Data Carpentry: workshops to increase data literacy for researchers

Aleksandra Pawlik, University of Manchester

International Digital Curation Conference

London, UK

10th February 2015

What is Data Carpentry

*Researchers should be able to **retrieve**, **view**, **manipulate**, **analyze** and **store** their's and other's data in an **open** and **reproducible** way. Researchers learn best when material is taught in their domain, so workshops will be **domain**-**specific**.*

Training specifically designed for scientists

http://datacarpentry.org

for novices:
there are no prerequisites, and no prior knowledge about the tools is assumed

Hands-on workshops

Based on the Software Carpentry training model

DATA CARPENTRY
MAKING DATA SCIENCE MORE EFFICIENT

Software Sustainability Institute

elixir
UNITED KINGDOM

Who is Data Carpentry

- **Project lead:**

  Tracy K. Teal (Michigan State University)

- **Steering Committee:**

  Karen A. Cranston (National Evolutionary Synthesis Center, NESCent)
  Hilmar Lapp (National Evolutionary Synthesis Center, NESCent)
  Ethan White (Utah State University)
  Greg Wilson (Software Carpentry Foundation)
  Karthik Ram (University of California)
  Aleksandra Pawlik (University of Manchester)

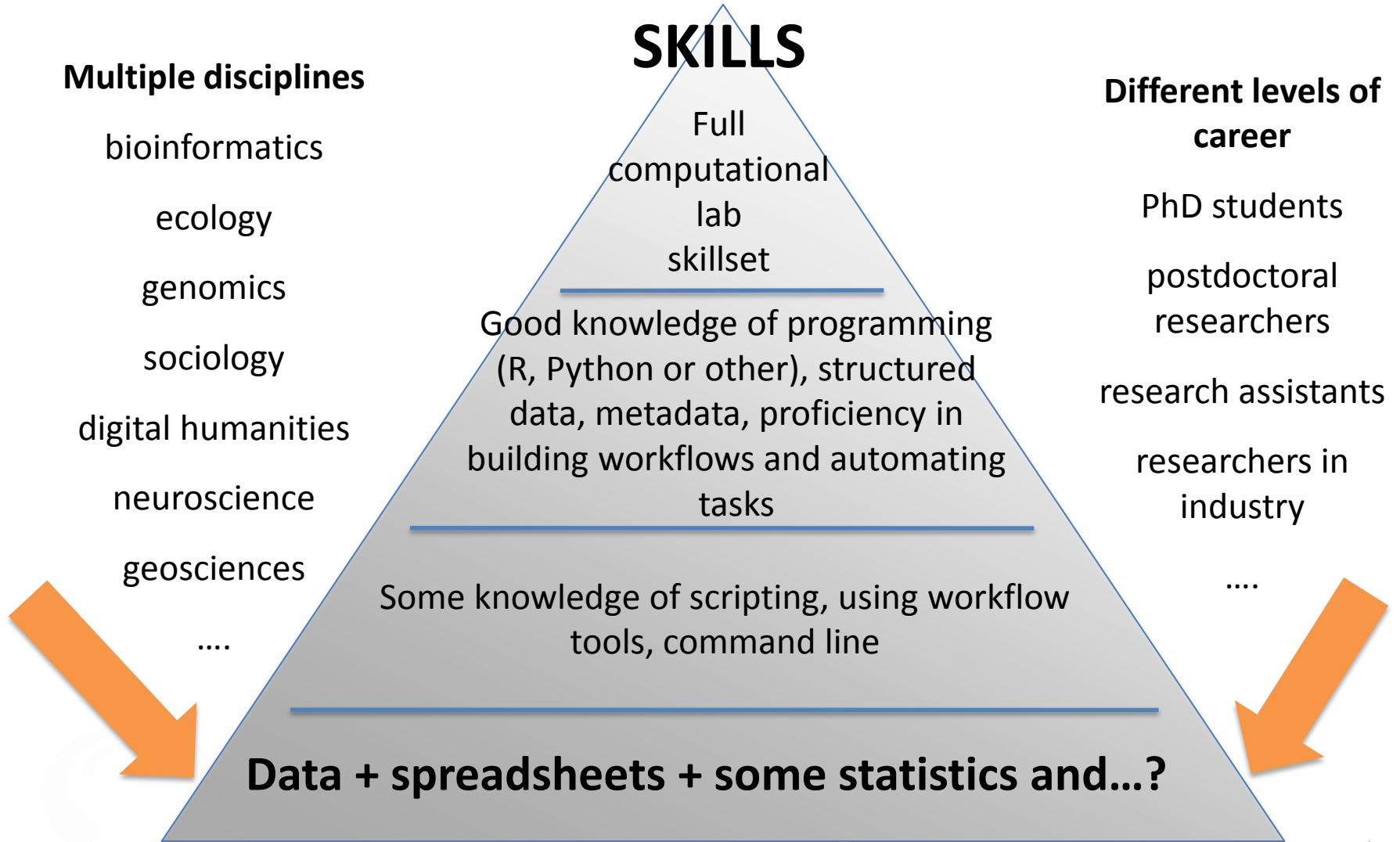- **Administrators:** Arliss Collins (MSL) and Giacomo Peru (SSI)

- **Assessment support:** Shari Ellis, iDigBio

- **Instructors:** international volunteer network

- **Funders and partnership organisations**

Target audience

Multiple disciplines

bioinformatics

ecology

genomics

sociology

digital humanities

neuroscience

geosciences

….

SKILLS

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

Data + spreadsheets + some statistics and…?

Different levels of career

PhD students

postdoctoral researchers

research assistants

researchers in industry

….

Workshops & Curriculum

- 2 days

- Hands-on

- Qualified instructors

- Helpers

- Post-it notes!

Image by Aleksandra Pawlik

Core curriculum

- Effective use of spreadsheets

- 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)

Progress so far

Workshops:

8 in total

4 at the NSF BIO Centers

- iDigBio
- BEACON
- SESYNC
- NESCent

1st Data Carpentry workshop in the UK in collaboration with ELIXIR UK

Hackathons for material development

Mozilla Science Lab sprints July 22-23, 2015

Open source materials

https://github.com/datacarpentry/datacarpentry/

Partnership with Software Carpentry Foundation

Challenges

Sustainability

- Business model and funding

- Partnerships to ensure

Instructors training

- Growing demand for workshops

- Working together with Software Carpentry Foundation

Assessment

- Feedback to improve Data Carpentry

- Evidence of impact making

Plans for the future

- Pilot project to set foundations for Data and Software Carpentry in ELIXIR

(coordinated by the ELIXIR UK Node)

  - **2** European-wide Hackathons for materials development
  - **2** DC workshops
  - **1** Instructors training

- Genomics Workshop Hackathon: Curriculum & Assessment Development, Cold Spring Harbor Laboratory

- Aim for **24 workshops** in the first calendar year (**2015**)

Workshop at IDCC2015!

Data Carpentry: using APIs to access life science data available on the web

Tomorrow, 11th February, 9-5pm
IDCC2015

- demonstrate the power of programmatically accessing life science data
- hands-on learning using API and ROpenSci packages
- directly experience the Data Carpentry teaching style
- discuss the applicability of Data Carpentry lessons in participants' own working context
- hands-on experience on collaborative lesson development

Thank you!