

Apr 5th, 12:00 AM

Data Science in Libraries

Matthew Burton
University of Pittsburgh

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

 Part of the [Scholarly Communication Commons](#)



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

Repository Citation

Burton, M. (2018). Data Science in Libraries. *University of Massachusetts and New England Area Librarian e-Science Symposium*.
<https://doi.org/10.13028/d3hv-8b36>. Retrieved from https://escholarship.umassmed.edu/escience_symposium/2018/program/6

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Data Science in Libraries

B. Tijerina, C. Erdmann, M. Burton, L. Lyon

<http://d-scholarship.pitt.edu/33891/>

IMLS RE-43-16-0149-16

PITTSSCI



Data&Society



Data Science in Libraries Workshop

May 16th & 17th, 2017

*Digital Scholarship Commons, Hillman Library, University of Pittsburgh
Pittsburgh, Pennsylvania USA*



Participants:



Data&Society



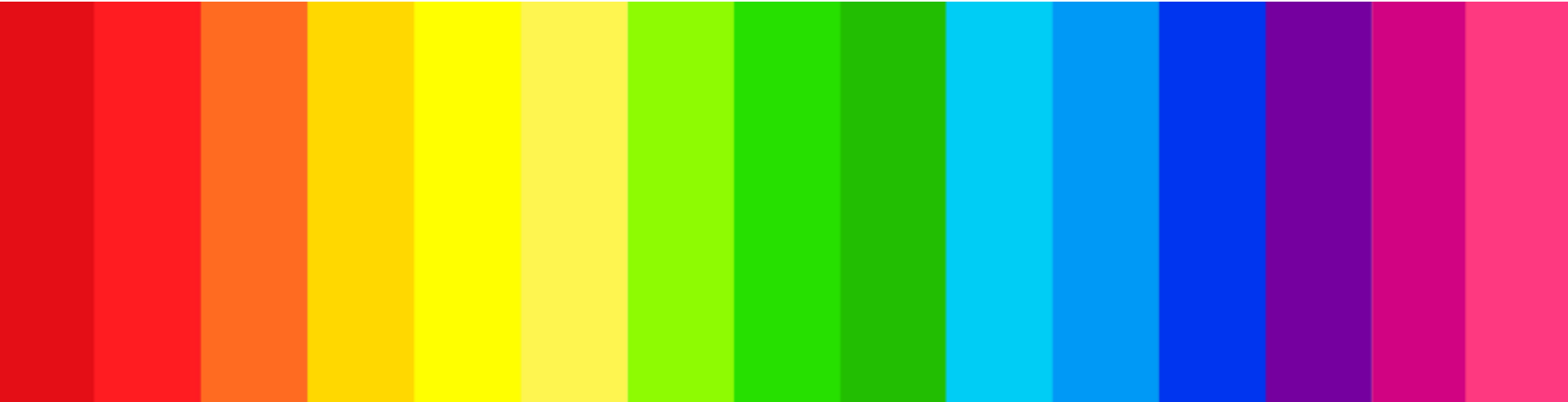
Framing: The Skills & Management Gaps

As librarians increasingly use, build, and maintain Data-centric systems and services, we need the skills to manipulate, analyze, and manage data are crucial. Librarians will need to meaningfully engage the tools and techniques of data science, but they currently face **two significant challenges**:

1. **The Skills Gap:** While practicing mid-career librarians are learning some data science skills, it is through ad-hoc, uncoordinated continuing education programs.
2. **The Management Gap:** Library administrators need toolkits and frameworks to strategically use data science for data-driven decision making and management of library operations.

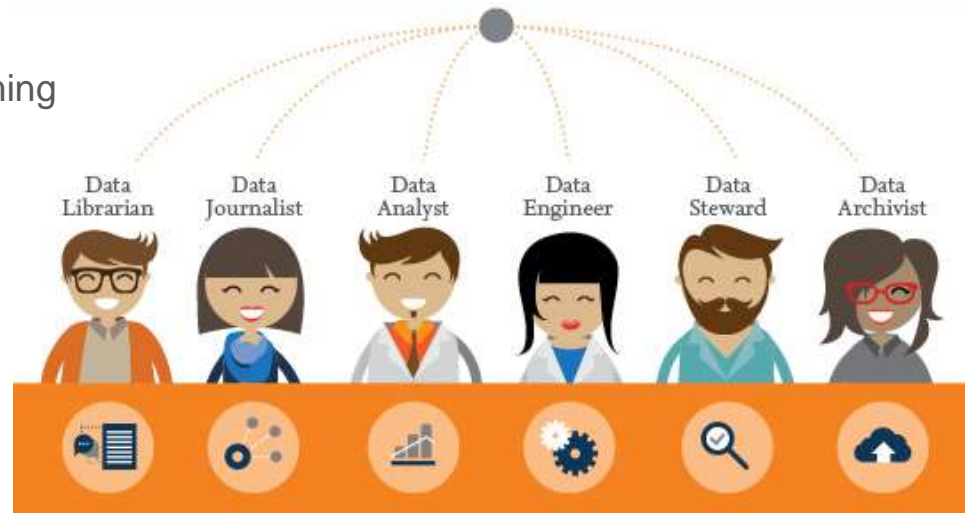
What is Data Savvy?

Data Science exists more or less on a spectrum, and spans work requiring the deep statistical and software engineering skills, to work that focuses on advocacy, policy development, and data management planning.

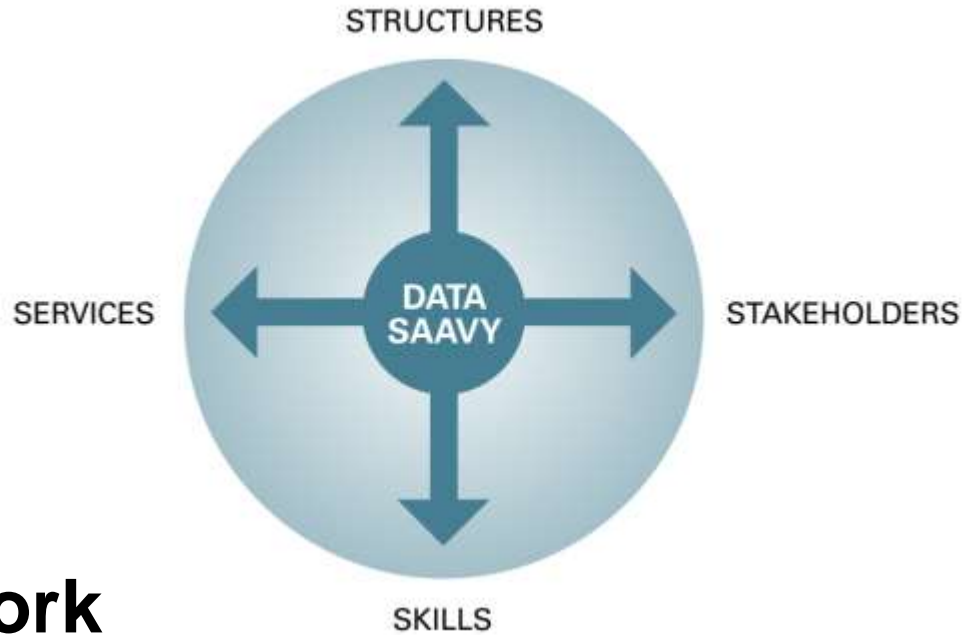


Data Savvy Roles

- Data savvy librarians gain familiarity with the datasets, understand technical methods and techniques, and speak multiple disciplinary languages allowing them to work more closely with researchers or the public.
- Some librarians engage more deeply, becoming technically proficient in data preparation and analysis, allowing them to work with data, automate workflows, and become fully embedded in research teams.



Lyon, L. Learning about research data in the lab at the Pitt iSchool.
<https://libraryconnect.elsevier.com/articles/learning-about-research-data-lab-pitt-ischool>



Multi-Faceted Framework

Skills: professional and informal training

Structures: organizational and managerial

Services: internal and external data savvy services

Stakeholders: researchers, IT, students, public, admin

Structures & Skills: Vicky Steeves

- Librarian for Research Data Management and Reproducibility (NYU Libraries & Center for Data Science)
- Part of team developing solutions that make it easier for researchers to share their computational work (ReproZip)
- Practitioner of reproducibility librarianship with explicit aim of helping researchers with the reproducibility and replication of their research
- Data-savvy librarian skills (understanding digital preservation policies and standards, unpacking R scripts, building tools and infrastructure)

Steeves, Vicky (2017) "Reproducibility Librarianship," Collaborative Librarianship: Vol. 9 : Iss. 2 , Article 4.
Available at: <https://digitalcommons.du.edu/collaborativelibrarianship/vol9/iss2/4>

Structures & Skills: The Carpentries

- Began with aim to teach research programmers better practices in software development
- Training programs that help the research community to learn new tools and methods
- Reduce tool overload, bridge gap between disciplines & roles, create community & shared learning materials

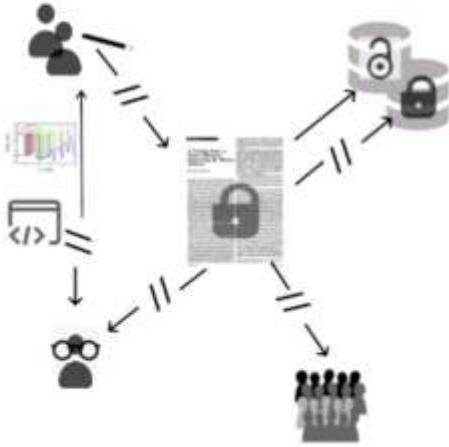
The Carpentries offers a unique space where multiple communities, from researchers to librarians, can come together, break down barriers, and learn new tools and methods in a lean, iterative fashion.

The Carpentries sits at the doorstep of libraries with a curriculum already in place, a worldwide network to learn from, and certificate-based instructor training programs.

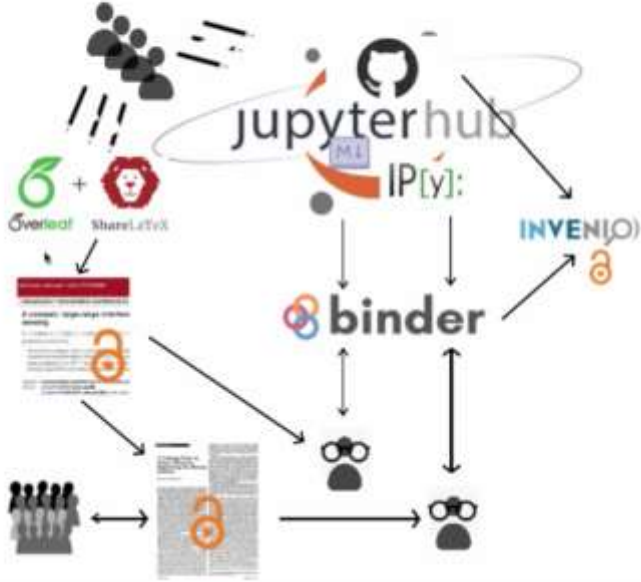


Services & Stakeholders: Caltech Library

NOW



FUTURE



Kremers, D., Antelman, K., Davison, S. From Stock to Flows. Coalition for Networked Information (CNI), Fall 2017 Membership. Video available at: <https://vimeo.com/250450136>

Services & Stakeholders: Carnegie Public Library

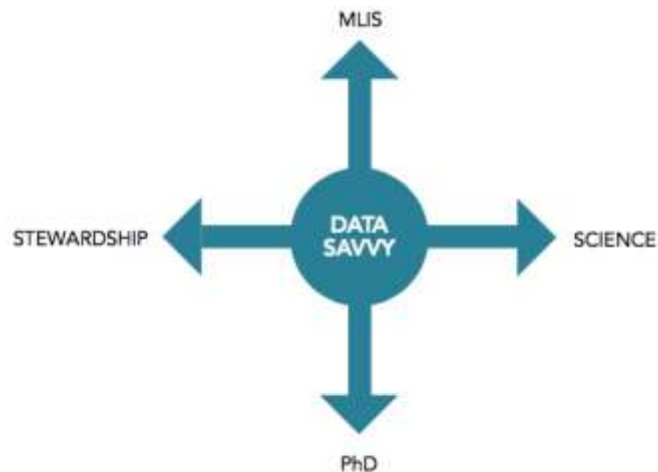


Source: <https://www.publicsource.org/mathematician-cathy-oneil-discusses-how-algorithms-can-perpetuate-inequity-why-math-needs-ethics-and-what-non-math-people-can-do/>

Tensions

Credentialing

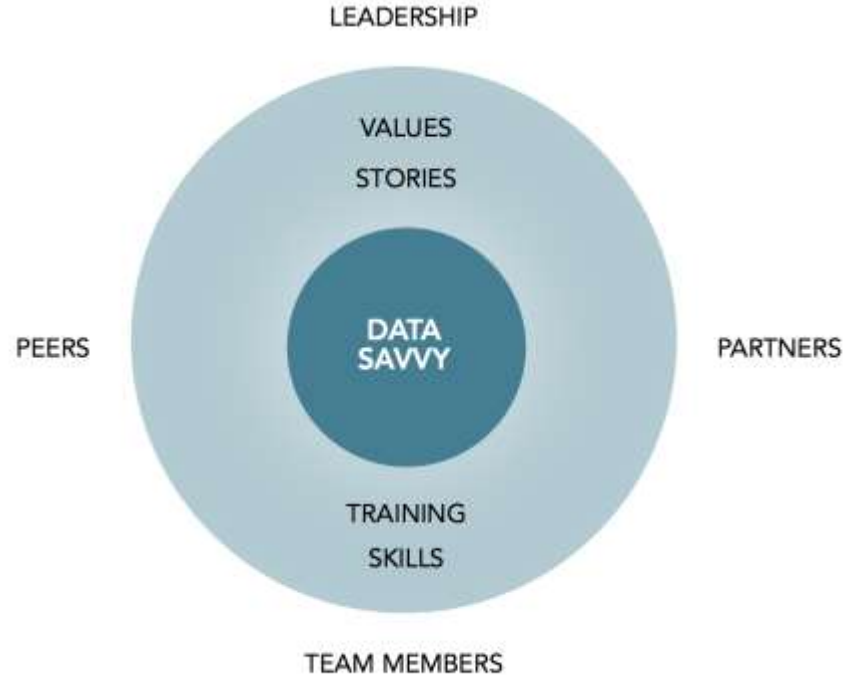
Re-engineering



Data Savvy 360 Model

Aspirational goals for a Data-savvy library organization, which can be characterized by:

1. Routine collection and strategic application of quantitative evidence
2. Effective communication and messaging driven by data-rich stories
3. Established professional education programs to extend and expand data savvy skills
4. An explicit understanding and articulation of the values and benefits of *science* and *stewardship* roles and responsibilities



THE DATA-SAVVY ROADMAP

FACET

Recommendation & Action

SCAN

Highlight success stories
Discover data lifecycle requirements

Perform data audit

Identify external datasets

STRUCTURES

Collaborate with leadership institutes

Utilize physical learning spaces

SKILLS

Utilize existing educational resources
Ethics and privacy

Advocate for software carpentry
Reposition the MLIS

Blended skills

SERVICES

Share assessment data
Libraries as amplifiers

Pilot data services
Triage services

Resources to scale-up
Re-engineer services

STAKEHOLDERS

Identify campus stakeholders
Outreach to researchers

Showcase capacity and capability
360-degree data-savvy model

Build data partnerships

Next Steps

Go on a **roadshow** to discuss the project and gather interest.

Report at <http://d-scholarship.pitt.edu/33891/>

Convene future **meetings** on data science in libraries.

E.g. Annual meeting, rotating locations

Explore opportunities for improving **discovery** of DS educational resources.

E.g. [The Journal of Open Source Education](#)

Share report findings with **leadership institutes**.

E.g. [Harvard](#), [UCLA](#)

Gather **training programs** and discuss shared, community program.

E.g. [Software Carpentry](#), Library Carpentry

Questions & Discussion

Dr. Liz Lyon

Visiting Assistant Professor

School of Computing and Information (iSchool),
University of Pittsburgh

Dr. Matthew Burton

Visiting Assistant Professor

School of Computing and Information (iSchool),
University of Pittsburgh

Chris Erdmann

Director of Library Carpentry

The Carpentries & California Digital Library

Bonnie Tijerina

Librarian, researcher

Data & Society Research Institute

<http://d-scholarship.pitt.edu/33891/>

Thanks to the Institute of Museum and Library Services.

PITTSSCI



Data&Society