Practices in Code Discoverability

Peter Teuben\textsuperscript{1}, Alice Allen\textsuperscript{2}, Robert J. Nemiroff\textsuperscript{3}, and Lior Shamir\textsuperscript{4}
\textsuperscript{1}Astronomy Department, University of Maryland
\textsuperscript{2}Calverton, MD
\textsuperscript{3}Michigan Technological University
\textsuperscript{4}Lawrence Technological University

Abstract.

Much of scientific progress now hinges on the reliability, falsifiability and reproducibility of computer source codes. Astrophysics in particular is a discipline that today leads other sciences in making useful scientific components freely available online, including data, abstracts, preprints, and fully published papers, yet even today many astrophysics source codes remain hidden from public view. We review the importance and history of source codes in astrophysics and previous efforts to develop ways in which information about astrophysics codes can be shared. We also discuss why some scientist coders resist sharing or publishing their codes, the reasons for and importance of overcoming this resistance, and alert the community to a reworking of one of the first attempts for sharing codes, the Astrophysics Source Code Library (ASCL). We discuss the implementation of the ASCL in an accompanying poster paper. We suggest that code could be given a similar level of referencing as data gets in repositories such as ADS.

1. Introduction

The importance of scientific codes has increased; indeed, this importance is considered a fact of life \cite{Weiner2009} and is continually being discussed in the literature\footnote{A larger scoped paper on this topic is also under preparation by the authors}. Many examples of public codes now exist that have become industry standard software, such as Sextractor, CLOUDY and GADGET to name a few.

In some fields (e.g., bioinformatics) journals include software used to generate results with their articles or require it be submitted. \cite{Gray2011} claim the astrophysics community is not there yet, but scientists are encouraged to release their codes, that their codes are good enough to release even if messy or rough \cite{Barnes2010}. Scientists may see their codes, or their research teams’ codes, as proprietary and thus refrain from publishing them.

Appropriate software is often equally important \cite{Grosbol2010} as data are to research, and Weiner, et al. (2009) state that useful public software packages have enabled easily as much science as yet another large telescope would have. Though the...
NSF was specifically addressing cyber-infrastructure with the statement that its strategic plan defines research infrastructure as including investments in experimental tools, we believe a case can be made for scientific codes fitting within this strategic goal (Stewart et al. 2010).

2. A Brief History of Source Codes

In the early years of computational astrophysics, several important codes were developed but made available only to the social communities surrounding their developers. These social communities would typically include close collaborators, graduate students, postdoctoral fellows, and the graduate students of close collaborators.

An example of this is the "Wilson-Devinney" code that models eclipsing binary star systems and their observable light curves. A first version of this code was written by Robert Wilson in or before 1971. The Wilson-Devinney code has been upgraded and adapted numerous times. Only recently has this code been made available via anony.ftp.

Another example is the "Aarseth" code that models gravitational N-body interactions started around 1960 as one of the participants of the IAU 25 body problem "contest". Several versions of the Aarseth code are now made publicly available by its primary author, Sverre Aarseth through his web site.

3. Previous Online Efforts

With the rise of the internet, packages such as AIPS, IRAF, GIPSY quickly became available to the community.

In the 1980s and 1990s, several prominent astrophysics codes were released to the public over specific web sites. Most of these were not associated with a specific scientific paper. Many of these sites still exist today. The primary way one found out about codes like these was through a mentor or collaborator. In the general field of computing, websites sprang up providing registration or repository service (e.g. freshmeat, sourceforge, github, google.code)

Two of the earliest code collections in astronomy were AstroWEB and ASDS, but neither are maintained anymore (though AstroWEB is still available at NRAO).

In 1999, Nemiroff and Wallin founded the online Astrophysics Source Code Library (ASCL) to house codes of use to the community, eventually resulting in a library of 37 codes, all of which had been described in the literature and used to produce research published in or submitted to refereed journals. The last code was added in 2002. The ASCL site also linked to other code libraries, most of which no longer exist or have not been updated in years.

SkySoft was created in 2001 by C.Baffa, E.Giani, and A.Checcucci. This site is intended to be a site which is community-supported, accepting codes and comments from coders and code users. It also features recent news on topics of interest to astronomer programmers, such as notices of upcoming conferences and workshops for this community. The majority of its code entries date from 2003, with some additional codes from later years.

The Astroforge project was modeled after the wildly successful SourceForge for open source software but focused on the needs of astronomers (Remijan, Brunner, Tillery, & Haider, 2003) and existed for three years.
We know of three other independent code information repositories, though cer-
tainly there are many project groups and individuals who pull together such information
for their own work, team, or subspecialty. One such subspecialty repository is AstrO-
matic for astronomical pipeline software; another is the codes wiki for computational
fluid dynamics.

The Astro-Code Wiki created by AstroSim - European Network for Computational
Astrophysics contains 54 codes and has been updated recently. AstroSim was intended
to be a five-year project to bring together European computational astrophysicists run-
ing from October, 2006 until September, 2011; its focus is on comparison of codes for
suitability for specific tasks.

Another repository called Astro-Sim houses about thirty codes, and provides fo-
rooms for discussion and links to other tools and libraries. Similarly, AstroShare, dis-
cussed by Shortridge (2009), also houses about thirty codes and allows for discussion
of topics such as releasing software, social media, and middleware.

4. To release or not to release source code

Previous endeavors, including the first incarnation of the ASCL, had not grown as codes
have proliferated. Indeed, some scientists are not in the habit of, are reluctant to, or
openly resist making their codes available to non-collaborators. A look at popular and
academic literature, our correspondence and conversations with scientist coders, infor-
mal surveys, and our experiences demonstrate the variety of reasons some scientists
have for not releasing codes.

**Intellectual property issues:** The workplace or granting institutions may place restric-
tions on sharing code.

**Codes reflect the reality of their creation:** Code is often “quick and dirty”; because
it is messy, a coder may be reluctant to release it. Codes also can have a narrow focus,
and the author(s) doesn’t seem it suitable for anything else.

**Releasing codes is not standard practice, useful to ones career, necessary, nor de-
sired:** it is a fact of life that coding does not get you brownie points.

**Releasing codes places demands on the coder, and released codes may be exam-
ined too closely and used inappropriately:** programmers may be worried there might
be bugs in their code (Barnes, 2010). This is again a problem of the lack of time to
check code results to some subjective level of increasingly greater tolerance when this
time could be used to write more papers and again advance in the “publish of perish”
dilemma.

4.1. Why codes should be released

Despite these arguments and absent any national security concerns, we believe it is in-
cumbent upon scientists to release their codes. If a code does the job it was designed
to do, it does not matter that the code may be messy, undocumented, or cobbled to-
gether and inelegant; as a tool used to produce results, the code should be available for
examination and study, just as any research protocol is.

We are not alone in this belief. Timmer(2) laments that the reliance on computational
methods in the sciences has scientists giving up on a key component of the scientific

[http://arstechnica.com/science/news/2010/01/keeping-computers-from-ending-sciences-reproducibility.ars](http://arstechnica.com/science/news/2010/01/keeping-computers-from-ending-sciences-reproducibility.ars)
method: reproducibility. A lack of transparency and reproducibility undermines public confidence in science as well as slowing scientific progress, engendering a credibility crisis, according to Stodden [3] who is working to develop the Reproducible Research Standard.

The NSF has made a recommendation that reproducibility should be promoted, and states that data and software used in the development of a scientific publication should be escrowed or archived where they can be examined and re-verified when needed (Stewart et al. 2010).

5. ASCL

We have implemented a new way to provide a large set of peer-reviewed described codes from the community in an easily accessible place. The details are described in our accompanying paper (Allen et al. 2012). An additional outcome of such a repository could be a referencing database for astrophysics code, similar to the newly established one for data that has been added to the ADS, cf. discussions during a BoF session (Accomazzi et al. 2012).

NOTE ADDED IN PROOF: ASCL codes are now incorporated into ADS.

References

Accomazzi, A., Derriere, S., Biemesderfer, C., & Gray, N. 2012, in ADASS XXI, edited by P. Ballester, & D. Egret (San Francisco: ASP), vol. TBD of ASP Conf. Ser., TBD

Allen, A., Teuben, P., Nemiroff, R. J., & Shamir, L. 2012, in ADASS XXI, edited by P. Ballester, & D. Egret (San Francisco: ASP), vol. TBD of ASP Conf. Ser., TBD

Barnes, N. 2010, Nature, 467, 753

Gray, N., & Mann, R. G. 2011, in Astronomical Data Analysis Software and Systems XX, edited by I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, vol. 442 of Astronomical Society of the Pacific Conference Series, 655

Grosbol, P., & Tody, D. 2010, ArXiv e-prints. [1004.4430]

Shortridge, K. 2009, in Astronomical Data Analysis Software and Systems XVIII, edited by D. A. Bohlender, D. Durand, & P. Dowler, vol. 411 of Astronomical Society of the Pacific Conference Series, 3. [0902.0255]

Stewart, C. A., Almes, G. T., & Wheeler, B. C. 2010, Cyberinfrastructure Software Sustainability and Reusability: Report from an NSF-funded Workshop held 29 and 28 March 2009.

Weiner, B., Blanton, M. R., Coil, A. L., Cooper, M. C., Davé, R., Hogg, D. W., Holden, B. P., Jonsson, P., Kassin, S. A., Lotz, J. M., Moustakas, J., Newman, J. A., Prochaska, J. X., Teuben, P. J., Tremonti, C. A., & Willmer, C. N. A. 2009, in astro2010: The Astronomy and Astrophysics Decadal Survey, vol. 2010 of Astronomy, 61P. [0903.3971]

[http://www.stanford.edu/~vcs/AAAS2011/](http://www.stanford.edu/~vcs/AAAS2011/)