Conception and Development of a Tool to Abstract Access to and Merge Multiple Astronomical Data Stores

Sinh Thong, G. Becker, R. D'Abrusco, M. McCollough, D. Van Stone, S. Winkelman
Center for Astrophysics | Harvard & Smithsonian

Abstract

The amount of astronomical data available has increased rapidly in recent years and is expected to accelerate dramatically with the advent of next generation instruments planned for the near future. Even now, it's possible to conduct high quality studies using archival data alone. Along with the growth in data, there's naturally a corresponding increased need for easier ways to find and view existing observations. Additionally, the established importance of multiwavelength approaches presents another challenge that access tools must address. In response to these and other related needs, we present an initial conception and zeroth order implementation of one example of such a tool. At the Chandra Data Archive, we've started developing the prototype of a web service to access a constructed database of Chandra observations annotated with SIMBAD objects located within the field of view of the observations. The idea is to leverage the well vetted, encyclopedic compilation of SIMBAD objects and corresponding properties to expand what we can currently offer in terms of Chandra data and how they can be searched. Recognizing that the interface may be comparable in importance to the data store, we also explore ideas for search input features that help bridge the gap between the natural thought processes of a user and the input requirements of the tool as well as feedback and output that facilitate exploration, including the serendipitous type, of the search space.

Chandra Observations Annotated with SIMBAD Objects

This experimental tool is currently being developed in Python 3 with a Sybase ASE back end. SIMBAD data are pulled from SIMBAD servers and combined with Chandra observation information from Chandra X-ray Center databases and inserted into our back end. SIMBAD objects covered by Chandra observations are linked together to support rapid region searches on Chandra footprints. We may eventually develop a web interface allowing users to search on SIMBAD object properties and other metadata as well as Chandra observation metadata.

Current Back End Work and Search Concepts Under Consideration

- **Data fusion**: Create database containing Chandra observations correlated with SIMBAD objects via footprints and object location, store all metadata of interest
- Refresh SIMBAD metadata every ~6 months
- Query parser

- Less Chandra-specific knowledge needed
- Search options
  - SIMBAD object properties and other metadata
  - Chandra observation metadata
- Link to ChaSeR for detailed observation information and data products
- Dynamic searching
  - See results while inputting parameters to visualize the search space
- Graphical input controls
  - Ultimately, more natural search input expression
- Visualizations for input-search space, interactive via mouse and keyboard

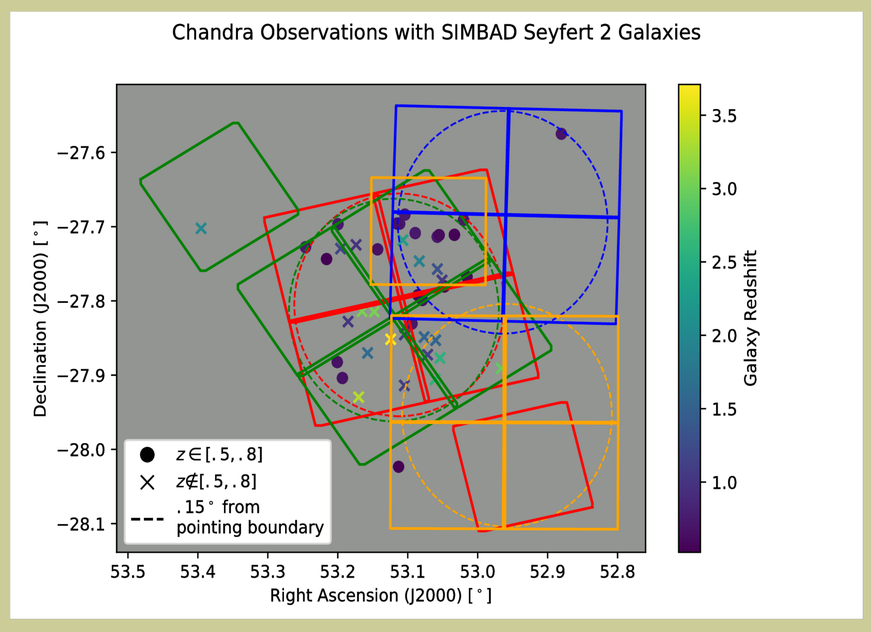
Data Representation and Algorithm Considerations

Chandra and SIMBAD data combined comprise an enormous dataset. One issue is that searches over arbitrary regions of the sky along with the vast number of SIMBAD objects can be computation intensive and require careful consideration of data representation and corresponding search algorithms. This is particularly a problem when attempting to support **dynamic searching**. We're currently using a relational database system that will at some point require expensive SQL joins but are investigating continuing with a **graph based system**. Many parameter spaces can be **discretized by space partitioning** and searched hierarchically in potentially O(log n) instead of O(n) time.

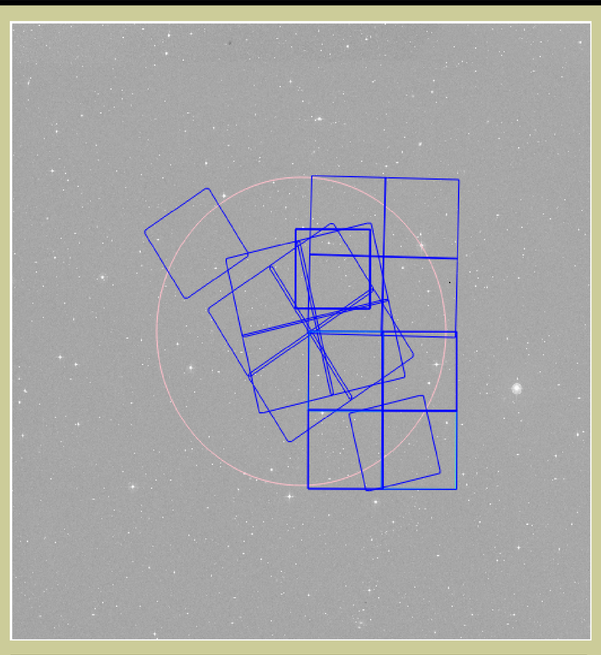
Search Examples

Searches for quasars and Seyfert 2 galaxies using Chandra observation and SIMBAD object properties. Chandra footprints with ObsIDs 441, 1672, 5017, and 5019.
We currently do not have an end-user interface. These visuals are not representative of the eventual search output format.

What ACIS observations contain Seyfert 2 galaxies with redshifts between 0.5 and 0.8 that are located less than 0.15 deg from the observation pointings with exposure times greater than 25 ks?

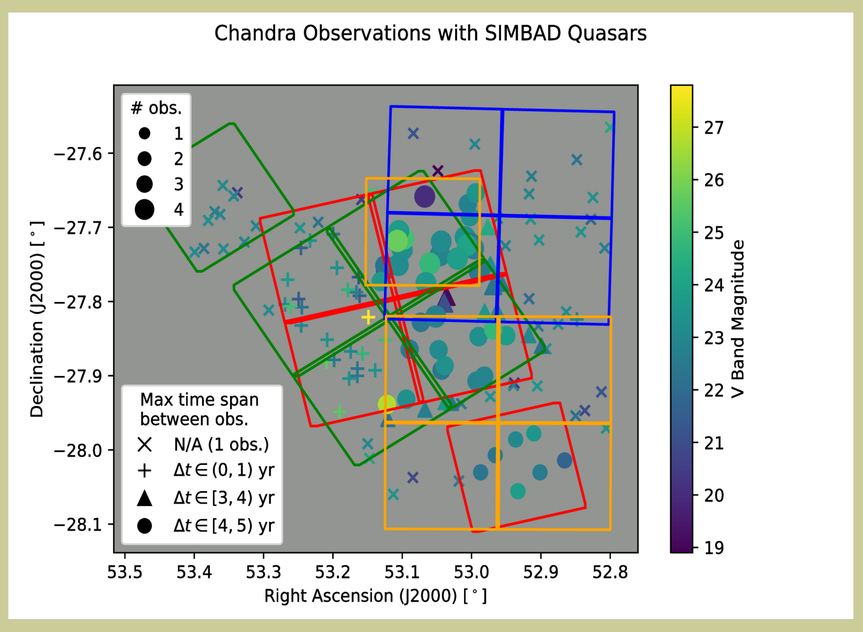


A simple search for Seyfert 2 galaxies using basic Chandra observation and SIMBAD object properties.



Chandra Footprint Service SDSS sky image of a patch of the Chandra Southern Deep Field overlaid with a small subset of Chandra footprints (tangent plane projected).

What Chandra observations contain quasars with V band magnitude between 22.3 and 24.6 that have been observed at least 3 times with a maximum time span between observations of at least 4 years?



A somewhat more complicated search for quasars using aggregate observation metadata and SIMBAD object properties.

Left to Right: A small subset of search results (see above panel) for Seyfert 2 galaxies and quasars respectively. Bolded text in rows denote objects that satisfy all search constraints specified.

SIMBAD Primary ID	RA (J2000) [deg]	Dec (J2000) [deg]	Redshift	Total Exp. Time [ks] (object within .15 deg of pointings)
[DSS2017] 1951	**53.216**	**-27.744**	**0.52**	**151.10**
2XMM J033227.1-280124	53.113	-28.023	0.53	0.00
[SWM2014] GOODS-S 41913	**53.033**	**-27.711**	**0.55**	**306.55**
[LBX2017] 119	**53.015**	**-27.768**	**0.57**	**306.55**
[SWM2014] GOODS-S 28060	**53.048**	**-27.781**	**0.58**	**306.55**
[LBX2017] 367	**53.085**	**-27.792**	**0.60**	**151.10**
[SWM2014] GOODS-S 41898	**53.055**	**-27.711**	**0.61**	**306.55**
[SWM2014] GOODS-S 45795	**53.115**	**-27.696**	**0.67**	**151.10**
[LBX2017] 340	**53.079**	**-27.799**	**0.67**	**151.10**
[SWM2014] GOODS-S 2546	**53.193**	**-27.904**	**0.67**	**151.10**
…	…	…	…	…

SIMBAD Primary ID	RA (J2000) [deg]	Dec (J2000) [deg]	V Band Mag	# Obs.	Max Timespan Between Obs. [yr]
[AZW2014] 5	53.208	-27.851	23.37	2	0.55
[CBH2010] 053.084042-27.573084	53.084	-27.573	21.18	1	N/A (1 obs.)
[DDG2006] 053.0893506-27.8644165	**53.089**	**-27.864**	**22.59**	**3**	**4.48**
[DDG2006] 053.1941096-27.7545166	53.194	-27.755	23.308	2	0.55
[DSS2017] 1429	52.987	-28.030	22.37	2	4.48
[DSS2017] 4168	**53.048**	**-27.865**	**23.16**	**3**	**4.48**
[DSS2017] 4241	**53.073**	**-27.828**	**22.43**	**3**	**4.48**
[DSS2017] 80151	52.914	-27.631	22.57	1	N/A (1 obs.)
[DSS2017] 80171	53.371	-27.679	22.65	1	N/A (1 obs.)
[FPG2012] M3320	53.209	-27.881	25.63	2	0.55
…	…	…	…	…	…

This work was supported by NASA under contract NAS 8-03060 to the Smithsonian Astrophysical Observatory for operation of the Chandra X-ray Center. SIMBAD is developed and maintained by CDS, Strasbourg.