Capturing Software Citations in Astronomy and Planetary Sciences. Edwin A. Henneken and the ADS Team, Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138, USA, ehenneken@cfa.harvard.edu.

Introduction: Soon, if not already, the research lifecycle (figure 1) will be fully digital. Every stage can be captured by one or more digital objects, each of which carries specific knowledge with regards to that stage. The most obvious inhabitant of this digital cosmos is the scholarly publication, but there are many more species, ranging from observation proposals, datasets, software to posters and presentations. With repositories like mission data archives, Figshare and Zenodo, all of these objects can, in principle, be made discoverable and citable. The project Asclepias [1] focuses on one of those objects: software. Asclepias, "Enabling software citation and discovery workflows", is a project funded through a grant from the Alfred P. Sloan Foundation to the American Astronomical Society (AAS), and it involves collaboration between the AAS, Zenodo, the NASA Astrophysics Data System (ADS). Why this collaboration? It reflects the fact that no one stakeholder can "solve" the problem of software citation..

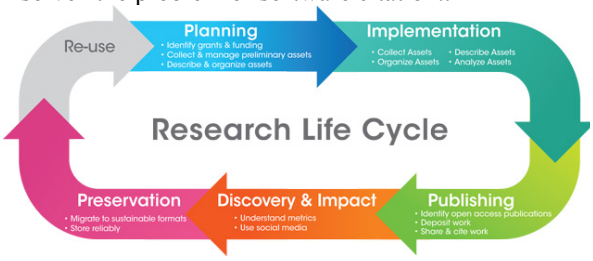

Figure 1 "Research Life Cycle" image from UC Irvine [Library Digital Scholarship Services](https://www.lib.uci.edu/dss)

What is so special about software citation that it takes so many parties to deal with it properly? This becomes clear when you think about how to move from a position where the curation and preservation of software products has been taken care of, to one where these products are discoverable, together with their links to the scholarly literature (by means of citations or otherwise). One challenge is that currently, there is no established standard or policy for citing software in scholarly publications. Writing a "software paper" may seem like a work-around, but it is not a very good one. It can work if you just write one, static piece of software. You describe how it works, acknowledge contributors and, perhaps, where people can get a copy, and you're done. In all other cases, even though it is better than nothing, a "software paper" is just a bad proxy for representing all aspects. Software is a highly dynamical object, especially when it evolves from version to version, with different contributors and different characteristics. For this reason, it is important to capture the acknowledgement of actual software products. Since reworking the entire, established process of capturing scholarly acknowledgement (read: "citations") is extremely hard, to say the least, adapting the existing mechanism of finding citations to include software, will be the next best thing ([2], [3], [4]).
This is exactly what the Asclepias project focuses on. Its goal is to promote scientific software into an identifiable, citable, and preservable object. The project focuses on the needs of two of the most important roles researchers play in the scholarly ecosystem: authors of scholarly manuscripts and developers of scientific software. This goal can only be reached if the research community participates, and the research community can only be expected to participate if there are tools and workflows in place that make the citation of software products just as easy as citing a publication. What needs to be in place to make this happen? The publishers will need to implement practices and tools that will support the integration of software citations into the scholarly publishing workflow. From the viewpoint of bibliographic indexing services (ADS), the implementation means the development of first-class support of a software framework to detect software citations and take all necessary steps to attribute and expose these citations. For repositories (Zenodo), implementation means interoperability with Github and similar codebases (like Bitbucket). The minting of DOIs and metadata registration is one step in this integration, and being able to send out notifications that software has been released, including version updates, is another. All these steps are only possible if metadata standards for software packages are adopted. Just slapping a DOI on a digital object is not a structural solution. It may help an object location be persistent, but that is not how you discover something. That is where the metadata is essential! Within the Asclepias project, the development of metadata is guided by the software citation practices put forward by the FORCE11 Software Citation Implementation Group. When all of this is in place, the three stakeholders will have put in place the framework that will provide the research community with simple instructions for how to make software discoverable and how to include software citations in the process of writing a scholarly paper.

Implementation: The collaboration in the Asclepias project (Muench et al., 2017) has worked on a so-

lution consisting of a technical framework and the promotion of a set of social practices which will "fix" the problems associated with software citations. Releases of software products and associated documentation will be deposited in a trusted repository (Zenodo), so that individual versions are archived as separate entities and proper authorship information is collected for each. Identification will be made possible using a persistent identifier, specifically a DOI. Archived software releases will be assigned unique, persistent identifiers and associated metadata so that precise, persistent connections can be made between papers and software. The metadata is registered with DataCite. Citations will include either a version-specific DOI or a "concept DOI", representing all versions (but always resolving into the record for the latest version). Software developers will be in full control of the process which determines the proper list of contributors for software packages on a release-by-release basis. This takes care of proper attribution. Cited software and its impact will be represented in discipline-specific indexing systems which are used by the targeted community for discovery and evaluation of scholarly content. For astronomy and planetary sciences, this is the NASA Astrophysics Data System (ADS). The metadata for software registered with DataCite, together with the publisher author guidelines for how to cite software, have allowed the ADS to develop a citation capture pipeline for software. If this pipeline finds a citation for an unknown software, a record is created and the citation is assigned. Otherwise the citation is added to the existing record. At the same time a citation event is sent to the Asclepias Broker system at Zenodo. The Asclepias Broker is a web service that enables building and flexibly querying graphs of links between research outputs. It's aiming to address a couple of problems in the world of scholarly link communication, with a focus on software citation:

- Governance of the scholarly links data and metadata
- Meaningful counting of software citations
- Sharing of scholarly links across interested parties

The citation events that the ADS submits to this Broker will appear in the citation data displayed in Zenodo records.

Results: Astronomy and Planetary Sciences have a large community of developers, a culture of openly sharing code and data, and established and centralized infrastructure for communicating results. As a result, authors have already been citing software, even before the Force11 guidelines were put in place. Now a system is in place to support systematic citation of software.

When the ADS citation capture pipeline was first switched on, it detected 2,260 citations of 1,125 software records. Close to 400 of these were from core astronomy journals. Citations were also detected from journals like *JGRE*, *JGRA*, *Geophysical Research Letters*, *Earth and Space Science* and *Geophysical Journal International*. Examples of software being cited in planetary science literature are the *NASA Ames Stereo Pipeline* [5] and the *Planetary-Code-Collection: Thermal And Ice Evolution Models For Planetary Surfaces* [6].

References:

[1] Muench, A., Accomazzi, A., & Holm Nielsen, L. (2017). *Asclepias: Enabling software citation & discovery workflows*. Zenodo. doi:10.5281/zenodo.803474 [2] Allen, A., Berriman, G.B., DuPrie, K., Mink, J., Nemiroff, R., Robitaille, T., Shamir, L., Shortridge, K., Taylor, M., Teuben, P., and Wallin, J. (2015). *Improving Software Citation and Credit*. eprint arXiv:1512.07919 [3] Rios, F. (2016). *The Pathways of Research Software Preservation: An Educational and Planning Resource for Service Development*. D-Lib Magazine 22, 10.1045/july2016-rios [4] Hwang, L. J., Fish, A., Soito, L., Smith, M., & Kellogg, L. H. (2017). *Software and the scientist: Coding and citation practices in geodynamics*. Earth and Space Science, 4, 670–680. doi:10.1002/2016EA000225 [5] Beyer, R., Alexandrov, O., & McMichael, S. (2018), *Neogeography-toolkit/Stereopipeline: Nasa Ames Stereo Pipeline* v2.6.1, Zenodo, doi:10.5281/zenodo.1345235 [6] Schörghofer, N. (2016), *Planetary-Code-Collection: Thermal And Ice Evolution Models For Planetary Surfaces*, v1.1.1, Zenodo, doi:10.5281/zenodo.48851

Acknowledgement: The Asclepias project is funded through a grant from the Alfred P. Sloan Foundation to the American Astronomical Society, 2016. The ADS is operated by the Smithsonian Astrophysical Observatory under NASA Cooperative Agreement NNX16AC86A.