Data Carpentry year 1 Narrative Report

Description of Data Carpentry's Progress (August 2015 - July 2016)

With the support of the Moore foundation, Data Carpentry has been able to greatly amplify our reach and build infrastructure and programs for sustainability over the past year.

Workshops

Since August, we have run 58 workshops in ten countries, serving approximately 1,100 learners. By careful growth of our instructor pool, we have been able to increase the frequency of our workshops (from 2/month in 2015 to 9/month in the past quarter), as well as expand our geographic reach. We now run workshops in the Americas, Europe, Asia and Oceania.

We have continued to assess the effectiveness and impact of our training developing and using pre and post-workshop surveys. We have over 600 completed surveys that show the demographics of our audience as well as workshop outcomes. Workshop attendees are primarily graduate students (31%), but there is also research staff (25%) and postdoctoral researchers (12%) and PIs (12%). Workshop attendees were also primarily from the life sciences (45%), with other (21%) and social sciences and library sciences (9%). The life science emphasis is likely due to our initial focus on lessons for ecology/biology.

From demographic data (collected only in the United States) we see that in the US, we are primarily reaching Caucasian (63%) and Asian/Pacific Islander (18%) participants. Black or African American (3%), Hispanic (6%) and American or Alaskan Native (0.2%) are under-represented. We are working to address this, primarily by working with diversity efforts in CyVerse and with the SACNAS community. We are running workshops in South America, Asia and Africa where a greater proportion of learners are likely Black or Hispanic. We offer workshop fee waivers to universities in underserved communities.

Over 50% of learners were new to programming having never programmed or less than once per year. A large majority (88%) frequently used Excel or another spreadsheet program to manage data, and one third were unsatisfied or very unsatisfied with their current data management strategy and analysis workflow. There was very good gender balance in these workshops (51% female / 47% male). 39% of learners said their level of data management and analysis skills was higher or much higher after the workshop, and another 51% said somewhat higher. As an overall outcome, 62% responded "agree or strongly agree" to their ability to apply their skills immediately after the workshop.

Instructor training

We have partnered with the Software Carpentry Foundation to develop an instructor training program - providing our new instructors with a solid foundation in evidence-based teaching practices. Since establishing this program in October 2015, we have run 30 instructor training events. Our Executive Director, Tracy Teal, is a certified Instructor Trainer, helping to develop the instructor training program and has run or taught at five of these instructor training events. We now have more than 200 certified Data Carpentry instructors prepared to teach workshops around the world. Our instructor community is highly engaged, with more than 80% teaching at least one workshop within their first year.

Workshop curriculum

Along with expanding our instructor pool, we have also grown our curricular reach. Our first workshops were targeted towards researchers in the life sciences (specifically ecology). We have since worked with CyVerse (formerly iPlant) to develop lessons focused on genomics and have partnered with the National Ecological Observatory Network (NEON) to develop a curriculum serving the geosciences community. These materials are now an active part of our curricular offerings. Data Carpentry is also a part of groups developing full-length curricular materials targeted towards researchers in the social sciences community, as well as for the library and information sciences community. These materials are currently being tested with targeted groups of learners and are expected to be released as part of our curricular core in the coming months. Work is ongoing to develop materials in image analysis, High Performance Computing and Astronomy.

All of materials continue to be developed collaboratively in Github under a CC-BY license. As one example of this collaborative effort, our Ecology lessons have 166 unique contributors and 276 github forks. Materials are used in all our workshops. We also have extensive anecdotal evidence that these lessons are being used by people outside of Data Carpentry instructors to teach in their classes, workshops or seminars or for self-guided learning. These use cases demonstrate the value of available, accessible and discoverable lesson materials. This type of use is difficult to track quantitatively however, as hits to the web sites are not representative of use for learning.

Strategic Planning, Sustainability, Communication and Advocacy

We have also worked on strategic planning and developed several important strategies and policies for the sustainability of the organization. These include our Lesson Roadmap for developing new lessons (http://www.datacarpentry.org/lessons-incubation/) and our Partnership program (http://www.datacarpentry.org/partnerships/), both developed in collaboration with and adopted by Software Carpentry. We also developed a policy for how workshops can be branded 'Data Carpentry' (http://www.datacarpentry.org/workshops/) and the criteria for self-organized workshops (http://www.datacarpentry.org/self-organized-workshops/). We created a Data Carpentry instructor FastTrack program to train existing Software Carpentry instructors to teach Data Carpentry (http://www.datacarpentry.org/blog/fasttrack/). We have been strongly involved in improving the instructor training program and developing a Mentorship subcommittee to support new and current instructors.

Our Executive Director, Tracy Teal, has been involved in promoting Data Carpentry and advocating for training in data skills and good data skills and perspectives for effective and reproducible research through invited talks and publications (see Events and Outputs list). She has been invited to participate in working groups and give presentations at these meetings. From these working groups there have been several papers (in preparation and in review) and grant opportunities. A grant to the NSF where Teal was a PI was recently not funded, but got high scores and will be resubmitted. Teal also has been interviewed by Nature and PLOS for inclusion of Data Carpentry in articles or podcasts. She is an Editor of the newly formed Journal of Open Source Software and on the Board of Directors at NumFOCUS. These activities engage Data Carpentry with the open source scientific community and link scientific software production to training. She was involved in authoring a funded grant to the Sloan Foundation on "Building Capacity for NumFOCUS Open Source Projects to Develop and Sustain Industry Relationships". This grant will help support Data Carpentry's efforts to develop relationships with industry for funding support and become more effective at developing business and sustainability plans.

Finally, we have formalized the Data Carpentry Vision and Mission statements to reflect what our organization aspires to achieve as well as how we are working to make the vision possible. We place importance on the way that we work as well as the outcomes of our activities and efforts.

Vision: Building communities teaching universal data literacy

Mission: Data Carpentry trains researchers in the core data skills for efficient, shareable, and reproducible research practices. We run accessible, inclusive training workshops; teach openly available, high-quality, domain-tailored lessons; and foster an active, inclusive, diverse instructor community that promotes and models reproducible research as a community norm.

Challenges and Opportunities Encountered

With our expansion into new content areas, we now have an extended community with members spanning a broad range of technical skill levels. This has required us to reevaluate our established mechanisms for curricular development and community interaction, both of which currently make heavy use of a specific technology not widely used outside of the computer sciences community. We are working towards integrating a more domain-neutral technology, to enable researchers from all fields to be fully active participants in our community. Similarly, rapid expansion of our instructor pool has raised questions about how best to ensure a consistent and high-quality educational experience for all of our workshop participants. Although learners consistently rate our instructors very highly in our post-workshop survey (with an average score of 4.4 out of 5), we are currently developing processes to ensure that we are able to maintain this high level of teaching quality as we continue to grow. Developing and establishing a quality assessment and assurance process for our instructors will be a core goal of the new Deputy Director of Assessment, who has just been hired and will start on August 22, 2016. This work will be carried out in conjunction with our Associate Director, who has a background in instructor training and evaluation of teaching effectiveness.

We are continuing to work on strategies for delivering content online. We have delivered workshops remotely to multiple locations effectively, but this format still does not scale to meet demand. We have been talking with DataCamp about the potential for providing access to lessons in a hands on, interactive way through their platform. As their platform provides for this delivery model and is based on lessons through Github, as we have already developed, there is good potential alignment.

While we have made good progress on developing financial sustainability with workshop fees and Partnerships, we continue to work to develop a comprehensive business plan that includes projections for costs and expenses. We have had challenges getting accurate financial information from NumFOCUS, partly due to their transition to a new accounting system. We continue to work with them for effective financial tracking, but are now also handling the tracking of our own accounting and coordinating and developing an invoice tracking system that NumFOCUS can use to keep us updated.

Personnel Changes

In October, 2016, we hired Maneesha Sane as our Program Coordinator. She was the current Software Carpentry Program Coordinator, and Teal had been involved in her hire for Software Carpentry. She is a Software and Data Carpentry instructor, had experience in program management and in volunteer community coordination. Since her hiring, we now have partnerships with the Software Sustainability Institute and Compute Canada, where their personnel coordinate workshops in these regions. She oversees these other workshop coordinators.

In May 2016, we completed recruitment and hiring of our new Associate Director, Dr. Erin Becker. With a background in both microbial genomics and educational research, Dr. Becker brings to the team a host of relevant skills. Dr. Becker's background enables her understand the needs of both our established learner base (primarily life scientists) and our growing social sciences community. She is actively working to simplify and improve communication channels, including website communications, and to develop and strengthen our instructor community.

We have hired Dr. Kari Jordan as our Deputy Director of Assessment to refine and continue our assessment efforts. Dr. Jordan has a background in engineering and STEM Education with an undergraduate degree in mechanical engineering and PhD and postdoc in STEM Education with her work focused on evaluating the factors that influence faculty adoption of evidence-based instructional practices and self-efficacy in learning. She has served on the board of the National Society of Black Engineers and the Ohio Diversity Council.

In our hiring we have recruited and interviewed diverse candidates. For both the Associate Director and Deputy Director of Assessment position, interview rounds had both men and women candidates from different regions and research domains. With the full staff now hired, the staff is all women with a variety of backgrounds. The Executive Director and Associate Director are Caucasian, Program Coordinator is Asian and Deputy Director of Assessment is African American. Two have a background in biology (ED and AD), one in engineering (DDA) and one in Humanities (PC).

Evidence of Awareness and Recognition

Our learners routinely report that workshops were useful in improving their ability to conduct research in their fields. The vast majority (>90%) of those responding to our post-workshop survey say that participating in the workshop was worth their time and led to improvements in their data management and data analysis skills. Further, about two-thirds felt ready to immediately apply what they learned in the workshop, while only eight percent felt unprepared to do so. These results are particularly impressive given that fully half of our learners are completely new to programming, having previously used these skills less than once per year.

Data Carpentry has a positive impact not only on our learners, but also on our workshop instructors. In March 2016, we surveyed Data and Software Carpentry instructors to learn what they valued about their involvement with the Carpentry community. We found that being a part of the Carpentry community served varied purposes for instructors beyond providing them with teaching skills and experience. This included helping them to develop better communication skills, improve their confidence and expand their professional networks. Instructors also reported gaining personal fulfillment from helping researchers in their field improve the quality of their science.

Since we have developed a Partnership program in May, we have signed seven new partners including University of Michigan, Michigan State, University of Florida, University of Leeds, UC San Diego, University of Miami and Netherlands eScience. We are also developing Platinum level partnerships with ELIXIR and the USDA, and are in the process of signing additional universities as Gold and Silver partners.

We have had two papers published, one on Data Carpentry and another on outlining skills needed for good computational practices. Data Carpentry: Workshops to Increase Data Literacy for Researchers. Tracy K. Teal, Karen A. Cranston, Hilmar Lapp, Ethan White, Greg Wilson, Karthik Ram, Aleksandra Pawlik. IJDC 2015, Vol. 10, No. 1, pp. 135-143 doi:10.2218/ijdc.v10i1.351 and Computing Workflows for Biologists: A Roadmap, Ashley Shade and Tracy Teal in PLOS Biology http://dx.doi.org/10.1371/journal.pbio.1002303. Teal and Becker have also continued to publish

in their fields (Teal: https://scholar.google.com/citations?user=EJZlW0cAAAAJ&hl=en; Becker: https://ucdavis.academia.edu/ErinBecker)

Expenditures

Primary expenditures have been for staff. The Executive Director salary has been as budgeted and paid throughout the first year of the grant. The Associate Director hire took longer than expected and was hired with a start date of May 9, 2016. As such, this salary was not spent for the first 9 months of the project. The salary budgeted for this time frame was $95,000, but only $26,500 has been spent. The AD salary was increased to $105,000 per year, with the additional salary coming from other Data Carpentry revenue. The Program Coordinator (Workshop Coordinator in the grant) budget was for $39,000/year for a half time position. The Program Coordinator was hired at a rate of $26.50/hour, approximately $27560.0/year, so we are under budget on this salary. We delayed the hiring of the Deputy Director of Assessment (Assessment Specialist in the grant) to year 2, so that we could hire for that position full time. That salary will therefore be spent in year 2. The salary was budgeted at $90,000/year and the position was accepted at $80,000/year, so we will be under budget on this salary.

Travel was budgeted at $1500/month. We have underspent on this as well, as most travel was for the Executive Director to attend meetings, but as an invited speaker many of her costs were covered by the meetings. In year 2, we will have more staff using this travel budget and Teal is already committed to attend meetings where travel support is not available. Infrastructure and supplies was budgeted at $800/month. We initially were underspending per month, with infrastructure costs primarily for servers and software. With more staff, and staff co-located in the Davis region, we are now renting co-working space at $700/month, so will be using that budget item going forward. Budget for Train-the-trainers ($18,000) has gone for salary support for the Director of Instructor Training at Software Carpentry who is running and coordinating these efforts. All 30 instructor training workshops conducted in the past year train both Software and Data Carpentry instructors, and an event at the University of Florida specifically focused on Data Carpentry instructors. $15,000 was budgeted for a hackathon in the first year. Some of this budget went to the hackathon to develop a Data Carpentry focused on working with spatial data, but the collaboration with NEON on the development of these lessons offset much of that cost. Similarly the collaboration with BEACON, SESYNC and CyVerse on the genomics lessons offset most of the costs for that hackathon. We therefore still have the budget for 2 hackathons in year 2 of the grant.