The *Chandra* Bibliography

Sherry Winkelman, for the CDA Team

As a member of the *Chandra* community, you may have noticed that Chaser provides links to publications which analyze *Chandra* data and that ADS provides links to *Chandra* data analyzed within a publication. If you have ever wondered where those links come from and how are they maintained, the answer is the *Chandra* Bibliography, which is curated and maintained by the *Chandra* Data Archive (CDA). Since September 2000, the CDA has been curating the *Chandra* Bibliography with two goals in mind: 1) aiding astronomers in their research by providing links between *Chandra* data and the literature on both ADS and Chaser and 2) provide tools to measure the science productivity of *Chandra* data in the astronomical literature.

The Metadata

The bibliography has four basic categories of *Chandra*-related publications: 1) *Chandra* Science Publications (CSP), where the *Chandra* data contributes significantly to the science presented within the publication; 2) *Chandra* Observatory Publications (CXO), for publications about *Chandra* instruments, software, or operations; 3) Miscellaneous *Chandra* Publications, when a publication refers to the *Chandra* observatory in some way but does not rise to the level of CSP or CXO publication; and 4) Not *Chandra* Related publications, for instances when '*Chandra*' or 'CXO' is in the text of the publication, but the publication does not fit within any of the other categories. These categories provide the framework for the *Chandra* Bibliography.

In addition to these categories, we collect a rich set of metadata that describes the *Chandra* relation to a publication in greater detail. Some of the flags we include for each publication describe: data use, e.g., direct/indirect analysis, multi-observatory analysis, computational analysis; essential publisher information; data origin verification; location of the data, e.g., main text, table, figure; *Chandra* instruments and software; associated catalogs and surveys; additional observatories and wavebands; acknowledgment of data, tools, services, and grants. A simplified interface to the bibliographic database which provides simultaneous browsing of the archive and the literature can be found at *http://cxc.harvard.edu/cgi-gen/cda/bibliography*.

How are *Chandra* Data Being Used?

We took a look at trends related to: 1) direct analysis, which is data analysis that begins with *Chandra* data products provided by the CDA, versus indirect analysis, which is data analysis that uses published results that were based on *Chandra* observations and 2) multi-observatory analysis of data. Our trends are based on refereed CSPs from 2013 to the present because complete tagging is available only for this small interval. When looking at direct and indirect data analysis in publications, we see that 20–30% of the refereed CSPs contain only analysis of the *Chandra* data provided by the CDA (direct analysis) and 15–30% contain only analysis based on previously published results (indirect analysis). The other 50–60% of refereed CSPs contain direct analysis of *Chandra* data combined with multi-observatory analysis, theory and/or computational analysis, or further analysis of previous results (direct + additional analysis). Of the CSPs with direct + additional analysis of data, 93% have a multi-observatory component in the analysis.

Which broad electromagnetic regions (or wavebands) and observatories are most often used in conjunction with *Chandra* data? When looking at these trends, we split the CSPs between those which focused on *Chandra* surveys and their catalogs and those which did not in order to see if there were any differences. When looking at how many wavebands *Chandra* data were combined with (X-ray from another observatory is considered another waveband here), we see no significant difference between CSPs using *Chandra* survey data and those without *Chandra* survey data. We do see some significant differences when we examine which observatory data are combined with *Chandra* data but there are no general trends. *Chandra* survey data were used with 34 other observatories, while non-survey data were used with 79. Not too surprisingly, *HST*, *Spitzer* and *XMM-Newton* were most often used with *Chandra* survey data and they are used with equal frequency (12.1%). On

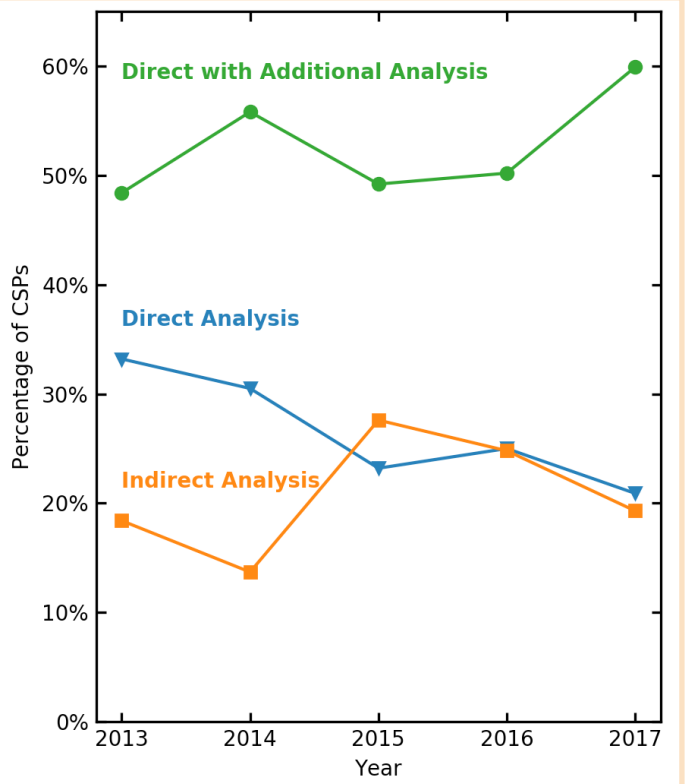

Figure 1: Percentage of CSPs with direct or indirect analysis of data.

the other hand, *XMM-Newton* (17.6%) was used most often with non-survey data with *HST* (9.1%) and Swift (8.5%) being used at about half the rate as *XMM-Newton*. The high rate of combined use of *XMM-Newton* and *Chandra* data is likely due to the complementary strengths of the two X-ray missions, while the combination of *Chandra* data with NASA's other Great Observatories is possibly a consequence of the fact that they were designed to have matching spatial resolution to maximize the scientific return of the missions. To give a sense of the multi-observatory use of *Chandra* data in the literature we summarize the frequency of usage of 20 observatories' data combined with survey and non-survey *Chandra* data.

We noticed an emerging trend in recent CSPs with regards to multi-observatory analysis: CSPs include a growing number of observatories. From 2013 to 2017, the fraction of CSPs that combine *Chandra* data with at least one other observatory has steadily grown. Amongst the CSPs that include data from at least one other observatory, the trend has been towards combining *Chandra* data with more and more observatories in a single CSP. Since 2014, the fraction of CSPs combining *Chandra* data with three or more other observatories has grown steadily while the fraction of CSPs combining *Chandra* data with only one other observatory has decreased. The rate of change from

Table 1: Multi-Observatory Usage of Chandra Data

Observatory	% with Survey Data	% with Non-Survey Data
HST	12.1	9.1
Spitzer	12.1	4.1
XMM-Newton	12.1	17.6
VLA	5.9	6.2
Herschel	5.2	1.2
Subaru	3.5	1.0
Keck	2.8	1.3
VLT	2.8	1.7
WISE	2.8	1.2
UKIRT	2.1	0.5
NuStar	1.7	4.6
Swift	1.4	8.5
ROSAT	1.4	1.1
Gemini	0.3	1.6
Suzaku	0.3	3.6
Fermi	0.0	1.6
GMRT	0.0	1.7
Integral	0.0	1.4

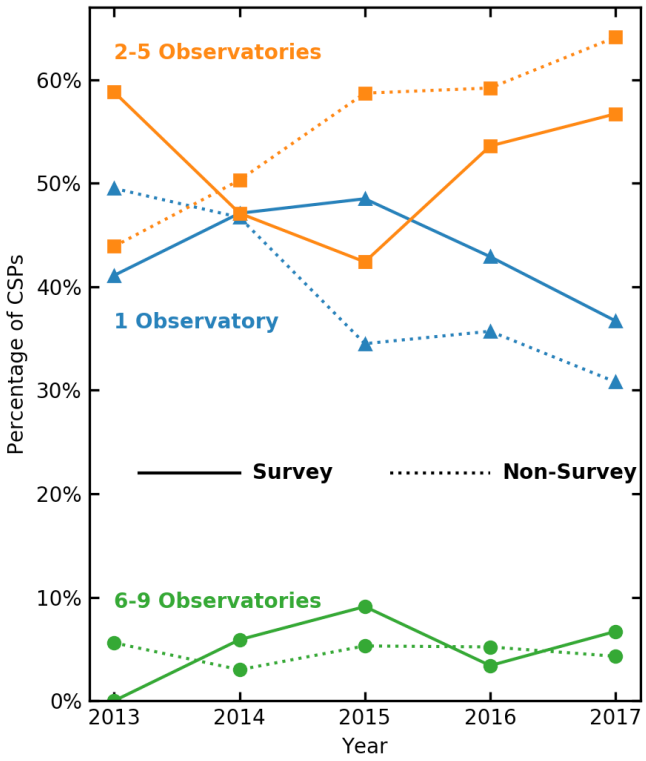

Figure 2: Percentage of CSPs using data from multiple observatories

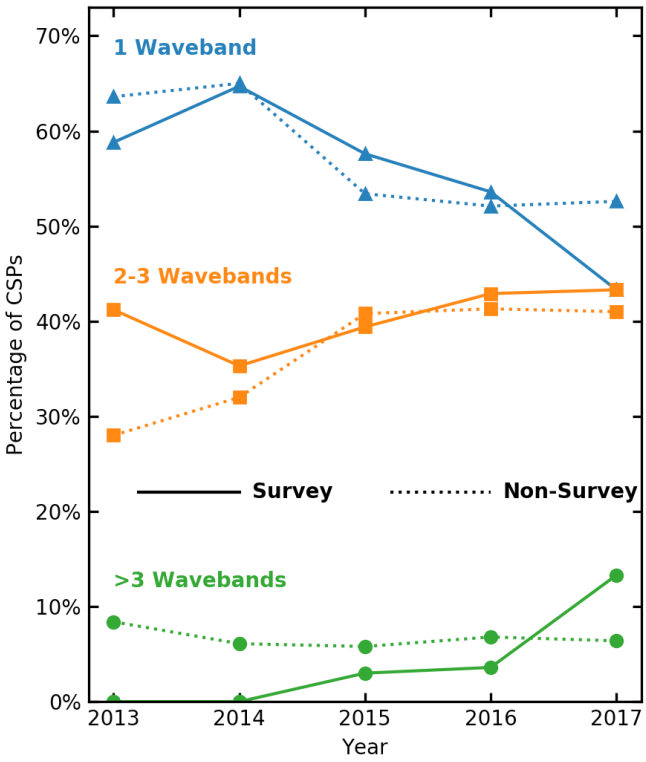

Figure 3: Percentage of CSPs using multiple waveband data.

one observatory to multiple observatories is greater for CSPs unrelated to *Chandra* surveys. We are also seeing an increase in CSPs that include data from five or more observatories, but it is too early to determine if this is a trend.

History and Effort

Such deep analysis of how *Chandra* data are used within the astronomy community is only possible by collecting a wide range of metadata. Over the years the metadata included in the *Chandra* Bibliography has undergone two major expansions. Each expansion in metadata was accompanied by new classification tools; added complexity to the bibliography database; and an intense backfill effort to bring the entire bibliography up to the new version. We have accomplished much with < 1 FTE for most of the mission. Major versions and their major advancements are listed below:

Version 1 (September 2000–June 2004) covered only full journal and proceeding articles with direct analysis of *Chandra* data and links to those data. The curation and dissemination of the bibliography were performed by a single person on a weekly basis.

Version 2 (April 2003–present) expanded the bibliography to include the four categories of *Chandra* publications we use today; covered abstracts and theses as well as all types of astronomical articles; data links were made for direct analysis of data only; and included minimal flags to describe the *Chandra* science context of CSPs. Curation and dissemination was now performed by two to three people and the backfill effort took about a year.

Version 3 (October 2013–present) included a vast expansion of metadata; data linking was extended to all CSPs, including those with indirect analysis of data; and curation was separated from dissemination. Curation is performed by two people, while dissemination is performed by a single person.

Due to the size of the bibliography and the complexity of the metadata and data linking in Version 3, the backfill efforts for this new version will be much greater than past expansions. We have complete metadata coverage for all four categories of publications from 2014 to present. All refereed CSPs from 1999–2000 and 2013 to present have complete metadata coverage, with completion of refereed CSPs from 2001 and 2002 expected by the end of June 2018.

The *Chandra* Bibliography has become an integral part of the CDA and is frequently used to assess the science output of the observatory. It has become a precious resource for management of the mission, demonstrating the necessity for gathering metadata for grants and acknowledgments as well as the use of *Chandra* tools. At the same time, long-term trends observed by curators have led to the addition of metadata describing the *Chandra* connection in publications, which suggest complex and interesting questions about the science impact of *Chandra*. And of course, we are always investigating new ways to present the metadata to the research community as well. ■