

Organizational Implications of Data Science Environments in Education, Research, and Research Management in Libraries

Erik Mitchell | Associate University Librarian & Associate CIO | U of California, Berkeley

Jenny Muilenburg | Data Services Coordinator | University of Washington

Vicky Steeves | Research Data Management and Reproducibility Librarian | NYU

December 14, 2015 | CNI Fall 2015 Meeting | Washington, DC

Scope

1. Introduce the Data Science Environments program
2. Explore perspectives around data science impact in libraries
3. Talk through potential positions/roles for data scientists in libraries
4. Next steps

Data Science Environments overview



ALFRED P. SLOAN
FOUNDATION

GORDON AND BETTY
MOORE
FOUNDATION

In 2013, the Gordon and Betty Moore Foundation and Alfred P. Sloan Foundation announced a new partnership with NYU, UC-Berkeley, and UW to “harness the potential of data scientists and big data for basic research and scientific discovery.” This is a five-year, \$37.8 million “cross-institutional effort to bring data science to the forefront of cross-disciplinary academic research.” From their materials:

“The Data Science Environments are working to bring about institutional change via campus-wide experimentation to catalyze a new era of research: cross-disciplinary efforts working towards new approaches to data-intensive discovery” (<https://www.moore.org/programs/science/data-driven-discovery/data-science-environments>).

Three campuses, three goals

This project has three core goals:

- **Develop meaningful and sustained interactions** and collaborations between researchers with backgrounds in specific subjects (such as astrophysics, genetics, economics), and in the methodology fields (such as computer science, statistics and applied mathematics), with the specific aim of recognizing what it takes to move each of the sciences forward;
- **Establish career paths** that are long-term and sustainable, using alternative metrics and reward structures to retain a new generation of scientists whose research focuses on the multi-disciplinary analysis of massive, noisy, and complex scientific data and the development of the tools and techniques that enable this analysis; and
- Build on current academic and industrial efforts to **work towards an ecosystem of analytical tools and research practices** that is sustainable, reusable, extensible, easy to translate across research areas, and enables researchers to spend more time focusing on their science.



NYU

**Center for
Data Science**

BERKELEY

Institute for
Data Science



eScience Institute

ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Moore/Sloan DSE working groups

Cross-university teams organize their efforts around six focal areas:

- strengthening an ecosystem of **tools and software** environments,
- establishing academic **careers** for data scientists,
- championing **education and training** in data science at all levels,
- promoting and facilitating accessible and **reproducible research**,
- creating **physical and intellectual spaces** for data science activities, and
- identifying the scientists' data-science bottlenecks and needs through directed **ethnography**.

The Center for Data Science at NYU

The Center for Data Science at NYU was established to give researchers a facility in which to work with big data in a multi-disciplinary setting. By advancing data science training and creating new research infrastructure, NYU's CDS collaborates with many departments across the global university, allowing for a diversity of programming for students and faculty as well as fostering a culture of active participation and collaboration amongst data scientists.



DSE at UW

The UW eScience Institute is made up of individuals with backgrounds in physics, astronomy, bio-engineering, bioinformatics, data management techniques, and computer science, who act as matchmakers, helping researchers apply the most appropriate technology to their research. It is located in the



Washington Research Foundation Data Science Studio, the space formerly occupied by the Physics/Astronomy Library.

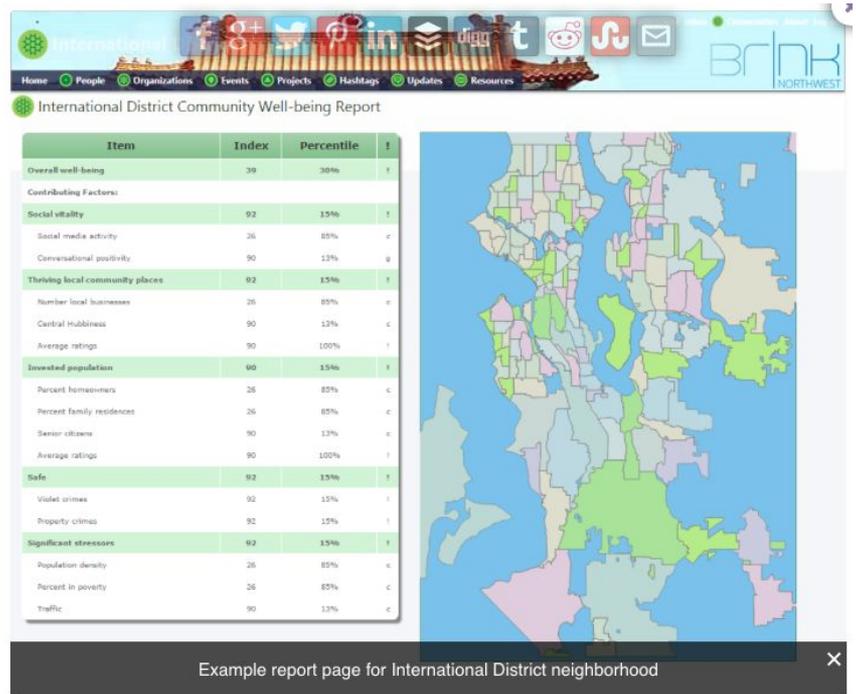
The WRF DSS is designed around the principle that innovative data science is advanced at universities through the creation of high quality physical spaces that successfully cultivate the “water cooler” effect, raise the level of prestige for data science and scientists, and are adaptable to a range of activities that can promote research collaborations and learning. The WRF Data Science Studio brings together eScience Data Scientists and researchers who reside in academic units spread across our large campus.

The Institute has formed a working group that is creating a template for data science education at the undergraduate level. A new Master of Science in Data Science has also been created at UW, as well as PhD programs in big data and data science in various departments as well as an integrative program that crosses department boundaries.

Early impact: Data science for social good

2015 projects @ UW eScience Institute:

- Assessing Community Well-being through Open Data and Social Media (photo)
- Open Sidewalk Graph for Accessible Trip Planning
- Predictors of Permanent Housing for Homeless Families
- Rerouting Solutions and Expensive Ride Analysis for King County Paratransit



DSE at UCB

A library space was transformed to host the Berkeley Institute for Data Science. This space serves as a work, collaboration, meeting and event space for all BIDS activities. BIDS often hosts related activities from groups across campus.

In parallel with the DSE, UC Berkeley has launched an undergraduate data science initiative (databears.berkeley.edu) that seeks to provide training for all undergraduate students in the coming years. The University has also launched a Data Science Planning Initiative that is asking broader questions around Data Science instruction and research in the University and which is co-chaired by leadership from the School of Information and the Department of Computer Science.

BIDS has served as an incubator for a number of projects - including Jupyter, iPython, R-Open Science and others. It serves, in part as a home for these research groups that did not have a home before.



How are our universities training data scientists?

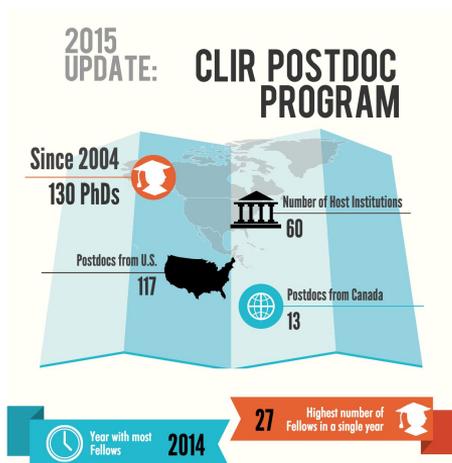
Institution	Educational programs	Position classifications
NYU	Master's degree program Several other programs with specialization in data science	Postdoctoral fellows Research engineer Tenure-stream faculty
UW	Undergraduate transcriptable option Master's degree/PhD programs	Postdoctoral fellows Tenure-stream faculty Data scientists Research scientists (dual appointment)
UCB	Undergraduate program Data Science Planning Initiative	Postdoctoral fellows Academic researcher roles Tenure-stream faculty

Our Questions:

1. How do DS and Librarians view the current state of connections between DS and Library and Information Science (LIS)?
2. What career paths exist for DS in Libraries?

Example programs / career paths

CLIR fellowship program



Digital Scholarship Centers
CNI report: <http://bit.ly/2014digschol>

UNIVERSITY OF MARYLAND

MITH MARYLAND INSTITUTE FOR TECHNOLOGY IN THE HUMANITIES

About Research Community

About

Maps & Directions

Contact

Policies

Home > About

The screenshot shows the top navigation bar with the University of Maryland logo and the MITH name. Below is a navigation menu with 'About', 'Research', and 'Community'. The 'About' page is active, showing a sidebar with 'Maps & Directions', 'Contact', and 'Policies'. The main content area has a breadcrumb trail 'Home > About' and a photograph of a person at a computer workstation.

Specialized consultants

UCSF Library

SEARCH & FIND

Today: Parnassus: 7:45am - 11pm MB Library: 9am - 9pm MB Hub: 24 hours

Find articles, journals, books and more

Home > Classes, Consulting & Help > Specialized Consulting Services

SPECIALIZED CONSULTING SERVICES

Various specialized Library contacts are available to meet with you.

REQUEST A CONSULTATION

The screenshot displays the UCSF Library website. At the top, there's a search bar and a navigation menu. Below the search bar, there's a search results area with a placeholder text 'Find articles, journals, books and more'. A breadcrumb trail reads 'Home > Classes, Consulting & Help > Specialized Consulting Services'. The main heading is 'SPECIALIZED CONSULTING SERVICES', followed by a sub-heading 'Various specialized Library contacts are available to meet with you.' and a prominent green button labeled 'REQUEST A CONSULTATION'.

Where are the existing connections between LIS + DS?

1. Data Science Specialization in Master of Library Science

Indiana University, Bloomington

Academic libraries are hungry for librarians who can work with and manage big data projects. With a specialization in data science, you can work on the forefront of this new science and support the work of academic data scientists.

2. Master of Information and Data Science (MIDS) Program

BIDS, University of California, Berkeley

The MIDS program is an innovative part-time fully online program that trains data-savvy professionals and managers. The MIDS program is distinguished by its disciplinary breadth; unlike other programs that focus on advanced mathematics and modeling alone, the MIDS degree provides students insights from social science and policy research, as well as statistics, computer science and engineering.

3 Job descriptions

Goal:

Think through potential roles for data science skills and interests in library environments, focusing on four areas (skills, roles, career path and impact)

Staff appointment

Dual appointment

Library academic appointment

Dual appointment



Skills:

- MLIS + DS

Impact:

- Program Development
- Improved Library Services
- Research Infrastructure

Roles/Career Paths:

- Research Infrastructure Developer/Manager
 - digipres for DS datasets
 - integration of HPC
- RDM Librarian!! ;)
 - reproducibility advocate
 - active outreach programming
- Data Science Subject Specialist
 - Library Liaison to DS School
- Library-ITS Services

Staff appointment

Idealized career path:

“I want to make progressive and transformative changes to an organization whose **mission is to serve scholarship broadly**....over time I want to serve in **product ownership, systems development and perhaps even leadership roles** in service of that mission”

Skills / qualifications

Certificate in data science methods

Specialized masters in DS

PhD with methods focus in DS

Expected roles

Technology / expertise translation

Information systems analysis and integration

Technology innovation / deployment

Impact

Contributing to original scholarship

Advancing efficiency / information impact in libraries

DS methods in library performance issues / metrics

Consulting role around data methods or DS issues

Career path

Advancement through IT

Advancement through business processes (e.g. domain expert or manager)

Strategic leadership (e.g. Chief Analytics Officer)

Library academic appointment

Skills / Qualifications

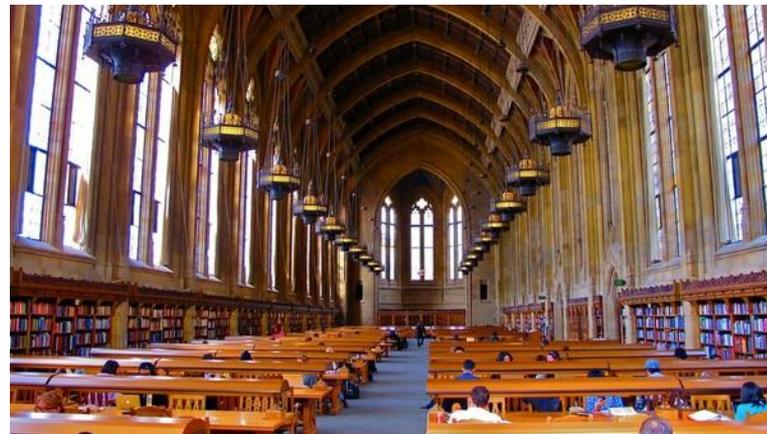
- MLIS; secondary Masters or PhD a bonus
- Background in data-focused research

Expected Roles

- Promote best practices for RDM
- Ensure research outputs are curated & archived

Impact

- Develop RDM and curation education
- Customize existing tools for RDM
- Improve grant-writing process



Career Path

- RDM librarian
- Domain expert for DS
- Grants/DMP manager
- Embedded librarian

Career path Lens

Data scientist as....	Challenges	Opportunities
Library-only staff appointment	'Siloed' advancement options Roles tied to library mission / services Fitting DS expertise and career into library mission	Non-research advancement Fit with library information system model New skill sets and perspectives
Academic library appointment	Fit with other librarian academic roles Collaborators / research	'Academic' style advancement Expanded roles/opportunities Opportunities for collaboration
Academic multi-departmental appointment	Work required to ensure impact fits library mission Addressing 'home department' issues	Expanded research areas Work with and/or develop new & interesting LibTech Better fit with some DS career paths

Next Step: Survey

Demographic questions: what is your current position and department?

What skills do you believe belong to a librarian? What are the top 3 skills?

What skills do you believe belong to a data scientist? What are the top 3 skills? How would you describe a data scientist?

For librarians:

What roles do you believe Data Science roles would play in the library?

For Data Scientists:

What job elements would you value in your career (e.g. service; original research; fame)

What kind of position is appealing to you (faculty, private sector, academic staff, research staff)

Have you ever considered a position in a library | Why or why not?

Questions & Discussion Points

How do the position descriptions match with your vision of the role of data science / data scientists?

How might the inclusion of data scientists impact your library/organization?

Where might data scientists help us answer old problems with new solutions?

Erik Mitchell | @mitcheet | emitchell@berkeley.edu

Jenny Muilenburg | @jmuilenburg | jmuil@uw.edu

Vicky Steeves | @VickySteeves | victoria.steeves@nyu.edu