Data Science in Libraries

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

http://d-scholarship.pitt.edu/33891/
IMLS RE-43-16-0149-16
As librarians increasingly use, build, and maintain the National Digital Platform, 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.
Data Science in Libraries Workshop

May 16th & 17th, 2017

Digital Scholarship Commons, Hillman Library, University of Pittsburgh

Pittsburgh, Pennsylvania USA
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.
Lyon, L. Learning about research data in the lab at the Pitt iSchool.
Environmental Scan: Training & Case Studies
Multi-Faceted Framework

(1) Organizational & Managerial Structures
(2) Stakeholders: Researchers, IT, Students, Administrators, Public…
(3) Professional & Informal Skills Training
(4) Data Savvy Services
Structures & Skills: Drivers

- Increasing computational, data driven needs of research community (i.e. Python data pipelines, Jupyter Notebooks)
- Recognition of library as resource for data management (i.e. Funder mandates)
- Not just a service provider, a collaborator with the research community, IT, industry...
- Shortage of data savvy employees in the library workforce (i.e. Informal training programs)
Structures & Skills: Barriers

- Branding - *Traditional vs DS Resource*
- Formal LIS Education - *Adaptable*
- Incentive Structure - *Purpose, Resources*
- Information Overload - *So Many Tools!*
- Drive By Workshops - *Bootcamps*
- Leadership - *Need Stories, Exemplars*
- **The Brick Wall** - *Inflexible Org Structures/Job Desc*

“I learn new skills, but I still need to do my old job.”
Structures & Skills: Exemplars

Training Programs:
- 23 Research Data Things
- Software/Library Carpentry
- Data Scientist Training for Librarians
- Data Science & Visualization Institute for Librarians

Librarians & Collaborative Work:
- Lauren Di Monte, Data & Research Impact Librarian
- Victoria Steeves, Librarian for Research Data Management and Reproducibility
Services & Stakeholders: Drivers

- Inclusive, cost-neutral, shared physical space
- Data for informed planning and problem-solving
- Campus use of metrics to demonstrate impact
- Increased use of data science in the classroom
- Facilitate collaboration across disciplinary team
Services & Stakeholders: Barriers

- Silo effect - *Seen as a fortress*
- Scale - *One-on-one vs full scale services*
- Resources - *Fitting in new services*
- Credibility and Image
- Experience - *Embedded knowledge of research process*
- Culture - *Break rules might be necessary*
Services & Stakeholders: Exemplars

Infrastructure

- Library of Congress (LoC) Lab Technical Pilot
- George Washington University Jupyter Notebooks Discovery & Use
- UIUC-Hathi Trust Research Center Textual Analysis

Community Programming

- Carnegie (Public) Library Data Science-Civic Tech Events
## FACET
- Short timeframe
- Recommendation & Action

## SCAN
- Highlight success stories
- Identify external datasets
- Discover data lifecycle requirements
- Perform data audit

## 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
Discussion & Reflections

- Professional Culture
- Organizational Culture
- Infrastructure
- Scale & Assessment
- Ethics & Values
Next Steps

Go on a **roadshow** to discuss the project and gather interest.
Mailing list at [http://d-scholarship.pitt.edu/33891/](http://d-scholarship.pitt.edu/33891/)

Convene future **meetings** on data science in libraries.
E.g. Annual meeting, rotating locations, ER&L

Explore opportunities for improving **discovery** of DS educational resources.
E.g. [The Journal of Open Source Education](https://thejose.org/)

Share report findings with **leadership institutes**.
E.g. Harvard, UCLA

Gather **training programs** and discuss shared, community program.
E.g. [Software Carpentry](http://swcarpentry.org), Library Carpentry
Acknowledgements

Matt Burton (University of Pittsburgh), Liz Lyon (University of Pittsburgh), Chris Erdmann (North Carolina State University), Bonnie Tijerina (Data & Society), Tracy Teal (Data Carpentry), Vernica Downey (Harvard University), Andrew Odewahn (O'Reilly Media), Lauren Di Monte (University of Rochester), Honora Eskridge (North Carolina State University), Carl Stahmer (University of California Davis), Dan Chudnov (Chudnov Consulting), Suchana Seth (Data & Society, Mozilla), Harriet Green (University of Illinois Urbana–Champaign), Alix Keener (University of Michigan), Steve Brewer (EU Edison), Kathryn Unsworth (Australian National Data Service), Jeannette Ekstrøm (Technical University of Denmark), James Baker (Library Carpentry), Josh Greenberg (Sloan Foundation), Amanda Rhinehart (Ohio State University), Ashley Sands (Institute of Museum and Library Services), Gail Clement (California Institute of Technology), Ian Cook (Highmark Health), Keith Webster (Carnegie Mellon University), Eleanor Tutt (Carnegie Library of Pittsburgh), Aaron Brenner (University of Pittsburgh), Tim Dennis (University of California Los Angeles), John Chodacki (California Digital Libraries), Jian Qin (Syracuse), Alistair Croll (Harvard Business School), Jan Brase (SUB Gottingen), Nic Weber (University of Washington), Meredith Schwartz (Library Journal), Renae Barger (University of Pittsburgh), Kornelia Tancheva (University of Pittsburgh), Jessica Kirchner (University of Pittsburgh), Chris Mills (University of Pittsburgh), Lauren Murphy (University of Pittsburgh)
Questions & Discussion

Dr. Liz Lyon
Co-Chair Department of Information Culture and Data Stewardship
School of Computing and Information (iSchool), University of Pittsburgh

Chris Erdmann
Chief Strategist for Research Collaboration
NCSU Libraries

Dr. Matthew Burton
Visiting Assistant Professor
School of Computing and Information (iSchool), University of Pittsburgh

Bonnie Tijerina
Librarian, researcher
Data & Society Research Institute

http://d-scholarship.pitt.edu/33891/
Thanks to the Institute of Museum and Library Services.