

April
2014

Measuring the success of the Software Sustainability Institute

The Software Sustainability Institute was founded to cultivate world-class research with software. To measure progress towards this goal, this report investigates the work of the Institute over the period 1 June 2010 to 31 March 2014. Success is measured across four broad objectives, as evidenced by metrics collected over the lifetime of the Institute.



About the Institute

The Software Sustainability Institute cultivates world-class research with software.

We help people build better software by working with researchers, developers, funders and infrastructure providers to identify key issues and best practice. The Institute is funded by the Engineering and Physical Sciences Research Council through grant EP/H043160/1.

Some of the work in this report has been supported by additional funding from the EPSRC, BBSRC, Jisc, NSF and the Sloan Foundation.

For more information, visit www.software.ac.uk.

About this report

This report was written by Simon Hettrick and Neil Chue Hong, based on work conducted at the Institute by themselves and Malcolm Atkinson, Mario Antonioletti, Rob Baxter, Les Carr, Stephen Crouch, David De Roure, Carole Goble, Alexander Hay, Duncan Hull, Malcolm Illingworth, Devasena Inupakutika, Mike Jackson, Mark Parsons, Aleksandra Pawlik, Tim Parkinson, James Perry, Michal Piotrowski, Arno Proeme, Ruyman Reyes, John Robinson, Terry Sloan, Dominic Sloan-Murphy and Shoaib Sufi.



This work by the University of Edinburgh on behalf of the Software Sustainability Institute is licensed under a Creative Commons Attribution 2.5 UK: Scotland Licence.

Objectives



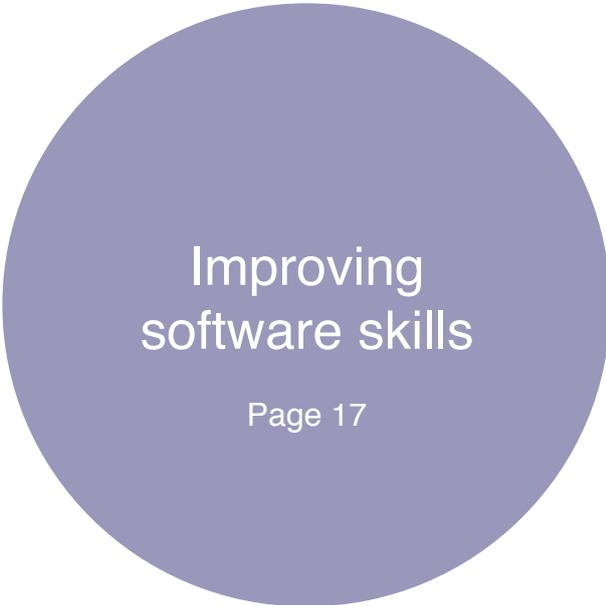
Getting software
on the research
agenda

Page 9



Supporting
communities that
want change

Page 11



Improving
software skills

Page 17



Improving
software

Page 21

Contents

About the Institute	2
About this report	2
Objectives	3
Executive summary	5
Scope and organisation of the Institute’s themes	5
Replicating the success metrics	6
Objectives	7
Getting software on the research agenda	9
Build a platform	9
Engage the research community	10
Achieve authoritative reputation	10
Supporting communities that want change	13
Identifying partners across disciplines and organisations	13
Engage high-value communities	14
Leveraging the community	14
Improving Software Skills	17
In-person training	17
Self-led training	17
Driving change	18
Improving Software	21
Consultancy	21
Testimonials and quantitative results	21
Lessons for the future	23

Executive summary

No-one can predict what the next breakthrough in research software will be. In just a few decades, software has become one of the foundations that underpin much of research in the UK, and internationally. Software not only supports the traditional areas of simulation and instrument control, but also data analysis. New research is being built independently on top of existing results. Tools and libraries outlive the lifespan of the hardware and projects that initiated them, making the job of sustaining this software harder to plan and perform. Nevertheless, we believe that we can strive to make the development, maintenance and use of research software better, and as an Institute we have invested hundreds of man months towards this cause.

Measuring the success of any project is challenging. Whilst it is easy to identify outputs, it is difficult to identify the outcomes derived from the application of these outputs - let alone measure the benefits that result from the outcomes. However, if this is not done, it is impossible to say if the project is delivering what it set out to do.

The Software Sustainability Institute has been running for four years, and over that time it has delivered a number of individual successes, many of which are detailed in this report. What this report also shows is how the interaction between the different themes of the Institute helps to deliver on its objectives: getting software on the research agenda, supporting communities that want change, increasing skills, and improving software.

Drawing on evidence, analysis and responses from collaborators, the different metrics show how the Institute has worked to deliver beneficial change to the UK research community. They also show how the Institute's priorities have evolved from its inception: after concentrating on building a solid platform, work in training and policy were added to the existing consultancy and community work. Above all, they show the breadth of interactions that the Institute necessarily instigates from one-to-one meetings to guides with tens of thousands of readers we've never met.

There is still work to be done. The evidence gathering required to produce this report currently consumes a large amount of effort. Sometimes this effort can be reduced by collecting information earlier in the process, in other cases we simply have to assume the burden in order to understand the magnitude of our impact. Particularly for our training and consultancy initiatives, we need to understand both the satisfaction levels of the people we work with, and the return on investment of each interaction, to allow us to prioritise the areas in which we work.

Going forward, this report shows both the success of the Institute and the requirement for continued work in this area to support change in the research community. With continued investment in this area, the UK can remain at the cutting-edge of international science and be flagbearers for the development, maintenance and use of research software.

Scope and organisation of the Institute's themes

In order to meet the Institute's broad remit, its work is separated into four *themes*: Community, Consultancy, Policy and Training. To understand the work of the Institute, it is necessary to know something about the scope of its themes.

The Community theme brings people together into networks, which makes it easier for the Institute to gather intelligence from the research community. This theme is responsible for the

organisation of the *Collaborations Workshop*, the Institute's annual conference, and the *Fellowship programme*, which grants researchers a year-long bursary in exchange for their knowledge and expertise.

The Consultancy theme directly addresses problems with research software by providing researchers with expertise and effort from the Institute's software developers. This theme is responsible for the *Open Call*, which is a regular competition that allows researchers to bid for time with the Institute's software developers.

The Policy theme identifies the major issues that affect the research community and runs campaigns to both raise awareness of the issues and determine solutions to them. This theme is responsible for the campaign for *Research Software Engineers*.

The Training theme identifies the skills and knowledge that researchers need to benefit from research software, and provide this training in person or by creating training materials. This theme is responsible for *Software Carpentry* in the UK.

The four themes are merely a way of breaking down a broad remit so that can be more easily managed. The Institute's staff work across all themes to ensure that ideas and information from one theme are fed into the others.

Replicating the success metrics

The data represented in this report are available online: bit.ly/SuccessMetricData.

Objectives

The Institute has delivered a number of achievements over the last four years: from being the key figure in the rise of Software Carpentry in the UK, to adding *Research Software Engineer* to the lexicon of the research community.

Taken individually, none of the achievements indicate a significant shift in the way that researchers perceive or deal with research software. However, the combined weight of the achievements has progressed four objectives which we believe are fundamental to cultivating world-class research with software.

1. Getting software on the research agenda

There can be no progress with any of the objectives unless the research community is aware of the importance of research software and the issues that restrict it.

2. Supporting communities that want change

By working with the communities who most want change, we can harness useful allies and significantly increase the impact of the Institute's limited resources.

3. Increasing skills

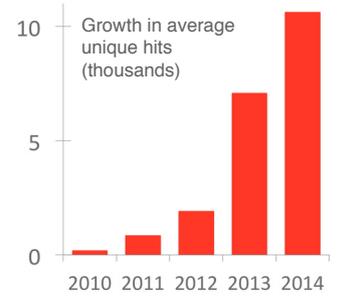
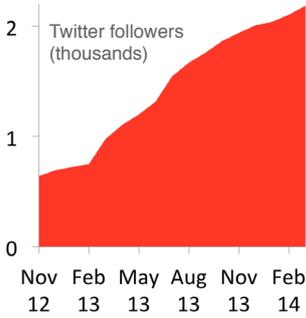
Ultimately, it is up to researchers to ensure that their software is of suitable quality for their research, so it is of fundamental importance to provide the skills researchers need.

4. Improving software

The Institute cannot work with all research software, but we can provide our expertise and effort to a select group who will become examples of best practice for the rest of the community.

Each of these objectives is discussed in a separate section of this report, and graphically represented by its own success metric diagram.

Getting software on the research agenda



Social media

Website

Build a platform

46%

Percentage of all traffic that lands on the blog

1341

since Nov 2012

Retweets

Average unique visits in 2014

10,625 per month

Engage the research community

155

since June 2010

Blog posts written by external authors

Conferences attended by staff

76

17

since June 2010

Workshops and events organised

Representation on boards and steering committees

14

Gain authority

38

Completed software projects

Case studies

21

3

Keynotes in 2013

Papers published

10

Getting software on the research agenda

Before we could challenge the problems that affect research software, we needed to make researchers aware of the important role that software plays in modern research: we had to get the topic of software on the research agenda.

We have built a huge audience around our website - over 140,000 unique visitors since June 2010 - and have maintained a significant growth in popularity every year since 2010. We have also created a vibrant and responsive Twitter community. This community has engaged with us through our social media and through the 17 events we have organised leading to, among other successes, 38 successful consultancy projects, a blog where almost 50% of the posts are written by external authors, invitations to present keynotes and our advisory role for the NSF's sustainable software plans.

Build a platform

Before the Institute was founded, there was no single group that represented researchers who use software. Isolated groups of researchers found it difficult to voice their concerns or benefit from the combined experience of the - at the time, completely unformed - research software community. For this reason, we set about creating a communications platform around which a research software community could congregate. Since its creation in June 2010, the Institute's website has been visited by 142,737 unique visitors, and has experienced an average annual growth of 295%, which led to an average of just over 10,000 unique visitors each month in 2014. The majority of web traffic originates in the UK (37%) and the US (21%), a significant contribution derives from India (5%), Canada (3%), Germany (3%) and Australia (2%) and the remainder is split fairly evenly over other countries. The popularity of the website proves the existence of a vibrant research software community, and represents a significant success in harnessing it.

We split our communications platform into two parts: the blog, which was designed to bring in visitors, and the remainder of the website, which would educate those visitors about software sustainability.

Rather than focussing exclusively on the work of the Institute, the blog publishes interesting news from across the field of research and software. With 46% of all traffic (that is, the 350,000 page views since June 2010) landing on the blog, we believe that this strategy has successfully drawn in visitors that might not have been easily enticed by more standard website content, such as guides.

The remainder of the website is made up of the guides, software evaluations, case studies and information about the Institute. Although this may be seen as the less exciting part of the website, it brings in the majority of visitors: 54% of all traffic.

Although the Institute created a Twitter account in 2010, we only began to focus on using Twitter to draw in visitors in late 2012. This saw our follower numbers leap from around 200 in August 2012 to around 2200 in March 2014. Twitter is sometimes viewed as a frivolity, but we found that increasing our Twitter followers from 200 to 1000, combined with tweeting about all new content, caused website traffic to double. Twitter is now the second most popular referring agency for our website (the first is Reddit), which justifies the effort that has been expended on increasing the Twitter community.

Engage the research community

If software sustainability is to gain traction, we must engage the research community that has been created around our communications platform. This will allow us to persuade the community to adopt sustainable practices, to entice them to provide intelligence from their domain and - importantly for a small organisation - to benefit from contributed effort.

A large part of our community engagement comes from the Fellowship and organising workshops, such as our annual Collaborations Workshop. In total, 17 workshops and events have been run by the Institute since June 2010, which has provided us with direct engagement with over 700 attendees. This face-to-face interaction is invaluable for establishing the message of software sustainability, which can then be carried back to the attendees' communities.

The blog has proven useful for engaging the research community, simply because we have made a concerted effort to secure authors from outside the Institute. In fact, over half of the posts (155 of the 300 posts since 2010) are written by external authors, which cements the position of the blog as the voice piece for the research software community, and provides us with community-wide authority - rather than simply representing the views of Institute staff.

Our Twitter following has become an excellent way to engage the community. Since September 2013, 19 authors have offered their services in response to Tweets for help, and this has led to the creation of 8 new guides. What's more, a number of questions posed by our followers, have subsequently been answered by other followers. In the experience of the Institute staff, this level of engagement is rare for a research organisation. A responsive community can only grow around an organisation if it feels genuine engagement with the goals of that organisation, which indicates that we have successfully communicated the importance of software sustainability to our community.

In addition to the engagement discussed above, we have also engaged the research community in conventional ways by attending around 80 conferences and by using our position on 14 boards and steering committees to represent the needs of the research software community.

Achieve authoritative reputation

If we are to provide confidence in the advice that we provide, we must develop a authoritative reputation.

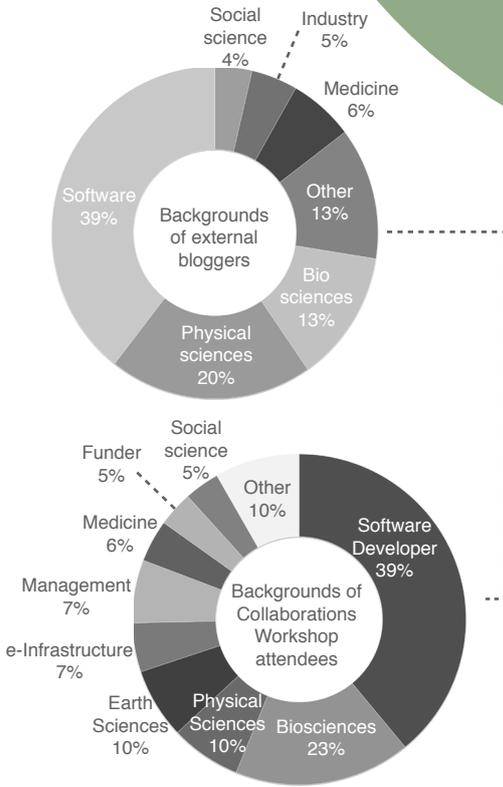
We gain authority by showing that the Institute has expertise in a range of subjects that are important to research software. The Consultancy theme provides this expertise, and by successfully completing 38 projects since 2010, we have displayed expertise in fields as diverse as medical imaging and particle physics. This subject will be discussed in more depth on page 21 which discusses the objective *Improving Software*, but for the time being it is important to note that we have completed these projects not just to improve software, but to give the Institute an authoritative reputation on the subject.

The work of our consultants is shared through 21 case studies, which have been viewed by 2600 unique visitors since June 2010. It is to be expected that considerably fewer people will view the case studies than the more popular blog, but encouraging so many visitors to show interest in the actual mechanics of the work we have completed is a great success.

When working with the research community it is important to gain authority using established

means. The Institute has benefitted from the reputation of the Director and Co-Investigators who were invited to present 3 keynote addresses in 2013 alone. This has led to international attention for the Institute. Notably, we have played a significant part as a co-organiser of the NSF and Sloan Foundation supported *Working towards Sustainable Software for Science: Practice and Experiences* (WSSSPE), which will inform the NSF, and projects from around the world, in their approach to sustainable software. The Institute is not a research organisation and as such it has not focussed on writing papers. Regardless of this lack of focus on papers, the Institute has authored 10 papers in isolation or with partners from consultancy projects.

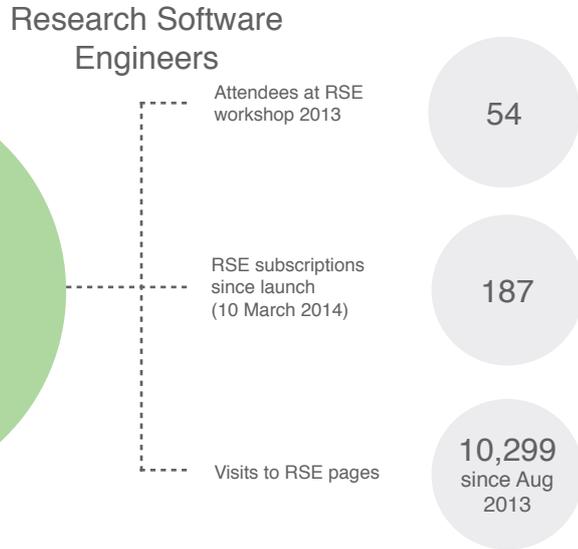
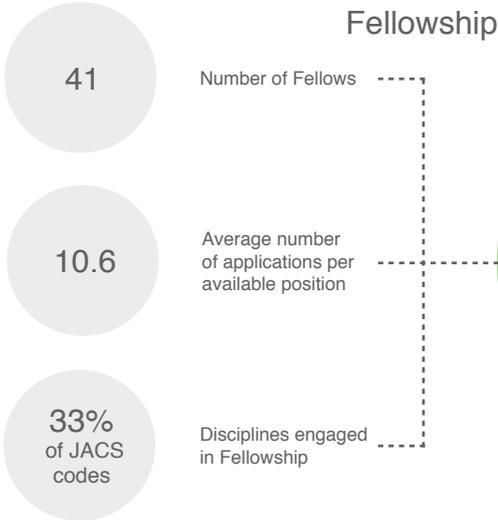
Supporting communities that want change



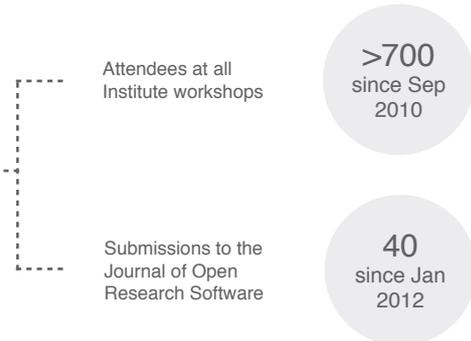
Identify partners across disciplines and organisations



Engage high-value communities



Leveraging the community



Supporting communities that want change

The Institute's objectives cannot be achieved without the support of partners. We have used various methods to identify people and groups whose goals align with the Institute, and have partnered with them to provide support and expertise.

Identifying partners across disciplines is a daunting task, but the creation of a Fellowship programme and a novel annual workshop has allowed the Institute to access representatives from across the UK and from all major disciplines. By focussing engagement on two high-value groups, the Institute has leveraged volunteer effort. This led to 10 workshops taking place in 2013 without effort from the Institute, and direct access to the people who develop much of the research community's software.

Identifying partners across disciplines and organisations

The Institute has worked with 52 organisations over the last four years, either as part of our consultancy work, through the Fellowship programme or as part of our lobbying. The Institute has benefitted from working with organisations across the length and breadth of the UK, which provides a geographical spread that is necessary for a national organisation to achieve authority.

It is difficult to identify communities across disciplines without the help of partners who are embedded in those disciplines, so two major programmes were initiated to engage cross-disciplinary partners: the Fellowship programme and the Collaborations Workshop.

The Fellowship has provided the Institute with representatives from a third of all the top-level disciplines as represented by the JACs classification (used by the Higher Education Authority to classify research disciplines). Using representatives from within disciplines to promote new ideas - such as software sustainability - is a good way of overcoming scepticism to them.

The Collaborations Workshop has successfully attracted a wide range of researchers. As would be expected, the major share of the attendees come from a software background (39%), the remainder come from different research backgrounds and other positions of importance to research, such as management, publishing and funding. The breakdown of the attendees' backgrounds provides insight into the disciplines that are most engaged with software sustainability. The biosciences account for the largest share (23%). Next are the physical sciences (10%) and the Earth sciences (10%). These three disciplines are well known to be ahead in promoting the use of software within their domain, so it is heartening to see them represented at the Collaborations Workshop.

The Institute's blog has also provided a way of identifying new communities by eliciting views from authors within these communities. Looking at the backgrounds of the external bloggers shows a preponderance for authors from a software background (39%), and the physical sciences and biosciences are again well represented with 20% and 13% of the articles authored from those disciplines respectively.

Engage high-value communities

Although a number of partnerships have been formed between the Institute and other groups, two of these communities have proven especially important to the Institute's goals: the Fellows and the Research Software Engineers.

The original goal of the Fellowship programme was simply to gain intelligence from a cross-disciplinary group of researchers, the real success of the Fellowship is that it has provided much more than just intelligence. The popularity of the programme has led to a high number of applications (around 10 on average for each place available), which has meant that the Institute has recruited a group of exceptional researchers as Fellows. Of their own volition, the Fellows have taken part in Institute events, run their own events, represented the Institute at their home institutions and created guides and blog posts for the Institute's website, and this interest has continued into the Fellows alumni years. The Fellowship is a successful community programme, which has created a large and active group of researchers who feel motivated by the goals of the Institute to continue to contribute even after their Fellowship funding has come to an end.

The concept of the Research Software Engineer - someone who is engaged in research but writes code, not papers - first came to prominence at the Collaborations Workshop 2012. It has since become a focus for the Policy theme: a good example of one Institute theme providing intelligence for another. Research Software Engineers produce much of the software that is used by researchers, so by creating a community to represent them, the Institute has direct access to the people behind research software and a way to influence the software that is produced.

Leveraging the community

The success of the Fellowship programme is well represented by a workshop at the American Geophysical Union (AGU) 2013 - a conference that attracts over 20,000 attendees. After meeting at a Collaborations Workshop, four of our Fellows applied to run a workshop at the AGU on *the role software plays in innovation, code sharing, ways this can be productive and rewarding*, and best practice in general. This was well received and attracted around 50 attendees from within a new community for the Institute. That the Fellows came together of their own volition and organised a workshop at a prestigious conference on the subject that is the focus on the Institute's work is compelling evidence for the success of the programme.

Although the AGU workshop is the prime example of the Fellows working together to publicise issues about research software, it is not the only one. In 2013, 10 events were organised by the Fellows and there are many examples of the Fellows keeping in contact with their colleagues in the network - a poll of the 2012 and 2013 Fellows reported that 8 out of 15 keep in contact. The impact of the Fellowship is best described by one of the Fellows "It never occurred to me to think about sustainability until I saw your call for Fellows. It's dramatically changed the way I approach my code and my data. It's also rubbing off on [other people in] my field."

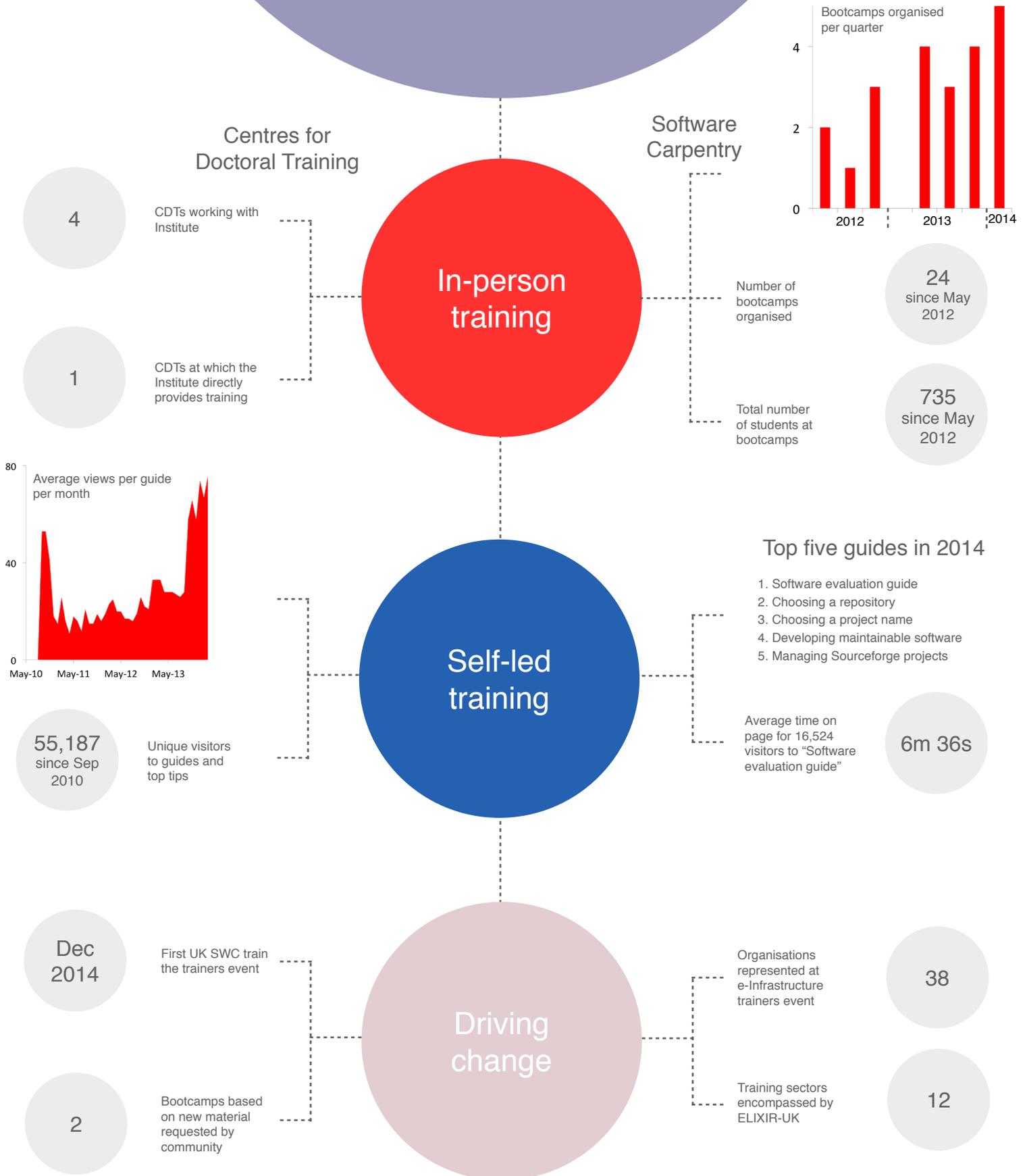
In September 2013, the first workshop for Research Software Engineers was held to test the viability of engaging this community. The response to the campaign has been exceptional: over 10,000 visits were registered on pages related to the campaign and the workshop sold out of the 60 places that were available. This shows that there is a significant interest in the principles of the Research Software Engineers campaign. By taking the time to attend the workshop - even though it has no history or reputation, and some attendees were unfunded - people showed that they are willing to invest time to contribute to our campaign.

The final and most recent step in engaging the Research Software Engineer community was the creation of a website that reflected the principles of the campaign. From the launch of the website in mid-March 2013 to the end of that month, 187 people have subscribed to the community. To gain such a volume of subscribers in only a couple of weeks indicates that there is a significant community of Research Software Engineers with whom the Institute can work.

Workshops in general provide a means to use the experience of the attendees to provide intelligence on the research software community. Since June 2010, over 700 people have attended Institute organised workshops. This creates a well-informed community of researchers who act to disseminate to their peers the information that they learned at the workshop.

To increase the reuse of software, there is a need to disseminate information about software that has a high reuse potential. This led the Institute and Uniquity Press to create the Journal of Open Research Software, which publishes papers on reusable software that are citable and allow reuse to be tracked. We have used our community events to publicise the journal, which has seen a significant increase in submissions from 6 in 2012, 13 in 2013 and 21 in the first quarter of 2014.

Improving software skills



Improving Software Skills

Training was not originally within the remit of the Software Sustainability Institute, but it soon became clear that a lack of software skills was preventing researchers from following best practice for sustainability. For this reason, in April 2012 we created the training theme and started work to improve the software skills of the research community.

We have been instrumental in the rise of Software Carpentry in the UK. From our first attendance at an event in April 2012, we have since run 24 bootcamps and taught over 700 researchers basic software engineering skills. We have also provided guides on many aspect of software sustainability, which have been read by over 55,000 people, and continue to show a strong growth in popularity.

In-person training

The obvious way to improve the skills of a community is to run training events. Rather than replicating existing work, we chose instead to align ourselves with Software Carpentry which shares many goals with the Institute and provides training that helps researchers “get more done in less time, and with less pain”. In response to our work, we were made the UK coordinators for Software Carpentry in February 2013.

It was wise to partner with Software Carpentry. We have found the bootcamps to be very popular with researchers, which shows that the content appeals to them. Since May 2012, we have received 68 requests to run bootcamps from organisations around the country and Europe. It is difficult to find instructors who will provide their effort for free, but we have become skilled at persuading them of the benefits, and have seen a steady increase in the number of bootcamps organised each quarter. In total, we have organised 24 bootcamps since May 2012 and brought Software Carpentry skills to over 700 researchers. What’s more, new material has been created by Institute staff for 2 bootcamps in response to requests from the community for specific types of training (namely on R, and Github). More of these *responsive* bootcamps are expected to occur in the future.

In 2013, the Institute was listed as a training provider for the EPSRC call for Centres for Doctoral Training. This provided access to potentially hundreds of research students at an important time in the development of their research skills. There was less enthusiasm than we hoped for the direct training course (where Institute staff teach at the CDT): only two CDTs included our course in their proposal and only one of those was funded. We believe that this is representative of the desire within CDTs to keep training in-house and the worry that adding costs to a CDT proposal would make it less likely to be successful. Another factor is that not all disciplines are aware of the importance of software in research, which is a mindset that the Institute was created to change.

In addition to the direct training, the Institute has been recruited as consultants for 4 CDTs, which entails advising on curricula and, importantly for Software Carpentry as described below, helps to train more instructors.

Self-led training

In-person training does not scale well without significant investment in trainers, but this short-coming does not affect self-led training based on resources made available on the Institute's website. We produce two main sources of training: guides, which are in-depth documents, and top tips, which provide concise advice.

The majority of our 74 guides have been written in response to requests for help from the research community, something which increases the likelihood that the guides will be popular. This theory has certainly been vindicated: the guides and top tips have been viewed by over 55,000 people since first published in September 2010. What's more, there has been healthy growth in the number of views each guide receives per month, leading to 76 views per guide being registered in March 2014. (We note that there is a statistical anomaly, which can be seen in the graph on page 16 where a peak occurs at the start of the data due to the relatively few guides available at that time.)

Looking at the breakdown of the most popular guides provides some insight into the interests of the research software community. It is heartening that the most popular guide on the website provides instructions on how to review the sustainability of one's own software. The 16,524 visitors to that guide, and the average time on page of 6.5 minutes, indicates that there is significant interest in software sustainability. Interestingly, the next four most popular guides (cumulatively viewed by over 20,000 visitors) are of application to new projects - choosing a repository, choosing project names - and to established projects - developing maintainable software and managing SourceForge projects. This may indicate that our guides are equally applicable to beginners and experts.

Driving change

To drive change in training, the Institute is working to increase the number of Software Carpentry trainers, and is partnering with other training organisations.

The growth of Software Carpentry is limited by one factor: the availability of trainers. We have taken an organising role in the international Software Carpentry movement to create a *train the trainers* community. On this subject, the Institute is sending two representatives - one staff member and one Fellow - to an international event run by the Software Carpentry founders. We have partnered with The Genome Analysis Centre (TGAC) to bring this training to the UK with a first event in December 2014. The Institute's involvement with four CDTs is also focussed on leveraging their training and enticing them to take part in Software Carpentry train the trainers events. Both of these efforts should make it easier for willing trainers to gain the skills necessary to run Software Carpentry bootcamps, which will significantly increase the number of bootcamps that can take place.

The Institute is a partner in ELIXIR-UK (and has a seat on the ELIXIR board), which will coordinate training over 12 sectors, from ICT and Software to Bioimaging. ELIXIR-UK will train scientific technologists to provide "trusted and sustainable software", which is an obvious shared goal with the Institute. The Institute's training theme will help coordinate delivery of the ELIXIR training across the UK, which will also provide access to new potential trainers for Software Carpentry.

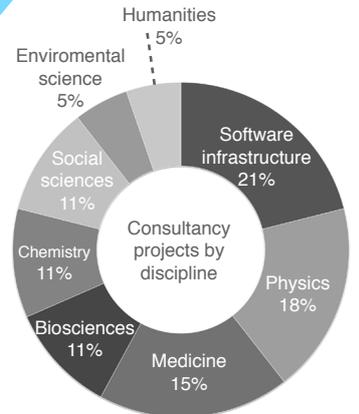
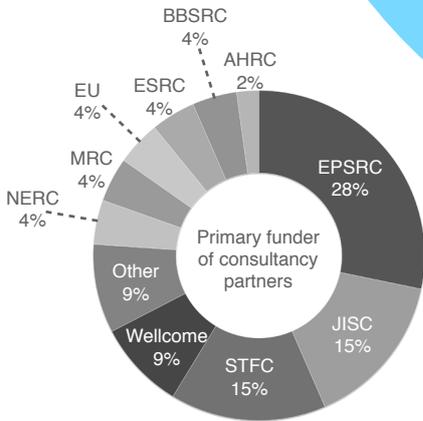
No single training organisation has the resources required to transform a novice into an expert,

nor can any single organisation cater for all disciplines and specialisms encountered in research. A number of training organisations have been created in the UK to fulfil these needs, but there is surprisingly little interaction between them. Coordination between training organisations would lead to significant advantages, by reducing the replication of effort that is inevitable when so many organisations exist, and providing a broader range of training to each organisation's users.

A good example, within the remit of the Institute, is the large number of organisations responsible for e-Infrastructure training in the UK. In August 2013, the Institute brought together 38 different training organisations with the goal of increasing cooperation. This was the first time to our knowledge that so many organisations have met simply to discuss how to work together, and the outcome of the workshop was a unanimous vote for the creation of a special interest group (SIG) that would provide a forum for this cooperation. We have combined this work with an informal partnership with *e-training* - a new website that provides the means for sharing resources and publicising training to a wider community.

It is unfortunate that limited resources have caused the SIG for e-Infrastructure trainers to be put on hold whilst work on other campaigns progressed. However, the enthusiasm for the SIG and the evident benefits of its creation means that it will be progressed later in 2014 once resources become available.

Improving software



Consultancy

38 since June 2010

Completed software projects

Applications to online software evaluation

92 since Apr 2012

"The Institute... taught me how to become a software manager rather than just a developer... I can't speak highly enough of them"
- Christopher Woods, AMRMMHD

"The Institute... turned what had been an extremely painful process into a highly productive and rewarding one"
- Joanna Wardlaw, Brain Research Imaging Centre

"This was a complex code that had been through many generations... the Institute has created a more sustainable and professional software that is accessible to other users."
- Sally Price, CPOSS

Testimonials

"The meetings and discussions I had with [the Institute] had an important impact on the design, usability and functionality of JournalTOCs. We implemented all the recommendations"
- Santiago Chumbe, JournalTOCs

"Working with the Software Sustainability Institute helped me to develop my own project management skills.. and I have adopted a similar process for other aspects of the project"
- Chris Rogers, MICE

"The report... provided valuable insights into how we could take our research prototype forward into open source software that would be more robust and suitable for the wider community"
- Mark Weal, LifeGuide

Quantative

18x

Increase in processing speed of ForestGrowth-SRC simulation codes

75%

Recommendations implemented by ICAT consortium

DMACRYS code variants now in existence from original 5 variations

1

Increase in Synthsys circadian rhythm simulations

10x

Improving Software

Our consultants partner with researchers and provide software development expertise and effort to improve their software. We can only work with a limited number of projects, but this provides examples of best practice that others can follow, and gives the Institute an authoritative reputation.

Over the last four years, the Institute has completed 38 projects with partners from a wide range of disciplines - from Physics to Humanities. These projects have been successful: every project partner would work with us again (taken from a survey with an 84% response rate). We have received excellent testimonials about the quality and usefulness of our work, and have evidence that the lessons taught to consultancy partners have been adopted by others in their communities.

Consultancy

We have worked to identify software that is limited in its growth or application which, if overcome, could benefit a significant user base. Over the last four years the Institute has completed around 10 projects a year with only a handful of in-house developers, which is testament to their skill and experience. The work is highly valued by our consultancy partners: when surveyed, 100% of the respondents said they would be keen to work with us again - many emphatically so (the survey had a response rate of 84%). What's more, working with the Institute was viewed as strengthening the case for funding, as one respondent stated "BBSRC very favourably viewed having SSI as a partner... and I feel that Wellcome (and other funders) are likely to as well".

The Institute is a cross-disciplinary organisation so we have targeted software from across the disciplines. Although the majority of our work falls within the remit of our main funder the EPSRC - 50% in total made up of software infrastructure (21%), Physics (18%) and Chemistry (11%) - the remaining projects derive from the fields such as Medicine (15%), Biosciences (11%), Social Sciences (11%), Environmental Sciences (5%) and Humanities (5%). This success in working with partners from across the disciplines is further evidenced by the fact that we have worked with groups that are funded by all of the research councils, plus the Wellcome Trust and the EU. Working across disciplines in this way supports our cross-disciplinary role and provides authority for the Institute across the disciplines.

The Institute also improves software by providing an interactive, online software evaluation service. This allows visitors to assess their software and receive tailored advice on how to improve its sustainability. Since this service launched in April 2012, it has been visited by 1210 unique visitors, 92 of whom have completed the survey. Although this equates to only an 8% conversion rate, this is to be expected because the survey is detailed and comprehensive, which means it requires a significant amount of effort to complete. Regardless of the conversion rate, providing 92 researchers with tailored advice on sustainability - without requiring the effort of a developer from the team - can only help improve the sustainability of research software.

Testimonials and quantitative results

Measuring the success of the consultancy projects is best viewed with a combination of quanti-

tative results, to measure the impact of our work on research, and testimonials, to measure how our work is judged by our consultancy partners.

Our projects fall into a number of categories. Some projects require us to refactor code to simplify it and make it more efficient: in one case leading to a 10-fold increase in processing speed for the Synthsys circadian rhythms code. Others projects require the code to be adapted to make use of new resources (such as HPC), which led to an 18-fold increase in processing speed for the Forest Growth software which predicts the yield of biomass that an area of land can produce.

Without good management, research software can proliferate into different versions which makes it difficult for researchers to identify the code they should use. The CPOSS software, which models intermolecular forces, is a good example. Before our work with the project, 5 different versions of the code were available, each slightly different to the next. We refactored and collated the code into a single version and created a version control repository in which to store it. This work changed the software into “sustainable and professional software that is accessible to other users”, as described by the leader of the CPOSS project.

Our consultants conduct sustainability reviews, which advise on improving the sustainability of software. The ICAT project ensures that large-scale scientific research projects will remain available, searchable and usable in future and represents 30,000 researchers. Our review of their code included 33 recommendations, 75% of which have been implemented and, even more importantly, these have been communicated to the ICAT partners which include 1 partner from the US and 11 partners in the EU. This is a good example of how working with one project can proliferate advice to a community of related projects.

The ultimate goal of the consultancy projects is to create exemplars of best practice which will entice others to follow them. In the words of one consultancy partner “The Institute... taught me how to become a software manager rather than just a developer... I can’t speak highly enough of them”.

Lessons for the future

Measuring success requires a balance to be found between two opposing motivations. Metrics are absolutely necessary to show that the work and practices of the Institute have an impact on the research community, yet collecting them requires a significant amount of effort that reduces the time that we can spend working towards our objectives. Whilst balancing these motivations during the creation of this report, we have learned lessons that will be of use to other organisations like our own.

Like software sustainability, metric collection should be built into projects from the outset. It is significantly easier to collect metrics as work progresses than to collect them retrospectively. This is easier said than done, because limited resources and time favour the completion of work over the collection of metrics. Ensuring that all partners in a project are aware of the necessity of metrics, and the timeline for their collection, will ensure that they are collected with the least amount of effort. The necessary metrics and their timeline should be included in project initiation documents (PIDs) to ensure that a baseline is recorded and that the project is not viewed as complete until the metrics are collected. Likewise, by collecting demographic information at the point of registration at events, we can more easily build a picture of the community we are serving.

Quantitative metrics can be collected relatively easily in some cases: the number of attendees at a bootcamp, the visitors to a page, the number of applications for the Fellowship. Sometimes these metrics are compelling, but often they require a longitudinal study to find the true impact of the work. Training is a good example. A metric that shows that many people attended a training event shows the success of the event, but it does not speak of the impact of that training. This can only be measured by investigating how many people implemented the training in their work, and this implementation often has little chance of occurring immediately - it may take months before an opportunity arises to put the training into practice. Longitudinal studies are costly, so we did not conduct any at this stage in the success metric collection. Instead we chose to release a report based on metrics that are less costly to collect, and after collecting feedback on this report we will determine which aspects of our work would most benefit from the cost of a longitudinal study. In this way, we will provide a wide range of metrics that are compelling, and a limited number of in-depth metrics that provide a strong foundation to our arguments.

We have found that short, targeted surveys are a useful way to provide depth to a metric without investing significant resources. An email survey which asks only a handful of easy to answer questions can be invaluable, because it is far more likely to receive a significant number of responses than a lengthy survey.

Finally, being able to measure a return on investment (the amount of effort saved, against the effort invested in a software package) is important to allow comparison of our work. It is important to remember that the most significant measure of the success of the Institute is whether it has led to a change of culture regarding the way research software is seen by the community.

