

A Programme for the Development of Computational and Digital Research Capacity in South Africa and Africa - version 1

Project Summary

One of the biggest challenges in research these days is to ensure that our students and researchers are equipped with skills to handle the ever increasing size of their datasets and use technology efficiently to perform reproducible analysis.

To address this issue, several international initiatives have been established over the last few years.

This proposal aims to address the computational and data handling (also termed Research Data Science) skills scarcity that exists amongst researchers and postgraduate students in South Africa and Africa. The programme will run over 12 months and will train 24 new instructors, up to 400 workshop participants, and build skill sharing user communities which can include even more researchers and students.

The programme will partner with established international organisations such as Software Carpentry (SWC), Data Carpentry, and the Mozilla Science Lab (MSL) to harness resources that are already developed and are successfully used across the globe to address these same issues. The materials that will be used in this programme are all published under the Creative Commons license and we'll specifically focus on building skills around tools where license fees may not be a hindrance to adoption in resource-poor environments.

Since 2014, groups such as University of Cape Town Information and Communication Technology Services (UCT ICTS) and UCT eResearch, North-West University Information Technology department (IT@NWU), Talarify, the South African Environmental Observation Network (SAEON), Stellenbosch University Genetics Department, PyConZA and several researchers and other individuals have committed hundreds of thousands of rands and hundreds of hours to run or participate in Software Carpentry (SWC) workshops in South Africa. To date, more than 200 researchers, research support staff, and students from more than 18 institutions in South Africa have participated in locally hosted workshops. The proposed programme will build on investments already made and wishes to provide a framework to build even more momentum over the 2016/2017 period in terms of computational and digital skills capacity and community building.

A Programme for the Development of Computational and Digital Research Capacity in South Africa and Africa



[A Programme for the Development of Computational and Digital Research Capacity in South Africa and Africa](#)

[Project Summary](#)

[Project Narrative](#)

[Background Information](#)

[The Software Carpentry Foundation \(SCF\)](#)

[Data Carpentry \(DC\)](#)

[The Mozilla Science Lab \(MSL\)](#)

[Statement of Need](#)

[Goals](#)

[Objectives](#)

[Activities Timeline](#)

[Project Partners & Key Responsibilities](#)

[Budget Request](#)

[Instructor Training Workshop](#)

[Software/Data Carpentry Workshops](#)

[One-year Follow-up Event](#)

[References](#)

Project Narrative

Background Information

Over the last few years, research datasets in all domains have been increasing in size at an unprecedented rate due to advances made in the technology that generate these datasets. Not only have datasets for single projects been increasing in size, but researchers are also starting to combine a greater variety of datatypes in their interdisciplinary analyses.

Researchers and postgraduate students from diverse areas of study, including humanities, engineering, life sciences, physical science, and medicine, need to acquire new skills as data scientists that mostly do not currently form part of the formal curricula of undergraduate and even

postgraduate training programs. Communities of practice have been shown to enhance skill sharing and consolidation of newly acquired skills as supplement to formal and informal training activities.

Recent incidents of scientific paper retractions have highlighted the need for training in practices that support reproducible research.

Several organisations aim to address aspects of the aforementioned issues. In the following section we provide more information about organisations that have been identified as potential partners in the proposed programme.

The Software Carpentry Foundation (SCF)

The *Software Carpentry Foundation* teaches researchers basic skills for research computing [1]. The first version of Software Carpentry ran in 1998 at the Los Alamos National Laboratories [2]. Over the course of the last 18 years the organisation evolved and in 2014 the Software Carpentry Foundation was established [3]. Since 2011, more than 16'000 researchers and postgraduate students from six continents have gone through training and they now have more than 400 qualified instructors worldwide [4]. The SCF also started a new programme where instructor trainers are trained to remove bottlenecks caused by a lack of qualified instructors [5].

Software Carpentry lessons are collaboratively developed and published under the Creative Commons license. Current lessons include material on automation using the Linux shell and Make, tools such as git and GitHub for version control and collaboration, databases with MySQL, and programming with Python, R, and MATLAB [6].

Workshops are mostly run in a two-day hands-on format with instructors live coding and helpers circulating in the classroom to help learners who fall behind or have specific questions [7]. More recently, members of the Software Carpentry community have been experimenting with running the workshops over four half-days or even as semester courses [8].

It is very clear that activities in the Software Carpentry community have a huge positive impact on researchers and postgraduate students in terms of gaining computational skills and even beyond that [9, 10].

Data Carpentry (DC)

Data Carpentry aims to make data science more efficient. Whereas Software Carpentry lessons are subject agnostic and workshops typically include participants with various research backgrounds, Data

Carpentry specifically focuses on high-quality, domain-specific training for researchers working in data-intensive research areas [11].

In 2015, Data Carpentry received USD 750'000 in funding from the Gordon and Betty Moore Foundation to support activities such as (i) expanding into new domains; (ii) training more researchers; (iii) increasing the number of Data Carpentry instructors; and (iv) developing strategies for delivering their material online [12].

Current material on offer include workshops with a biological focus looking at the whole lifecycle of data processing and include tips for organising data properly in Excel, managing data in SQL, and using either R or Python in combination with these tools. A Genomics workshop is also available and include lessons on data organisation, cloud computing, command line usage, and data visualisation in R [13]. Lessons currently under development include material for Geospatial Data Carpentry, Social Sciences Data Carpentry, Data Visualisation, and Working with APIs [14].

The Mozilla Science Lab (MSL)

The *Mozilla Science Lab* is working towards making research more collaborative, accessible, and usable through the Internet [15]. One of their key initiatives is the Study Groups programme. Mozilla Study Groups are informal learning communities where peers share skills or discuss challenges experienced in specific analyses (or more broadly) and assist each other in finding solutions. The Study Groups can either be focussed on a specific study area (e.g. genomics) or on the use of specific sets of tools (e.g. R). To date, there are 19 Study Groups around the world, with more joining the network and community each day. In total, this group led over 140 events and 46 open lesson plans at over 20 sites in less than a year.

Mozilla provides monthly calls, mentorship, and scaffolding for creating event sites that come with Codes of Conduct and a handbook and lessons for getting started [16]. A website template is included as part of the resources in the handbook and study groups can merely customise the information to display the aim of their own group and showcase their events [17]. The Genomics Hacky Hour Study Group is an example of a study group running at the North-West University [18]. The group has been active since September 2015 and group events are attended by between five and 10 participants weekly [19].

In February 2016, Mozilla Science Lab invited several researchers from across the globe to participate in the Working Open Workshop and subsequently become part of the first ever Open Leadership Cohort [20]. Mozilla Science Lab supports study groups through mentoring, small grants, running follow-up events, and more.

Statement of Need

A large proportion of emerging and established researchers as well as undergraduate and postgraduate students and postdoctoral research fellows in Africa do not have the skills to deal with the research data deluge. The phenomenon is a reality in diverse research domains such as social, physical, and life sciences, as well as engineering, medicine, and beyond and is not isolated to a specific domain, institution, or career stage. The problem is not confined to developing countries or the African continent, as can be seen from the development of initiatives such as Software and Data Carpentry and the Mozilla Science Lab's Study Groups.

In order to equip our researchers to participate in reproducible research and continue to be internationally competitive and collaborators of choice for multidisciplinary, multi-institutional projects, a concerted effort will have to be made to elevate researchers' data handling and computational skills.

Skills and knowledge acquired through participation in workshops run abroad or away from researchers' and students' home institutions are often not transferable, as individuals do not have access to environments aligned with the workshop environment and are often very isolated and not able to maintain their new skills. It is imperative that the broader university community are informed of and involved in training initiatives of this extent. This includes university IT departments, research offices, libraries, as well as research management. IT departments should specifically be aware of ways to support the needs of the 21st century researcher to perform cutting-edge research and should consider new ways to deploy infrastructure and hardware to support these newly required and acquired skills.

Not only is it essential to have access to the necessary computational infrastructure and hardware, but it is also vitally important to help foster active communities of practice at universities to provide informal though regular opportunities for those who are actively participating in projects that need support beyond what can be provided in the more formal university structures. In South Africa, very few such informal study or user groups currently exist in academic environments and it is expected that mentoring and demonstration of successes by existing study or user group communities will be hugely important in adoption of these practices.

Goals

The programme aims to address the skills scarcity in data scientists, who can apply their domain knowledge to the handling and analysis of data. The target audience includes researchers from all career stages, including postgraduate students, as well as emerging and established researchers. The

programme will be open to researchers from any research institution in Africa with an initial limit on non-South African applicants until more funding can be sourced from other organisations.

Instead of running isolated workshops to teach basic data analysis and data handling skills to researchers and postgraduate students and sending participants back to unsupportive home environments, we aim to offer a 12-month programme that will include:

- a kick-off event to train 20 new instructors from a pool of people who've already been exposed to the Software Carpentry model;
- support for these newly qualified instructors to run workshops at their home institutions with up-front buy-in from their home institutions;
- access to resources and mentoring through Mozilla Science Lab to support establishing and running of communities of practice at newly qualified instructors' home institutions;
- a one-year follow-up event to provide participants in the original instructor training event, as well as highly active community members from newly formed study groups, an opportunity to share their experiences and challenges and develop a proposal for the way forward.

The instructor training workshop will consist of the typical two-day instructor training preceded by an ice-breaker event the evening before the workshop starts. The third day will focus on community building practices including discussion of Mozilla Science Lab Study Group resources and feedback from community leaders about their experience in community building.

To maximise the possibility of successfully training instructors and building communities of practice around them, the programme will build on work done over the past 12 months and specifically solicit applications from institutions and individuals who have already expressed interest to support these initiatives. Several workshops were ran in South Africa and elsewhere in the period 2014/2015 and more than 200 participants have already participated in a Software or Data Carpentry event. Applications will specifically be solicited from that audience and applicants will be expected to provide proof of support from their institution to run a workshop within 12 months of completing their training. As Software Carpentry workshops typically involve at least two instructors, applicants will have to apply in teams of two to four to ensure the lonely applicant does not return to their home institution without a teaching buddy with whom to run a workshop.

If this initiative can be funded and delivered successfully, we will increase the number of qualified instructors on the African continent from nine to 29. These instructors will have access to support from their institutions to run workshops, as well as from the team initiating the grant application, the broader Software and Data Carpentry communities, and the Mozilla Science Lab staff responsible for mentoring Study Group leaders.

As mentioned earlier, newly qualified instructors will be expected to run a Software Carpentry workshop within 12 months from completing their training. Workshops can typically accommodate 20 to 40 participants. The total number of learners that would be able to participate in this initiative can thus be as many as 400 over a 12 month period, should the ten pairs of newly qualified instructors each run a workshop for 40 participants.

Combining the workshops with study groups where participants can get support after attending a workshop will increase the reach of this programme. Not only will workshop instructors and participants be able to attend study group meetings, but people who did not participate in the workshop may also be drawn to the study groups or be prompted to start their own study groups possibly with a more specific focus on subjects of their choice.

Lastly, the programme aims to bring together all newly qualified instructors, as well as two or three of their most active study group participants one year after the initial instructor training event to discuss lessons learned, progress, and new ideas to improve future collaborative learning skills. This will not only serve as a networking event, but also to follow-up effectively and continue nurturing future affiliations which provide value. Incidentally, the drive for computational and digital skills capacity and community building will be reinforced, thus maximising the event impact.

References

1. Software Carpentry – <http://www.software-carpentry.org>
2. Software Carpentry: lessons learned version 1 (Feb 2014) - <http://f1000research.com/articles/3-62/v1>
3. Software Carpentry: lessons learned version 2 (Jan 2016) - <http://f1000research.com/articles/3-62/v2>
4. Recent Statistics - <http://software-carpentry.org/blog/2015/10/recent-stats.html>
5. Our New Instructor Pipeline - <http://software-carpentry.org/blog/2016/02/new-instructor-pipeline.html>
6. Software Carpentry Lessons - <http://software-carpentry.org/lessons/>
7. Software Carpentry Workshops - <http://software-carpentry.org/workshops/>
8. Software Carpentry as University Course - <http://software-carpentry.org/blog/2016/02/swc-as-a-university-course.html>
9. Software Carpentry get more done in less time (2015)- <http://arxiv.org/abs/1506.02575>
10. Value of Software Carpentry to Instructors Report (2016) - <http://software-carpentry.org/blog/2016/01/instructor-survey.html>
11. Data Carpentry - <http://www.datacarpentry.org/>
12. Data Carpentry receives grant from Moore Foundation - <http://www.datacarpentry.org/blog/moore/>
13. Data Carpentry Existing Lessons - <http://www.datacarpentry.org/workshops-offered/>
14. Data Carpentry Workshops Under Development - <http://www.datacarpentry.org/workshops-development/>
15. The Mozilla Science Lab - <https://mozillascience.org/about>
16. Mozilla Science Lab Study Group Handbook - <http://mozillascience.github.io/studyGroupHandbook/>
17. Mozilla Science Lab Study Group Website Template - <https://github.com/mozillascience/studyGroup>
18. The Genomics Hacky Hour Study Group at NWU - <http://anelda.github.io/GenomicsHackyHour/>
19. Building Genomics Capacity through a Study Group - <http://www.nwu.ac.za/eresearch/news/building-Genomics-capacity-through-a-study-Group>
20. Working Open Workshop & Open Leadership Cohort - <https://www.mozillascience.org/wow-introducing-working-open-workshops-and-the-open-leaders-cohort>