Working up and building the foundation for Data and Software Carpentry within ELIXIR
ELIXIR SWC/DC Pilot Project

**Building the Foundation for Data Carpentry and Software Carpentry within ELIXIR**

(lead by ELIXIR UK; collaborating and supporting Nodes: ELIXIR NL, ELIXIR FI, ELIXIR CH, ELIXIR SE, ELIXIR SI)

3 events:

- Data Carpentry hackathon & workshop hosted by ELIXIR Finland, 16-19 March 2015

- Data Carpentry hackathon & workshop hosted by ELIXIR Netherlands, 22-25 June 2015

- Data and Software Carpentry Instructor training hosted by ELIXIR Switzerland, early 2016 (exact dates TBC)
Purpose

1. Leverage the work of the existing Data and Software Carpentry initiatives.
2. Launch the Data and Software Carpentry initiatives for ELIXIR, closely working with the existing international Carpentry networks.
Training needs

Researchers view the major limiting factor in research progress as a lack of expertise in how to handle and analyze data effectively and efficiently.

This can only be addressed through high quality, widely available training and is the resource most highly desired by researchers.

Why Data & Software Carpentry

- Training in computational skills designed specially for researchers
- Free and open source training material
- Updated and improved lessons, examples and exercises
- Well established and tested training model
- International community of instructors, material contributors and maintainers
- Community led and community oriented
- Organizational Membership – foundations for long-term collaboration
Teaching basic lab skills for scientific computing so that researchers can do more in less time and with less pain, working more effectively with source code.

Core curriculum:
- Automating tasks using the Unix shell
- Structured programming in Python, R, or MATLAB
- Version control using Git or Mercurial

Teach basic concepts, skills and tools for working more effectively with data. Workshops are designed for people with little to no prior computational experience.

Core curriculum:
- Limitations and caveats of data analysis using spreadsheets
- Moving from spreadsheets into more powerful tools (R or Python)
- Using databases (introduction to SQL)
- Workflows and automating repetitive tasks (command line shell and shell scripts)
Data Carpentry - Target audience

Multiple disciplines
- bioinformatics
- ecology
- genomics
- sociology
- digital humanities
- neuroscience
- geosciences
- ...

SKILLS

Different levels of career
- PhD students
- postdoctoral researchers
- research assistants
- researchers in industry
- ....

Data + spreadsheets + some statistics and...?

Full computational lab skillset

Good knowledge of programming (R, Python or other), structured data, metadata, proficiency in building workflows and automating tasks

Some knowledge of scripting, using workflow tools, command line
Software and Data Carpentry - comparison

**Software Carpentry**

- Focus on software development skills, tools and methods
- Training designed assuming prior knowledge
- Training materials re-used across disciplines
- Software Carpentry Foundation Organizational Membership

“Training software modules/libraries developers”

**Data Carpentry**

- Workshop model
- Open source materials
- Infrastructure based on GitHub
- Community
- Instructor pool & instructor training
- Shared, consolidated curriculum across the centers to increase scalability of the materials

“Training software modules/libraries users”

- Focus on skills essential for data wrangling, munging, or manipulation
- Training designed with no prior knowledge assumed
- Training materials customized for different domains

“Training software modules/libraries users”
Workshops

- 2 days
- Hands-on
- Qualified instructors
- Helpers
- Post-it notes!
- Materials available online after workshop for free use
- Pre- and post-workshop standard surveys

Administration fee ($1250, then $750) but:
- No fee for self-organised workshop
- Organizational Membership for reducing / re-budgeting the cost
Organizational Membership

Software Carpentry Foundation Organisational Membership Program

**Partner**

- at least 3 qualified instructors; coordination of 20 workshops/year; $10,000 or staff time dedicated instructor training; no admin fee; Advisory Council rep;

**Affiliate**

- at least 2 qualified instructors; coordination of 3 workshops/year; $5,000 or staff time reserved places @ instructor training; no admin fee; Advisory Council rep;

**Sponsor**

- underwrites cost of at least 2 workshops/year; staff time or US$2000 / year; reserves places @ instructor training; publicized as Sponsor's; assistance with recruiting

Partnerships transferable to Data Carpentry

ELIXIR UK is a Software Carpentry Foundation Affiliate
Training materials and licensing

- Collaborative lesson development
- Creative Commons License CC-BY
Instructor training

“Basics of educational psychology and instructional design, and how to use these ideas in both intensive workshops and regular classes.”

• Curriculum developed by Greg Wilson, Executive Director of Software Carpentry Foundation
• Principles of best practices for teaching, education and pedagogy – evidence-based and research supported

• Either online (extending to 12 weeks but in your own time)...
• ...on in 2-days intense workshop face-to-face

Plans on running the instructor training remotely to reduce costs and help scale it up (currently Greg runs the training supported by Tracy Teal and Aleksandra Pawlik)
First ELIXIR Pilot event at ELIXIR Finland

Hackathon 16-17 March 2015

• 2 days
• 11 Nodes represented
• Materials developed/improved:
  ✓ dplyr module for the R lessons;
  ✓ NGS data analysis lessons;
  ✓ RNAseq data analysis using BRB Digital Gene Expression;
  ✓ improvements on the shell lessons;
  ✓ improvement of the spreadsheet lessons.

Workshop 18-19 March 2015

• 2 days, hands-on Data Carpentry
• 29 participants
• Selected material from the hackathon used
• Hackathon participants as observers & helpers
Further work

- Developing a pool of certified instructors across ELIXIR.
- Supporting ELIXIR Nodes in becoming Software Carpentry Foundation Partners/Affiliates.
- Launching Software and Data Carpentry workshops across ELIXIR.
- Adopting Software and Data Carpentry teaching model for training within ELIXIR (eg. recasting the existing training materials in Carpentry-style).
Thank you

Questions?