Library Carpentry: software skills training for library professionals

Jez Cope

The University Library, University of Sheffield

James Baker

Sussex Humanities Lab, University of Sussex

Much time and energy is now being devoted to developing the skills of researchers in the related areas of data analysis and data management. However, less attention is currently paid to developing the data skills of librarians themselves: these skills are often brought in by recruitment in niche areas rather than considered as a wider development need for the library workforce, and are not widely recognised as important to the professional career development of librarians. We believe that building computational and data science capacity *within* academic libraries will have direct benefits for both librarians and the users we serve.

These technical skills can be of direct value across a wide range of common library activity. For example, many roles involve updating and processing metadata in various systems, which can be tedious and error-prone when done manually but lends itself to automation when the skills to do so are available. In many cases all that is required is enough knowledge to recognise such opportunities and raise them with more technical colleagues for implementation.

Reporting on complex, inconsistent and often messy datasets (such as usage statistics from different publishers) is another common task. Cleaning and normalising ("wrangling") this data using manual techniques in spreadsheet software is time-consuming, but there are many tools commonly used by data scientists to make the process not only easier but more reproducible and hence more transparent.

Further, liaison and research support librarians are increasingly expected to facilitate researchers' data needs. There is no expectation that we should become researchers ourselves (at least in the UK where librarians are not typically expected to hold doctorates), but learning some of these skills and applying them in our own work is an excellent way to learn the language of researchers and build rapport.

Library Carpentry is a global effort to provide training to librarians in technical areas that have traditionally been seen as the preserve of researchers, IT support and systems librarians. Established non-profit volunteer organisations such as Software Carpentry and Data Carpentry offer introductory research software skills training with a focus on the needs and requirements of research scientists. Library Carpentry is a comparable introductory software skills training programme with a focus on the needs and requirements of library and information professionals.

Library Carpentry has four main guiding principles.

1. It is COMMUNITY LED. Library Carpentry has been community led and open access from the outset. The initial run at City University London established the tenor and content required to deliver software skills training to library and information professionals. Since then development and maintenance of modules has been driven by the community.
2. It is RESPONSIVE. Feedback mechanisms are embedded within lesson plans to ensure that lessons are responsive to community need. In turn, the technologies we teach and build our modules around were chosen to ensure as lesson materials are responsive.
3. It is DISTRIBUTED. As module development has become more distributed over time, a nominated individual has been assigned responsibility for maintaining each lesson. GitHub Issues and Gitter are used to log and coordinate issues and developments, further enhancing this tranparent development model.
4. It seeks to create a SNOWBALL EFFECT. In order to expand the Library Carpentry community and embed the software skills within the library community, attendees are asked to pledge to pass on something they learn to someone else when 6 months. This transmission of skills through professional networks seeks to turn trainees into trainers, thereby also deepening the skill set of the community.

In its initial exploratory run, Library Carpentry took the form of four three-hour sessions held at the City University London Centre for Information Science across four successive Monday evenings. These sessions – held in November 2015 and funded by the Software Sustainability Institute – attracted 59 participants from 14 institutions in London and its environs. Lessons were team developed and team taught, and were published CC-BY, building on earlier material developed for the British Library Digital Scholarship Training Programme, Software Carpentry, and the Programming Historian. They covered:

1. regular expressions (patterns that match character combinations)
2. the Unix shell (a command line user interface)
3. git (a version control tool)
4. OpenRefine (an interactive data cleanup tool)

Since then, 13 Library Carpentry workshops have been organised, training well in excess of 250 librarians in 7 countries across 4 continents: in Australia, Canada, Denmark, Norway, South Africa, Switzerland, and the USA. During the Mozilla Science Lab Global Sprint (2-3 June 2016), an international team developed module materials, added a new module on SQL (a relational database management

language), assigned administrative roles required to support future development, and republished the materials using the Data Carpentry lesson template. Library Carpentry now has a distributed management and maintenance structure.

This paper will describe how the material was developed and delivered, and report on challenges faced, lessons learned and future plans. There will also be an update on the latest activity and an open invitation to participate in this growing and global community of librarians!