Raising the Profile of Research Software: Recommendations for Funding Agencies and Research Institutes in the Netherlands

This paper was open for comments via a public document since September 2019,[1] and a community feedback session was held during NL-RSE 2019[2]. This version of the paper incorporates feedback received through these channels.

Authors (alphabetically):
Anton Akhmerov, Maria Cruz, Niels Drost, Cees Hof, Tomas Knapen, Mateusz Kuzak, Carlos Martinez-Ortiz, Yasemin Turkyilmaz-van der Velden and Ben van Werkhoven

Incorporating comments via public document from:
Michael Doube, Alejandra Gonzalez-Beltran, Yan Grange, Vincent van Hees, Scott Henwood, Patricia Herterich, Daniel S. Katz, Andreas Mueller, Esther Plomp and Heidi Seibold

comments via NL-RSE event from:
Tania Allard, Victor Azizi, Patrick Bos, Jan C. Brammer, Gerasimos Chourdakis, Jeremy Cohen, Stephan Druskat, Yifat Dzigan, Peter Kalverla, Vedran Kasalica, Meta Keijzer - De Ruijter, Paul Konstantin Gerke, Dafne van Kuppevelt, Johan Philips, Pablo Rodríguez-Sánchez and Shoaib Sufi

and special comments from: Tom Bakker and Joris van Eijnatten.

Disclaimer: All opinions expressed in this paper are the authors' own opinions and do not necessarily reflect the official policy or position of their institutions.

Abstract

Research software is fundamental to today's research, yet it does not receive the recognition it deserves. For example, research software is not yet adequately recognised in the scholarly record through citation. This situation contrasts starkly with the fact that open science and reproducible research is given ample attention at the policy level. We argue that if at the policy and institute level, everyone is convinced that open science leads towards better, more transparent and reproducible research, then visible and practical steps forward should be made that ensure that research software is treated on equal footing with publications and research data.

[1] "Request for comments -- Raising the Profile of Research Software," 17 Sep 2019, https://nl-rse.org/2019/09/17/RaisingResearchSoftware.html.
[2] "NL-RSE19: Program," 20 Nov 2019, https://nl-rse.org/events/NL-RSE19.html.

In this paper, we present a concrete follow-up to the Declaration of Research Assessment (DORA) in the form of recommendations for raising the profile of research software. These recommendations provide steps for achieving recognition for research software as a fundamental and vital component of research. This paper focuses on the Netherlands, but the issues and recommendations could be adapted and adopted elsewhere. We divide our recommendations into four categories: Software availability and quality, Software sustainability, Training, and Human capital.

Introduction

"Without data it's difficult to validate results. But without code, we waste the opportunity to advance science."

Neil Chue Hong
Software Sustainability Institute

Research software is fundamental to contemporary research. In many fields it is no longer possible to conduct research without software, whether software libraries or small data analysis scripts.[3] The announcement in April 2019 of the first image of a black hole (figure 1),[4] making front-page news around the world, is a case in point. This landmark image could not have been made without research software, both the standard packages already in use in the field as well as specially developed algorithms to process and analyse the data coming from several radio telescopes around the world. Research software enabling academic research is by no means exclusive to astronomy; many examples can be found in life sciences,[5] chemistry,[6] and history[7] to name a few.

[3] "It's impossible to conduct research without software, say 7 out of 10 UK …," 4 Dec 2014, https://www.software.ac.uk/blog/2014-12-04-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers.

[4] "Focus on the First Event Horizon Telescope Results - IOPscience," Apr 2019, https://iopscience.iop.org/journal/2041-8205/page/Focus_on_EHT.

[5] C. Geng et al. "iScore: A novel graph kernel-based function for scoring protein-protein docking models," Jun 2019, https://doi.org/10.1093/bioinformatics/btz496.

[6] L. Ridder et al, "Substructure-based annotation of high-resolution multistage MSn spectral trees," Sep 2012, https://doi.org/10.1002/rcm.6364.

[7] H. Wijfjes et al, "PIDIMEHS: Pillarization and depillarization tested in digitized media historical sources," 2016, https://doi.org/10.17026/dans-xzj-vhgd.

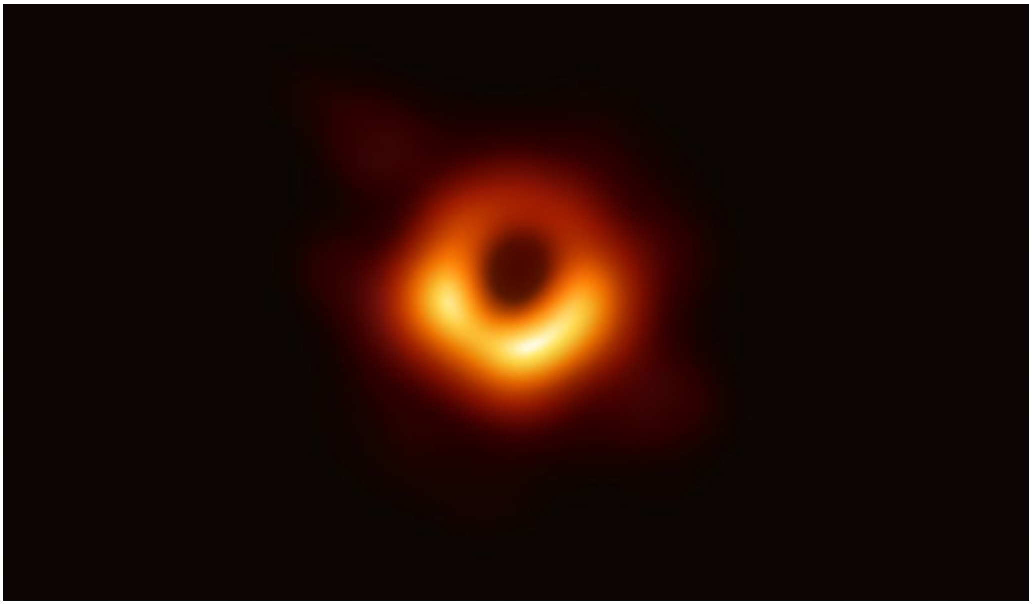

Figure 1: The black hole at the center of the galaxy Messier 87 as revealed by the Event Horizon Telescope. This image would not be possible without research software in its various guises. Image credit: Event Horizon Telescope Collaboration.

Yet research software does not receive the recognition it deserves. Among other things, research software is not adequately recognised in the scholarly record through citation.[8,9] There is a lack of funding and incentives for those who develop research software.[10,11] Acknowledgement of the importance of research software as an independent research output lags behind that of research data and research publications.[12] In terms of the incentives and rewards structure, both software and data contributions are considered second-rate relative to publications.[13] According to a 2018 survey,[14] a significant percentage of Research Software Engineers (RSE) in the Netherlands are not acknowledged in publications about research to which their software was crucial.

[8] J. Howison and J. Bullard, “Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature,” Journal of the Association for Information Science and Technology, vol. 67, no. 9, pp. 2137–2155, May 2015 https://doi.org/10.1002/asi.23538.

[9] L. A. Barba et al., “Giving software its due through community-driven review and publication,” Apr 2019, https://doi.org/10.31219/osf.io/f4vx6.

[10] A. Siepel, “Challenges in funding and developing genomic software: roots and remedies,” Genome Biology, vol. 20, no. 1, Jul 2019, https://doi.org/10.1186/s13059-019-1763-7.

[11] “How to support open-source software and stay sane,” 01 Jul 2019, https://www.nature.com/articles/d41586-019-02046-0.

[12] “Making Software a First-Class Citizen in Research,” 28 Nov 2018, https://software.ac.uk/blog/2018-11-28-making-software-first-class-citizen-research.

[13] “Why we need a hub for software in science,” 17 Dec 2015, https://blogs.lse.ac.uk/impactofsocialsciences/2015/12/17/why-we-need-a-hub-for-software-in-science/.

[14] B. van Werkhoeven, “RSE Survey 2018 Results,” Jul 2019, https://nl-rse.org/2019/07/01/RSE-Survey-2018-results.html.

This situation is at odds with the discussions around open science and reproducibility.[15] These discussions, which gained much prominence in recent years, have all largely focused on open access publications and FAIR data.[16] As a result, open data and open access publishing are becoming mainstream in many fields of research. In many disciplines, enriching and storing research data is now seen as a valid form of research output.

We argue that **if open science is to lead towards better, more transparent, and reproducible research, then research software needs to be treated on an equal footing with research data and research publications, both at the policy level and in the practice of researchers and research institutes**. There cannot be fully open, transparent, and reproducible research without open and sustainable research software. Such software requires, in turn, appropriate funding, incentives, recognition and rewards to be in place, not only for the software itself, but for the people who created it. Without these, it is difficult to sustain and maintain the algorithms and research software that are crucial for scientific advances. If not sustained, software is likely to have to be "reinvented" many times, leading to duplication of effort and costs rather than collaboration and reuse.

The Declaration on Research Assessment (DORA) recognizes the need to improve scholarly research evaluation. It recognizes data and software as being just as important as research articles. By signing DORA, NWO (Dutch Research Council), ZonMw (Dutch organization for health research and healthcare innovation), KNAW (Royal Netherlands Academy of Arts and Sciences), and others have committed to taking steps in changing the scholarly research evaluation criteria.

The Netherlands is already one of the leading countries in its efforts to embrace digital transformation in research: NWO, ZonMw, KNAW signed DORA; TU Delft was a pioneer in creating dedicated data stewards and RSE groups[17,18,19]; SURF and NWO set up the Netherlands eScience Center to create an effective bridge between research and digital technologies; the National Program Open Science focuses on professional education for Open

[15] "Improving Reproducibility in the Empirical Sciences, Koninklijke Nederlandse Akademie van Wetenschappen," 2018, ISBN 978-90-6984-720-7.

[16] M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," Sci. Data , vol. 3, Mar 2016, https://doi.org/10.1038/sdata.2016.18

[17] "Role of Data Stewards and Data Stewardship Community," 4 Jan 2017, https://openworking.wordpress.com/2017/01/04/role-of-data-stewards-and-data-stewardship-community/.

[18] M. Teperek, M. J. Cruz, E. Verbakel, J. Böhmer, and A. Dunning, "Data Stewardship addressing disciplinary data management needs," International Journal of Digital Curation, vol. 13, no. 1, pp. 141–149, Dec 2018, https://doi.org/10.2218/ijdc.v13i1.604.

[19] M. Teperek, A. Dunning, "Data (and code) roles of the future at TU Delft," 25 Jun 2019, https://doi.org/10.5281/zenodo.3256576.

Science and data stewardship,[20,21,22,23] and already recommends considering research software in the track of researchers.[24]

In this paper, we present a concrete follow-up to the DORA declaration in the form of recommendations for raising the profile of research software. These recommendations aim at maintaining leadership in this critical transformation of research. These recommendations and their rationale are based on the Dutch research landscape, but they could be adapted and adopted elsewhere.

Recommendations and rationale

We provide recommendations that funding agencies and research institutions can directly implement; not all recommendations apply to both categories, but they are provided together for simplicity. We divide our recommendations into four broad categories and provide the rationale that motivates the recommendations in each category.

Although research policy is a complicated area, we focus on the four categories defined below. We believe that modest improvements in these areas will have a large impact on the research software landscape, and that they can be achieved with relative ease.

Recommendations

Software availability and quality

- Require that software generated as part of research is archived in compliance with the FAIR principles (as applicable to software[25]), in repositories that facilitate citation, long term preservation and accessibility of the software. This should stimulate software findability and reuse. This recommendation is in alignment with the FAIR software recommendations from https://fair-software.nl/.
- Publish all software produced as part of research projects under an open-source license, unless there are valid (ethical, legal or commercial) reasons not to do so.

[20] "KNAW, NWO and ZonMw to sign DORA declaration," 18 Apr 2019, https://www.nwo.nl/en/news-and-events/news/2019/04/knaw-nwo-and-zonmw-to-sign-dora-declaration.html.

[21] "Evolution or revolution?," Apr 2019, https://publicaties.zonmw.nl/evolution-or-revolution/.

[22] "Netherlands eScience Center," https://www.nwo.nl/en/about-nwo/organisation/nwo-domains/netherlands+escience+center.

[23] "Project F: Professionalisering van opleidingen en training op het gebied van Open Science en Datastewardship," https://www.openscience.nl/en/node/171.

[24] "Notitie – Erkennen en waarderen van onderzoekers", May 2018, https://www.openscience.nl/files/openscience/2019-02/notitie-erkennen-en-waarderen-van-onderzoekers.pdf.

[25] A.-L. Lamprecht et al., "Towards FAIR principles for research software," *Data Science*, vol. 3, no. 1, pp. 37–59, Jun 2020, https://doi.org/10.3233/DS-190026.

- Provide the necessary means to store and share software during the active research phase and provide clear guidelines for the long-term storage and preservation of software. These storage facilities should facilitate as much as possible openness and findability of software.
- Encourage the development of software that adheres to good practices developed by specialist knowledge centers.

Software sustainability

- Implement calls to fund the revitalization, dissemination, long-term maintenance, and further development of existing open-source software important to academic research.
- Stimulate community building and organisation of workshops around important topics such as key software packaging and publishing, computational reproducibility, etc.
- Survey the research software community to identify the needs for maintaining specific software packages.
- Require grant proposals to make a strong case for specific existing software to be used in the proposed research project.
- For grant proposals where new software is proposed to be developed, explain why the existing software is not sufficient either for use or collaborative development, and how generated software will be managed (for example by using software management plans. [26,27]

Training

- Include basic digital competency skills, such as algorithmic thinking and data management, in all core curricula of academic institutions at both undergraduate and graduate level. Digital competence is fundamental, like reading and writing, to all disciplines.
- Provide training opportunities for researchers to develop digital skills, research software best practices, data analysis skills, etc.
- Provide researchers with support and guidance on digital competency. For example, building local training capacity through joining and supporting initiatives like the Carpentries and Code Refinery.
- Stimulate the training of specialists on digital competency skills (Research Software Engineers and Data Stewards).

[26] "Writing and using a software management plan," https://www.software.ac.uk/resources/guides/software-management-plans.

[27] J. Spaaks, J. Maassen, "Netherlands eScience Center Software Sustainability Protocol," 8 Oct 2018 https://doi.org/10.5281/zenodo.1451751.

Human capital

- Stimulate the formalisation of Research Software Engineer (RSE) and Data Steward roles. These roles should have formal job descriptions and have proper career progression paths.
- Stimulate the recognition and rewarding of all research outputs (including datasets and software) in addition to research publications by requiring CVs of RSEs to list software (including software maintenance, contributing to software communities, etc.), alongside other research outputs, when evaluating research grants and for hiring and promotion.

Rationale

Software availability and quality

For research to be truly reproducible, it should be possible to re-run research analyses using the same data and the same software to produce the same results. In principle, all software used to produce results should be preserved.

First and foremost we argue that the software should be available, preferably under an open source licence.[28] Furthermore, software must be developed in such a way that it can be reused in the future. To this end, research organisations should provide (either directly or indirectly) access to storage facilities adequate to making research software findable. This could be, for example, by instructing on the use of GitHub, or by providing an institutional or (supra-)national storage facility (GitLab) which still allows for software to be findable.

Following good software development practices helps with building software that is easier to preserve and/or reuse. Funding agencies should point to guidance for such good software development practices and should encourage projects that develop software to adhere to those practices. These should be sets of recommendations that make sense for individual communities and have been agreed upon by those communities. Funding agencies should work together with specialist knowledge centers (Netherlands eScience Center, Software Sustainability Institute, Research Software Alliance, etc.) to develop such guidance and should facilitate researchers to collaborate with specialist knowledge centers to follow their guidance.

Existing guidance:

- DLR guide
- eScience Center Software Development Guide
- SSI
  - Software Evaluation Guide
  - Online sustainability evaluation

[28] "The Open Source Definition," 22 Mar 2007, https://opensource.org/osd.

- CLARIAH Software quality guidelines
- CESSDA's Software Maturity Model

Software sustainability

Due to its nature, software requires continuous maintenance in order to function properly. With technology constantly evolving — new hardware, new operating systems, new versions of software libraries, etc. — software needs to be updated periodically to ensure that it still operates as expected, and to fix compatibility issues that emerge. Frequently, these changes require very specific knowledge about that particular piece of software. However this type of maintenance is seldom taken into account during the initial development of the software and it is typically funded in an ad-hoc fashion. Not all software developed is worth maintaining in the long term; research communities have the implicit knowledge of which software is more frequently reused and therefore worth the effort of keeping alive and for how long -- some software may need to be sustained for only a couple of years while others might need to be sustained for much longer.

Funding mechanisms for continuous maintenance of relevant software packages are necessary. Efforts for building communities around these software packages should be stimulated, as this has been shown to be an effective way to keep software alive.

For software which cannot be kept alive, proper curation mechanisms should be available, allowing ‘old’ software to be documented and kept accessible for reference purposes. Standard practices of, and assistance with, code deposition and documentation would be helpful.

Training

Programming skills have become a vital part of academic research. Its importance is comparable to that of the skills required in an experimental laboratory (e.g., pipetting or making dilutions) or in fieldwork (e.g. oral interviews). While today programming skills are often concentrated in a few individuals with specialist training, future researchers should acquire these skills early on as part of their training. Practising researchers should have the opportunity to acquire these skills as part of their career development. Digital skills are becoming fundamental to future research in all disciplines, in the same way reading and writing (or mathematics and statistics) currently are. For these skills to become more commonplace, it is important to stimulate knowledge and skills exchange activities (study groups/hacky hours etc.) inside and outside the institutions.

Human capital

The Netherlands has been one of the leading European countries in training, hiring, and embedding Data Stewards in research and academic institutions. Data Stewards are indispensable in supporting researchers in activities, which has resulted in the establishment of data stewardship programmes, such as the one at TU Delft. As put in a recent report,[29] “professional data stewardship needs professional data stewards”. There are ongoing efforts[30,31,32] to define data steward profiles within the UFO, which is the job profile function ordering system used at Dutch universities.[33]

Considering that data and software are intrinsically linked[34] and that software is fundamental to research, having data skills and competencies alone is not enough. Research institutes need Research Software Engineers (RSEs)[35] in addition to Data Stewards. RSEs “closely collaborate with researchers to understand the challenges they face, and then develop research software to provide the answers”,[36] “combine expertise in programming with an intricate understanding of research”[37] and “apply the skills and practices of software development to research to create more robust, manageable, and sustainable research software”.[38] Together, RSEs and Data Stewards have an important role in ensuring research reproducibility.

Analogous to the efforts to define data steward job profiles, there is a need to define RSE job profiles. Professional RSE function profiles will also make RSE jobs more attractive, helping to meet the growing demand for such people.

DORA puts an emphasis on changing how research outputs are evaluated. This includes the recognition of non-traditional research outputs such as research data and research software. A recent report by VSNU[39] highlights the need for a change in the academic recognition and reward system; the same report proposes the redesign of academic career paths, and modernisation of quality assessment of research and research proposals.

[29] I. Verheul, M. Imming, J. Ringerma, A. Mordant, J.-L. van der Ploeg, and M. Pronk, “Data Stewardship on the map: A study of tasks and roles in Dutch research institutes,” 2019. https://doi.org/10.5281/zenodo.2669149

[30] “Towards a community-endorsed data steward profession ... - Zenodo,” 25 Jan 2019, https://zenodo.org/communities/nl-ds-pd-ls/.

[31] S. Scholtens et al., “Life sciences data steward function matrix,” 11 Feb 2019, https://doi.org/10.5281/zenodo.2561722.

[32] S. Scholtens et al., “Towards a community-endorsed data steward profession description for life science research - Poster: Function Matrix,” 29 Mar 2019, https://doi.org/10.5281/zenodo.2616476.

[33] “Functie_ordeningsystem_UFO - VSNU,” https://www.vsnu.nl/functie_ordeningsystem_ufo.html.

[34] P. K. Doorn and P. Aerts, “A Conceptual Approach To Data Stewardship and Software Sustainability: Scientists in charge, with a little help from their friends,” Data Archiving and Networked Services (DANS), 2016. http://hdl.handle.net/20.500.11755/59c24848-9cf7-437c-b2d5-e943e9e4a35e

[35] M. Teperek and A. Dunning, “Data (and code) roles of the future at TU Delft,” 25 Jun 2019, https://doi.org/10.5281/zenodo.3256575.

[36] “What is a Research Software Engineer?,” 13 Jun 2017, http://nl-rse.org/2017/06/13/what-is-rse.html.

[37] “What is an RSE?,” https://rse.ac.uk/what-is-an-rse/.

[38] “What is an RSE?,” https://us-rse.org/what-is-an-rse/.

[39] “Room for everyone’s talent,” Nov 2019, https://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Position%20paper%20Room%20for%20everyone%E2%80%99s%20talent.pdf.

This is of fundamental importance to RSE roles, as one of their core activities is developing new software and contributing to existing software and they would not usually produce as many journal publications. Involvement in open software development projects should be considered as a core activity of RSEs and recognized accordingly. Research software related activities should be recognized and considered as equally important for hiring and promotion committees. In accordance with the VSNU repost, RSE roles should be part of the redesign of academic career paths, and recognition of research software should be considered as an important component of quality assessment of research.

Final remarks

Software has become a crucial part of research, but it still does not receive the same recognition as other research outputs. There is a need to fully acknowledge that research software is as important as research data and scientific publications, as expressed in DORA. We share the Research Software Alliance's vision that research software should be "recognised and valued as a fundamental and vital component of research."[40]

We have provided recommendations that funding agencies and research institutions in the Netherlands can implement to achieve this goal. These recommendations do not offer a full solution to addressing this issue, but provide a further step in the direction of achieving recognition for research software as a fundamental and vital component of research. These recommendations are focused on the Netherlands, but the issues and recommendations could be adapted and adopted by other countries.

We believe that even minor policy improvements in the domain of research software will lead to visible improvement in academia. Therefore, we urge funding and research institutes to address and adopt at least some of our suggestions.

Acknowledgements

This paper follows on from a meeting the authors had with representatives from NWO (The Netherlands Organisation for Scientific Research) on 28 March 2019.[41] Our aim in that meeting was to highlight the importance of research software in contemporary research and its relationship to research data, open science, and reproducibility in research. We thank NWO for receiving us on 28 March 2019 and for the discussions we had then, which led to the creation of this paper.

[40] "Research Software Alliance," https://www.researchsoft.org/hestia-front/.
[41] A. Akhmerov et al., "Making Research Software a First-Class Citizen in Research," 19 Apr 2019, http://doi.org/10.5281/zenodo.2647436

The authors are also aware of and pleased with the many ongoing efforts in the direction of these recommendations since these were originally written. We encourage funding agencies and research institutes to continue working in this direction and are glad to contribute to the continuous improvement of the research landscape in the Netherlands and beyond.