Software Sustainability Institute Status Report
March 2013 to August 2013

The Software Sustainability Institute continued its third year building on the new organisational structure built around four teams:

- Community (led by Shoaib Sufi)
- Consultancy (led by Steve Crouch)
- Policy (led by Simon Hettrick)
- Training (led by Mike Jackson)

Changes in the organisation include the promotion of Simon Hettrick to Deputy Director (effectively replacing Rob Baxter though in a different role).

The approximate staffing for each of the teams over the reporting period is:

- Community: 1.5 FTE
- Consultancy: 4 FTE
- Policy: 2 FTE
- Training: 1.5 FTE

There will be further changes ahead in the next six months. We are continuing to attempt to recruit a Consultancy post in Southampton. Mike Jackson will be moving from Training Lead to Consultancy. Aleksandra Pawlik will take a larger role in Training leadership, and acting as liaison with ELIXIR. A project officer support role is being recruited at Edinburgh to support the administration tasks for Community and Training, which were previously undertaken by Mike and Aleksandra.

Significant highlights in the last six months include:

- Campaigns on e-Infrastructure Training and the Research Software Engineer which have drawn on internal and external collaboration to foster communities using a mixture of dissemination, outreach through workshops and coordination of diverse inputs.
- Training Collaborations. Software Carpentry has continued from strength to strength with new workshops being organised and run each month. We have also had a significant input into the international strategy for SWC. This has also led to several collaborations (CDT, ELIXIR, NCAS).
- Open Call: has attracted high quality submissions.
- Fellows: have collaborated on ideas such as a book and a workshop at the AGU Fall meeting.

Changes from revised plan

There are no significant changes from the staffing organisation and fellows programme presented at the last Advisory Board. The Open Call has been assessed, and additional lightweight assessment mechanisms and a longer call cycle (half-yearly rather than quarterly).

Collaborations

Highlights

The last six months has seen a focus on collaboration that either build on areas of strength for the SSI (such as computational training), or new projects with existing collaborators (such as the software development projects). This has been useful to ensure we utilise our available effort efficiently, but highlights our need to increase the amount of project and business development support available to the team.

- There have been a number of collaborations furthering our training activity
    - The ELIXIR-UK node will focus on training for the biosciences and will include Software Carpentry
    - Connected to ELIXIR-UK, the SSI is contributing to some of the GOBLET bioinformatics training task forces.
    - We are partnering with 10 CDT consortia to provide or develop training based around Software Carpentry.
    - We are collaborating with NCAS on a bid to NERC to provide training for postgraduates.
- The SSI has been invited on to the Advisory Board / Steering Committee of a number of related UK and US organisations including:
    - rOpenSci (Neil Chue Hong)
    - National e-Infrastructure Project Directors Group (Neil Chue Hong, Carole Goble)
- The SSI has pushed forward the campaign for better scientific software development training in the UK, with presentations and major conferences and representations to EPSRC. It has also been active in the areas of reproducible research, citation of research outputs, and career paths.
- The SSI has conducted a number of surveys to inspect different aspects of the UK research software landscape.
- Carole Goble's talk on "Don't Publish, Release!" won the Agent of Change award at Beyond the PDF 2.
- The Journal of Open Research Software is now attracting a submission every two weeks.

Dissemination

- Participation in TSB *Multi-Disciplinary Software Development* workshop, London.
- Participation in PREPARDE *Peer Review of Data* workshop, London.
- Presentation *"Software Sustainability and the Importance for Research Sustainability"* at the ScalaLife workshop, Munich.
- Presentation *"Don't Publish, Release!"*, panel *"The Business Case"* and poster *"Information Circuits"* at Beyond the PDF 2, Amsterdam.
- *Collaborations Workshop 2013*, Oxford.
- Participation in *GOBLET/ELIXIR Train the Trainers* Conference, Norwich.

- Invited speaker *"Communicating Trust, Enabling Criticism"* at Research Without Border event on Communicating Computational Research, Columbia University.
- Presentations *"Building sustainable software for science: why good code is only the beginning"* and *"Software Information and Scientific Publications"* at the EGI Community Forum, Manchester.
- Keynote on *"The UK Software Sustainability Institute"* at the Software Tools and Sustainability workshop, Amsterdam.
- Presentation on *"Supporting our Researchers"* at the UK e-Infrastructire Academic Users Community Forum, London.
- Presentation *"Supporting our Researchers"* at the National e-Infrastructure Project Directors Group meeting,
- Presentation at PLGrid Seminar, Krakow.
- Participation at e-IRG meeting, Dublin.
- Invited Presentation on *"Tracking Software Contributions"* at ORCID-DRYAD Symposium on Research Attribution, Oxford.
- Aron Ahmadia visiting and talking about High Performance Python.
- Presentation on *"Supporting research software developers"* at CMS seminar, Reading.
- Contributed to CC-BY hackday, London
- Invited Participation at *"What to teach Biologists about Computing"*, Annapolis.
- Contributed to Repository Fringe 2013, Edinburgh.
- Dhavide Aruliah visiting and talking about undergraduate teaching curriculum.
- Presentation *"Software carpentry: or, how to persuade ecologists that they actually want to learn how to do this stuff properly"* and Workshop *"What Makes Good Code Good?"* at INTECOL13, London
- Presentation *"WORKFLOWS, PROVENANCE AND REPORTING: A LIFECYCLE PERSPECTIVE"* at Biodiversity Informatics Horizons 2013, Rome.
- Invited speaker on *"Time well spent. Workflows for environmental Omic analysis"* at International Environmental 'Omics Synthesis conference, Cardiff.
- Panellist on *"Research Technologist" panel* at Digital Research 2013 conference, Oxford.
- Presentation on *"A Call to Action: Removing Barriers to Capability/Research in the Wold"* at UK e-Infrastructure Academic User Community Forum, Oxford.
- Co-chairing two workshops *"Workshop on Sustainable Software for Science: Practice and Experience"* and *"Workshop on Software Engineering for High Performance Computing in Computational Science & Engineering"* at Supercomputing 2013, Denver.
- Blog post on *"Why you shouldn't be ashamed of releasing your code"*.
- Conducted software section of National e-Infrastructure *Software Support and Training Survey.*
- Conducted Jisc *Software Usage Survey.*
- Contributed piece *"Cough up money for software or fall behind in science ranks"* for The Conversation.
- Submission to House of Lords inquiry into Scientific Infrastructure.

Leads

- Information provided to all successful CDT stage 1 bid PIs resulted in 10 collaborations.
- NCH brokered discussion between ORCID, ODIN and GitHub to look at ways of identifying software by researchers in GitHub via ORCID
- Various new Training and networking leads from ELIXIR/GOBLET workshop, particularly GARNET (Charis Cook), and EBI training.
- Tim Storer from Glasgow University got in touch about collaboration on software development model bids - we will take this up in the future
- Wifak Gueddera got in touch about a Digging into Data bid looking at software repositories, however this was declined due to closeness of deadline – potential to take this work forward as a responsive mode bid to ESRC.
- We are supporting another biochemistry bid from Bristol following on from the successful work with Adrian Mulholland's group and are exploring a potential repository related project with Bristol.
- We are supporting a bid from one of our previous Fellows (CHARYBIDIS) looking at coastal dune modeling software.
- We are supporting a bid from one of our current Collaborators to enhance BoneJ.
- Following a very successful guest blog post, we are collaboration with Ian gent on his Recomputation Manifesto.

Community

Meetings & community activities

We prepared and ran the Collaborations Workshop 2013 CW13 In March 2013 in Oxford. There has been some discussion at the last advisory board meeting about the value of this meeting,; please see additional synopsis appendix covering how the workshop went, resources produced, satisfactions ratings and whether those attending thought it worthwhile to attend in the future. Most of March was focused on CW13. After the event outcomes were published on the SSI Website; in the following months we had request for the material from both ITaaU and EGI. It was safe to say that CW13 received much praise and the Fellows were actually looking forward to the next one. We announced our Plans for CW14 in August 2013; this year it will be themed as 'Your reproducible research'; theming was suggested as possible way to mitigate confusion about the intent of the meeting for attendees during the previous advisory board meeting.

A meeting with the 2013 Fellows was held at CW13 to discuss their suggestions, plans and ideas; this was an action from the advisory board. The Fellows suggested 'Fellows Forever' i.e. always being Fellows and not Alumni, they also were keen on discussing whether they could combine funding. The thought was that this would then be retrospectively applied to the Agents/SuperPals of 2012 also; increasing the co-hort of Fellows who were not actively funded by still able to apply for a communal fund and have their opinions heard.

Work was started in April 2013 for a "Spreadsheets in Science workshop" to be held early 2014; additional funding for this is being sought from the Alfred P. Sloan Foundation.

The community team took part in discussions with the IT as a Utility Network (ITaaU+) coordinator, Steve Brewer. Steve asked us to be on the steering group of an IT Utilities workshop he was planning in October. The Workshop will be held on the 8th Oct 2013 at the University of Southampton conference facility. With the help of further discussions we were able to re-shape the name of the meeting as "IT utilities and other digital services: what is the optimum discovery model for developers, users and providers?".

In August we began preparation for the Fellow 2014 programme, by getting feedback from the current Fellows about experiences, updating promotional material, opening applications and holding a webinar launch event.

Fellows activities

In March we started discussing the details of the "Software in Polar Research Workshop" suggested by Allen Pope; the workshop is planned for 17th September 2013, this had now been through various stages of clarification pertaining to what is acceptable to fund and what is not; Aleksandra Pawlik from the team will be attending and representing the SSI and helping facilitate group discussion sessions.

One of the suggestions from the Fellows this year was on writing a book/anthology about Software in Research; in the form of personal experience, a domain introduction for new starters or other aspects. At the 6-monthly Fellows meeting (which 10 of the Fellows attended) it had an alternative title of 'The little book of software struggles'. SSI was managing the production of the book but we wanted this to be a Fellows lead; they themselves saw the value of this so it was interesting that although all of the Fellows are contributing, the ones who have taken on the editorial control (Nick Pearce and Adam Crymble) are both from the humanities. The aim is for the book to be ready in the spring; currently we have half of the envisaged material in near final form; we have monthly meetings with the editors and the community team to make progress on the anthology.

We held a 6-month Fellows meeting in June, this was mainly structured around chapters for the book, and it was also an interesting networking activity for the Fellows. Cefn Hoile was has a specific focus on microcontroller based devices and human computer interaction had a number of interesting discussions with people around devices that would help their science, an example of this was a more reasonably priced snow depth meters for field work. Also Richie Abel had some interesting initial discussion on characterisations of snow crystals with Mel Sandells when they discussed the problems around trying to define what snow was and thus validating the computer models used to represent snow. Another interesting outcome was that we asked them whether we should have more Fellows rather than fewer but with sliding scales of awards and they were not that keen on that; they wanted to deal with each other as peers. They also were strongly in favour of the continuing Fellowship programme and wanted to attend the Collaborations Workshop 2014 as part of this. They were keen for us to help them support more of a community amongst the Fellows and thus they pushed for another meeting for Fellows perhaps focused this time on IP and possible commercialization routes for software/research; this is planned for November 2013.

In May we had a Skype telecon with Anna Powell-Smith (Shoaib, Simon and Aleksandra): Anna provided several suggestions about contacts with the developers' (and policy e.g. data.gov) community that might be useful. Being a freelance developer, time was one of the luxuries she actually had very little of; we had already made the decision internally to keep her as a bridge to the more general developer community and to make use of her knowledge, expertise and links even though she did not fit the mold of a research based Fellow.

In June there was a discussion on the Fellows e-mail list about IP and whether people knew who owned their code; it was clear that was a lot of confusion and in the subsequent Fellows 6 monthly meeting there was a suggestion that this should be a topic in the additionally requested Fellows meeting towards the end of the year.

Staffing

Aleksandra Pawlik joined as Research Software Community Specialist in March. She is a real asset to the Community theme. Her role started to change in August as she took more of a lead in the Training theme. She is now only 50% on community. Community will get 50% of a new hire that is being sought at Edinburgh.

Dissemination

- Shoaib gave a talk "Engaging the software in research community" at the Communities and Communication session at the EGI Community Forum in Manchester, 8-12 April 2013
- Aleksandra presented at the Champions Workshop EGI Community Forum in Manchester, 8-12 April 2013 ("SSI Fellows and Agents" and "SSI Agent")
- Fellowship Programme pages updated in May allowing interested people to register their interest in the Fellowship Programme 2014.
- Robin Wilson's (Fellow) report from the Wavelength Conference 2013 in March
- We published the outcomes of the Collaborations Workshop 2013 (in full by April)
- We published a blog by Fellow Stephan Lautenschlager's on “A Digital (R)evolution in Palaeontology”
- Aleksandra gave a seminar about SSI at Cyfronet Krakow on 20th May - Cyfronet would like to develop collaboration with SSI
- Aleksandra attended the e-Infrastructure Reflection Group meeting in Dublin 22-23 May
- Published Fellow Kayla Iacovino's blog post “Cutting edge tools for cutting edge results”, in May.
- Published Agent's report (Kristy got a special consent to extend her Agent activities) Association of American Geographers Annual Meeting 2013
- Published Fellow Caitlin Bentley's report on International Conference on Social Implications of Computers in Developing Countries, 2013
- Published Fellow Robin Wilson's report on DART (Discrete Anisotropic Radiative Transfer Model) Workshop
- Aleksandra attended ISMB 2013 conference in Berlin. She was one of the panelists at the discussion panel " What makes for a successful professional bioinformatics network?"
- Published Kayla Iacovinos report on Instrumentation and Software for the Monitoring of Active Volcanoes 2013
- Published Fellow Nick Pearce’s report on Web scraping for arts, humanities and social sciences workshop 2013
- Published Fellow Barry Rowlingson's report "Research Computing User Group Inaugural Meeting"
- Fellow report: CEDAR Workshop 2013 by Alex Chartier
- Fellow report: National Astronomy Meeting 2013 by Alex Chartier
- Fellow report: Information Gathering trip to Australia 2013 by Nick Pearce
- Blog post on Intelligent Systems for Molecular Biology (ISMB) Conference

Leads

- Richard McLennan; Richard manages the EGI Champions network and is interested in learning how we run SSI's networks (Agents and Fellows). He saw Aleksandra's presentation at EGI Champion's Workshop and Shoaib's presentation at EGI. We are now having regular quarterly meetings with him over Skype where we exchange ideas around best practice on running community networks e.g. stating minimal expectations for outreach for the sponsor (SSI/EGI) when funding travel .
- PL-Grid Plus / Cyfronet Krakow; a nation-wide project supporting software and HPC for research; introduced them to Neil at EGI Forum. Eryk Ciepiela from Cyfronet attended the Manchester bootcamp as a helper/observer and will help (together with a few other Cyfronet people) at SWC in Krakow; Cyfronet are interested in an potential SSI plans for Europe.
- Physiome Group; a multidisciplinary team from Charles University in Prague; talked with one of the researchers at EGI Forum, Tomas Kulhanek (an EGI Champion). They're interested in SSI's activities and goals. Would like to discuss potential "European-wide SSI".
- Contacted Alexandra Simperler from Imperial and David Palmer from University of Strathclyde (see above) to learn more about them and their interest in the Fellowship programme. Alexandra was an experienced training hand so we put her in touch with Greg Wilson from Software Carpentry to see if there were possible fruitful discussions on training for scientists.
- Other contacts have been made, but although not of direct relevance to community we route them to Training and Consultancy; this is evidence of our attempt to not let the themes become siloed e.g.:

    - Peter Strutton who is the Business Development Manager from Loughborough University, Aleksandra met him at the eITR in Hartree in August; he asked if SSI could provide software consultancy for some of Loughborough University commercial partners (for example, Rols Royce). Aleksandra put Peter in touch with Mike and Steve. Also, Peter is interested in HPC training that SSI could potentially provide (again, for commercial users of their HPC at Loughborough University).

Consultancy

Highlights

The Open Call

The Institute's Open Call for projects[1] accepts submissions from researchers who wish to improve their software, their development practices, and their community of users and contributors. Our consultants work closely with successful applicant research teams to assess these aspects, providing recommendations for improvement, and assisting with the implementation of those recommendations. An Open Call project typically runs between 2-6 months, where we work together with successful applicants to develop a tailored work plan suitable for the project.

There have been two completed rounds of the Open Call to date, the first from September 2012 – January 2013, and the second from February 2013 – April 2013.

To summarise the progress to date of the first round of Open Call projects:

- **BoneJ** - Royal Veterinary College, University of London Vet School (9th May 2013 – 30th September 2013)

  Formerly from a BBSRC-funded project, the BoneJ software[2] is a suite of plugins for bone image analysis, developed and maintained by a single developer. It provides free, open source tools for trabecular geometry and whole bone shape analysis. Widely used by scientists worldwide, the software has been cited in over 70 academic journal and conference papers and used in fields such as materials, food and soil science.

  The goals of the project were to improve engagement with the BoneJ community, identify areas for improving the software and its management of code contributions, and assist in lowering the high overhead of maintaining the hard-coded website. Achieving these goals will ensure a lower bus factor for the software (currently 1), increased communal responsibility and effort for its development, and freeing the current developer to tackle more strategic concerns with the software. In turn, this will provide a firm foundation for securing future funding, with necessary project tasks already identified.

  In the last four months, a community survey was developed and sent to the community to gather opinions on suggested future work and to attract community members to assist development and documentation authorship. This has been positive, revealing BoneJ's importance to its users' research and the areas in which they would like to contribute, and has helped to steer the project and future goals. It has also identified

[1] http://www.software.ac.uk/open-call
[2] http://bonej.org/

community interest in assisting with testing BoneJ, integrating contributions, improving and maintaining the website, and providing user case studies. A comprehensive development review was also conducted generating 47 recommendations for improvement in a number of areas, including the website, use of GitHub, the user and developer experiences, releases, source code, documentation and testing. In addition, coding standards guidelines and contribution policy for external developers were created.

This work is continuing to contribute to the case for future BoneJ funding, with the lead developer's proposal to the BBSRC Bioinformatics and Biological Research Fund now reaching the full proposal stage. As noted by the lead developer, 'SSI's involvement was noted in a very positive light'.

- **Imperial College Lower Limb Model** – Osteoarthritis Centre of Excellence, Imperial College London (17th May 2013 – 30th September 2013)

  Funded by EPSRC and the Wellcome Trust, the Imperial College Lower Limb Model (ICLLM)[3] is a musculoskeletal modelling software package. It allows the prediction of subject-specific muscle and joint forces at the hip, knee, and ankle by employing motion data obtained through gait analysis techniques. These outputs may be used to assess disease progression, or evaluate both surgical and non-surgical treatment strategies. Currently, the model is used in age-related, sports and blast injury analysis.

  The project goals were to improve the internal software, documentation and development practices of the team, increase the manageability of their multiple variant code bases, and assist in the provision of appropriate infrastructure and processes to share their code. Once these goals are achieved, the software will be more manageable and in a position to be exploited in future collaborations and projects, which was previously not the case and a major blocker to such endeavours. In addition, the technical issues and increased effort they have experienced dealing with multiple code bases will be significantly reduced.

  Since the project began in May, a development review was conducted which included 42 recommendations including documentation, development practice, dissemination, developer experience, version control options and testing. Based on these recommendations, the ICCLM team have moved to a single code base and improved their documentation and coding policies, and a Subversion revision control system hosted on the public Assembla[4] infrastructure has been selected to host the code. The ICLLM's project lead has stated 'The Institute's team were invaluable' in assisting in these areas[5].

[3] http://www3.imperial.ac.uk/osteoarthritis/research_objectives/lower_limb_musculoskeletal_model
[4] https://www.assembla.com/home
[5] http://www.software.ac.uk/blog/2013-08-01-leg-lower-limb-model

The final stages of the project will see the code ported to that repository and the team's software development practices adapted to using that infrastructure, with guidance from the Institute. This work will position the team and the software to exploit the software in other potentially lucrative research areas where there is collaborative interest, such as cancer and surgery within the Imperial College Healthcare NHS Trust.

- **ForestGrowth-SRC** – Centre for Biological Sciences, University of Southampton (?? June 2013 – December 2013)

  Funded by NERC and Forestry Research (the research arm of the Forestry Commission), ForestGrowth-SRC is a model for predicting the yield of woody biomass (such as popular and willow) from an area of land. Developed jointly by the Centre for Biological Sciences at the University of Southampton and Forest Research, the software takes into account the hundreds of different biological processes that contribute to yield, such as photosynthesis, light interception and the way the plant uses carbon. Forest Growth SRC models all of these processes and uses them to predict their effect on yield.

  The project goals were to address the software's scalability for handling much larger simulations which can only be run on Windows, and improving code development practices and usability for external collaborators. Currently, scaling up to 100,000+ square kilometres (the area of the UK), with ranges of starting inputs for temperature, precipitation and CO2 levels means a run takes many days to complete. Increasing the software's scalability and addressing the complexity of using the software will lead to increased research outputs and open up new avenues for collaboration.

  Starting in June, the project has seen the ForestGrowth-SRC FORTRAN code ported from Windows to Linux and test simulations have successfully been run on the University of Southampton's IRIDIS4 cluster[6], a very significant step for the project. Further work on improving the software's esoteric and complex command line interface will greatly assist in its integration with other ecosystem models such as the ECOSSE[7] soil model developed by the University of Aberdeen.

  Of particular note is that the ForestGrowth-SRC PI has submitted a joint bid with the SSI into the University's Institute for Life Science's fund, which provides ~20k seed funding to prepare for full research grant funding with industry. This bid includes 3 months FTE effort for an SSI developer, with the Institute potentially being involved in any full grant proposal if successful.

[6] http://cmg.soton.ac.uk/iridis
[7] http://www.scotland.gov.uk/Publications/2007/03/16170508/0

From the second round of Open Call projects which closed on April 30th, three have been selected to begin in October 2013, taking into account available consultancy effort and the summer vacation period. Planning of these projects was begun in June 2013, to be finalised in the coming weeks:

- **LabBook** - Bioinformatics Support Service, Imperial College London

  LabBook[8] is a specialised electronic Android-based note-taking system that aims to improve the accuracy and efficiency of recording experimental processes in the laboratory. The ultimate aim is to aid the digitisation of the complete research workflow from bench to publication, and dramatically simplify collaboration, data archival and reproducibility. The work so far has been unfunded, and despite a solid foundation and enthusiasm from potential users progress has slowed significantly due to an inefficient development process and the lack of a strategic plan.

  The target audience for LabBook has so far been lab-based life scientists working in academic institutions, but there is also broader interest from many other researchers, ranging from pharmaceutical R&D to archaeology – the potential market for such a solution across the academic spectrum is huge.

  The Institute will help ensure the sustainability of the system by assisting the team in turning it into an extensible, open-source platform, with development driven by user feedback and code contributions from external developers. This will involve hardening the existing application, which requires a thorough review for reliability and usability in line with software engineering best practices, and marketing the software to other research domains by leveraging the Institute's Fellows and considerable Policy & Communications contacts. The ultimate aim is to help the team produce a sustainable and well managed project that can justify funding applications for further development.

- **DawnScience** – Diamond Light Source, Rutherford Appleton Laboratories

  The Data Analysis WorkbeNch (DAWN)[9] is an eclipse based workbench for doing scientific data analysis. It implements sophisticated support for visualization of data, python script development, debugging and execution, and workflows for analyzing scientific data. It is developed by and for the synchrotron community foremost but has strong overlap with other communities like neutron scattering, photon science and any scientific communities with the above or similar needs.

  The DawnScience team already have a productive and well-established development process and infrastructure which has led to a well-engineered implementation. A technical review of their testing and documentation would be useful, however they more importantly need

[8] http://labbook.cc
[9] http://www.dawnsci.org/

assistance with how to market the software and also require better engagement with existing users to better understand what they need from the software in order to inform future development decisions. A key measurement of the success of DawnScience will be the extent to which they have established a solid user base, and increasing the collaborations beyond those they already have (e.g. with the University of Sheffield) will form the main motivation for the project.

- **BASIL/FABBER** – FMRIB Centre, University of Oxford

  Components of the existing FMRIB software library[10], the BASIL/FABBER tools allow FMRI, MRI or DTI brain imaging sources to be analysed. These tools employ a core method that enables the fitting of any given non-linear model to data based on the principles of Bayeisan inference, and has been applied to a number of neuroimaging applications.

  The team are increasingly developing it within this area to both include new functionality and also to apply it to new data types/problems (i.e. new non-linear models). However, the current codebase is becoming unwieldy and has the potential to degrade further as the software is extended in the future. The current library, developed for use by collaborators in Germany, is functional but as a library it needs improvement. Additionally the algorithm itself has much broader application that the MRI/4D data for which it is currently used.

  Potential work items include a development review, improving the development code to make it more usable internally and externally, and making it available for wider use via appropriate infrastructure. The goal of this work is to provide a solid foundation that will lead to greater use and adoption of the library, and more portable to other application areas.

Another submission based on the Atomic Simulation Environment (ASE) from the Center for Atomic-scale Materials Design at Denmark Technical University, which was a resubmission from the previous round, was rejected since it did not sufficiently address the UK impact of the work from its previous submission. In addition, there is a strong UK bias towards using a competitor, ChemShell, which is more broadly used.

During the review of the second round of Open Call projects in May, we decided to invite two SSI Fellows (Nick Pearce, Michael Croucher) as external reviewers, with review information treated in confidence. Michael Croucher was also invited onto the review panel to engage in review discussions. Not only did this give a valuable external perspective on the candidate projects, but Michael also provided minor but useful suggestions on improving the process for future reviews (mainly explanation of acronyms). He remarked in general that guidance on coming into the process as an outsider was well provided, and praised the process overall, recommending that Fellows should be involved in the future. This is currently being considered for the third round review.

[10] http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/

Based on our experiences running the call to date, the use of a quarterly period for accepting Open Call proposals has moved to a more flexible deadline model, with following deadlines being agreed by SSI Operations. This allows us to take into account factors such as available consultancy effort, vacation periods and other opportunities. The next (third) round of the Open Call is due to close on October 31st 2013, at which point the first round of project work would have finished with the second round already begun. This means the third round projects are expected to start January/February 2014, leaving a customary and necessary 2-3 months of planning before these projects begin.

Externally Funded Projects

The Institute has also been involved in a number of externally funded projects within the reporting period:

- **Libhpc** – with London e-Science Centre, Dept. of Aeronautics at Imperial College London (Stage 1: July 2011 – June 2013; Stage 2: July 2013 – June 2015)

  Funded by EPSRC, the Libhpc project[11] aimed to develop a generic component-based framework for developing and running High Performance Computing applications on remote, distributed platforms such as the Cloud. The Libhpc framework enables scientists to build and deploy HPC applications in a platform-agnostic manner while taking advantage of platform and hardware-specific optimisations.

  The Institute provided the libhpc project with access to parallel computation expertise and assisted in the engagement with research communities to promote uptake of the software.

- **AMRMMHD** – with the EPCC at University of Edinburgh, University of Bristol (September 2011 – August 2013)

  The Adaptive Multi-Resolution Massively-Multicore Hybrid Dynamics (AMRMMHD) project[12], funded by EPSRC, was part of a group of projects surrounding Sire[13], a molecular simulation framework developed at the University of Bristol. The codes developed by AMRMMHD have been used to study the drug resistance mechanisms of influenza strains, including this year's outbreaks.

  The Institute provided project management and architectural expertise, which helped the project to overcome the project's big challenge of moving from a one man, small scale software project into a multi-developer programme. The Institute assisted in compartmentalising development of their molecular simulation software across the

[11] http://www3.imperial.ac.uk/lesc/projects/libhpc
[12] http://www.epcc.ed.ac.uk/projects-portfolio/amrmmhd-adaptive-multi-resolution-massively-multicore-hybrid-dynamics
[13] http://siremol.org/

collaboration, whilst ensuring it remained compatible overall.

The Institute also provided guidance on capture the developer's technical knowledge so that the software could move into a self-contained, well-documented state for future use. As noted by the AMRMMHD PI, 'The management and community side to what the Institute does is what makes it stand out for me. The ability to offer that help makes them different - I can't speak highly enough of them'[14]. This Institute is also continuing to provide support. for the follow-on project, INQUIRE.

- **APES** – with the EPCC at University of Edinburgh (April 2013 – April 2016)

  The APES (Advanced Potential Energy Surfaces) project[15] aims to help researchers advance their understanding of the structure and function of molecules by improving the models used to represent their force fields.

  With the EPCC, the Institute is helping to parallelise the codes that underlie the AMOEBA force field software (such as TINKER) so that AMOEBA can be run on high-performance computing clusters such as HECToR and its successor ARCHER, as well as NFS-provided resources in the US.

  The Institute's work so far has enabled TINKER's FORTRAN code to be compiled on HECToR, and a preliminary scalability analysis to be conducted to work out how best to take advantage of HECToR's capabilities. The code has also been profiled for optimisation, and investigations into a C/C++ version of TINKER have been conducted.

- **Jisc Software Hub** – with OSS Watch at Oxford University (January 2013 – October 2013)

  The aims of the Jisc Software Hub[16] are to catalogue all the existing software that Jisc has funded over the last decade or so, and also attempt to promote some of this software to encourage uptake within the UK academic community and further afield.

  The Institute's role has been to define the metadata that will be used in the Software Catalogue to describe any software outputs that have been produced, and to populate the Catalogue using information already stored in Jisc's PIMS (Programme Information System) and other sources.

  To date, PIMS has been loaded up on to the Hub and its data is continuously being enriched. Criteria to promote items from the internal catalogue to the public showcase have also been developed, and the functionality of the Hub has undergone a period of testing, resulting in

[14] http://www.software.ac.uk/resources/case-studies/getting-grips-molecules
[15] http://www.software.ac.uk/who-do-we-work/apes
[16] http://www.epcc.ed.ac.uk/blog/2013/05/13/jisc-software-hub

refinements to the metadata model. The first meeting with the project's Advisory Board in May was largely positive, with suggestions for improvements, and an Advisory Board meeting convened over Skype led to a favourable review of the current Hub by the Board. Second phase funding options are currently being considered.

- **DMACRYS** – Sally Price's Research Group, University College London (January 2011 – July 2013)

  The DMACRYS software[17] simulates the likely crystal structures an organic molecule can adopt. DMACRYS is very successful in pharmacology, having a well established and continually growing developer community that extends across academia and industry.

  Building on the Institute's successful engagement reported in the last consultancy report, version 2.0.8RC1 of the DMACRYS software has been made available, incorporating two threads of development from inside UCL and one from the University of Cambridge. This represents the first major release of DMACRYS in over three years. In addition, the release is also the first to include an acceptance test suite which was developed by the Institute.

Software Evaluation: Self Assessment

For those researchers wishing to conduct an evaluation of their own software, the Institute also provides an online Sustainability Evaluation Service[18] (as well as information on how we conduct evaluations ourselves[19] in case researchers wish to use our processes for self-evaluating their software). To date, 16 users have made use of the online evaluation facility in the reporting period.

Publicity

Following the advice from the last Advisory Board, a strategy for publicising the Open Call was developed with the Policy and Communications team in February 2013, for the Open Call in general and for highlighting the progress made in the individual Open Call projects. This has been successfully implemented. As part of this, later in September 2013, the Open Call will be heavily publicised towards its next submissions deadline of 31st October 2013. This will involve working closely with the Policy and Communications team to leverage their contacts network, as well as developing appropriate publicity text to go out to Institutional departmental newsletters in October.

Working with the Policy and Communications team, a further four case studies have been written and publicised. In addition, six blog posts on the various consultancy activities have been published (see *Dissemination* later). In the

[17] http://www.cposs.org.uk/
[18] http://www.software.ac.uk/online-sustainability-evaluation
[19] http://www.software.ac.uk/software-evaluation-guide

future, we aim to increase the number of consultancy blog posts and case studies, as the current round of Open Call projects are yielding good results and two of these are drawing to a close.

Web Site Redesign

Our consultancy web pages have been revised as part of our web site redesign activities, including a consultancy landing page[20] that more cohesively describes our consultancy services on offer, and a comprehensive page listing our completed and active projects[21].

Miscellaneous

- ICAT – Following our collaborative evaluation of their project last year, the project reported in February that they have implemented 75% of the evaluation's recommendations.
- SeIUCCR Summer School, August 2013 - The SeIUCCR project is developing and supporting a network of Community Champions to advocate the use of e-Infrastructures in their research. In their most recent Summer School, we were involved in the Ask the Experts session where we discussed and provided advice on the problems faced by 8 researchers with their software. We're currently awaiting survey feedback on this activity and the course as a whole.

Leads

- Mark Miller (San Diego Supercomputer Center) from CIPRES Science Gateway project (NSF), is considering submitting into the Open Call. Although a US-based phylogenetics HPC resource, 500,000 core hours have been provided to 77 UK researchers and students.
- Peter Strutton, the Business Development Manager at Loughborough University whom Aleksandra Pawlik met at the e-Infrastructure Workshop at Hartree in August, is interested if SSI could provide some one-to-one software consultancy to some of their business HPC users.
- Ian Gent, author of the Recomputation Manifesto[22], is looking to develop recomputation.org, a site which aims to accept and provide recomputable experiments. Initial discussions have yielded potential for the Institute to assist with how the experiments could be provided online, and this will be followed up in September.

Effort

[20] http://www.software.ac.uk/consultancy
[21] http://www.software.ac.uk/who-do-we-work
[22] http://www.software.ac.uk/blog/2013-07-09-recomputation-manifesto

Due to expenditure of effort on other Institute activities by staff, primarily on Software Carpentry training activities during February to April, this caused some effort issues for consultancy activities despite additional management and consultant effort being brought in which was invaluable. Although all key organisational and management tasks were completed in this period, the main consequence was that the planning of the first round of Open Call projects was delayed, and subsequently the projects started later than anticipated. As noted elsewhere, the lack of effort for training was resolved, partly with external assistance.

To ameliorate any such recurrence in the future, and to expand our capability to conduct consultancy activities, Southampton has been involved in recruiting two additional Software Consultants. This has proved problematic to date, largely due to Certificate of Sponsorship and lack of communication issues caused by Southampton's Human Resources. However, our second round of recruitment yielded an appropriate candidate to whom we have recently made a conditional offer, and we're pursuing other rounds of recruitment to secure a second Software Consultant. Both positions are until the end of the SSI project, in May 2015.

Dissemination

Our blog posts

- 20/02 The process of cataloguing software – Neil Chue Hong and Mario Antonioletti
- 18/06 Monkeying around with chemistry: the APES project – Arno Proeme
- 15/07 Strengthening bone shape analysis – Mike Jackson
- 26/07 Breaking the barrier of software maturity - three Open Call projects – Steve Crouch
- 01/08 A leg up for the Lower Limb Model – Angela Kedgley
- 08/08 Predicting biomass yield to enhance energy security and meet emissions targets – John Robinson

Blog posts by others promoting, referencing or reblogging us or our posts

Case studies (with Policy and Communications)

- 07/05 Ensuring a future for the digital arts and humanities community (Arts-humanities.net)
- 08/05 Analysing texts in digital archives (TextVRE)
- 17/05 Achieving unexpected goals from an ambitious social science project (NeISS)

- 27/06 Getting to grips with molecules (AMRMMHD)

Other publications, news articles etc.

- 23/04 The APES Project – Arno Proeme, EPCC blog
- 19/06 Survey on Software Discovery and Usage – Mario Antonioletti, EPCC blog
- 06/06 The Jisc Software Hub – Mario Antonioletti, EPCC blog
- 08/07 The Software Sustainability Institute at INTECOL13 – Mike Jackson, EPCC blog

Policy and Communications

Highlights

- Two campaigns, both in important and topical areas, achieved prominence and were accompanied by well-attended workshops
- Significant exposure for the Institute
- First publication in Times Higher Education
- Excellent uplift in statistics for the website and Twitter

Quick statistics for Communication: 1 February-31 August 2013

- 38,491 unique visitors to the website
- 79 blog posts contributing 74% of the unique visits to the site
- 951 new Twitter followers
- 210 tweets, 329 retweets and 415 mentions leading to 506,476 Twitter impressions

Overview of the last six months

The last six months have seen the Institute go from an organisation with only an informal approach to policy, to having led two popular campaigns that have attracted significant interest – including press interest – across the country and worldwide. Both campaigns have seen the creation of a new community, have increased the Institute's visibility across the research community and have raised our status as an effective policy group.

As a direct result of our focus on Twitter, the blog and syndication, there has been a significant increase in traffic to the website and activity on Twitter. The website has for the first time seen spikes in traffic of 4000-5000 unique hits in a day. Twitter is no longer a tool for broadcasting information, but has become an effective tool for communicating with our community and requesting information from it.

Research Software Engineer campaign

Outcomes

- Created a new community of research software engineers: people who have a direct influence on the success of the Institute's goals.
- Created a committee of prominent RSEs who have been very successful access point to the RSE community.
- Raised the Institute's profile with a successful article and blog posts, and increased our status as a vital organisation within the field of research software.

Overview of the campaign

Research Software Engineers (RSEs) are the people behind research software. They play a fundamental role in ensuring high quality research software, but lack

recognition and reward for their contribution to research. Campaigning to improve the status of RSEs was an obvious choice of a campaign for the Institute, because any improvement in their status should have a direct beneficial effect on research software.

This campaign have been significantly enhanced by choosing a well connected and experienced RSE committee. This consists of Chris Cannam, Dirk Gorissen, James Hetherington, Caroline Johnston and Mark Woodbridge. The EPSRC's Louise Tillman has an interest in this campaign, so we have kept her abreast of developments and we have also benefitted from her insight.

The RSE campaign has been split into three parts: raising visibility, community building and statistics gathering.

Raising visibility

The visibility of the RSE role has been increased by both the Institute and the RSE Committee raising the issue at conferences and meetings, by writing popular blog posts, and by successfully targeting the Times Higher Education for an article (see links at end of this section). The visitor statistics for the blog posts, the popularity of the RSE workshop and the very fact that we persuaded a national magazine to feature our cause is testament to the increased visibility of the RSE.

RSE workshop

Feedback from the workshop is attached at the end of this section.

Community building kicked off with a workshop for RSEs on 11 September 2013. The original 50 registrations sold out within three weeks, so a further 10 places were added and these sold out in the next 9 days. A further 8 people who could not attend the workshop asked to be added to the workshop's mailing list. We are confident that the workshop could have been significantly bigger given more time to organise pre-publicity – and a larger venue.

At the workshop, a near unanimous vote was cast for the creation of an RSE community. Plans need to be confirmed, but it would appear that this community would have two purposes: one would be a *weight of numbers* group that could lobby the research community and improve the situation of the RSE, and the second would be a group that could swap skills and contacts. For example, mentoring was discussed, as was the creation of a training swap where RSEs with one skill set could provide informal training to others in the community.

If an RSE community is created, the Institute's role would be that of a mentor and an organisation friendly to the community's goals. We should take a seat on the community's running committee, but the plan is not to run the community ourselves. Ultimately, the community must self-sufficient and self-steering. Plans for the creation of the community are yet to be formalised (the workshop was held the day before this document was written), but we are pleased to announce that the workshop has already led to the creation of a local RSE community in Oxford.

Statistics gathering

Taking the campaign further will not be possible without concrete numbers about the RSE community. We lack fundamental statistics: such as the size of the community. We will run a simple survey in the next 6 months in which we will try to capture the size and make up of the RSE community. We will take advice on the survey from the RSE community and from specialists in social science, such as Nick Pearce, who is one of our Fellows.

Future plans

The next 6 months will see the RSE campaign create a national RSE community, collect numbers on the size and make up of that community, and begin to share these numbers and the reasons why RSEs should be rewarded and recognised with controlling organisation in the research community.

e-Infrastructure trainers campaign

Outcomes

- Created a community of e-Infrastructure trainers which has increased our profile across the e-Infrastructure community
- Raised the Institute's status with a successful workshop and reinforced our reputation as an organisation that brings communities together
- Potential to further raise the Institute's status with a report that will bring together the training community's recommendations for improving training

Description of the campaign

In February 2013, Simon Hettrick was invited to present at the e-Infrastructure Academic User Community Forum (e-IAUCF) on the provision of skills and training within the e-Infrastructure community. This presentation led to discussions with some of the main training providers in the UK, namely the DCC, EPCC, NAG, STFC, Hartree Centre and HPC Short Courses, and these discussions led to the creation of a committee. A consensus was formed that there was a need for the creation of a training community across the UK, and that certain other changes to training should be made. The Institute had the expertise and contacts needed to organise such a campaign, and hence has led it over the past six months.

Providing the skills for e-Infrastructure report

The training committee (formed from the training groups listed in the last section) found that they agreed on many recommendations for improving training, namely:

1. We must improve communication between training providers.
2. The true cost of training should be communicated to the research community.
3. We must recruit trainers rather than rely on volunteers.
4. Trainers must work together to publicise training.
5. Incentives are needed to motivate the production and sharing of good training materials.

A plan was created to draw these recommendations into a report that could be distributed to people with an interest in training: the Research Councils, funding organisations, training organisations and policy makers.

Each of the training committee members have written a submission for the training report (and a new submission will be submitted from the EBI at the end of September). Simon Hettrick and Claire Devereux are now editing these submissions to ensure both consistency and sufficient evidence is included.

The EPSRC's Susan Morrell is interested in the report, so we have kept her abreast of developments and will release the report to Susan before it is released.

The report is intended to be released at the end of October 2013.

Workshop for e-Infrastructure trainers

On 14 August 2013, we held a workshop for e-Infrastructure providers. The fifty available places sold out in just over three weeks, so a further 10 places were added. In total 56 registrations were made and 50 people arrived at the event. Relatively little publicity was needed to gain these registrations, which indicates that there was significant interest in the event.

The workshop split the attendees into groups that discussed issues that had been nominated by the attendees themselves. Discussions ranged from how to incentivise production of training materials, certification and – importantly – what people would expect from a training community.

There was a near unanimous vote to create a training community that could work together to share resources, publicise training and prevent duplication of effort. Similar to the RSE community described above, the intention with the training community is for the Institute to help set up the community and take a position on a running committee, but for the community to be self-sufficient and self-steering.

A report on the outcomes of the workshop will be written in the next two months and this will be distributed to all attendees and made available on the Institute's website.

Website redesign and reorder

Outcomes

- Website content has been rearranged into four themes
- New widgets have been implemented
- Significant amount of content has been rewritten
- Theming of pages has been completed

Reasoning behind the redesign and reorder

At the Advisory Board meeting in February, plans for the an overhaul of the Institute's website were described. These plans focussed on improving the links

between material, reordering content to improve the logical connection between materials, and the rewriting of some pages to change the feel into that of a national facility that is actively seeking to collaborate with projects, rather than just a research project that had completed work.

Changing the language of the website

One of the suggestions of the advisory board was to change the language of the website so that it felt less like an academic project and more like a national facility that welcomed collaborations with research project. In short, the website should pull in collaborators rather than simply pushing out information about its completed projects.

Most of the content that needed to be rewritten has now been rewritten. Although progress has been good, some of this work is still ongoing and a final review of the website content has yet to be completed. This review, and any final edits, will be completed before the end of the year.

Improving links between material

The links between material on the site has been implemented in a number of different ways, almost all of which rely on *widgets*. These are small boxes of links that appear on the right-hand side of each webpage. Some widgets, such as *Most popular* and *Latest News* attempt to link material by enticing users to click on links to content that is either popular or fresh.

A set of landing pages has been created: one for each of the four themes (community, consultancy, policy or training). The landing pages link together content that is relevant to that theme, with the intention of increasing linking based on the theory that people who visit the policy page will be interested in our other policy content. This is combined with a new theming system that forces all new webpages to be marked as belonging to one of the Institute's four themes. A new widget used on the landing pages displays *Latest from <relevant theme>* which ensures that the latest content pertaining to the theme is easily accessible on the theme's landing page.

In addition to the theming above, a tagging system is currently under development. This is very similar to the theming system, but relies on a wider set of categories. Rather than being limited to one of the four themes, the tagging system will allow a page to be additionally tagged with any of a reasonable set of topics, for example: *software development, workshops, funding opportunities, etc.* The tagging system will be especially useful for improving linking in the blog, which is traditionally where one would expect to find such a system. The tagging system will be combined with a widget that will allow all pages with the same tag to be displayed, and will be available on the website before the next Advisory Board meeting.

Reordering of content

Since the first Institute's website was created in 2010, both the volume of content and its breadth has grown significantly. This is, of course, to be welcomed, but shoe-horning the new content into the organisational structure of the old website has created some less than logical ordering.

A new system for organising content was devised and presented at the last Advisory Board. Since all of our work is based on the themes, it seemed logical that the material we generate should naturally fall into these themes. Hence the new system is based on the four themes plus menus for *about* (which is the logical place to group background information on the Institute), *blog* (which gets it's own menu in an attempt to increase its ease of location and further increase its popularity) and *resources* (which is the logical place visitors look for guides and related materials). Hence the website's new menu bar now lists: About, Blog, Community, Consultancy, Policy, Training and Resources (with additional space for a temporary menu item for special events like workshops or the Fellowship launch).

Each of the menu items is provided with its own landing page, which draws together the content and explains why it is logically connected. The intention here is to provide a page on which people would expect to find logically related material.

Smaller campaigns

CDT campaign

Although the CDT campaign falls into the training theme, the promotion of the campaign was led by the policy theme. The main outcome of this work was many new contacts across UK universities and the harvesting of difficult to obtain contacts, such as those people within universities who are responsible for directing and financing groups within research departments.

Strategy meetings

The agenda for the quarterly operations meeting has always been overloaded. However, most of the issues that people wanted to discuss required only a short discussion and then sign off from the Director and each of the theme leaders. For this reason, the policy theme began running a fortnightly strategy meeting where all staff could meet and discuss issues that required consensus decision.

Five strategy meetings have been held, which have dealt with issues ranging from the creation of a timeline for the Collaborations Workshop 2014 to the type of content that should exist on the theme landing pages. The result of these meetings is that the quarterly operations meeting can now focus exclusively on the larger issues that are best solved face-to-face.

Consolidation and rationalisation of contacts list

As Simon Hettrick devolves responsibility for all communications coming from the Institute, it has been necessary to consolidate the contacts that he uses into a single list. This is far from easy, but it has proven to be a useful endeavour as it has also led others within the Institute to add to the list their own contacts.

We now have a large list of contacts, which has made it far easier for other people in the Institute to raise awareness of their work. However, the list still needs to be rationalised with the information that is necessary for effective use of the contacts: who is interested in what, who should be contacted only with

important information, etc. This rationalisation will be completed by the next Advisory Board meeting.

The Great Licenceathon

The Great Licenceathon was an idea for a campaign where the Institute would publish blogs from researchers about their experiences of learning who owned their code. The idea was to highlight the lack of knowledge about this important area, to raise awareness of the need for clear instructions on licensing issues and to apply some pressure to institutions to make their ownership and licensing policies accessible and clearer.

An early blog post about the Great Licenceathon garnered significant interest and around ten volunteers who were willing to investigate their own institutions and write about their experiences. However, there simply was not enough manpower to run the campaign correctly in parallel the RSE and e-Infrastructure trainers campaign. Hence, the Great Licenceathon has been postponed until 2014.

Links

Research Software Engineer campaign

- Times Higher Education article: http://bit.ly/1ed4Pjc
- The Craftsperson and the Scholar (268 unique hits): http://bit.ly/W6vbqj
- Top ten reasons to be a Research Software Engineer (4,740 unique hits): http://bit.ly/1emSQiT
- If we value software, we need "a fundamental sea-change" in academia (320 unique hits): http://bit.ly/1azrZP9
- Workshop for RSEs webpage (761 unique hits): http://bit.ly/rseworkshop
- Topics discussed at the workshop: http://bit.ly/1da7CFL

e-Infrastructure trainers campaign

- Blog post about the problems with e-Infrastructure training: http://bit.ly/10MhFMO
- List of organisations that attended the workshop: http://bit.ly/16pjxhg
- Workshop for e-Infrastructure trainers webpage (621 unique hits): http://bit.ly/eIWorkshop
- Topics discussed at the workshop: http://bit.ly/1da7CFL

The Great Licenceathon

- Blog post about the concept: http://bit.ly/189BX1R

Training

Highlights

Software Carpentry UK coordination

Software Carpentry[23] is an international initiative to give scientists and researchers the software development skills they need to achieve more, in less time, and with less difficulty. Software Carpentry's primary vehicle is the "boot camp", an intensive, highly-interactive 2 day course in software development skills. Software Carpentry coincides with our strategic goals of promoting reproducible research and improving software development skills within research. Carole Goble is on the Software Carpentry advisory board.

Over the past 6 months we have continued our strategic engagement with Software Carpentry in our role as administrators for Software Carpentry within the UK. This is in addition to our ongoing activities in delivering boot camps and contributing to discussions about how boot camps are run, how their impact can be assessed, how much they cost to run, how they can be funded and how instructors can be identified and supported.

Our contributions have been recognised by the awards of an instructor badge to Aleksandra Pawlik, and organiser badges to Carole Goble and our fellows Alex Chartier, James Hetherington and Robin Wilson[24]. Our involvement has been blogged or re-blogged in at least 21 articles by third-parties or on third-party blogs.

In our role as UK administrators, we have helped to organise 5 boot camps for Manchester, Oxford, Southampton, Bath and Bristol and are helping to organise 7 boot camps for the Institute of Cancer Research (to be held at Greenwich with Software Carpentry's founder, Greg Wilson, instructing), the University of Exeter, the Scottish Universities Physics Alliance (SUPA) (to be held in Glasgow), the University of Edinburgh, the University of Cambridge (the first UK R-based boot camp) and the University of Manchester (the first UK MATLAB-based boot camp), and the GARNet plant sciences network. We have also discussed boot camps for 11 other organisations[25].

In August, EPCC's Luke Tweddle joined the Institute to handle our Software Carpentry administration.

Software Carpentry boot camps

We have helped to deliver 9 boot camps to over 360 post-graduate and post-doctoral researchers, directly contributing to our strategic goal of promoting

---

[23] http://software-carpentry.org

[24] http://software-carpentry.org/badges/index.html

[25] Department of Physiology, Anatomy and Genetics, University of Oxford; Centre for Ecology and Hydrology, University of Oxford; Pirbright Institute; Department of Genetics, University of Cambridge; Department of Geosciences, University of Edinburgh; Department of Computing, Imperial College London; Rutherford-Appleton Laboratory; London School of Economics; Open University; University of Highlands and Islands; National Oceanography Institute, Southampton.

software development skills within research. Every boot camp we've been involved in has been over-subscribed.

We instructed on 3 out of the 5 boot camps run within the UK from March to August. The other 2 boot camps in the UK were organised by our fellows. We also instructed on the first Polish boot camp and on 4 in the USA. This included instructing on the first boot camp for Women in Science and Engineering, in Boston. This was a significant event, attracting over 100 attendees, and was sponsored by Mozilla, Alfred P Sloan Foundation, JP Morgan, NumFOCUS, Python Software Foundation, Microsoft, KAUST and Intel.

We delivered boot camps to varying audiences including both national or regional, open to all or closed to a specific institution/department, and generic or domain-specific. The boot camps were located at:

- 07-08/03 The University of Virginia. A boot camp for researchers from across the university (~33 attendees). Steve Crouch (instructor).
- 18-19/04 The University of Manchester. A boot camp for bioinformaticians (~24 attendees). Carole Goble (organizer), Mike Jackson and Aleksandra Pawlik (instructors), Shoaib Sufi (helper).
- 09-10/05 The University of Oxford. A boot camp for students of the Oxford Doctoral Training Centres (~30 attendees). Mario Antonioletti and Shoaib Sufi (instructors).
- 18-19/05 Jagiellonian University, Krakow. A boot camp for researchers from across the university (~28 attendees). Aleksandra Pawlik (instructor).
- 03-04/06 The University of Southampton. A boot camp for the Institute for Complex System Simulation and Computational Modeling Group (~42 attendees). Steve Crouch and Tim Parkinson (helpers).
- 24-25/06 Women in Science and Engineering, Boston. Three concurrent boot camps for beginner intermediate and advanced programmers (~118 attendees). Aleksandra Pawlik (instructor).
- 01-02/07 University of Oklahoma. A boot camp for researchers from across the university (~40 attendees). Aleksandra Pawlik (instructor).
- 11-12/07 Indiana University. A boot camp for researchers from across the university (~25 attendees). Aleksandra Pawlik (instructor).
- 15-16/07 The University of Bath. A boot camp for attendees at the Beacons Satellite Symposium, which included attendees from UK Meteorological Office and the Plymouth Marine Laboratory and from the universities of Bath, Bristol and Liverpool (~24 attendees). Mike Jackson (instructor), Michael Croucher (helper).

Our fellows organised 3 boot camps:

- 04 and 08/04 University College London. A boot camp for high-energy physicists (~31 attendees). Organised by James Hetherington who also instructed.
- 03-04/06 The University of Southampton. A boot camp for the Institute for Complex System Simulation Institute and Computational Modelling Group (~42 attendees). Organised by Robin Wilson.

- 15-16/07 The University of Bath. A boot camp for attendees at the Beacons Satellite Symposium (~24 attendees). Organised by Alex Chartier.

We also organised and delivered a boot camp "highlights" day at the EGI Forum in Manchester on 11/04. This consisted of 3 standalone sessions (~20 attendees each), each covering a skill taught on a boot camp.

In the coming months our fellows and ourselves will be involved in the following boot camps:

- 24-25/10 Institute of Cancer Research, Cancer Research and other researchers at the Digital Enterprise Hub, Greenwich. James Hetherington (instructor).
- 03-04/12 PRACE Advanced Training Centres at The University of Edinburgh. Mike Jackson (organiser, instructor) and Mario Antonioletti (instructor).
- 07-08/01 Department of Applied Mathematics and Theoretical Physics, The University of Cambridge. Aleksandra Pawlik (instructor). This will be the first UK R boot camp and is very relevant to our collaboration with ELIXIR.
- 13-14/01 The University of Manchester. Aleksandra Pawlik and Shoaib Sufi (instructors). This will be the first UK MATLAB boot camp and will be run in association with MathWorks. Our fellow Michael Croucher is organizer.
- 08-10/04 GARNet plant sciences network. We will provide an instructor as part of our collaboration with ELIXIR.
- Centres for Doctoral Training as part of our EPSRC-approved training offering, see below.

Software Carpentry UK sustainability

We have been encouraging UK researchers to engage with Software Carpentry and join the volunteer pool of instructors. One of our past Institute collaborators, Christopher Woods from the University of Bristol, volunteered to become involved. Chris instructed at the Bath boot camp, was awarded a badge at our recommendation, instructed at Bristol and will be instructing on the upcoming Exeter boot camps. James Morrison, from the University of Highlands and Islands, who attended our boot camp in Manchester, stepped in at short notice to instruct at the Southampton boot camp (after an instructor had to pull out) and will instruct on the upcoming SUPA boot camp. David Martin, from the University of Dundee, who helped at our boot camp in Bath, will be an instructor at Exeter. Elliot Marsden and Fergus Cullen from The University of Edinburgh who attended the Edinburgh boot camp in 2012, served as helpers at the Bath boot camp.

Software Carpentry have introduced a $1500 administration fee to cover the overheads of identifying instructors, setting up registration and mailing lists and

other costs. For UK boot camps this cost is waived, as this cost is covered by the Institute. UK hosts are made aware of this.

Software Carpentry and Centres for Doctoral Training

At the advisory board meeting of February 2013, it was suggested that the Institute should consider using the EPSRC call for Centres for Doctoral Training as an opportunity extend our software training portfolio. We offered two training packages to potential applicants: a Software Development for Researchers package in which we provide a boot camp for applicants; and, a Trainers package in which we train attendees in how to organise and deliver a boot camp. Each packages was offered to applicants for £12,000. Full details of our offering were posted on our web site[26] and information e-mailed to 79 contacts at current CDTs (from the EPSRC web site) and 23 Directors of Research at Russell Group universities on 04-05/03. A follow-up email was sent on the 26/03, before the stage 1 applications.

4 groups contacted us directly, 2 of whom were already Institute collaborators (at Oxford and Southampton) and one at an Institute partner institution (at Edinburgh). The one external contact was from UEA. At this stage, Oxford and UEA wrote us into their bids and we provided letters of support.

Once the successful Stage 1 proposals were announced, we sent another email to all 176 PIs who passed to Stage 2 on 06/06. We also attended the CDT meeting in Swindon on 07/06 and distributed fliers. We were approached by 20 CDT consortia, including 4 as a result of communication from the first round (although only 2 of the bids we supported in Stage 1 were successful). Of these 20, 3 wrote to confirm that they will offer boot camps themselves, 3 we passed on to other training providers (as they sought specialised courses outwith the Software Carpentry remit) and 7 came to no result. Consequently, we will be offering our Software Development for Researchers package or providing advice to 10 consortia:

- Theory and Modelling in Chemical Sciences - Bristol
- Resilience of Infrastructure and Communities through Multidisciplinary Fire Science and Engineering – Edinburgh
- Water Informatics: Science and Engineering – Exeter
- Cross-Disciplinary Approaches to Non-Equilibrium Systems – Kings College London
- Cloud Computing for Big Data – Newcastle
- Biomedical Imaging – Oxford
- Systems Approaches to Biomedical Science – Oxford
- Synthetic Biology – Oxford
- Next Generation Computational Modelling – Southampton
- Web Science Innovation – Southampton

Training for e-infrastructure

[26] http://www.software.ac.uk/Training-For-CDTs

Since April, we have been leading an activity to build a community of UK e-infrastructure training providers. This has its roots in a talk on "The provision of training for e-infrastructure users and potential users" given by Simon Hettrick at the UK e-Infrastructure Academic User Community Forum on 15/04. This talk garnered significant interest amongst UK e-infrastructure training providers. There was a recognition that, by pooling knowledge and expertise, training providers would be ideally placed to provide recommendations on training provision in the UK to EPSRC, the e-Infrastructure Leadership Council, the e-Infrastructure Project Directors Group and the RCUK's National e-Infrastructure Council. The goal is to contribute to improving UK e-infrastructure training.

Under the direction of Simon Hettrick and Claire Devereux of STFC, representatives of training providers will write essays on their experiences, their challenges and recommendations as to how these challenges can be addressed to improve UK e-infrastructure training. These essays will together form a training report to be submitted to the bodies named above. The current authors of the report, and the training providers they represent, include Claire Devereux (STFC, NES, SeIUCCR, EGI), Ian Bush (NAG), David Henty (EPCC, PRACE), Mike Jackson (Software Sustainability Institute, Software Carpentry), Dave Cable (Hartree Centre), Harpreet Dhanoa (DiRAC), Joy Davidson (DCC) and Mark Rodger (HPC Short Courses, MidPlus Regional HPC Centre).

The Institute has also run a training workshop[27] at the Hartree Centre on 14/08. This is planned as the first of a series of workshops to build a community of training providers, to implement recommendations from the training report, and to discuss best practice. Demand for this first workshop was so great that the original 50 places sold out within 3 weeks so an additional 10 spaces were made available. The next steps for the training community include identifying the right way of running future workshops (perhaps through a Special Interest Group).

DiRAC driver's licence

In collaboration with Jeremy Yates (DiRAC project director, UCL), Harpreet Dhanoa (DiRAC project manager, UCL), James Hetherington (Research Software Development Team Leader, UCL) and Andrew Turner (Project Manager, EPCC), we completed development a "driver's test" for the DiRAC (Distributed Research utilising Advanced Computing)[28] consortium. This test is a software development skills aptitude test.

We held a 3rd dry run of the test with Dr. Yates' researchers at University College London, on 15/05. James Hetherington and the representatives of other DiRAC sites attended this dry run so they could be introduced to the test, its aims and its execution. Following this dry run, we worked with Jeremy and Harpreet to plan the roll out the driving test across the DiRAC consortium. The roll out is now under way. Training coordinators at the DiRAC regional sites in Durham, London, Leeds, Edinburgh and Leicester plan to collectively run through 70 post-doctoral research associates, 130 PhDs and 40 other new users by the end of this

[27] http://www.software.ac.uk/workshop-e-infrastructure-trainers
[28] http://zuserver2.star.ucl.ac.uk/dirac/home.html

year between 16th September and early December[29]. The ongoing cohort is estimated to be 30-40 a year.

Both the Institute and Software Carpentry will continue to work with Jeremy, and DiRAC, in evaluating the effectiveness of the driving test. We have introduced Jeremy to other organisations who are interested in driving tests for their communities. This includes the Australian National Data Service, the US National Energy Research Scientific Computing Center (NERSC) and the University of Washington.

ELIXIR

The Universities of Manchester and Oxford have led the submission to BBSRC of a proposal to set up an ELIXIR[30] training node within the UK. This proposal includes the development of training materials and delivery of boot camps for bioinformatics, to meet the pressing demand in this domain. This activity will also contribute to scaling up the UK instructor pool for boot camps and promoting sustainability of Software Carpentry within the UK. The Institute is continuing to work closely with ELIXIR in the planning and delivery of this capability.

Carole, Neil and Mike participated in the first ELIXIR-UK/GOBLET Trainers workshop at The Genome Analysis Centre, Norwich, 25-26/03 which was attended by over 50 ELIXIR/GOBLET members. A follow-up meeting is proposed for early November to elaborate a training strategy for ELIXIR, to which we will contribute.

As part of our involvement in discussing a Software Carpentry boot camp for the GARNet [31] plant sciences network, we introduced the organisers to our ELIXIR collaborators. It is now planned that ELIXIR will provide, along with ourselves, an instructor to the GARNet boot camp in April 2014.

Training services

At the invitation of Matthew Smith from the BES Computational Ecology Specialist Interest Group and Microsoft Research, and Greg Wilson from Software Carpentry, we ran a "What Makes Good Code Good?" discussion workshop at INTECOL13. Over 100 delegates attended the workshop. We also gave a talk on "Robust extensible ecological models and methods" as part of a symposium on "Not just for geeks: broadening scope and participation in predictive ecology". Over 30 delegates attended the talk.

We delivered a talk on "Why should scientists understand how to write better software?" and ran a workshop on "Managing Sustainability into Software" at the SeIUCCR Summer School 2013 (27-30/08) to 24 attendees. This is the fourth time we have delivered the "Managing Sustainability" workshop for this summer school series.

[29] https://www.dirac.ac.uk/training.html
[30] http://www.elixir-europe.org/
[31] http://www.garnetcommunity.org.uk/

As part of our web site redesign, our training pages[32], which describe the training services that we offer the UK research community, have been updated.

Challenges

Since the last report we have brought 3 more Institute members on-stream as Software Carpentry instructors. However, the demand for boot camps has expanded in parallel. However, staff were less overloaded with boot camps for the last 6 months and the need (rather than the desire) for staff to instruct on more than two boot camps per quarter was reduced. The contributions of James Hetherington, Christopher Woods, James Morrison and David Martin to boot camps across the UK have also helped with satisfying demand.

Effort

The effort available for training activities is fine. The only issue, as noted above, is how this effort is distributed across staff. Committing to instructing on too many boot camps can and does limit the effort we can provide to the Institute's other activities at specific times.

Dissemination

Our blog posts

- 15/03 Snowstorms and blackouts no match for Software Carpentry in Virginia! – Steve Crouch
- 16/04 Software Carpentry highlights from the EGI Community Forum – Mike Jackson
- 19/04 Software Carpentry makes its debut in Manchester – Mike Jackson
- 22/04 Don't forget the people – a fractured training landscape – Simon Hettrick
- 15/05 The Oxford Doctoral Training Centre gets Software Carpentry – Mario Antonioletti
- 31/05 Software Carpentry bootcamp in Krakow – Aleksandra Pawlik
- 10/06 Software Carpentry at Southampton – Mike Jackson
- 28/06 Software Carpentry Bootcamp for Women in Science and Engineering in Boston – Aleksandra Pawlik
- 16/07 Software Carpentry bootcamp at the University of Oklahoma – Aleksandra Pawlik
- 24/07 Software Carpentry in Bath – Mike Jackson
- 24/07 Software Carpentry bootcamp at Indiana University – Mike Hansen, Jeff Shelton, Aleksandra Pawlik
- 25/07 Workshop for e-Infrastructure trainers. Sold out! And more places added... – Simon Hettrick
- 29/07 Why scientists and engineers must learn programming – extract from article by Philip Guo, Assistant Professor of Computer Science, University of Rochester
- 30/07 DiRAC driving test ready to roll – Mike Jackson
- 31/07 Should we be scared of choosing an OSS licence? – Simon Hettrick

[32] http://www.software.ac.uk/training

- 23/08 What makes good code good at INTECOL13 – Mike Jackson
- …plus 14 training-related news items

Our top tips and guides

- 13/05 Top tips for packaging software - Steve Crouch
- 16/05 Top tips for choosing a computing infrastructure - Mike Jackson
- 22/05 Top tips for porting an application to a computing infrastructure - Mike Jackson
- 07/06 Top tips for working with a technical writer - Craig Haiss, HelpScribe
- 03/07 Building a good working relationship - top tips - Steve Crouch and Neil Chue Hong
- 19/08 Choosing an open source licence - Neil Chue Hong and Tim Parkinson

Our Software Carpentry blog posts

- 16/04 Feedback from the EGI Forum – Mike Jackson
- 17/05 Experiences with the Oxford DTCs – Mike Jackson
- 30/05 Krakow bootcamp experience – Aleksandra Pawlik
- 02/06 Software Carpentry at INTECOL13 – Mike Jackson
- 02/06 From a helper to an instructor – Aleksandra Pawlik
- 15/06 Boot Camp in Bristol, September 12-13, 2013 – Mike Jackson
- 18/07 DiRAC driving test ready to roll – Mike Jackson
- 18/07 Feedback from Bath – Mike Jackson
- 23/08 What makes good code good at INTECOL13 – Mike Jackson

Blog posts by others promoting, referencing or reblogging us or our our posts

- 19/04 Software Carpentry Bootcamp Manchester, PJB, ironic_cog blog. Cites Aleksandra, Carole and Mike
- 24/04 Manchester Once Again – Greg Wilson, Software Carpentry blog
- 24/05 Feedback from the Oxford DTCs – Jonathan Cooper, Software Carpentry blog. Cites Mario, Shoaib, and the Institute with links
- 04/06 Six useful things I have learnt at Software Carpentry boot camp – Richard Edwards, Cabbages of Doom.
- 05/06 Feedback and experiences from Southampton – Robin Wilson, Software Carpentry blog. Cites Institute and fellowship
- 20/06 Software Carpentry: Lessons Learned – Greg Wilson, Software Carpentry blog. Contributions from Neil, Steve, Mike
- 27/06 Boston WiSE Software Carpentry boot camp – Terri Yu, terriyu@/media/blog. Cites Aleksandra and the Institute with links
- 04/07 British Ecological Society Computational Ecology Group, Newsletter 14. Cites INTECOL13 “What Makes Good Code Good?” workshop and “Robust extensible ecological models and methods” talk, Steve, Mike and the Institute

- 05/07 WiSE Bootcamp Roundup – Greg Wilson, Software Carpentry blog. Cites Aleksandra and Institute with links
- 06/07 Workshop for e-infrastructure trainers – Greg Wilson, Software Carpentry blog. Cites Simon and Institute with links
- 16/07 Computational Competence for Biologists – Greg Wilson, Software Carpentry blog. Links to our DiRAC dry run two blog post
- 26/07 Report on the Indiana Bootcamp – Greg Wilson, Software Carpentry blog. Cites Aleksandra and Institute with links
- 19/08 How to get the most out of INTECOL – Rob Salguero-Gómez, Journal of Ecology and Wiley blogs. Cites "what makes good code good" workshop

Other publications, news articles etc.

- 01/03 EPSRC approved software training for Centres of Doctoral Training – Simon Hettrick, web page promoting our CDT offering
- 16/04 Software Carpentry highlights at the EGI Forum – Mike Jackson, EPCC blog
- 24/04 Don't forget the people - a fractured training landscape – Simon Hettrick, EPCC blog
- 09/04 Don't forget the people - a fractured training landscape – Simon Hettrick, GridCast blog
- 14/05 Software Carpentry @ Oxford DTC – Mario Antonioletti, EPCC blog
- 04/06 Hammering on with Software Carpentry, EPCC News 73, Summer 2013, Mike Jackson
- 19/07 Software Carpentry in Bath – Mike Jackson, EPCC blog
- 25/07 DiRAC driving test ready to roll – Mike Jackson, EPCC blog
- 23/08 What makes good code good at INTECOL13 – Mike Jackson, EPCC blog
- EPSRC-Approved Software Training for Centres of Doctoral Training – Greg Wilson, section on Software Carpentry's Partnerships web page