Got Data—Will Publish!

Arnold Rots, Sherry Winkelman,
Glenn Becker

The *Chandra* Data Archive's bibliography database (*http://cxc.harvard.edu/cgi-gen/cda/bibliography*) provides its users not only with a powerful research tool, it also allows us to track what happens to all the valuable *Chandra* observations.

Rather than asking how many papers were published on *Chandra* observations or how many citations they received, we started wondering how much of the *Chandra* data actually result in refereed journal articles and how often. After a fair bit of experimentation we constructed the graph in Fig. 1. It shows what fraction of the available *Chandra* exposure time was published—once, twice, thrice, or more often—as a function of the data's age, in steps of one year.

The figure represents the status on 1 August 2011 and includes all non-calibration and non-engineering observations that were made prior to 1 August 2010. More precisely, it shows what percentage of exposure time that has been available for *N* years or more, was presented in a paper in a refereed journal, *N* years after the observation was first released to the PI.

It turns out that this figure, notwithstanding its simplicity, yields a lot of information:

- Second publications start about a year after the first, indicative of the common one-year proprietary period.
- It takes about two and a half years for half the data to be published.
- After four years the curve starts to flatten.
- At seven years about 85% of the data have been published and we have nearly reached an equilibrium. This is a very high percentage.
- Also at seven years, 50% of the data have been published more than twice, indicative both of the popularity of the archive, and the utility of the data.
- Beyond ten years the percentage of data published exceeds 90%. However, since that range pertains pretty much to the data from the early days of the mission, we are curious to see whether such a percentage will be sustained for the later years, too.

Similar statistics are not available for other missions or observatories, but we are confident that the high percentage of data that gets published represents a high score, confirming that *Chandra* data are judged extremely valuable by its user community. Similarly, the statistics on multiple publications is indicative of the intensive use that the community makes of the *Chandra* Data Archive.

A more extensive paper on the Chandra publication statistics will appear in the April 2012 issue of PASP (Volume 124, No. 914).

This kind of information is available thanks to the extensive bibliographic database that the Archive Operations Team maintains. We are grateful for the help we receive from users who provide in their papers explicit information on which observations they use. We would like to encourage all *Chandra* Guest Observers and users of the *Chandra* Data Archive to follow their example or, even better, to embed Dataset Identifiers in their manuscripts (see *http://cxc.cfa.harvard.edu/cda/datasetid.html)*.

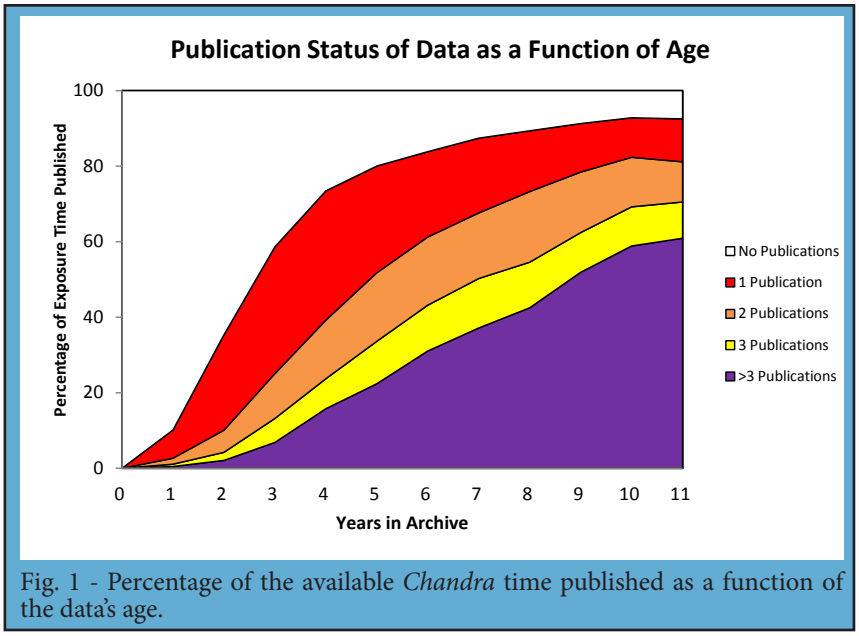

Fig. 1 - Percentage of the available *Chandra* time published as a function of the data's age.