DATA SCIENTIST TRAINING FOR
LIBRARIANS
#DST4L

C. Erdmann | DST4L @ Designing Libraries IV | @libcce
We started speaking the same language. A side conversation with a Harvard faculty member has developed into a partnership involving interactive data analysis in the classroom. In the faculty’s own words, the library moved from being “moribund” to a must have service.

Another Example:
http://lorenabarba.com/blog/

Prof. Barba gave keynote at PyData Seattle

Prof. Lorena Barba delivers a keynote at PyData Seattle 2015.
I was afraid of lost connections or opportunities, that our patrons saw us as old fashioned, that we weren’t seen as being tech or data savvy.

Breaking down the library fortress
- Christian Laursen
Experience the Research Data Lifecycle

Just by having some idea, background of where our researchers are coming from, we can make a connection when the opportunity presents itself and build on it, develop better services.

Jeffrey Heer
What are they saying?

> Guidance on Tools & Workflows

Sharing Scholarly Output

Linking to Data

Programming Skills (Training)

Librarians Working w/ Scientists
DATA SCIENTIST TRAINING
FOR LIBRARIANS
Core Goals

---

1. Allow librarians to experience and understand the research data lifecycle.

2. Train librarians to have skills and practical knowledge to build and use data tools (OpenRefine).

3. Forward a culture within libraries that enables data-centric approaches (change our mindset).

4. Grow a community of data savvy librarians.

<empower librarians>
Winning Recipe: Tools, Concepts & Context
#DST4L @ DTU

http://podcast.llab.dtu.dk/feeds/dst4l-2015/
Take away message
To build better services, librarians can help by engaging in communities, understanding their infrastructural needs, and to experiment with new scholarly objects from these communities.

Ian Mulvany
@IanMulvany  Follows you
Head of technology for eLife, editorial board member for Open Research Software, interested in scicom, altmetrics, science on the web and climbing.
The library perspective
Introduce core goals of the course, provide additional context from the librarian perspective and reassure everyone (that this stuff is really hard to learn).

Chris Erdmann
@libcce
Head Librarian, Harvard-Smithsonian CfA
- Working on open access, metadata & data curation, linked data, altmetrics and interactive tech in libraries.
Free your metadata

Much of the work in data science is spent on data wrangling. Using OpenRefine, you can extract data, clean, reconcile, link to data and explore the data.
Academia must learn from open source
Importance of reproducibility (explain, repeat, learn). CERN & the EU’s commitment to open access. Big data at CERN. Advance from 17th century publishing model. The benefits of GitHub w/ examples from CERN (by reducing friction). Using it in services such as Zenodo but changing how we do curation.
GitHub Fundamentals
Getting started (w/ GitHub for Desktop and Atom), introduction to Git(Hub), creating repositories, Markdown 101, using issues, demonstrating workflows, making commits, pushing changes, branching and merging. How GitHub can be used in libraries and academic settings.
Visualization
Demystifying visualization. Why visualization, basic guidelines, visualization in libraries, overview of techniques & methods, and tools to start with.

Alice Thudt
@al_ice_t  FOLLOWS YOU
Evolving Scholarly Communication

Familiarity with interactive notebooks (Jupyter/IPython), baby Python, text analysis, Naives Bayes and building a classifier (for automated curation). Demonstrate the use of online algorithm services (such as Algorithmia).
Feedback, Experiences & Next Steps
Blog/Student Perspective
http://altbibl.io/dst4l/category/blog-posts/

Projects & Presentations
http://altbibl.io/dst4l/category/data-stories/

Hackathons
http://goo.gl/Ly6dxU
DST4L Student Comment:
http://www.youtube.com/watch?v=U5ZYM085bNo&t=1m21s

...very helpful, preparing for long career, going to see it more and more, will keep using skills set, like doing it, fun problem solving...

Library Connect Article:
Data Scientist Training for Librarians: A course and a community
A Pythonic Solution

My job may be creating and editing MARC records, but I am capable of more than that.

Author Network for Query author:"kurtz, m"

This network visualization finds groups of authors within your search results. You can click on the segments to view the papers connected with a group or a particular author.

Group Activity Over Time (measured in papers published)
Find your data scientist training librarian & weave the training into your organizational fabric.
Training Astronomers
Thank you!
Now let’s get our hands dirty with data!