LOOKING BEFORE LEAPING: CREATING A SOFTWARE REGISTRY

Alice Allen
Astrophysics Source Code Library, editor@ascl.net

AND

Judy Schmidt
Astrophysics Source Code Library

ABSTRACT

What lessons can be learned from examining numerous efforts to create a repository or directory of scientist-written software for a discipline? Astronomy has seen a number of efforts to build a repository or directory of scientist-written software, one of which is the Astrophysics Source Code Library (ASCL). The ASCL (ascl.net) was founded in 1999, had a period of dormancy, and was restarted in 2010. When taking over responsibility for the ASCL in 2010, Allen sought to answer the opening question, hoping this would better inform her work. We also provide specific steps the ASCL is taking to try to improve code sharing and discovery in astronomy and share recent improvements to the resource.

1. INTRODUCTION

The Astrophysics Source Code Library (ASCL) was founded in 1999 to improve the transparency and reproducibility of research by making the software used in astrophysics research discoverable for examination (Nemiroff & Wallin 1999). We also encourage reuse of software and advocate for accurate software citation; citing codes increases transparency and gives software authors a way to demonstrate the value of their work, thus providing incentive to make their programs available. Software methods are vital to research (Joppa et al. 2013; Ince et al. 2012; Peng 2011); the importance of computational methods in astronomy is underscored by the existence of journals devoted specifically to these methods (Gray & Mann 2011).

For the ASCL, sustainability is not nearly so ambitious as that defined in The Blind Men and the Elephant: Towards an Empirical Evaluation Framework for Software Sustainability from last year’s WSSSPE: “a measure of a system’s extensibility, interoperability, maintainability, portability, reusability, scalability, and usability” (Venters et al. 2014). Our sustainability model is much smaller, with the minimum requirement being that the source is available to be read by a human. Our bottom line is Is the source code available to be read, and can it be found easily? Preservation and discovery are the most important attributes – if the code is not preserved and cannot be found, no other attributes of sustainability matter!

Several libraries or registries of astrophysics codes have been created over the past two decades. Each effort offered opportunities for researchers to share their software techniques and improve the reproducibility of research, but none reached the tipping point of widespread adoption by the community. What lessons can be learned from examining these efforts?

2. LOOKING BACK AND AROUND

The major astronomy efforts Allen looked at included:

- Astronomical Software Directory Service (ASDS) [Hanisch et al. 1994]
- Astroforge [Brunneg 2005]
- SkySoft [Baffa et al. 2004]
- Astro-Code Wiki [Moore 1999]
- Astro-Sim [Brinkmann 2007]
- AstroShare [Shortridge 2009]

She contacted several of those involved with these efforts to ask what is or had worked well for them, and what had not; for others, she sought out information online as to the fate of the resource. Some of these no longer exist; others carry on but are not very active. She also looked at the ASCL and discussed its early days with one of its creators.

3. LEARNING FROM THE PAST

What we learned from these endeavors was instrumental in our work to restart and grow the ASCL. Each of the efforts to aid communication and share knowledge of codes useful for astrophysics has offered valuable information to the community, yet a number of common factors inhibited the growth and use of these software directories.

The experience of ASDS demonstrates that metadata curation is an ongoing requirement for a resource to be useful and used, and that code authors do not want to take on that task (Hanisch 2011). The original ASCL was not successful according to Nemiroff because most people just didn’t know about it and had few ways to find it. He came to realize that consistent exposure is needed for a resource to become known and used (Nemiroff 2011a). AstroShare and SkySoft are also affected by a lack of marketing, replicating the ASCL’s experience in its earliest days: an initial announcement, a splash or two that generated activity, and then a gradual fading of interest.

Curation and marketing are not the most difficult hurdles; the greatest inhibitors relate to human nature, including the unwillingness of scientists to share their codes...
openly, the effect of the lack of an adequate reward system for software authorship, and the competitive environment in astronomy.

A depressingly small number of programmers are willing to share their work openly even on their own websites. Most emails that ASCL associate editor Kimberly Dure and Allen have sent to software authors to request either a download site for or an archive file of a code we have found in a research paper go unanswered; authors who do reply usually say their codes are not available. The common reasons code authors are reluctant to share their work are covered more fully elsewhere (Barnes 2010; Shamir et al. 2013; Zentner et al. 2014).

Nemiroff’s assumption that software authors would be eager to deposit their code proved false (Nemiroff 2011b). Indeed, experience shows that even authors with open code are reluctant to provide copies of their software to a repository site; many prefer to keep their work close to them, especially if they continue to refine and develop these codes. Most efforts have relied on software authors to submit codes, yet most authors are unwilling to submit information on their codes and maintain metadata stored in a software directory, especially as the payback for doing so may be uncertain.

Astroforge ceased to exist when its initial funding ended (Brunner 2011). The AstroSim project was funded for five years and is now closed; though its Astro-Code Wiki is still available, it is not updated often and is not actively advertised. Success of a software directory requires a change in community attitudes and behavior to one of routine code sharing. Absent outside pressures, change in how a discipline works often takes time, longer than a funding cycle of three or even five years. Without uptake by the community, a new software directory is unlikely to receive additional funding, and as the funding goes away, so too may the resource.

4. CREATING A SHARING COMMUNITY

Space does not allow a complete list of strategies that can be employed to inspire a community to new behaviors, but to start to move the astronomy community to one that routinely shares codes, we can:

1. ensure there is a way to share software (build an infrastructure) that
2. provides incentives for making codes available, and
3. enlist/involve others with appropriate credibility in the community to endorse the effort,
4. market the effort effectively,
5. engage the community to learn what barriers and incentives exist and
6. mitigate these barriers and nurture the incentives, and then finally,
7. be patient.

Let’s look at a few ways we implemented these steps when restarting the ASCL.

Build an infrastructure

In 2010 at Nemiroff’s behest, Allen moved the existing codes from the HTML pages on which they were stored to the phpBB discussion forum that houses, among other things, the discussion threads for Astronomy Picture of the Day (APOD), making the resource more visible and easier to use.

Enlist/involve others to endorse the effort

In 2011, an advisory committee was formed that includes several people prominent in the astronomical software community, computer science professors, well-known code authors, and the founders of the ASCL.

Provide incentives for making codes available

The ASCL started assigning unique identifiers to codes in 2011, providing a way to cite software even when it does not have a descriptive paper associated with it. ADS started indexing ASCL entries in January 2012, which makes codes in the ASCL more discoverable and also provides a way to track citations to ASCL entries.

Market the effort effectively

Peter Teuben, chairman of the ASCL Advisory Committee, and Allen developed a marketing plan for the ASCL that includes presenting the ASCL at conferences, writing guest pieces for blogs such as AstroBetter, and using social media. Further, APOD provides consistent and effective exposure for the ASCL by providing a link to it on the bottom of the APOD page every four to eight weeks.

Engage the community

The ASCL sponsors Special Sessions at American Astronomical Society (AAS) meetings and Birds of a Feather sessions at Astronomical Data Analysis Software and Systems (ADASS) conferences that split the allotted time between presentations and open discussion with attendees. These sessions have been invaluable to learn what additional barriers to code sharing may exist and what incentives may be useful to encourage sharing. These sessions also let the ASCL learn what the community expectations are, which helps us to meet or manage these expectations.

Mitigate barriers and nurture incentives

Because authors are often reluctant to deposit their codes, the ASCL dropped the requirement that codes be housed on the ASCL itself. Instead, we link to the software rather than store and serve the actual source codes, as several other astronomy software directories have done. (The ASCL can and does house software, but the majority of codes it has registered reside elsewhere.) Rather than rely on scientists, who are generally very busy, to provide information on their software, we seek codes and add information about them to the ASCL. In addition to aiding citations and citation tracking for software, the ASCL advocates for greater recognition for its authors.

Be patient

Ten years—that’s our timeline. At the 10-year mark, we should have a good idea whether the ASCL will reach a tipping point and become embedded in the way the community works. Our calendars already have
We continue to expand and improve the ASCL. In early 2014, Judy Schmidt, the primary administrator of the APOD discussion forum, suggested moving code entries to a MySQL database, and shortly after receiving a positive response from Allen to her suggestion, had created a database and fully populated it with ASCL entries. Working with Alberto Accomazzi and Carolyn Stern Grant from ADS, Schmidt and Allen normalized software author names and improved procedures for the flow of data from the ASCL to ADS. Several features were added to the ASCL, including automatic bibcode generation for ADS ingestion, an improved submission process, one-click author searching, and more flexible browsing.

PHP and mySQL were chosen for their ease of use and portability. Both are powerful and ubiquitous tools for web applications. Codeigniter by EllisLab is an open source PHP framework and was chosen because it enables rapid development and increased security for the simple but fully customized user and administrative experience required for the ASCL. The framework is also fully documented [http://ellislab.com/codeigniter/user-guide/] in the event development must be handed over to new hands.

Previously, the ASCL WordPress blog and the code records were disconnected and part of two seemingly different websites. WordPress is now fully integrated into the website both as a blog and for its use as a simple content management system. The discussion forum the ASCL used was incorporated into the new website and topic management for codes is now fully automated. With the ASCL, the blog, and the forum housed under a single, continuous format, user experience is now vastly less confusing and more respectable.

The new database infrastructure offers opportunities for collaboration through the exposure of stored data; we hope the capabilities and flexibility of the new infrastructure can be leveraged to increase software discovery and citation, elevate the status of those who create the programs that enable so much research, and help to make research more transparent and reproducible.

6. CONCLUSION

The ASCL’s requirement for research software sustainability is very basic: the software must be discoverable and available for examination. Ensuring even this basic requirement can be daunting, as there are barriers to and a lack of incentives for code sharing. Looking at several attempts to create astronomy software directories with an eye toward change management suggests specific steps can be taken to provide an environment that encourages software sharing and discovery. Some of these steps are outlined, and how the ASCL has implemented them is shared with the hope that other disciplines facing some of the same challenges may find this information useful. Finally, we have provided information on recent changes and improvements to the ASCL.

REFERENCES

Baffa, C., Giani, E., & Checcucci, A. 2004, in Astronomical Society of the Pacific Conference Series, Vol. 314, Astronomical Data Analysis Software and Systems (ADASS) XIII, ed. F. Ochsenbein, M. G. Allen, & D. Egret, 157
Barnes, N. 2010, Nature, 467, doi:10.1038/467753a
Brinkmann, S. 2007, astro-sim.org is up!
Brunner, R.J.; Remijan, M. 2005, Astronomy catalog searching and software hosting
Brunner, R. J. 2011, private communication
Gray, N., & Mann, R. G, 2011, in Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 655
Hanisch, R. J. 2011, private communication
Hanisch, R. J., Payne, H., & Hayes, J. J. E. 1994, in Astronomical Society of the Pacific Conference Series, Vol. 61, Astronomical Data Analysis Software and Systems III, ed. D. R. Crabtree, R. J. Hanisch, & J. Barnes, 41
Ince, D. C., Hatton, L., & Graham-Cumming, J. 2012, Nature, 482, 485
Joppa, L. N., McNerny, G., Harper, R., et al. 2013, Science, 340, 814
Moore, B. 1999, AstroSim - A five year European program which provides research support for activities in computational astrophysics
Nemiroff, R. J. 2011a, private communication
—. 2011b, Who is this R. J. Nemiroff? Some Favorite Astronomy Ideas
Nemiroff, R. J., & Wallin, J. F. 1999, in Bulletin of the American Astronomical Society, Vol. 31, American Astronomical Society Meeting Abstracts #194, 885
Peng, R. D. 2011, Science, 334, 1226
Shamir, L., Wallin, J. F., Allen, A., et al. 2013, Astronomy and Computing, 1, 54
Shortridge, K. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 3
Venters, C., Lau, L., Griffiths, M., et al. 2014, Journal of Open Research Software, 2
Zentner, L., Zentner, M., Farnsworth, V., et al. 2014, Journal of Open Research Software, 2

APPENDIX

LINKS

SkySoft http://www.skysoft.org/html/index.php
Astro-Code Wiki http://www.astrosim.net/code/doku.php
AstroShare https://www.astrosim.net/bin/view/AstroShare/WebHome
Astronomical Software Directory Service, Astroforge, and Astro-Sim are no longer available