Community Recommendations for Sustainable Scientific Software

Downs, R. R. [1], Lenhardt, W. C.[2], Robinson, E.[3], Davis, E.[4] , Weber, N.[5]

Abstract

A facilitated, roundtable discussion activity at the 2014 Federation of Earth Science Information Partners (ESIP) Summer Meeting elicited recommendations on community activities to improve practices for the sustainability of scientific software. These suggestions fell into three broad themes – (1) improving collaboration and community engagement through publications and presentations (2) developing workshops, training & documenting best practices and (3) creating incentives and motivation w/ awards, citation and a reviewed software repository. In addition to the recommendations coming out of the roundtable activity, this paper highlights how community-led groups such as ESIP are key to move a sustainable software effort in its various forms from concept to reality.

Introduction

Communities are integral to the development and sustainability of scientific software (Howison and Herbsleb, 2014; Katz et al., 2014). The Federation of Earth Science Information Partners (ESIP), a broad-based, distributed community of science data and information technology practitioners, has long recognized the risks of unsustainable practices related to data, informatics, and cyberinfrastructure. Over the last 15 years, ESIP has worked at the forefront of improving sustainable practices along the data lifecycle.

Given the ESIP community history and the natural connections between data management and software development (Lenhardt, et al, 2014), it is logical that more recently, the ESIP membership has turned to examine the issues related to software and the benefits that can be attained from the sustainability of scientific software. Starting in the summer of 2013, the ESIP meeting included a panel on the sustainable software topic. From this panel, ESIP formed a cluster devoted to science software. Over the last year, these efforts evolved to become the central theme of the ESIP 2014 Summer Meeting, in Copper Mountain, Colorado. As part of this meeting, the theme was carried from the plenary presentations to a roundtable, lunchtime discussion that leveraged the perspectives on sustainable software of over 300 Earth science community representatives, including data distributers, providers of data and information products, developers of tools for Earth science, modelers, and funders. The outcome of this discussion was captured and has been analyzed to identify recommendations from the community to improve practices for scientific software sustainability.

1. Center for International Earth Science Information Network (CIESIN), Columbia University, Palisades, NY, rdowns@ciesin.columbia.edu
2. Renaissance Computing Institute (RENCI), University of North Carolina at Chapel Hill, Chapel Hill, NC, clenhardt@renci.org
3. Foundation for Earth Science, Boulder, CO, erinrobinson@esipfed.org
4. Unidata, University Corporation for Atmospheric Research (UCAR), Boulder, CO, edavis@ucar.edu
5. University of Illinois, Champaign-Urbana, Champaign, IL, nmweber@illinois.edu

Methodology

The roundtable lunch discussion on the sustainability of scientific software was held during the main conference day after a series of plenary speakers focused on sustainable software issues. Prior to the roundtable lunch activity, 36 meeting contributors were asked to serve as discussion facilitators for the planned roundtable. The remaining attendees were each sequentially assigned, from an alphabetized list, to one of 36 tables. Facilitation included reading the questions to participants at the table and capturing ideas generated during the discussion. During the plenary session, the attendees were invited to join the roundtable lunch discussion and share their perspectives on the sustainability of scientific software. Each table contained instruction sheets and a form to collect the responses from the table.

Three sets of questions guided the discussion at each table. The first set pertained to the definition of sustainable scientific software and the second set elicited perspectives on various aspects of sustainable scientific software. The third set of questions requested recommendations for activities that the ESIP community might consider for the near future to improve practices for the sustainability of scientific software. The initial results, described here, reflect responses to the third question that participants have recommended for the ESIP community to improve scientific software sustainability practices.

Initial Results

We received responses from 25 of the 36 invited tables. Initial analysis revealed the following actionable activities recommended for the ESIP community to improve the sustainability of scientific software. It is anticipated that ESIP contributors, including ESIP's Science Software Cluster, will take up some of these recommendations going forward.

Collaboration

The participants recommended new activities and co-sponsorship of activities to encourage ESIP members to collaborate with other groups on the sustainability of scientific software, including groups such as the International Council for Science Committee on Data (CODATA), the World Data System (WDS), the Research Data Alliance (RDA) EarthCube, and COOPEUS. Participants also recommended inviting scientists and end users to attend ESIP meetings and share their perspectives on software sustainability.

Publications and Presentations

Activities were recommended for the ESIP community to increase awareness, visibility and understanding of scientific software sustainability issues within the Earth science community. Participants recommended producing publications and presentations to inform the Earth science community about these issues, suggesting that community members propose AGU sessions focused on software sustainability, offering conceptual information that is less technical. Likewise, submitting papers to *Eos* and to the WSSSPE also were recommended to inform the Earth science community about the importance of scientific software sustainability.

Workshops, Training & Best Practices

The participants recommended raising awareness of software sustainability and facilitating different levels of training. Suggestions included developing training modules for simple software lifecycle skills and learning modules to improve understanding about the sustainability of scientific software, similar to the Data Management Training Modules developed by the ESIP Federation (Duerr and Hoebelheinrich, 2012). Participants also recommended

conducting training on agile development techniques and convening software carpentry events, like those offered during the 2014 ESIP Federation Meeting (http://commons.esipfed.org/2014SummerMeeting).

Develop and Document Best Practices

Recommendations included examining incentives, policies, and practices and highlighting examples of good scientific software sustainability. Activities would include creating software management plans and recommendations for organizations and individuals for improving software sustainability, establishing criteria for the sustainability of scientific software, documenting use cases and good sustainable examples, and developing impact metrics for software. Promoting practices for provenance, modularity, and version control also was suggested. Participants recommended developing a science software sustainability model or even a simple checklist or matrix for scientific software sustainability. They also suggested establishing metadata standards and profiles for workflows and software to ensure that best practices are followed for the sustainability of software components and their dependencies. The community-developed and vetted, ESIP Data Citation Guidelines are such an example related to data.

Incentives and Motivation

The meeting participants suggested offering incentives including awards and citations to recognize contributions to the sustainability of scientific software. Offering awards would stimulate recognition for individuals who contribute to scientific software sustainability within their organizations. Participants also recommended improving attribution by developing templates and guidance for software citation, which could offer motivation for reusing such software. Incentives also were suggested to motivate scientists and developers to proactively produce good documentation and guidance for improving provenance and version control. Opportunities for funding also were recommended for refactoring of identified useful software as well as for research examining software sustainability issues.

Reviewed Software Repository

Participants recommended that ESIP members might create a curated and reviewed software repository that includes an ESIP "stamp of approval" for reviewed software. The repository could utilize a taxonomy of different types of software and measurable characteristics of sustainability to serve as a clearinghouse for scientific software and as a central 'vetter' of reusable standards and software. Such a repository also could serve as an inventory for software reviewed by expert users who rate and measure the sustainability of submitted software, applying tools, such as the Reuse Readiness Levels (NASA Earth Science Data Systems Software Reuse Working Group, 2010) and the Technology Readiness Levels (Mankins, 1995), to conduct such reviews.

Discussion & Conclusion

The recommendations offered by the participants suggest three broad themes that could improve the sustainability of scientific software: (1) Community and collaboration is crucial both within ESIP and beyond to partners to move sustainable software forward; (2) There is a need for training and best practices around sustainable software; (3) In order to enable sustainable software there must be recognition for the work through incentives like awards and citations.

As the ESIP community pivots towards examining issues related to sustainable software and devoting a semi-annual meeting theme to the topic of sustainability it will almost by definition move the sustainable software agenda forward. The content of the meeting sensitized the immediate ESIP meeting attendees and by extension

the broader research community to the importance of sustainable software. The roundtable activity created opportunities to operationalize sustainable software concepts.

In this paper our goal has been to describe recommendations observed for improving the sustainability of science software. These conclusions grow from the ESIP community focus on sustainable science software as reflected in the ESIP Summer 2014 meeting. Even though the participants largely represented the Earth science informatics community, these recommendations also apply to other communities and we look forward to making those connections.

Acknowledgements

The authors very much appreciate the efforts of the volunteer facilitators and attendees of the 2014 Summer Meeting of the Federation of Earth Science Information Partners (ESIP) who participated in the reported roundtable discussion activity and shared their perspectives. This work is based on the presentation by Lenhardt, Downs, Weber, and Robinson (2014). Approval to conduct the research was requested and received from the Columbia University Institutional Review Board.

References

Duerr, R. and Hoebelheinrich, N.J. 2012. Data Management for Scientists Short Course. Federation of Earth Science Information Partners: ESIP Commons. http://commons.esipfed.org/datamanagementshortcourse.

ESIP Data Stewardship Committee. 2012. Data Citation Guidelines for Data Providers and Archives. Federation of Earth Science Information Partners: ESIP Commons. http://commons.esipfed.org/node/308. doi:10.7269/P34F1NNJ.

Howison, J. and Herbsleb, J.D. 2014. The sustainability of scientific software: ecosystem context and science. http://james.howison.name/pubs/HowisonHerbsleb-Sustainability.pdf.

Katz, D.S., Choi, S.T., Lapp, H, Maheshwari, K, Löffler, F, Turk, M, Hanwell, M.D., Wilkins-Diehr, N, Hetherington, J, Howison, J, Swenson, S, Allen, G.D., Elster, A.C., Berriman, B and Venters, C. 2014. Summary of the First Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE1). Journal of Open Research Software 2(1):e6, DOI: http://dx.doi.org/10.5334/jors.an.

Lenhardt, W.C., Downs, R.R., Weber, N., and Robinson, E. 2014. ESIP 2014 Summer Meeting Roundtable. July 9, 2014. 2014 Summer Meeting of the Federation of Earth Science Information Partners. July 8-11, 2014, Copper Mountain, Colorado. http://commons.esipfed.org/node/2368.

Lenhardt, W.C., Ahalt, S, Blanton, B, Christopherson, L and Idaszak, R 2014. Data Management Lifecycle and Software Lifecycle Management in the Context of Conducting Science. Journal of Open Research Software 2(1):e15, DOI: http://dx.doi.org/10.5334/jors.ax.

Mankins, J.C. 1995. Technology Readiness Levels. National Aeronautics and Space Administration. http://www.hq.nasa.gov/office/codeq/trl/trl.pdf.

NASA Earth Science Data Systems Software Reuse Working Group. 2010. Reuse Readiness Levels (RRLs), Version 1.0. April 30, 2010. https://earthdata.nasa.gov/esdswg/software-reuse/reuse-readiness-levels-rrls.