Journal of
open research software

Downs, R R et al 2015 Community Recommendations for Sustainable Scientific Software.
*Journal of Open Research Software,* 3: e11, DOI: http://dx.doi.org/10.5334/jors.bt

ISSUES IN RESEARCH SOFTWARE

# Community Recommendations for Sustainable Scientific Software

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

[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

Science software has contributed to research practices, but the sustainability of scientific software presents challenges for the future use of research resources. Identifying improvements for science software sustainability practices can contribute to the re-use of science software. A focus group study was conducted to identify ways to improve science software sustainability practices of the Earth science community. 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, and documenting best practices and (3) creating incentives and motivation with 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.

**Keywords:** Science software; software sustainability; sustainable systems; focus groups

## Introduction

The creation and adoption of software has become an integral aspect of current research practices [1]. Today, software is commonly used to plan and design research, to obtain funds and approvals, to create instruments and tools, to collect and analyze data, and to publish and archive results and other research resources. Simultaneously, the new capabilities spawned from the development and adoption of software also offer new challenges for progress in science. Much of today's science infrastructure, or cyberinfrastructure, is dependent on software [2], which often is used throughout the research lifecycle to create and utilize the resources needed for science. Systems, programs, scripts, workflows, and processes are commonly made from software that is needed to discover, render, or use science artifacts in digital form. Designing systems that enable the use of research resources requires the use of similar or compatible software to access the data, instruments, and tools that were previously created [3]. Furthermore, the exact version of software may be needed to replicate or reproduce the results of previous research [4]. Without software, many current research practices could not be conducted for collecting, discovering, and analyzing data to produce results. For example, conducting research using geospatial data is largely dependent on geospatial information systems and remote sensing software. Software has become an essential aspect of science that is necessary to maintain and advance current research practices and infrastructure.

Science software must be sustainable and reliable to contribute to future science practices. The dependency of science on software necessitates expeditious efforts by the research community to ensure the sustainability of science software. Likewise, the research community needs to ensure that science software can be relied upon to reproduce research results [5]. With such dependencies, unsustainable and unreliable science software reduces the potential for using current research resources in the future and for the reproducibility of science. In addition, utilizing science software that is not sustainable has the

potential to increase risks, over time, to the use of scientific instruments, tools, and data. Using unsustainable science software also increases risk to the credibility of science [6]. Developing and managing science software to be sustainable contributes to the potential for enabling ongoing use of research resources and to the reduction of potential risks to the future of science.

Identifying improvements for science software sustainability practices can contribute to the sustainability of science software. The future use of current research resources will depend on capabilities to develop and manage sustainable science software. The research community will need to identify and adopt ways to improve upon current practices for science software development and management. To address these challenges, a focus group study was conducted to identify ways to improve the science software sustainability practices of the Earth science community.

Over the last 15 years, the Federation of Earth Science Information Partners (ESIP), a broad-based community of science data and information technology practitioners, has worked at the forefront of improving sustainable practices along the data lifecycle [7, 8]. Since it has been recognized that communities are integral to the development and sustainability of scientific software [9, 10] and given the ESIP community history and the natural connections between data management and software development [11], it is not surprising 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 semi-annual meeting included a panel and breakout session on the topic of sustainable software. From these activities, 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. The theme — "Linking It Together: Sustainable Software Advancing Science Data and Services" — was set forth and discussed during the plenary presentations and carried through to a lunchtime roundtable that engaged approximately 300 meeting attendees in 8-person focus group discussions. The outcomes of these discussions were captured and have been analyzed to identify recommendations from the community to improve practices for scientific software sustainability.

## Methodology

The focus group method was employed for data collection as it has been used in various fields of social science to enable the elicitation of in-depth perspectives and ideas on a topic of interest that emerge from interactive discussions among participants [12]. 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 each asked to serve as a discussion facilitator for an assigned table. Facilitation included reading the questions to participants at the table and capturing ideas generated during the discussion. The remaining 250 Earth Science community representatives (including data distributers, providers of data and information products, developers of tools for earth science, data users, and funding agency representatives) were each sequentially assigned, from an alphabetized list, to one of 36 tables, with eight participants at each table.

For this study, each table was considered a focus group. Focus groups are valuable for obtaining empirical observations on various topics, including complex issues, such as software engineering [13] and health science research [14]. The table assignments provided a reasonable sample size of eight attendees for each of the focus groups [15].

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 participants were not asked to identify themselves and any responses that contained the names of participants were de-identified prior to analysis. 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 28 of the 36 invited tables. Initial analysis revealed the following actionable activities recommended for the ESIP community to improve the sustainability of scientific software. ESIP contributors, including ESIP's Science Software Cluster, are actively working to define approaches to implementing the recommendations going forward.

### Collaboration

Participants in the summer meeting recommended that ESIP work with other science and informatics organizations to develop and co-sponsor new activities that encourage collaborations between members of the various communities that focus on ways to increase the sustainability of scientific software. A number of other community groups were mentioned including the International Council for Science Committee on Data (CODATA), the World Data System (WDS), the Research Data Alliance (RDA), Earth Cube, and COOPEUS. Participants also recommended working to increase the number of scientists and end users that attend ESIP meetings to gain more of an end user perspective 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 [16]. Participants also recommended conducting training on agile development techniques and convening software carpentry events, like those offered during the 2014 ESIP Federation Summer 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 [17]. These guidelines and other resources are developed and reviewed by teams of volunteers who are members of clusters, working groups, and committees that are open to the entire ESIP community for contributions. Completed resources are voted upon for approval during a Business Meeting of the ESIP Assembly.

#### 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 [18] and the Technology Readiness Levels [19], to conduct such reviews.

## Discussion & Conclusion

Organizing and facilitating multiple, informal roundtable discussions to elicit recommendations for improving scientific software sustainability provided opportunities for various perspectives, including those of Earth science researchers and data science practitioners, to be shared by and among the ESIP community members who participated in the focus group study. In addition, the semi-structured organization of the questionnaire on science software sustainability issues enabled each table of participants to provide responses in accordance with the interests and perspectives represented within their focus group. Since the participants who contributed to the roundtable discussions represent the Earth science informatics community, the recommendations for improving scientific software sustainability that were elicited from the participants could reflect perspectives that come from the practices and culture of that community. Organizing similar roundtable discussions or focus groups to elicit recommendations from other scientific communities may reveal different perspectives for improving scientific software sustainability that reflect the practices and cultures of the represented communities.

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, such as devoting a semi-annual meeting theme to the topic of software sustainability and creating a group to work on the topic, it has almost by definition begun to move the sustainable software agenda forward within the Earth science informatics community. 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.

## Competing Interests

RRD currently serves as a volunteer member of the Board of Directors of the Foundation for Earth Science, the nonprofit organization that provides managerial support for the Federation of Earth Science Information Partners (ESIP).

WCL previously served as the volunteer President of the ESIP Federation.

ER currently serves as the Executive Director of the Foundation for Earth Science.

ED currently serves as the volunteer Chair of the ESIP Information Technology and Interoperability Committee.

NW currently serves as the student intern for the ESIP Science Software Cluster.

## 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 [20]. Approval to conduct the research was requested and received from the Columbia University Institutional Review Board. The authors also appreciate the suggestions for improving an earlier version of this paper that were received from the anonymous reviewers as part of the peer review process for the 2nd Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE2). Support for Robert R. Downs was provided by the National Aeronautics and Space Administration under Contract NNG13HQ04C for the Socioeconomic Data and Applications Center (SEDAC). Partial support for W. Christopher Lenhardt was provided by National Science Foundation (NSF) award 1216817 Conceptualization of a Water Science Software Institute.

## References

1. **Goble, C** 2014 Better software, better research. Internet Computing, IEEE, 18(5): 4–8. DOI: http://dx.doi.org/10.1109/MIC.2014.88
2. **Venters, C C, Lau, L, Griffiths, M K, Holmes, V, Ward, R R, Jay, C, Dibsdale, C E** and **Xu, J** 2014 The Blind Men and the Elephant: Towards an Empirical Evaluation Framework for Software Sustainability. Journal of Open Research Software, 2(1): e8. DOI: http://dx.doi.org/10.5334/jors.ao
3. **Downs, R R** and **Marshall, J J** 2010 A Proposal on Using Reuse Readiness Levels to Measure Software Reusability. Data Science Journal, 9: 73–92. DOI: http://dx.doi.org/10.2481/dsj.009-007
4. **Stodden, V** and **Miguez, S** 2014 Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. Journal of Open Research Software, 2(1): e21. DOI: http://dx.doi.org/10.5334/jors.ay
5. **Merali, Z** 2010 Computational science: Error, why scientific programming does not compute. Nature, 467(7317), 775–777. DOI: http://dx.doi.org/10.1038/467775a
6. **Blanton, B** and **Lenhardt, C** 2014 A Scientist's Perspective on Sustainable Scientific Software. Journal of Open Research Software, 2(1): e17. DOI: http://dx.doi.org/10.5334/jors.ba
7. **Duerr, R E, Downs, R R, Tilmes, C, Barkstrom, B, Lenhardt, W C, Glassy, J, Bermudez, L E** and **Slaughter, P** 2011 On the Utility of Identification Schemes for Digital Earth Science Data: An Assessment and Recommendations. Earth Science Informatics, 4(3): 139–160. DOI: http://dx.doi.org/10.1007/s12145-011-0083-6
8. **Wilson, A, Downs, R R, Lenhardt, W C, Meyer, C, Michener, W, Ramapriyan, H** and **Robinson, E** 2014 Realizing the Value of a National Asset: Scientific Data. Eos Transactions of the American Geophysical Union, 95(50): 477–478. DOI: http://dx.doi.org/10.1002/2014EO500006
9. **Howison, J** and **Herbsleb, J D** 2014 The sustainability of scientific software: ecosystem context and science. Available at: http://james.howison.name/pubs/HowisonHerbsleb-Sustainability.pdf.
10. **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
11. **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
12. **Morgan, D L** 1996 Focus Groups. Annual Review of Sociology, 22: 129–152. Available at: http://www.jstor.org/stable/2083427. DOI: http://dx.doi.org/10.1146/annurev.soc.22.1.129
13. **Kontio, J, Bragge, J** and **Lehtola, L** 2008 The focus group method as an empirical tool in software engineering. In Guide to advanced empirical software engineering. Springer London, pp. 93–116. DOI: http://dx.doi.org/10.1007/978-1-84800-044-5_4

14. **Carlsen, B** and **Glenton, C** 2011 What about N? A methodological study of sample-size reporting in focus group studies. BMC Medical Research Methodology, 11(1): 26. DOI: http://dx.doi.org/10.1186/1471-2288-11-26
15. **Onwuegbuzie, A J** and **Leech, N L** 2007 A call for qualitative power analyses. Quality & Quantity, 41(1): 105–121. DOI: http://dx.doi.org/10.1007/s11135-005-1098-1
16. **Duerr, R** and **Hoebelheinrich, N J** 2012 Data Management for Scientists Short Course. Federation of Earth Science Information Partners: ESIP Commons. Available at: http://commons.esipfed.org/datamanagementshortcourse.
17. **ESIP Data Stewardship Committee** 2012 Data Citation Guidelines for Data Providers and Archives. Federation of Earth Science Information Partners: ESIP Commons. Available at: http://commons.esipfed.org/node/308. DOI: http://dx.doi.org/10.7269/P34F1NNJ
18. **NASA Earth Science Data Systems Software Reuse Working Group** 2010 Reuse Readiness Levels (RRLs), V1.0. April 30, 2010. Available at: https://earthdata.nasa.gov/esdswg/software-reuse/reuse-readiness-levels-rrls.
19. **Mankins, J C** 1995 Technology Readiness Levels. National Aeronautics and Space Administration. Available at: http://www.hq.nasa.gov/office/codeq/trl/trl.pdf.
20. **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. Available at: http://commons.esipfed.org/node/2368.

**How to cite this article:** Downs, R R, Lenhardt, W C, Robinson, E, Davis, E and Weber, N 2015 Community Recommendations for Sustainable Scientific Software. *Journal of Open Research Software* 3: e11, DOI: http://dx.doi.org/10.5334/jors.bt

**Published:** 18 November 2015



]u[ *Journal of Open Research Software* is a peer-reviewed open access journal published by Ubiquity Press

OPEN ACCESS