Better Software, Better Research: Providing Scalable Support for Scientific Software Development

Neil Chue Hong¹ (N.Chue Hong@software.ac.uk), Mario Antoniolletti¹, Les Carr⁴, Steve Crouch⁴, David De Roure³, Iain Emsley², Carole Goble⁶, Simon Hettrick³, Devasena Inupakutika⁴, Mike Jackson¹, Mark Parsons¹, Aleksandra Pawlik², Giacomo Peru², Olivier Philippe⁶, Arno Proeme¹, John Robinson⁷, Shoaib Su⁷

1. Software Sustainability Institute, EPCC, University of Edinburgh, JCMB, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, UK
2. Software Sustainability Institute, School of Computer Science, University of Manchester, Kilburn Building, Manchester, M13 9PL, UK
3. Software Sustainability Institute, Oxford e-Research Centre, University of Oxford, Keble Road, Oxford, OX1 3QG, UK
4. Software Sustainability Institute, School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK

ABSTRACT
Over the last five years, the Software Sustainability Institute (http://www.software.ac.uk) has refined a programme of activities to support better scientific software development in a scalable and cost effective way. We present a summary of these activities and the context in which we operate to inform the CSESSP workshop.

1. Introduction
Modern research is impossible without software. From short, throw-together temporary scripts to solve a specific problem, through an abundance of complex spreadsheets analysing collected data, to the hundreds of software engineers and millions of lines of code behind international efforts such as the Large Hadron Collider and the Square Kilometre Array, there are few research areas where software does not have a fundamental role.

![Figure 1: Results of UK Research Software survey (2014)](image)

In 2013-2014, the UK research councils invested £840m in research grants for work that directly relies on software [1], representing 30% of the total investment in research by RCUK and a figure which has risen by 3% on average over each of the last 4 years. In a 2014 survey (see Figure 1) of 15 Russell Group universities conducted by the Institute [2], 92% of researchers indicated they used research software (which we define as software that is used to generate, process or analyse results for publication). Of these, 68% said that their research would be impossible without software, representing 170,000 researchers in the UK who could not conduct their research without software.

2. The UK Research Software Ecosystem
We characterise the UK research software ecosystem as follows:

• **Research software is ubiquitous.** The “traditional” users of computational infrastructure (e.g. simulations and computational methods) have long histories of relying on software. But “Big Data” and Data Science are also only possible with software to access, analyse, visualise, send and store that data. Software use is not restricted to the hard sciences: the use of research software is across all disciplines. The capacity of all researchers, in the UK and worldwide, to generate new insights depends on the availability of research software and the ability of researchers to use it.

• **Research software is widely developed.** Researchers do not just use “off the shelf” software, the majority of researchers develop their own. A related challenge that we seek to address is that - although there is no significant difference in the use of software by men and women, men are significantly more likely to develop software than women (70% of men relative to 30% of women).

• **Research software training is underprovided.** The lack of suitable computational training is a problem that has previously been identified by the UK Research Councils, most recently by the MRC and BBSRC (“Informatics skills are applicable to many areas of both the biosciences and the medical sciences”; “Data analytics appear to be particularly vulnerable.”) [3]. The Institute’s own surveys supports this view, with 71% of researchers reporting that they have no formal training in software development.

• **Funding for software development is largely hidden.** The UK has made and continues to make massive investments into software, and is seen as a leader in the development of new software, for example: UK e-Science programme (£213m); Collaborative Computational Projects (£9.2m in latest round); EPSRC: Digital Economy programme (£150m), Software for the Future (£11m); BBSRC: Tools and Resources Development Fund (£27m), Bioinformatics and Biological Resources Fund (£30m); ESRC: Digital Social Research (£10m), Big Data (£64m); InnovateUK: Digital Economy (£42m), ICT (£2.4m). Although there is significant investment for developing new software, there is too little in maintaining and re-using software and much is “hidden” within research grants; to quote from [4] “Software is the infrastructure, and hardware the consumable…. A lack of investment in software maintenance does not allow best use to be made of the UK’s existing scientific infrastructure.”
3. Providing scalable support for software

The Software Sustainability Institute supports better research by helping researchers to build and use better software. We do this by providing a national focal point of information and resources to support the best practice of research software development.

The key objectives of the Institute are:

- Delivering skills: providing essential software skills to researchers in research institutions, CDTs and doctoral schools, enabling researchers to support themselves and their communities;
- Understanding challenges: working with research communities to understand topical challenges, and bring together the people needed to solve them;
- Supporting communities: to develop software that meets the needs of reliable, reproducible and reusable research;
- Communicating requirements: Research the software requirements of communities and communicate them to research stakeholders;
- Providing authority: Act as a central hub of representation, coordination and dissemination of best practice for research software, both nationally and internationally.

To deliver on these objectives, the Institute’s work is split into five teams which operate at increasing scale up to tens of thousands of researchers:

- Research Software: working directly with researchers who are developing software through our Open Call to ensure it meets the needs of reliable, reproducible and reusable research;
- Training: coordinating, defining, and delivering training on software development and data science skills to UK research organisations, and working to build a sustained training platform;
- Community: bringing people together via events, networks and our Fellowship programme to identify, understand and facilitate solutions for common challenges;
- Policy: research into the social, economic and technical drivers of the research software community, understanding needs, and working to enact required changes through campaigns;
- Communications: ensuring the work of the Institute is disseminated to the widest possible audience (220,000 visitors to our website, 20,000 regular readers of our blog), and working with collaborators to amplify the impact of our work.

These teams work together to deliver on the key objectives, resulting in the following outcomes:

- A skilled researcher base in the UK: through training events, guides and recruitment of Fellows.
- Recognition of the importance of software to research: through running workshops, facilitating working groups, and working with the scholarly communication community.
- Professionalisation of the research software area: through supporting the RSE community, creating a network for heads of RSE groups, and developing policy and guidance.
- Increased scientific integrity: through developing and piloting a research software accreditation framework and extending existing Data Management planning tools and guidance to cover research software.
- Protection of investment: working with stakeholders to develop and implement policies to support software sustainability in the research they fund to ensure that productivity, quality, cost effectiveness, and reusability are maximised.

A key factor in our ability to deliver authoritative advice to different communities is through our highly successful and very competitive (10+ applications for each place) Fellowship Programme. We have built and sustained an active and growing network of 60+ Fellows, including 9 from biosciences, 6 from social science, 8 from environmental and geosciences, 12 from mathematics and computing, and 22 from engineering and physical sciences. The Fellowship provides discipline-focused leverage for our work, enabling our activities to be transferred into different domains. The Fellow’s reach and credence as ambassadors in their research communities promotes software sustainability in discipline specific activities. Successes of the Fellowship programme include: having successfully organised 10+ domain workshops relating to software techniques in the social sciences [5], using, improving and combining models [6][7], working with clinicians [8], and creating new software and documentation for scientists [9]; 6 Fellows becoming accredited Software Carpentry instructors; contributing 52 articles to our blog; and attending 68 domain-specific events to determine community requirements and engage in dialogue [10].

4. Conclusions

The Software Sustainability Institute has had five years of experience creating a model for providing scalable support for scientific software development. Its success relies on the interlinking work of its teams, and its connection to research communities through its Fellowship programme, community events, guides, blog and website.

5. Acknowledgements

The Software Sustainability Institute is supported by EPSRC grant EP/H043160/1 and grant EP/N006410/1 from the EPSRC, BBSRC and ESRC.

6. References

[8] Clinstat: http://www.ucl.ac.uk/clinstat