IDCC | *Workshop*

Data Carpentry: using APIs to

Tracy K. Teal
Michigan State University, East Lansing, MI, USA

Karen A. Cranston
National Evolutionary Synthesis Center (NESCent), Durham, NC, USA

Hilmar Lapp
National Evolutionary Synthesis Center (NESCent), Durham, NC, USA

Ethan White
Utah State University, Logan, UT, USA

Greg Wilson
Software Carpentry Foundation, Toronto, Canada

Karthik Ram
Section of Evolution and Ecology, University of California, Davis, CA, USA

Aleksandra Pawlik
University of Manchester, United Kingdom

Workshop overview

It has been established that the growing availability of well-curated and publicly-available datasets can advance research progress in many domains. However, many researchers still lack the data analysis and management skills to most effectively use these datasets. Data must be accessed from remote repositories and large scale analyses conducted with command line tools or programming languages such as Python or R. Data Carpentry workshops aim to teach these concepts, skills and tools for working more effectively with data to researchers, particularly those with little prior computational experience. In this workshop, we will conduct an introductory hands-on lesson for using R and ROpenSci packages to access data from the web through APIs. We will follow this with a discussion of the lesson and the effectiveness of the materials and approach and finally conduct collaborative lesson development/improvement through github. This workshop is for anyone interested in learning more about using APIs and R for accessing and working

Draft from 30th October 2014

Correspondence should be addressed to Aleksandra Pawlik, Room 1.17 Kilburn Building, Oxford Road, University of Manchester, M13 9PL, Manchester, United Kingdom. Email: aleksandra.pawlik@manchester.ac.uk

The International Digital Curation Conference takes place on [TBC] in [TBC]. URL: http://www.dcc.ac.uk/events/international-digital-curation-conference-idcc

with data or if they are interested in learning more about Data Carpentry workshops to see if they would be useful for researchers they work with in their communities.

Objectives of the workshop:

- demonstrating the power of prgrammatically accessing data from the web;
- hands-on learning using API with ROpenSci packages to access data from the web;
- directly experiencing Data Carpentry teaching style;
- discussing the applicability of Data Carpentry lessons to the participants' own working context;
- hands-on experience on collaborative lesson development and improvement (via GitHub)

Additional information

Full-day workshop

no-fee

Speakers: Aleksandra Pawlik, University of Manchester & Data Carpenry

Technical requirements: Projector, room with sufficient number of power sockets, participants should bring their laptops

Audience: Anyone making use of digitalized assets (including those interested in others making use of the assets they created); anyone interested in teaching how to effectively use datasets etc.

Max number of attendees: 15

Draft agenda

- **9:00 - 9:30** Introduction and aims of the workshop
- **9:30 - 11:00** Using ROpenSci packages to access data from the web through API - part 1 (hands-on) *
- **11:00 - 11:20** Break
- **11:20 - 12:00** Using ROpenSci packages to access dta from the web through API - part 2 (hands-on)
- **12:00 - 12:30** Breakout discussion - in groups about the hands-on lesson **
- **12:30 - 13:30** Lunch
- **13:30 - 14:00** Reporting back and discussion ***
- **14:00 - 15:00** Collaborative lesson development and improvement (in groups) part 1 ****
- **15:00 - 15:30** Break

- **15:30 - 16:30** Collaborative lesson development and improvement (in groups) part 2
- **16:30 - 17:00** Reporting back and conclusions

* This part of the workshop will be taught as a Data Carpentry module. The participants will need their laptops with particular software installed (the details will be available online before the conference as a part of the workshop description; the setup won't be too complex).

** The participants will be split into small groups (of max 5) and will discuss the their experiences with regards to the lesson which they just participated in. Their discussion will be guided by a set of questions such as:

(1. How did this style of lesson feel? 2. What was good? 3. What was missing? 4. What would I improve? 4. How? 5. Would that be beneficial for what I do at work? If yes, then how?)

Each group will need to come up with "3 improvements we would like to make to the lesson" (these can be anything - sequence, fixing typos, adding more explanation etc.)

*** This discussion may alter the groups' "3 Improvements we would like to make"

**** The "collaborative lesson development/improvement" will be done via GitHub website interface. Leveraging GitHub infrastructure we will show the participants how even without knowing Git and GitHub it is possible to improve training material built by a community.