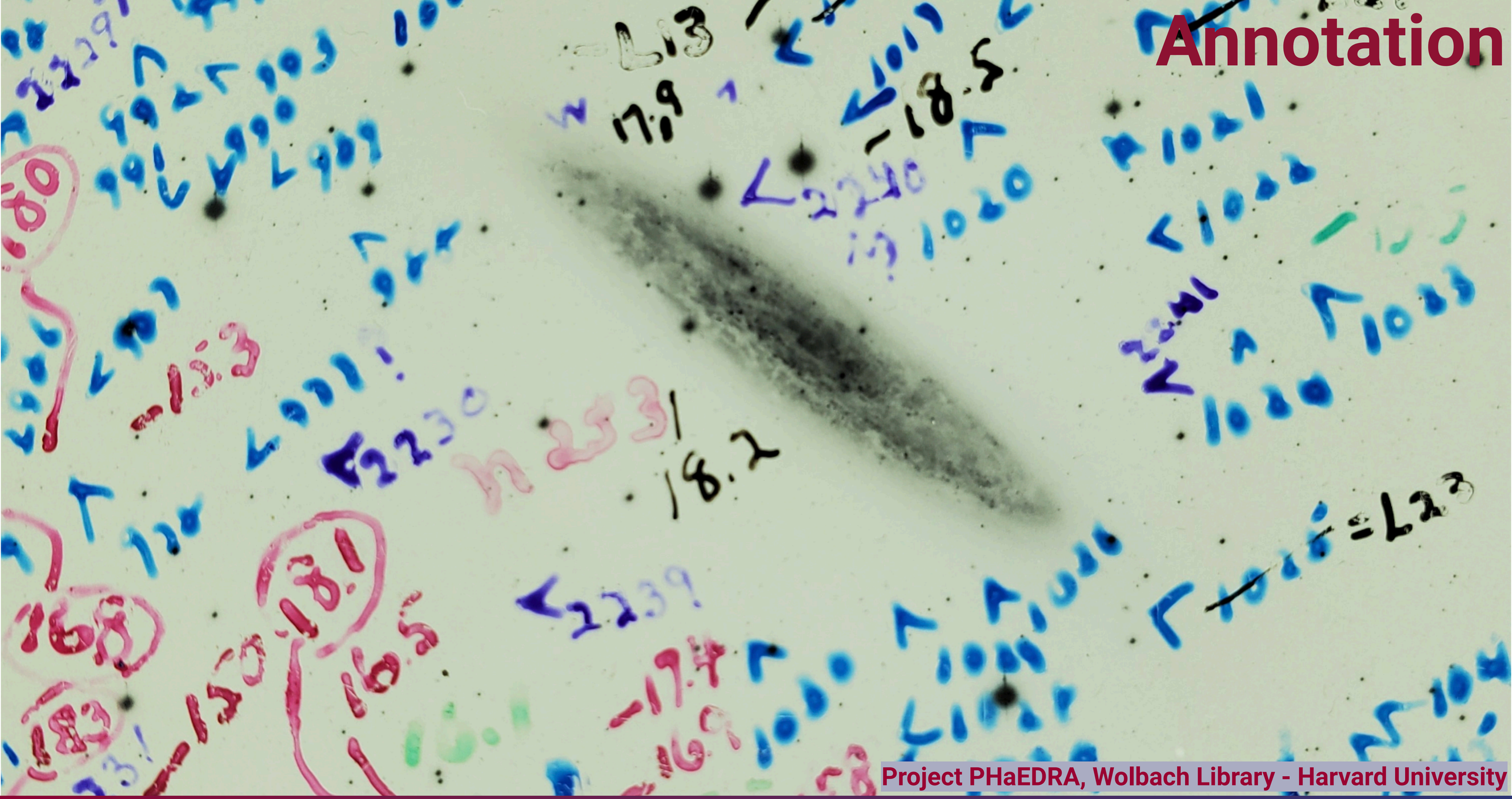
CDF-N



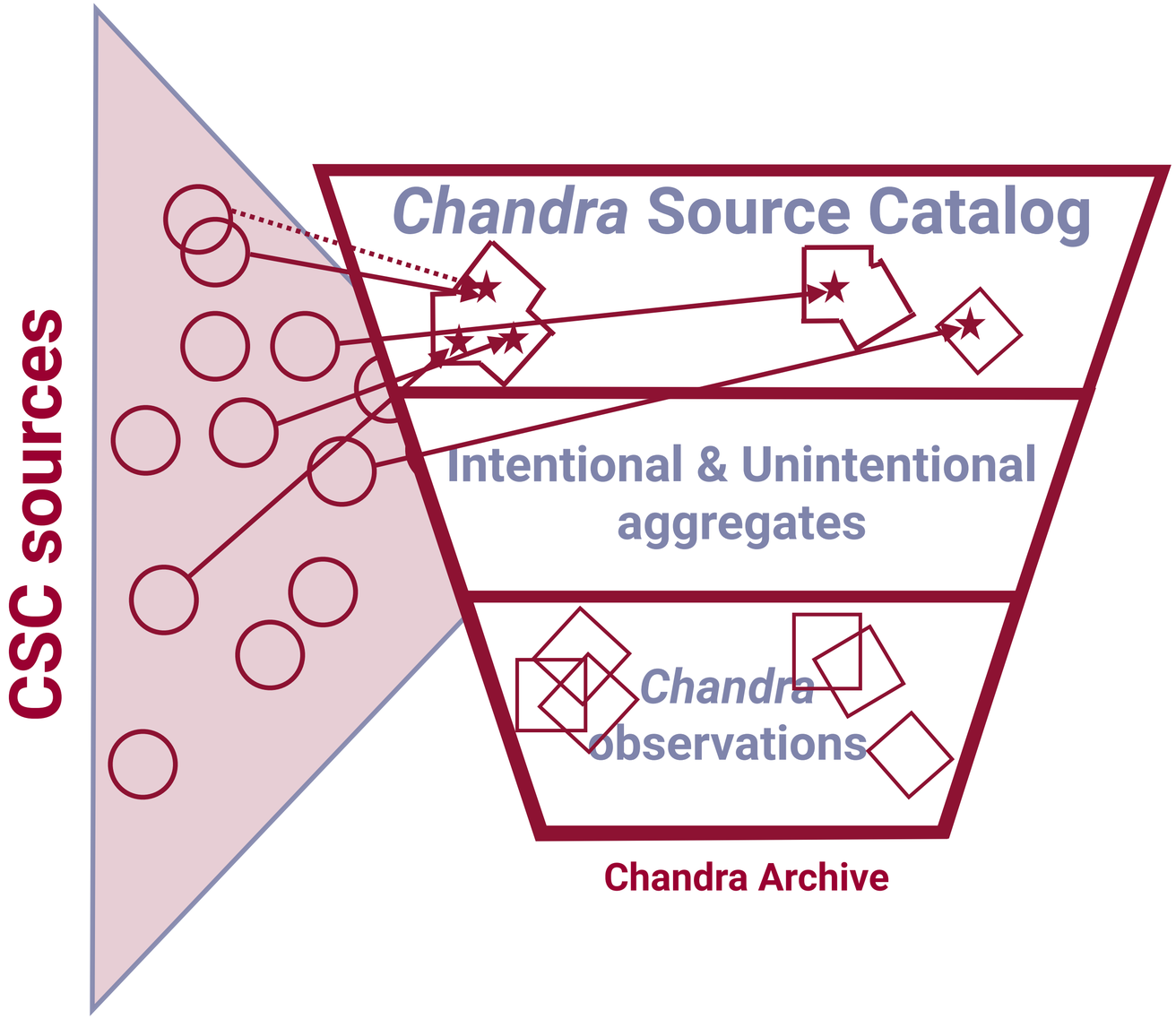
CDF-S



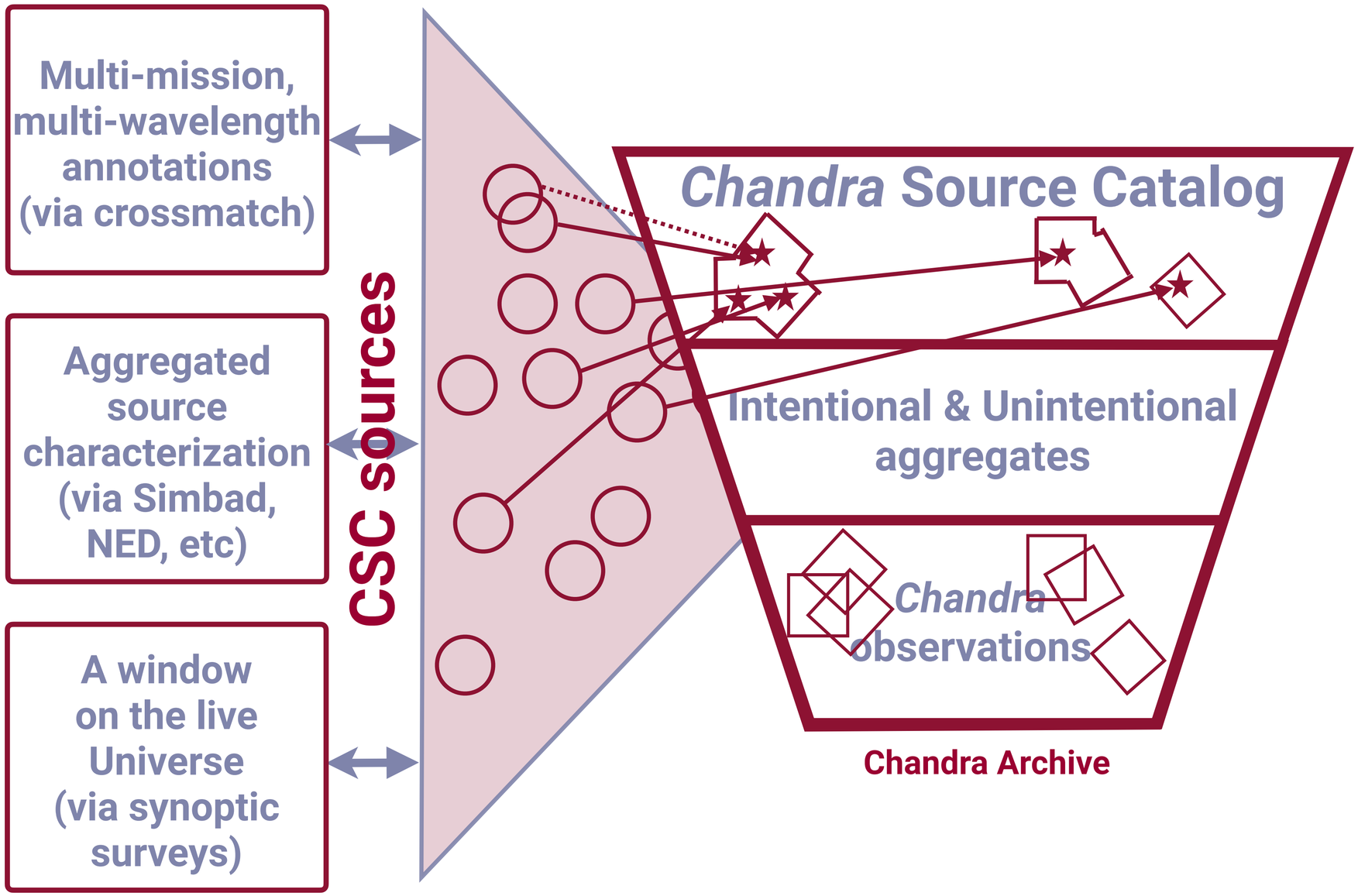
The **Chandra Source Catalog** is the **culmination of a data aggregation process** that has been possible thanks to *Chandra* longevity.


Annotation
Project PHaEDRA, Wolbach Library - Harvard University

Modern annotations



Modern annotations

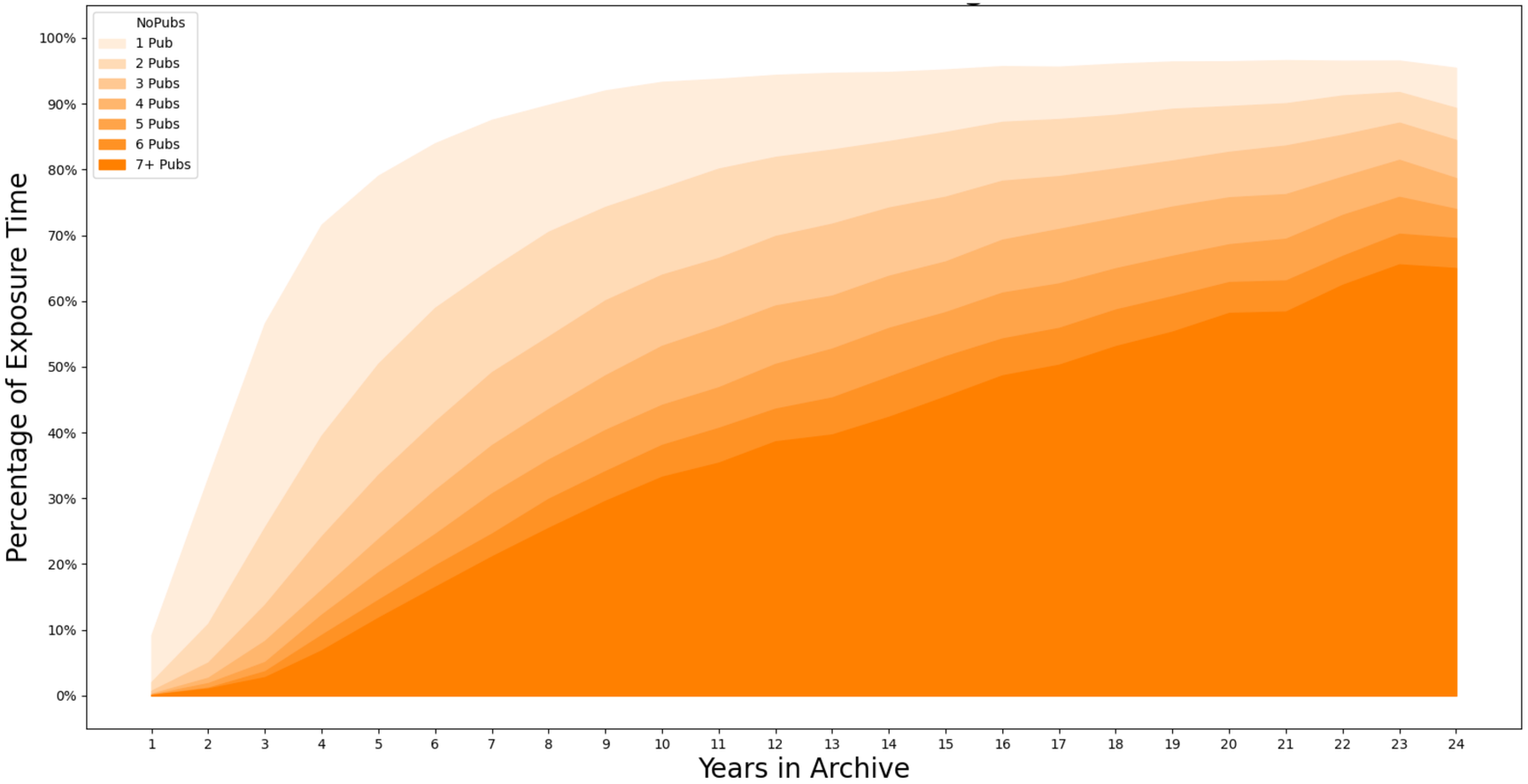


Chandra bibliography

25898 *Chandra*-related publications, with 141267 literature-observation links for 9175 papers referring to 17241 observations. 9848 "*Chandra* Science Papers" (CSPs), i.e. papers that wouldn't have written without *Chandra*.

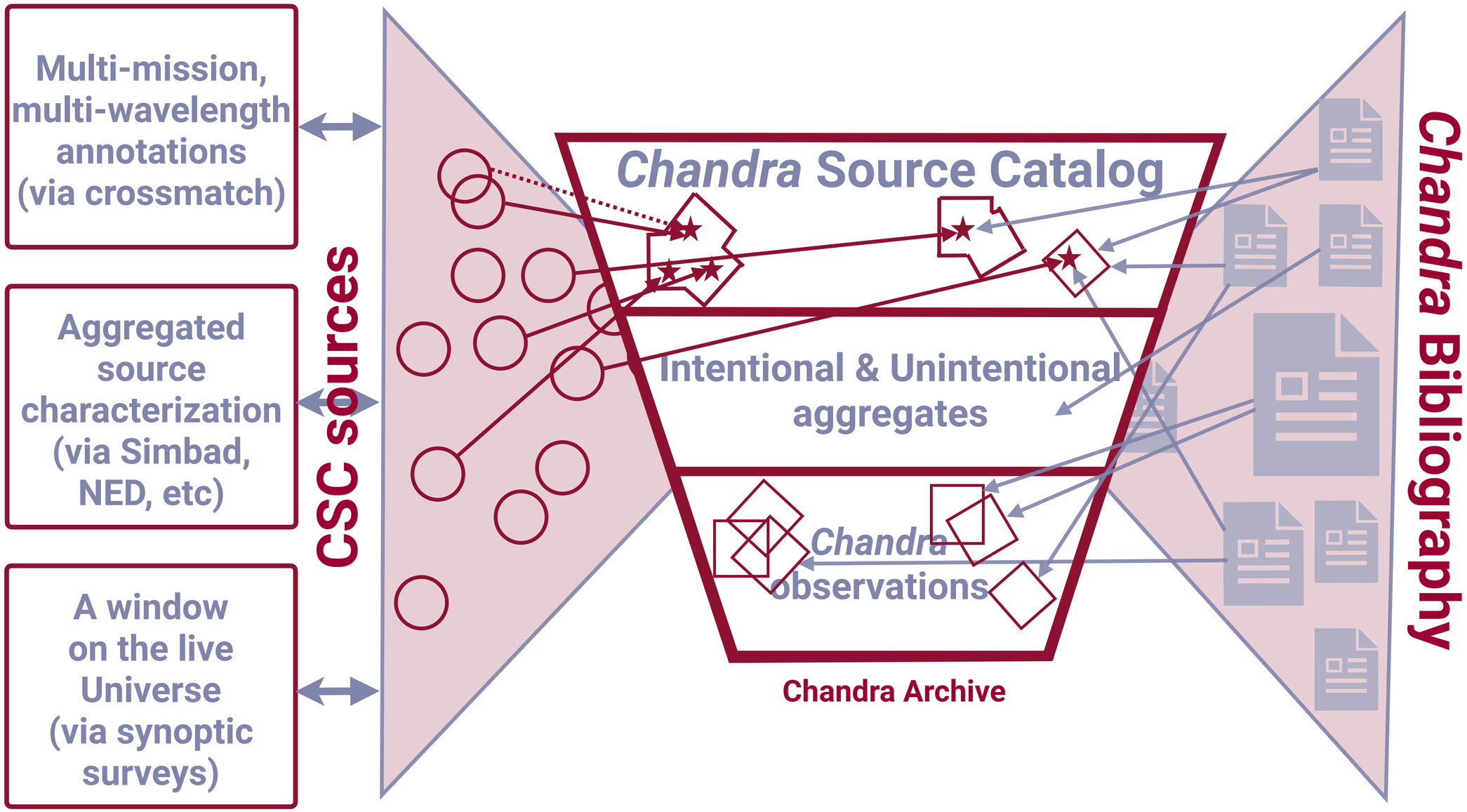
Chandra bibliography

25898 *Chandra*-related publications, with 141267 literature-observation links for 9175 papers referring to 17241 observations. 9848 "*Chandra* Science Papers" (CSPs), i.e. papers that wouldn't have written without *Chandra*.

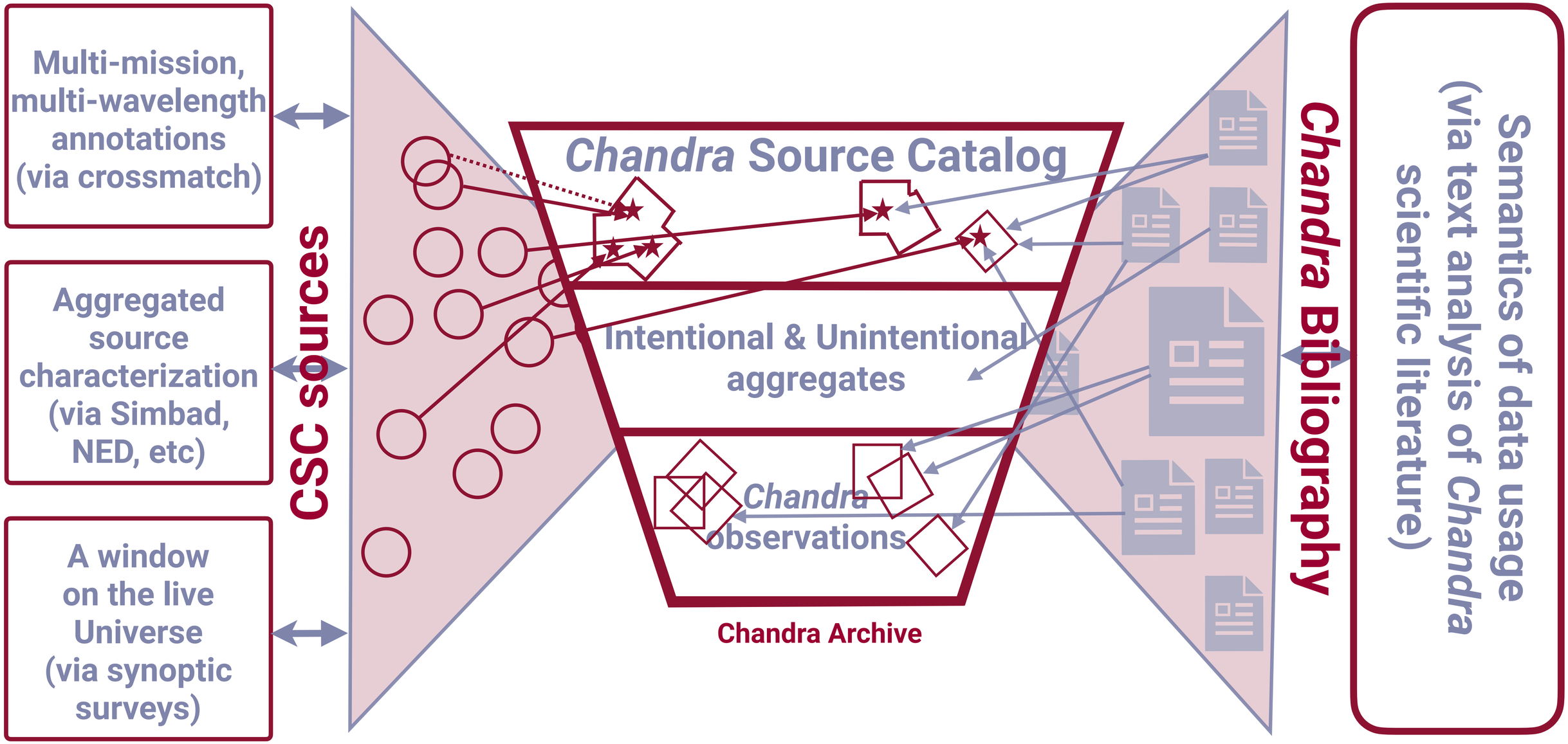


- Chandra data are **heavily re-used**
- "Archival science"
  - Mixed archival papers (new and old data) accounts for ~80% of CSPs.
  - Purely archival papers (all old data) accounts for ~45% of CSPs.

Modern annotations



Modern annotations



Introspection

Archive introspection

The *Chandra* archive is at the core of a complex, hierarchical, multi-variate dataset

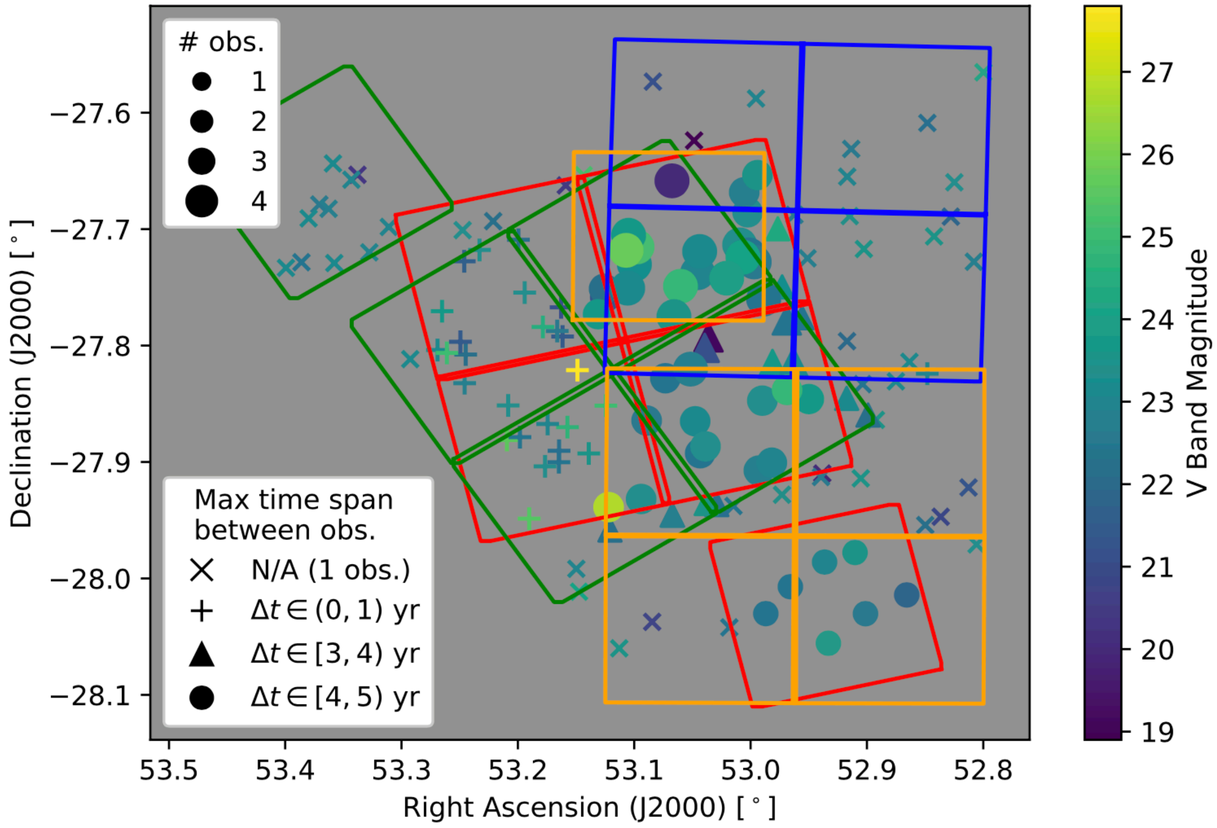
- Annotations (via catalog) enhance mission-specific domain with general concepts
  - Can we answer questions about *Chandra* data including non-*Chandra* constraints?
- Bibliographic footprint of *Chandra*-based literature
  - How and why are *Chandra* data used for science?
  - Gaining insight into re-usage that differs from original science
  - Patterns can be used to predict how missions science evolves with time

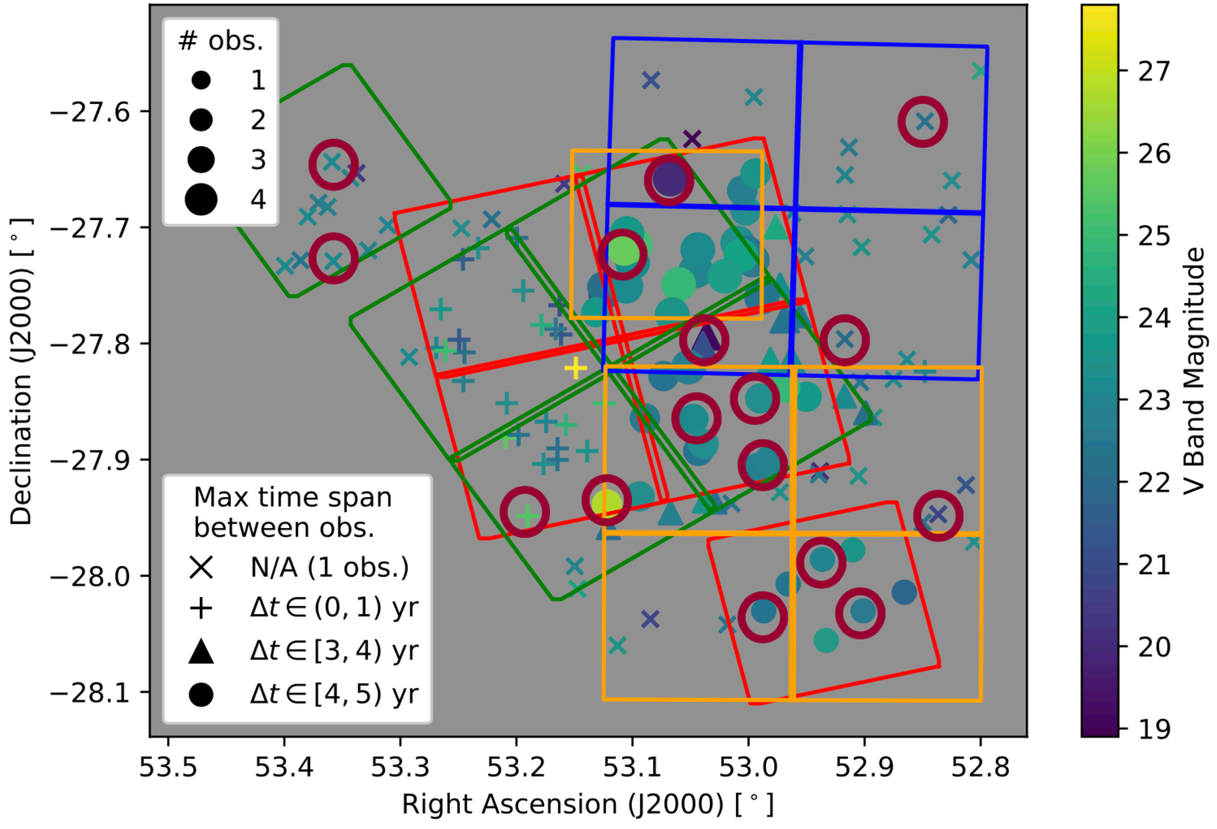
Investigate the full *relational network* of the mission archive

- Relationships between archival entities and their annotations are seldom mined.
- Movie of the mission legacy growth over time, not a picture.

Hard questions

"Find all optical (spectroscopically confirmed) quasars (with emission line EW measurements available), with a CSC counterpart with significance > 15 and covered at least 2 *Chandra* observations with time gap in the [Δt_{min}, Δt_{max}] interval, and having shown blazar-like variability (according to some broker) as observed by the Zwicky Transient Facility."





Harder questions

Can we optimize the scientific content of a catalog by tweaking the archive?

- Which are the under-sampled areas of the archival parameter space?
- Can the mission observational strategy be fine-tuned (GOs vs GTOs vs TOOs vs DDTs ratios, new types, etc) to improve balance of different classes of sources?
- Which observations should be prioritized to maximize the chance of discovery of rare, interesting sources?

Can we predict the long-term scientific legacy of new observations based on the usage pattern of similar data?

- Insight-based recommendation systems for archival observations ("*We noticed you downloaded this observation, may you are also be interested in* ***similar data****...*")
- Automatic prediction of legacy value for proposed observations

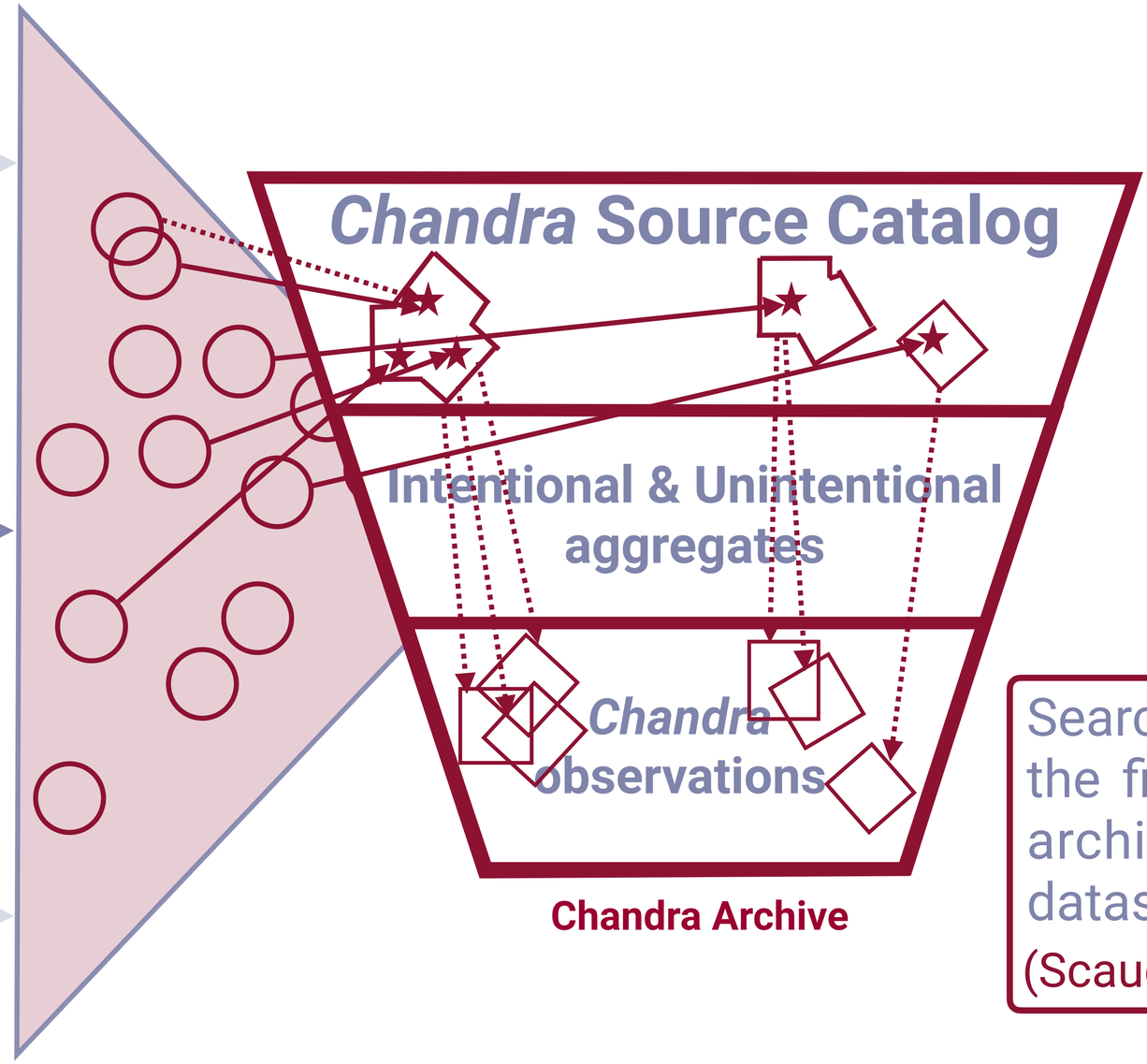
The first steps

Multi-mission, multi-wavelength annotations (via crossmatch)

Aggregated source characterization (via Simbad, NED, etc)

A window on the live Universe (via synoptic surveys)

CSC sources



Search for a strategy to optimize the final scientific legacy of the archive and CSC using both datasets.

(Scaudo, D'Abrusco et al.)

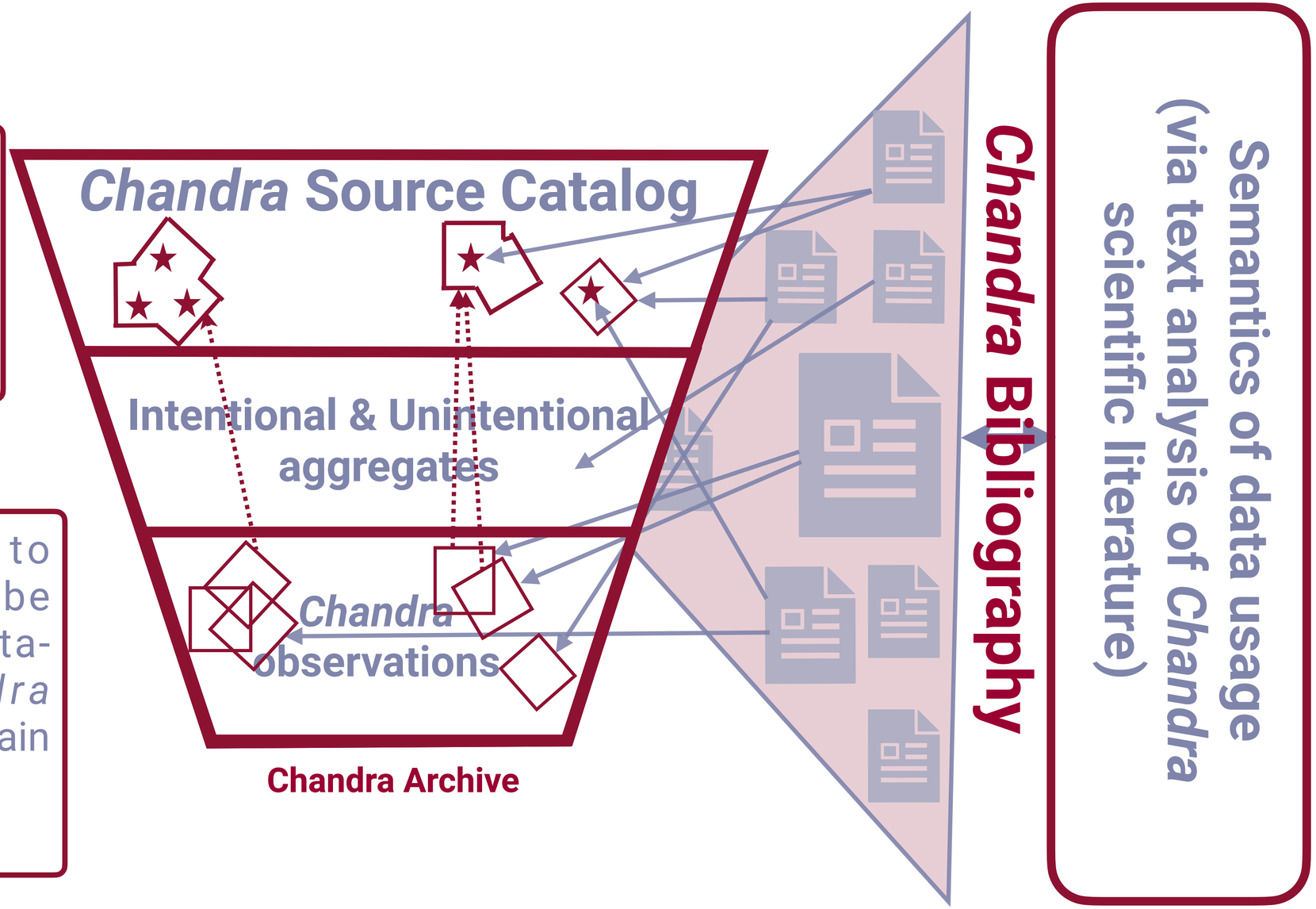
The first steps

Can AI tools help classifying astronomical papers? The *Chandra* bibliography is used as template dataset.

(ADS, CDA)

Can AI systems learn how to recognize what science can be produced based on data? Data-literature links from *Chandra* bibliography are being used to train the models.

(astroAI/CfA, ADS, R. Martinez-Galarza/CXC)



The current status

Pro's

- "Archive introspection" is already feasible for multiple long-lived missions with slightly different (but comparable) observational strategies
- Motivations and methods are here!
    - Data-driven decision-making will be ubiquitous in the future, even in astronomy.
    - ML/AI techniques, powerful hardware and know-how are available

Con's

- Ingrained practices are difficult to change for existing missions
    - They worked well (enough) to build rich archives, why changing them?
    - Difficult to factor in the sociological impact of operational strategies
- Data-driven decision-making can be perceived as controversial
    - Are we de-emphasizing key scientific questions to maximize the long-term "scientific legacy" of a mission?
    - Are we taking away the human factor?

Take-away points

Catalogs should be a day 1 deliverable

- Large multiplier for scientific output and impact of mission!
  - Open up mission-specific scientific domain to non-experts
  - Gateway to an incredibly vast annotation domain
- Take full advantage of the highly correlated archival structure of pointed observatories
  - Always prefer "hierarchical" to "flat" catalogs!

Trying out archival introspection...

- The archives of live, long-lived missions are great testbeds for methods, goals and expectations

..and use it to optimize future missions

- **Learning how data has been used is learning how the mission should evolve**
- ML/AI techniques and tools will be embedded in all aspects of mission operations
- Data-driven *nudging* of observational strategies can increase mission impact in the long-term


Thank you!