

Grant to The Carpentries

Sloan Foundation Grant No: G-2018-11120

Title: Development of data skills curriculum and scalable open and collaborative lesson development infrastructure

Narrative Annual Report

In year 2 of our 'Development of data skills curriculum and scalable open and collaborative lesson development infrastructure' grant, we focused on development of domain specific curricula and on expanding and utilizing the documentation and structures we created in year one. We strengthened our ability to continue supporting development of new curricula and to maintain existing curricula - both through our paid Core Team with two new hires, and through in-kind contributions from community members.

In our Narrative we've outlined progress on grant goals and deliverables, including areas where we have updated our strategic approaches.

Grant deliverables:

1. Lesson templates and guidelines for open and collaborative lesson development in general
2. Open curricula for two-day workshops focused on the foundational skills for data analysis and management in Economics and Image Analysis
 - o A minimum of 6 pilot workshops taught in different countries for each curriculum
 - o Development of strong networks in each domain for future workshops
 - o An assessment instrument for each of these curricula
 - o An assessment report on outcomes of workshops
3. An outline and set of milestones for the development of a curriculum in Chemistry

Key milestones achieved in Year 2:

- Continued to build up our Curriculum Development Team, to support community-driven lesson development and ensure the stability of our lesson infrastructure.
 - Hired two new full-time team members through funding from the Chan Zuckerberg Initiative:
 - A Lesson Infrastructure Technology Developer who will focus on creating accessible, well-documented, robust lesson infrastructure, and

- Simplified our templates and other infrastructure for our lessons and workshop websites, to make it more accessible for community members with a range of technical backgrounds.
 - Implemented “remote themes” for our lessons to facilitate infrastructure around lesson templates for Maintainers and lesson contributors.
 - Set up [Scaffolding](#) to help workshop participants who cannot overcome software installation issues to use cloud-based solutions for the software used in our workshops.
 - Improved discoverability of our lessons by adding systematic GitHub topic tags to all our official lessons.
 - Removed redundancy and added validation for setting up workshop websites, reducing errors in our data.
 - Expanding the curricula supported by our workshop website template to include newly developed lessons and reduce user error from manual entry.

Below we include more information on each of these milestones as they relate to our three primary deliverables.

1. Lesson templates and guidelines for open and collaborative lesson development in general

The foundation of this project is to improve and develop our lesson development and maintenance infrastructure. This infrastructure allows for more effective development of the specific curriculum in this proposal, and also establishes processes for the development and maintenance of all of our future and current lessons. Across The Carpentries Lesson Programs, we have more than 40 active lessons, with an additional 30 lessons under development in The Carpentries Incubator and Lab. Interest and momentum for open, collaboratively developed curriculum continues to increase, and we want to ensure that we can support and foster that interest and demand.

Milestone: Continued to build up our Curriculum Development Team, to support community-driven lesson development and ensure the stability of our lesson infrastructure.

Our experience over the past two years has clarified the types and magnitude of work that can be expected from voluntary contributors. Many people are interested in creating new lessons to teach topics of interest to them, but fewer are able to commit the time needed to make those lessons broadly teachable and to keep them up to date with new technology developments. While we have come a long way in developing the community structure to support lessons (see milestones below), supporting that community structure itself takes a significant amount of dedicated staff time. To continue developing our community infrastructure around lesson

development, we have [secured funding from the Chan Zuckerberg Initiative](#), enabling us to support two additional full-time Curriculum Team members.

Dr. Zhian Kamvar joined the team as [Lesson Infrastructure Technology Developer](#) in March. His work is focused on creating and supporting accessible, well-documented, robust infrastructure to support our lesson stack and to remove barriers to involvement from contributors with different levels of technical expertise. Dr. Kamvar brings to The Carpentries a well-attested ability to communicate technical subjects to non-technical audiences, multi-fluent experience with applicable technologies, and a passion for and commitment to open scholarship. His educational background is in biology, with a focus on plant pathology. Throughout his career, he has focused on developing technical solutions to existing problems within the field, and to educating members of his team and community to spread technical skills that are useful for their work. He has extensive experience working with open, community based initiatives.

Dr. Toby Hodges will join the team in August as [Curriculum Community Developer](#), focusing on supporting all of our community segments involved in lesson development and maintenance. He will create clear communication pathways to facilitate collaboration among our lesson Maintainers, Curriculum Advisory Committee members, lesson contributors, and those developing or proposing new lessons. He will also develop an open, peer-reviewed editorial process for community developed lessons. Along with his experience in curriculum development, Dr. Hodges brings to The Carpentries a well-attested ability to establish and grow communities. Dr. Hodges has also been a very active community member with The Carpentries for the last few years. Dr. Hodges's educational background is in biology, with a focus on biochemistry and bioinformatics.

Milestone: Improved and expanded curriculum on 'how to develop curriculum' to provide guidance for the lessons in this proposal, as well as other Carpentries lessons.

In year one, we released an alpha version of our [Curriculum Development Handbook](#). This early release focused on early stages in curriculum development (e.g. defining a target audience, setting learning objectives). In year two, we developed additional content around recruiting and training lesson maintainers, running pilot workshops, and building a broad community of instructors to teach the lessons. A brief overview of these new chapters can be found below.

- [Chapter 6: Community development roles](#) - describes the roles of lesson authors, reviewers, maintainers, curriculum advisors, and instructors. Documents the skill sets required for these roles and how to recruit and onboard for each role.
- [Chapter 7: The lesson life cycle](#) - defines in detail the stages (pre-alpha, alpha, beta, and stable) that we use to categorise our lessons. Provides a step-by-step outline for how a lesson progresses through each stage, including an approximate timeline. Distinguishes among three tracks on which lessons can progress - the official track, the community track, and Carpentries Lab track, and describes the process and criteria for each (see **Figure 1**).

In addition to the new chapters described above, we also expanded and updated the previous six chapters in response to feedback from lesson developers.

We continue to use this handbook as a resource for lesson developers. Before the creation of this resource, much of our community's knowledge and practices around lesson development was undocumented or out of date, making it difficult for newcomers to The Carpentries to replicate the pedagogical approaches they saw and admired in Carpentries workshops. The existence of this resource greatly reduces our need to dedicate staff time in communicating about the curriculum development process, making our approaches more scalable.

As discussed in year one, we continue to be excited about transforming this handbook from a written resource to a fully-fledged curriculum in the future (analogous to our [Instructor Training curriculum](#)). In June 2019, Dr. François Michonneau and Dr. Tracy Teal delivered a short workshop at a [regional Carpentries conference in Manchester, UK](#) to pilot some elements of this curriculum. The majority of workshop attendants had no prior experience with The Carpentries, and so the first half of the workshop focused on introducing core components of The Carpentries pedagogical model and our community values. General feedback was that this portion of the workshop was too theoretical, and indicated that participation in our instructor training workshop should likely be a prerequisite for this course. Making our instructor training workshop a prerequisite will also clarify how Instructor Training will focus on how to deliver a lesson while the Curriculum Development workshop will focus on lesson content design. Other feedback gathered from that workshop will inform our planning on transforming this handbook into a full-fledged curriculum. Delivering on this intention is a part of [our strategic plan](#) over the next 3-5 years.

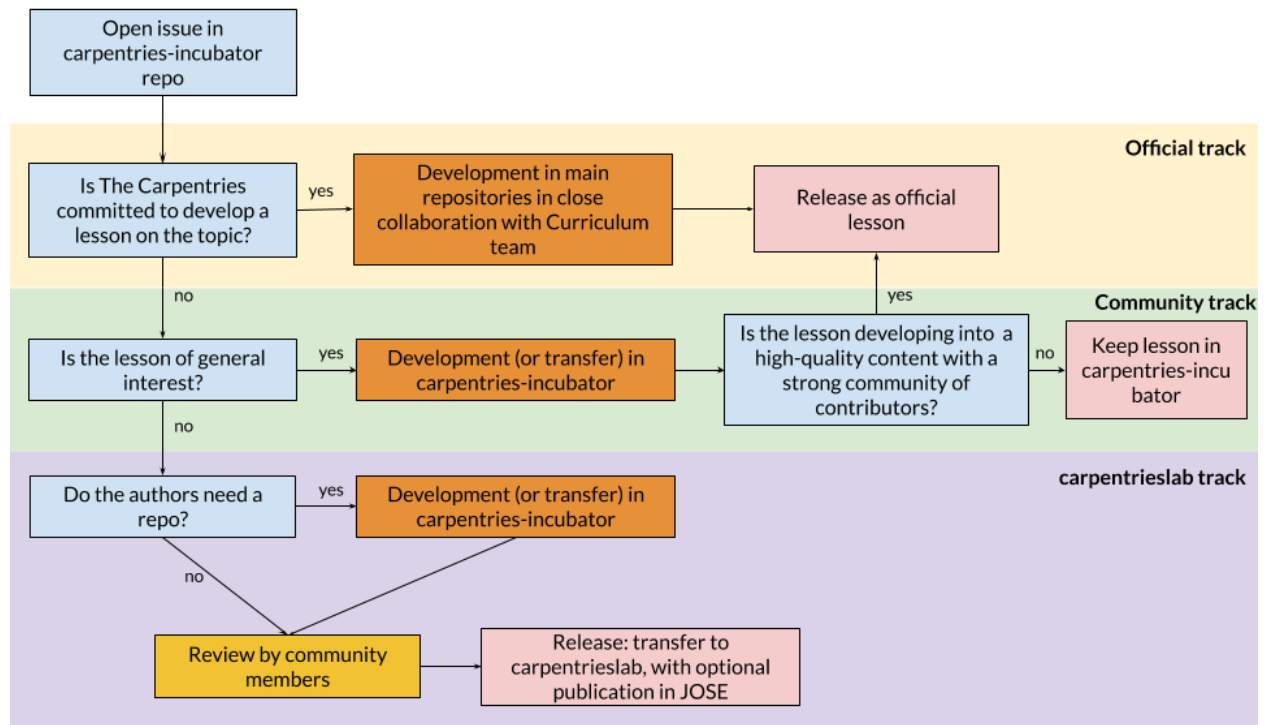


Figure 1: Flowchart showing process for development of lessons falling into the official track, community track, and Carpentries Lab track of lesson development.

Milestone: Established The Carpentries Incubator as a low-barrier place for community members to share and collaborate on Carpentries-style teaching materials at all stages of development.

In year one, we created [The Carpentries Lab](#), as a place to host community-contributed lessons. The Carpentries Lab was envisioned as a collection of high-quality, peer-reviewed lessons that follow best practices in pedagogy and the general teaching practices used in Carpentries workshops. Supporting this vision requires the creation of an editorial and peer-review process, a metadata tagging system to enable people to find lessons they are interested in using or contributing to, and a forum-like space for lesson users to communicate about their experiences with the lesson and to provide feedback. Although we have not yet been able, in year 2, to fully create these processes and structures, the community has continued to be highly enthusiastic about sharing and co-developing lesson materials. To “get out of the way” and enable the community to share resources, we have created a low-barrier place to post lesson materials - [The Carpentries Incubator](#). The Incubator is envisioned as a first-step, uncurated repository of community-shared lessons and lesson ideas - the only requirements are 1) use of The Carpentries lesson template, 2) use of an open license, and 3) conformation with The Carpentries [Code of Conduct](#).

Since introducing The Incubator in July 2019, we have received 36 submissions, ranging from lesson ideas to fully-developed lessons. Eight of these lessons are significantly developed, have been taught by the original developers, and are good candidates for transfer to the Lab.

- [SQL for Humanities](#)
- [OpenRefine for Humanities](#)
- [Python for Humanities](#)
- [Introduction to Geospatial Data with Python](#)
- [From a Spreadsheet to a Database](#)
- [Git Using RStudio](#)
- [Introduction to the Internet of Things](#)
- [Reproducible Computational Environments Using Containers](#)

Another seven submissions are at the “pre-alpha” stage, meaning significant development work has been undertaken, but it is not yet ready for instructional use. To have eight lessons reach the alpha stage, completely independently of staff involvement, is both a sign of how much enthusiasm there is for developing new lessons, and an indication of the utility of the resources we have spent the past two years developing.

To ensure the community is able to access these new lessons, we have created a [“Community Lessons” page](#), which pulls all lessons from The Carpentries Incubator organisation on GitHub, and displays some minimal information, including lesson life cycle stage and topics. An early priority for our new Curriculum Community Developer, once they are onboarded in August, will be fleshing out the functionality of this page, to enable users to sort or filter lessons by length, topic, or completion level. Until we are able to onboard the Curriculum Community Developer, the Community Lessons page serves as a low-tech resource listing for community members interested in teaching Carpentries-style lessons on different topics.

Milestone: Simplified our templates and other infrastructure for our lessons and workshop websites, to make it more accessible for community members with a range of technical backgrounds.

One of the challenges associated with the growth in the number of curricula we offer has been the maintenance of the lesson template across all the repositories that host the content of the lessons. Support from the Sloan Foundation has allowed us to transition to using remote themes where the template is not part of the lesson repositories themselves but instead hosted separately and only pulled at the time of the build. This model allows us to keep the template updated and propagate the updates quickly across all our lessons. It makes it easier for lesson maintainers to not have to worry about the technical aspect of the lesson and instead focus on the content of their lessons.

Another challenge for lesson Maintainers has been to preview the lessons on their own computer before changes they make or contributions they review from the community are made

public. To help with this, we worked on developing Docker containers that encapsulate all the dependencies needed to preview lessons. This technology makes it more accessible to Maintainers to preview changes. It ensures higher quality contributions and lessons.

To improve discoverability of our lessons outside of our existing community, we have created a hierarchical structure of [GitHub topic tags](#). Tag levels include lesson program (Data Carpentry, Library Carpentry, or Software Carpentry), repository type (lesson or workshop overview page), tool (e.g. python, sql), curriculum (e.g. ecology), skills (e.g. data management, metadata), human language (english or spanish), and life cycle. Establishing this consistent tagging system serves a secondary purpose, as these tags can be treated as metadata to enable community members to find lessons on particular topics, or at particular stages in the life cycle. We implemented a low-key tag sorting functionality on our [Community Developed Lessons page](#), as a proof of concept, and plan to significantly extend this functionality through the next year as part of our Carpentries Incubator project.

Our lessons are designed to be used in a synchronous workshop setting, which means that specific information about each workshop needs to be communicated to learners in a clear and consistent manner. To help workshop instructors communicate essential information to workshop participants, we use a [workshop website template](#) GitHub repository. Before their workshop, instructors are responsible for creating a copy of this repository and modifying it to match their workshop details (e.g. location or connection information, software installation instructions, schedule). This template was originally developed to support two different workshop types - Software Carpentry with Python and Software Carpentry with R. We now have twelve official workshop variations, and many others that our community teaches but are not yet officially part of our curricular offerings. Expanding the capabilities of this template to support instructors for these different curricular variations became a priority, as we saw high rates (44%) of [user error in workshop website setup](#), leading to confusion among learners about required software and scheduling details. We updated the template to be more flexible and to accommodate existing workshop variations. More work remains to be done to allow for unanticipated types of mix-and-match curricula. We also added validation steps to the template, to reduce errors in the data that we collect from instructor-created websites.

The recent COVID-19 pandemic has also required enhancements to our template infrastructure. We have pivoted to holding all of our workshops online, which unfortunately makes it more difficult for instructors to help learners troubleshoot software installation issues. To alleviate these issues, we have set up cloud instances (“[scaffolds](#)”) for each curriculum, containing all of the software and data that learners need to fully participate in that course. These scaffolds use either RStudio Server or MyBinder, depending on the curriculum, and do not require any setup from instructors or learners. In line with The Carpentries pedagogical philosophy of empowering learners, we are still recommending that learners attempt to install all software locally, so that they can continue learning after the workshop ends. We are monitoring these scaffold solutions to ensure that they serve their purpose for learners, while not decreasing the benefit to learners

of the workshop. We have also updated our workshop website template and workshop tagging system to enable online workshops.

Milestone: Further developed infrastructure for the maintenance of lessons and for building an Instructor community around new lessons.

Maintainer community

As described in our year one report, The Carpentries have had the [Maintainer](#) role in our community for several years, but until recently, Maintainers did not receive any training or guidance in their role. In early 2018, we developed and piloted a [Maintainer onboarding process](#) to introduce new and existing Maintainers to the social, curricular, and technical aspects of being a Maintainer. We also introduced [Maintainer guidelines](#) to provide a clear set of responsibilities for those interested in joining the Maintainer team. In early 2018, we onboarded 27 new Maintainers to our lessons using these materials.

To continue to support onboarding of new Maintainers, and facilitation of the Maintainer community on an ongoing basis, we onboarded a [Maintainer Community Engagement Lead](#), Angela Li, an R Spatial Advocate at the Center for Spatial Data Science at University of Chicago, Illinois. Angela is an active member of The Carpentries community, serving as an Instructor, Maintainer, Instructor Trainer, and Mentor. She also has experience in community organization through her role as co-founder of the R-Ladies Chicago chapter. Since January, Angela has led efforts in building and sustaining the Maintainer community. She has rebooted our Maintainer Onboarding program, recruiting 33 new volunteer Maintainers from nine countries around the world. These new Maintainers will complete onboarding in July 2020 and become full-fledged members of the Maintainer team. In rebooting the Maintainer Onboarding program, Angela is evaluating our current structures for integrating new Maintainers into the team, as well as taking stock of future need for more Maintainer support. Follow up projects on this front include establishing recurring cycles for Maintainer recruitment and onboarding, creating a Maintainer alumni / retirement program to enable inactive Maintainers to step down without guilt, and continuing to enhance opportunities for Maintainers to skill-share and support one another in their work.

This year we also introduced a new system for gathering asynchronous feedback from Maintainers on issues relevant to them. The introduction of this system was driven by feedback from Maintainers that many decisions were coming out of the monthly Maintainer community calls, which the majority of Maintainers are not able to attend on a regular basis. Starting in September 2019, we have been engaging Maintainers in a two-fold approach to feedback, bringing issues both to the synchronous monthly meetings, and to a [“Request for Comments” GitHub repository](#). Through this repository, we have been able to engage more members of the Maintainer community in the decision making process. This system also increases the transparency of our work by making these conversations publicly viewable and documented in writing.

Instructor onboarding and community building

In our year 1 report, we discussed the importance of building communities of instructors to teach newly developed curricula. Because Data Carpentry lessons are domain specific, they require a specific set of skills and background knowledge to teach, beyond knowledge of the technical tools being taught (e.g. R, Python). To ensure we have a body of instructors capable of teaching curricula in new domains, we are taking a two-pronged approach. First, we are actively prioritizing instructor applicants from individuals with relevant domain experience. For more information about this approach and successes in implementing it, see the section “**Increased training of instructors with backgrounds in economics and chemistry**” below.

In addition to bringing in more instructors with particular educational backgrounds, we are also drawing on our active instructor body by providing curriculum-specific onboarding for new lessons. Onboarding reduces the participation barrier by demystifying the curriculum, drawing parallels to existing lessons that instructors may have taught in the past, and noting specific areas that may be challenging to teach. Each of our onboardings has been held as a synchronous webinar, and is also offered in recorded format on [our Youtube channel](#), for scalability. In year one, we introduced onboarding for our new Geospatial and Social Sciences curricula. This year, we added onboarding for the newest version of our [Genomics curriculum](#), which has had nearly 500 views. Although many of these views are likely partial, we have had confirmation of onboarding completion for 50 instructors.

Milestone: Coordinated updating and release for 36 of our core lessons to provide publication credit for the 1100 contributors to, and maintainers of those lessons.

Serving as a lesson maintainer is a significant time-investment. As the vast majority of our lesson maintainers are volunteers, we need to ensure that they are rewarded for the work that they do, in a way that aligns with the incentivization structure for their primary occupations and career paths. Most of our community members function within the academic community, in which publication is a significant career product. We help support our Maintainers and reward their work by [publishing our lessons through Zenodo](#). This also provides reward for non-Maintainer contributors, as every person who submits an accepted change to a lesson is listed in the lesson’s authorship, in order of magnitude of contribution. Maintainers receive editorial status, and are also listed as authors. [Our most recent lesson release](#) acknowledged more than 1100 community contributors to 36 of our lessons.

Because lesson publication is an important part of providing professional recognition for the work of our volunteer community, we would like to have more regular, and frequent releases. (Before the current release, we had not done a release in 18 months.) Although partially automated, our current release process is time consuming and requires several manual steps.

As the number of lessons, and the number of contributors to each lesson increases, it will be increasingly important to automate as many steps of this process as possible. Adoption of remote themes for lessons, discussed above, has been one important step in this automation process. Further template improvements, currently being undertaken by Dr. Kamvar as our new Lesson Technology Infrastructure Developer, will serve towards this end, by making template updating centralized, rather than dependent on involvement from Maintainers from each lesson. We will continue to seek input and feedback from the Maintainer community and other stakeholders as this new template takes shape.

2. Open curricula for two-day workshops focused on the foundational skills for data analysis and management in Economics and Image Analysis

This year, we have been able to capitalize on our investment in infrastructure and process development, which was our focus in year 1 of this grant. Working closely with dedicated members of the community, we have facilitated development of an alpha stage curriculum for economists, as well as a beta stage lesson on image processing with Python. Both new curricula fit into our existing two-day workshop model. A major takeaway from this work has been the importance of forming collaborations with individuals who have independent incentive to develop a finished lesson product, and have dedicated resources to enable them to spend significant time on curriculum development.

Economics Curriculum

Milestone: Completed alpha version of Economics curriculum.

In the past year, we have continued to work with Dr. [Miklos Koren](#) to create an alpha version of this curriculum. Dr. Koren is an associate professor in the department of Economics at European Central University and a senior research fellow at the Institute of Economics. He is also an active member of the European Economic Association and an editor for the *Journal of International Economics* and the *Review of Economic Studies*.

Dr. Koren has used our Curriculum Development Handbook to create initial lessons for introducing economists to [Stata](#) and to the [bash shell](#). By following the process outlined in the Curriculum Development Handbook from the very beginning, Dr. Koren has been able to create lessons with a clearly defined target audience, minimising “scope creep”, and ensuring lessons are maximally effective for learners. As discussed in our year one report, although Stata is not an open-source tool (as are the other technologies that we teach), it is widely used in the Economics research community and learners will likely have some exposure to this software before attending a workshop. We decided to create an initial version of the curriculum using Stata, to “meet learners where they are” and reduce cognitive load. In line with our [community values around open access](#), we plan to translate these materials into R after beta piloting.

The alpha version of this curriculum has been piloted twice - in August 2019 at the European Economic Association's (EEA) Annual Congress in Manchester, UK, and again as part of a graduate course at European Central University (ECU) in October. Dr. François Michonneau from The Carpentries core team was in attendance at the August pilot, and collected feedback (see Appendix A and B). We saw substantial increases in learners' self-efficacy (e.g. "I can write a small program/script/macro to solve a problem in my own work") increased from a weighted average of 3.90 to 4.42 (out of a total possible score of 5.0), and "I am confident in my ability to make use of programming software to work with data" increased from 3.20 to 4.00. Eighty-one percent of learners were Windows users, and all but two learners had used Stata for a previous course or project - confirming our initial design choices in planning the curriculum. All survey respondents either "agreed" or "strongly agreed" with the statement "I can immediately apply what I learned at the workshop", and the average recommendation score for the workshop was 83 / 100.

A third alpha pilot will be held virtually in advance of the EEA's 2020 Congress in August, using lessons modified based on feedback from the previous two pilots. Dr. Koren is supported in this work by colleagues at ECU - Dr. Arieda Muço and Dr. Andras Vereckei, both of whom have served as co-instructors at pilot workshops and act as maintainers of the lessons. We had initially planned to start beta pilots (with instructors outside of the original development team) in September 2019. As we gain more experience in working with volunteer curriculum developers, we are learning to adjust our expectations of the pace of work volunteers can sustain in addition to their full-time jobs. Although this curriculum has made continuous, steady progress over the past year, it is not yet ready to be taught by novice instructors. We will continue working with Dr. Koren and his colleagues to incorporate lessons learned from the August 2020 pilot, and make this workshop broadly available in late 2020.

In the past year, we have badged 72 Instructors with expertise in Economics, with another 58 currently going through the certification process. Combined with our efforts from year one, we now have 88 badged Instructors qualified to teach the Economics curriculum. This represents a healthy instructor population size to support future workshops, as this curriculum becomes an official Carpentries offering and is requested by hosting organisations.

Image Curriculum

Milestone: Drove Image Analysis curriculum to beta stage in final preparation for publication.

At the time of our last grant update, we had partnered with Dr. Tessa Durham Brooks and Dr. Mark Meysenburg at Doane College, Nebraska. Drs. Brooks and Meysenburg had created a three-day curriculum for teaching image processing with Python to undergraduate students, as part of an NSF iUSE project ([DIVAS](#)). This curriculum was created using The Carpentries lesson template and pedagogical approaches and had been piloted in-house, as well as at the

University of Nebraska, Lincoln. Because these workshops were taught by the original lesson developers, they fell within the “alpha pilot” phase of our curriculum lifecycle.

In May 2019, Dr. Erin Becker, our Associate Director and Curriculum Team Lead, attended another alpha pilot workshop for this curriculum. As someone who was not involved in the initial lesson design, her observation of a pilot was critical in determining gaps in the documentation of the curriculum so that it could be further fleshed out to make it ready for teaching by new instructors. Based on her observations, Dr. Becker noted two primary barriers to the lesson being adopted as an official Data Carpentry lesson.

First, the lesson was being taught as a three-day workshop, as part of a five day sequence, in which it was preceded by a standard two-day Software Carpentry Python workshop. This meant that learners and instructors needed to dedicate five days to a workshop - a reasonable assumption in an immersive summer course such as DIVAS, but not a realistic expectation for most Carpentries workshops. Our primary audience of graduate students, postdoctoral researchers, and university professors and staff, have pressing job and course duties to return to after the workshop and prefer shorter formats. Instructors frequently have to travel to a workshop, and many are not able to be away for a full work week, for the same reasons. Because of this, we determined that we needed to make this course teachable as a two-day workshop, to fit in with The Carpentries other standard offerings and our instructional delivery model. To accomplish this, we decided to make familiarity with Python and some bash shell concepts a prerequisite for the course. We identified all of the Python and bash concepts and commands that were used in the image processing lesson, and created an official [prerequisites list](#) to clearly communicate these expectations. We also trimmed some redundant examples from the lesson, and made the capstone exercises optional.

The second barrier to adopting this lesson as an official Data Carpentry course was the software being used. The original course developers used a Python imaging library called OpenCV, which is notoriously difficult to install, and not backwards compatible. They managed this by hosting the course on a virtual machine, negating the need for learners to install software locally. Having learners work on their own computer is a central part of The Carpentries pedagogical model, as it enables learners to continue learning at their own pace after the workshop. We investigated alternative imaging libraries, and decided to adopt SciKit-Image, part of the well-documented, and actively supported [SciKit-Learn toolbox](#). Switching from OpenCV to SciKit-Image not only reduced cognitive load for learners and simplified installation, it also greatly simplified the coding examples, helping us to shorten the workshop from three days to two. The lesson is now officially hosted in the [Data Carpentry GitHub organisation](#), and listed on our [lessons page](#).

After these major changes, the lesson was piloted at two new institutions by instructors outside of the original development team - in January 2020 at the European Molecular Biology Laboratory, and February at the University of Arizona. Each of these workshops had a dedicated notetaker, who collected copious amounts of information about how the materials

were working in practice and where things needed to be clarified or smoothed out. The pre- and post-workshop survey responses from learners at the February 2020 University of Arizona pilot are attached as Appendix C and D. As with the economics pilot, we also saw an increase in learners' self-efficacy on several statements. However, the gains were smaller - possibly because this workshop is targeted towards intermediate learners who already have relatively high levels of self-efficacy towards basic programming tasks. For example, "I can write a small program/script/macro to solve a problem in my own work" increased from a weighted average of 4.04 to 4.18, and "I am confident in my ability to make use of programming software to work with data" increased from 3.68 to 3.96, and both statements started at a higher average than for the economics survey data.

We are in the process of working through the rest of this feedback and preparing the lesson for publication, at which point it will reach the "beta" stage of our lesson lifecycle and be available on request as an official Data Carpentry lesson. We had initially planned this publication in March 2020, but have been delayed due to the COVID pandemic.

3. An outline and set of milestones for the development of a curriculum in Chemistry

In year 1, we focused on creating infrastructure, processes, community structures, and resources to support curriculum development. In year 2, we applied these resources and structures to help committed community members create new curricula in Economics and Image Processing. The resources and processes we have developed have greatly facilitated the lesson development process, not only in these two areas, but in the many lessons that our community members have started developing in the Incubator. However, our experience this year has further emphasized the need for developing collaborations with committed individuals who have an independent investment in creating a high-quality lesson product, and who are able to bring their own resources to support their time spent on the project.

With economics and image processing, we were fortunate to find individuals in our extended networks who met these criteria and were able to productively collaborate on lesson creation. We have not yet been able to connect with a chemistry domain expert who has the time and motivation to engage with us in the lesson development process.

Milestone: Increased training of instructors with chemistry backgrounds.

Through our Instructor Training program, we have been able to proactively shift the composition of our community to include more chemistry professionals. In the past year, we have badged 106 Instructors with expertise in Chemistry, with another 150 currently certifying. Combined with our efforts from year one, we now have 126 badged Instructors qualified to teach a future curriculum in Chemistry.

These newly certified instructors bring with them their own active networks, as well as connections they will establish throughout their tenure as Carpentries Instructors. Diversifying the domain expertise in our community will naturally lead to an expansion in our curricular offerings, as our community-developed lessons pipeline (Carpentries Lab and Incubator) becomes further established. With the resources and infrastructure that we have established over the past two years, we are now in a position to support community development of curricula in any domain, including but not limited to chemistry.

Budget changes

In our year 1 report, we noted that we overspent on staff salary and underspent on hackathons, as we determined that spending more time on laying the groundwork for efficient infrastructure - although requiring more staff time upfront - would scale better overall and create a more positive experience for the community. We continued to utilize a portion of Dr. Erin Becker's time on the Curriculum Development Handbook and on Maintainer Community building and process development. Dr. Becker, with her background in pedagogical research, was also the primary resource for community members developing curricula, and led much of the second-stage cleanup efforts to align the new lessons with our pedagogical framework. Rather than moving back towards a hackathon development model, which we have found to be ineffective in creating polished deliverable lessons, we continued with our new model of having a small number of committed community members working with our staff, to develop and polish lessons. This model is reflected in our spending in year 2 of this grant, as most of the funding originally allocated to travel has been reallocated to staff salary. Both years did include travel for Dr. Becker and Dr. Michonneau to attend and assist at pilot workshops, giving them first-hand experience with the curriculum as delivered and able to use that experience in curriculum improvements. As also noted in year 1, Dr. Michonneau's salary was higher than planned, as at the time of our original proposal we had a different fiscal sponsor with different fringe overhead.