

## Evidence for the importance of research software

By [Michelle Barker](#), [Daniel S. Katz](#), [Alejandra Gonzalez-Beltran](#)

This blog analyses work evidencing the importance of research software to research outcomes, to enable the research software community to find useful evidence to share with key influencers. This analysis considers papers relating to meta-research, policy, community, education and training, research breakthroughs and specific software.

The [Research Software Alliance](#) (ReSA) [Taskforce for the importance of research software](#) was formed initially to bring together existing evidence showing the importance of research software in the research process. This kind of information is critical to achieving ReSA's vision to have research software recognised and valued as a fundamental and vital component of research worldwide.

### Methodology

The Taskforce has utilised crowdsourcing to locate resources in line with ReSA's mission to bring research software communities together to collaborate on the advancement of research software. The Taskforce invited [ReSA google group members](#) in late 2019 to submit evidence about the importance of software in research to a [Zotero group library](#). Evidence could be from a wide range of sources including newspapers, blogs, peer-reviewed journals, policy documents and datasets.

The submissions to Zotero submitted to date highlight the significant role that software plays in research. We decided to tag the submissions based on analysis of the submissions and how [some community research software organisations](#) categorise focus areas. Some of the documents were also tagged by country and/or research discipline to enable users to search from that perspective were relevant. This resulted in tagging of submissions with one or more of the following tags, which are explained in the following sections:

- Meta-research
- Policy
- Community
- Education and training
- Research breakthroughs
- Software

### Submission contents

Explanation of each category, and examples of some of the resources, are highlighted below. It should be noted that some of the resources mentioned below are important in a number of the categories; however, have only been cited here in one category.

**Meta-research** contains resources that include analysis of how software is developed and used in research:

- [Charting the digital transformation of science: Findings from the 2018 OECD International Survey of Scientific Authors \(ISSA2\)](#) includes evidence that 25% of research produces new code.
- [The Ecosystem of Technologies for Social Science Research](#) tracks increase in the use of software tools, along with characteristics of key tools. It is noted that whilst many commercial tools are available, the more innovative ones are coming out of academia.
- [The top 100 papers](#) analyses the top 100 *Nature* papers and finds that the vast majority describe experimental methods or software that have become essential in their fields.
- [UK Research Software Survey](#) considers responses of 1,000 randomly chosen researchers to show that more than 90% of researchers acknowledged software as being important for their own research, and about 70% of researchers said that their research would not be possible without software.
- [Understanding Software in Research: Initial Results from Examining Nature and a Call for Collaboration](#) reveals that “32 of the 40 papers examined mention software, and the 32 papers contain 211 mentions of distinct software, for an average of 6.5 mentions per paper.”

**Policy** is used to tag resources that focus on policy related to software, including the need for increased policy focus in this area.

- [An environment for sustainable research software in Germany and beyond: current state, open challenges, and call for action](#) examines how German funding bodies are increasingly acknowledging the importance and value of sustainable research software and related infrastructures, and makes recommendations on how to further improve this.
- [Community Organizations: Changing the Culture in Which Research Software Is Developed and Sustained](#) found that the US National Science Foundation made 18,592 awards totaling \$9.6 billion to projects that mentioned “software” in their abstracts between 1995-2016, which suggests that government funding policy needs to recognise the critical nature of research software.
- [RDA COVID-19 Working Group Recommendations and Guidelines](#) contains recommendations for policy makers, funders, publishers and research community members working to overcome the challenges of COVID-19.
- [Recognising the importance of software in research: Research Software Engineers \(RSEs\), a UK example - Study](#) is a case study that discusses the current challenges faced by RSEs and policy conclusions to further help support RSEs and to contribute to the progress of open science in Europe.
- [UK’s research and innovation infrastructure: opportunities to grow our capability](#) highlights the importance of software by recognising software and skills as one of the six computational and e-infrastructure themes.

**Community** considers work on the importance of research software communities in ensuring best practice in software development.

- [Computational Research Software: Challenges and Community Organizations Working for Culture Change](#) identifies the importance of sustained community in the development of high-quality software, and introduces efforts by grassroots organisations and projects to improve software quality, productivity, and sustainability. These endeavors ensure the integrity of research results and enable more effective collaboration.
- [Community Organizations: Changing the Culture in Which Research Software Is Developed and Sustained](#) provides an overview of the grass-roots organisations and projects that have evolved to address growing technical and social challenges in research software productivity, quality, reproducibility, and sustainability. This article then discusses opportunities to leverage their synergistic activities while nurturing work toward emerging software ecosystems.

**Education and training** identifies work that considers issues round skills, training, career paths and reward structures.

- [How do scientists develop scientific software? An external replication](#) considers how scientists acquire software engineering knowledge to suggest improvements in this process.
- [Software Use in Astronomy: an Informal Survey](#) finds that all participants use software in their research, and identifies the ten most popular tools. 90% of participants write at least some of their own software. while only 8% of them report that they have received substantial training in software development.
- [Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators](#) highlights the importance of skills in software development for data analysis, with 90% of respondents indicating they are currently or will soon be analyzing large data sets.

**Research breakthroughs** include papers on significant research accomplishments that acknowledge reliance on research software tools:

- [Case Study: First Photograph of a Black Hole, Enabled by NumFOCUS Tools.](#)
- [Pegasus powers LIGO gravitational wave detection analysis.](#)
- [Software framework designed to accelerate drug discovery wins IEEE International Scalable Computing Challenge.](#)

**Software** has been applied as a tag to resources that mention particular pieces of software, including when this is part of a broader focus in the resource. This category includes a sample of the tens of thousands of papers that rely on research software and that collectively build knowledge in a field, for example:

- [Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture.](#)
- [Challenges in funding and developing genomic software: roots and remedies.](#)
- [Researchers find bug in Python script may have affected hundreds of studies.](#)
- [Securing the future of research computing in the biosciences.](#)

- [These 7 CERN Spinoffs Show The Project Isn't Just Theoretical.](#)

### Other useful approaches

This analysis has been useful in elucidating some of the ways in which the value of software can be demonstrated. However, there are also other approaches that could be useful. For example, methods to evaluate economic value are providing valuable statistics for research data, but comparable examples for research software are rare. The recent European Union publication, [Cost-benefit analysis for FAIR research data](#), finds that the overall cost to the European economy of not having Findable, Accessible, Interoperable and Reusable (FAIR) research data is €10.2bn per year in Europe. A 2014 Australian study by [Houghton and Gruen](#) similarly demonstrates the economic value of data, estimating the value of Australian public research data at over \$1.9 billion a year. One of the few economic valuations of research software is a 2017 analysis by [Sweeny et al.](#) of the return on investment generated by three Australian virtual laboratories, which provide access to research software and data for researchers, was at least double the investment for every measure. This indicated that the services had a significant economic and user impact - by one measure the value of one virtual laboratory was over 100 times the cost of investment. It would be useful if more studies were undertaken to demonstrate the economic benefits of research software using different methodologies.

### Conclusion

This summary of evidence of the importance of research software to research outcomes illustrates increasing recognition of this fact. This summary could also be useful to encourage the community to consider where additional work could be useful (such as expanding existing surveys in specific countries and disciplines to get a more global scope), and to inspire the recording of more of this information (which could include striking examples that convey the impact of failing to understanding the costs and responsibilities of thoughtful software management).

We encourage readers to submit additional resources to the ReSA resources list, which is publicly available:

- Add it directly to the ReSA [Zotero group library](#) (requires Zotero account).
- Submit an issue in [GitHub](#) (requires GitHub account).
- [Email](#) it directly to ReSA.