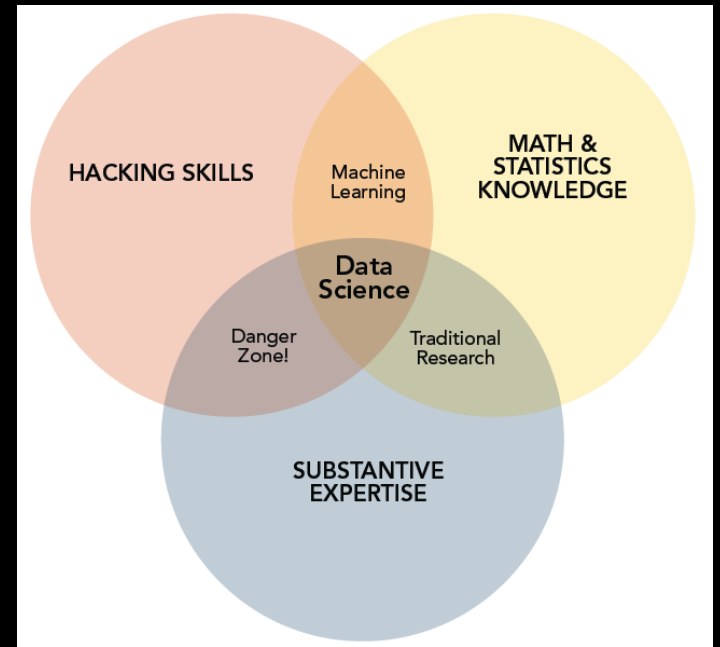# *Shifting to Data Savvy: The Future of Data Science in Libraries*

Liz Lyon and Chris Erdman

LIBER Conference July 2018, Lille France

Agenda

- Data Savvy Report
- The Management Gap
- Data Science Roles
- The Skills Gap

- Supporting Data Savvy Communities
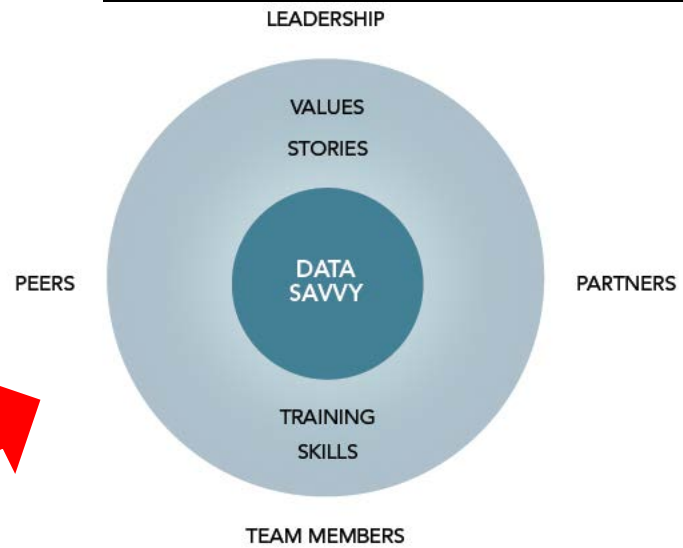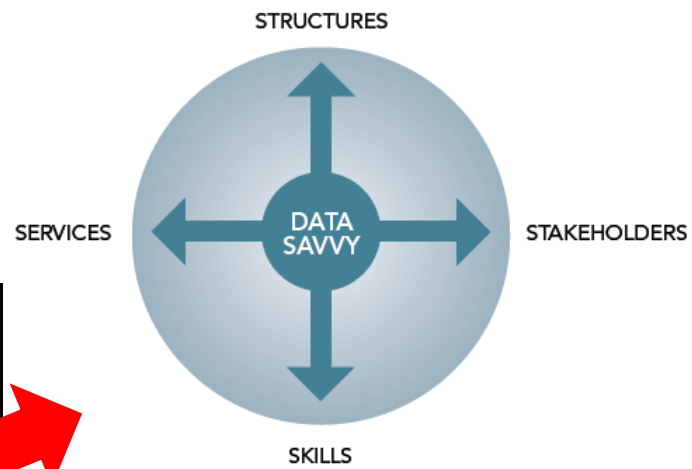- Training Strategies
- Library Carpentry
- Help needed!

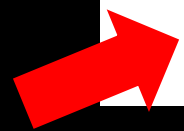# Data Science in Libraries White Paper

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

- IMLS-funded study
- Data-savvy concept

**Shifting to Data Savvy:
The Future of Data Science
In Libraries**

Matt Burton
Liz Lyon
Chris Erdmann
Bonnie Tijerina

STRUCTURES

SERVICES — DATA SAVVY — STAKEHOLDERS

SKILLS

LEADERSHIP

VALUES

STORIES

PEERS — DATA SAVVY — PARTNERS

TRAINING

SKILLS

TEAM MEMBERS

- Multi-faceted framework
- Re-engineering tension
- Credentialing tension
- Data-savvy 360 model organization
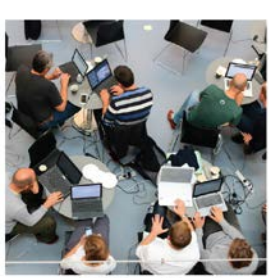
Data Science in Libraries White Paper 1

Shifting to Data Savvy: The Future of Data Science In Libraries

Matt Burton
Liz Lyon
Chris Erdmann
Bonnie Tijerina

- Skills gap
- Management gap

STRUCTURES

SERVICES ← **DATA SAVVY** → STAKEHOLDERS
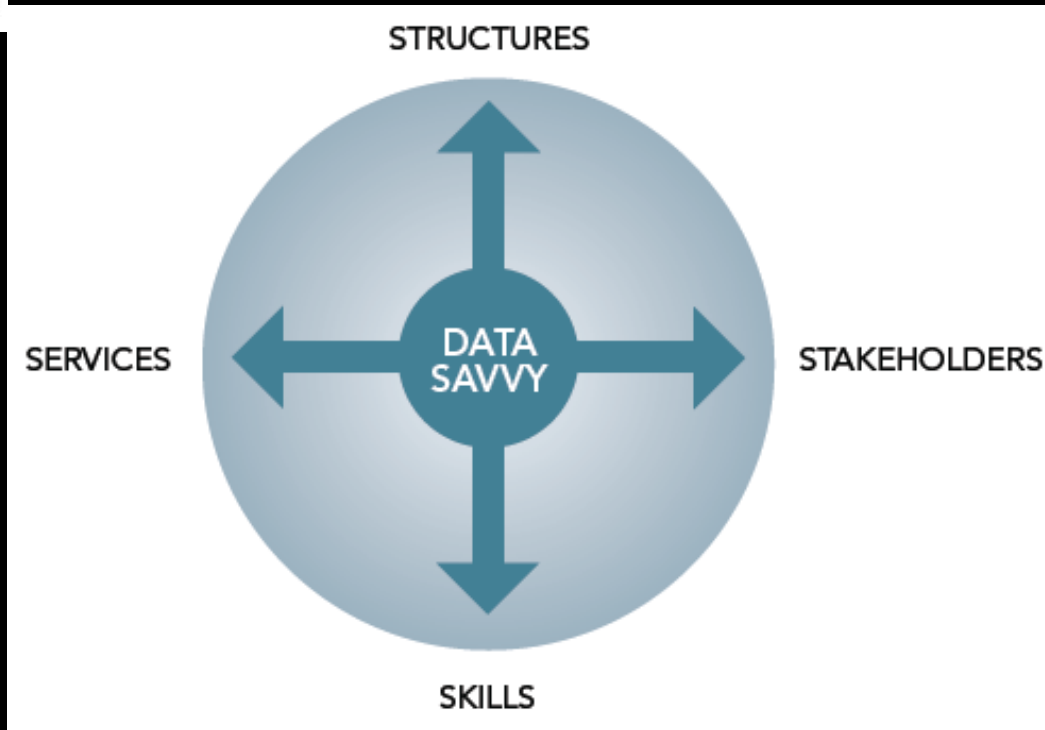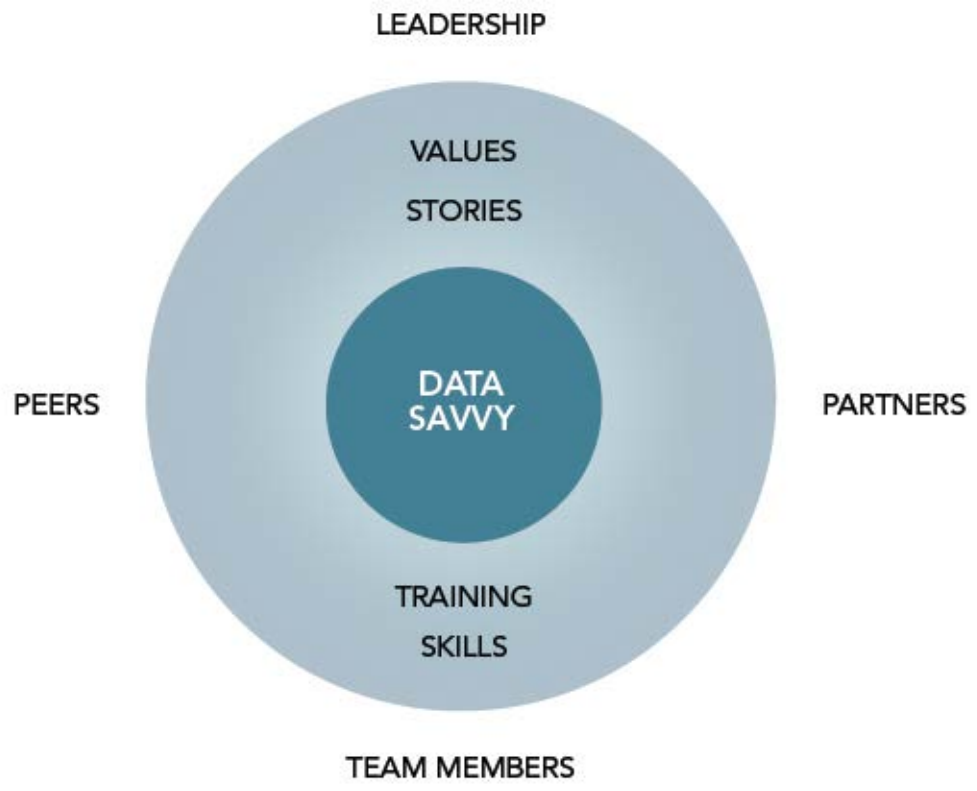
SKILLS

- ***Re-engineering tension***: data stewardship vs data science
- ***Credentialing tension***: PhD + deep disciplinary knowledge vs MLIS + broad information skills

Data Science in Libraries White Paper 2

- Skills gap
- Management gap

- Service opportunities, internal data, external (research) data
- Leadership, data-savvy teams, collaboration, partners
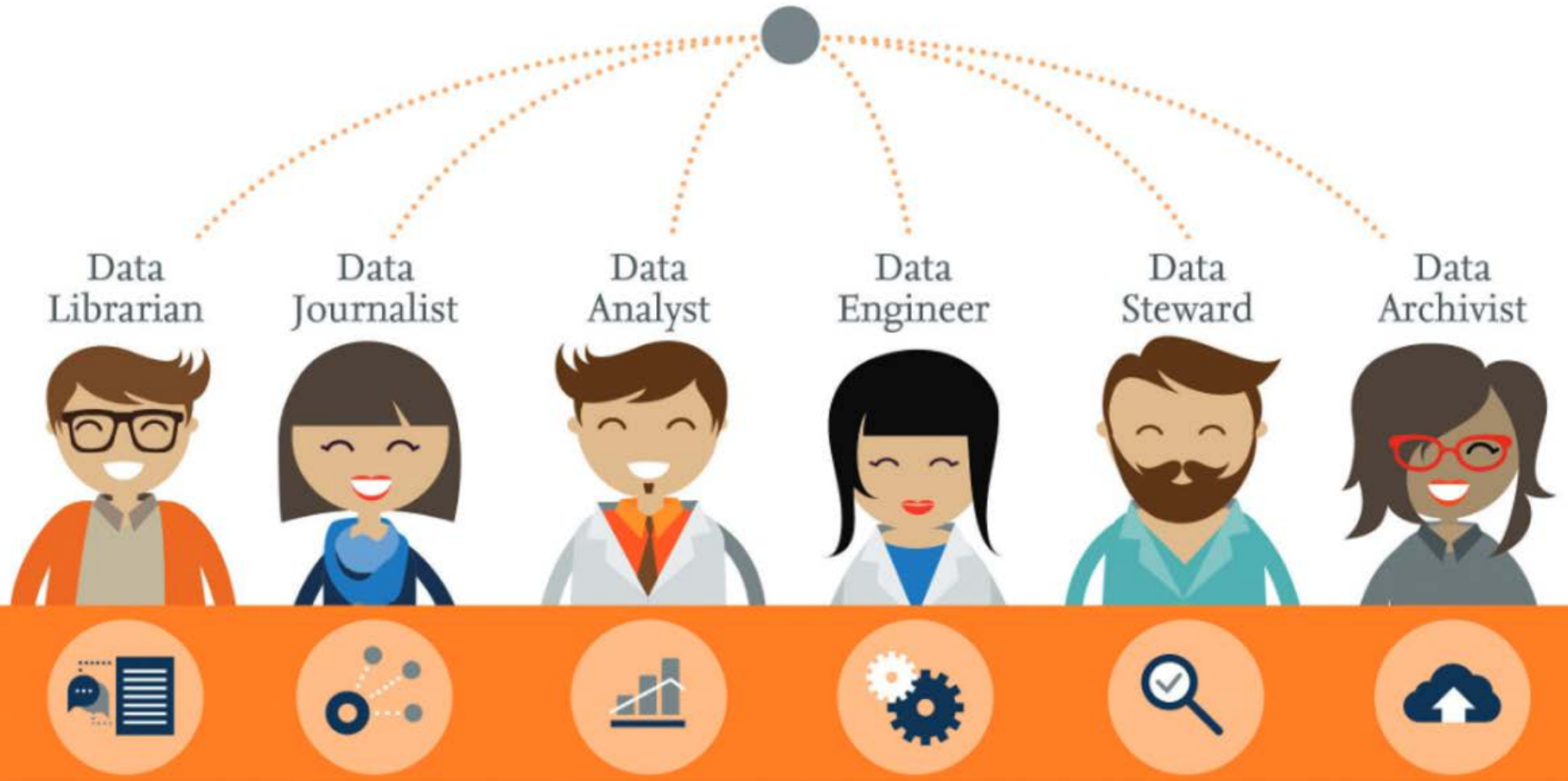- Embedded data roles, scaling challenges, credibility, skills

Luis Martinez-Uribe
Data Scientist
DataLab, Library
Fundación Juan March

IDCC – Barcelona 21 February 2018

Exemplar

The DataLab promotes a data centric vision to enable data exploitation via curation
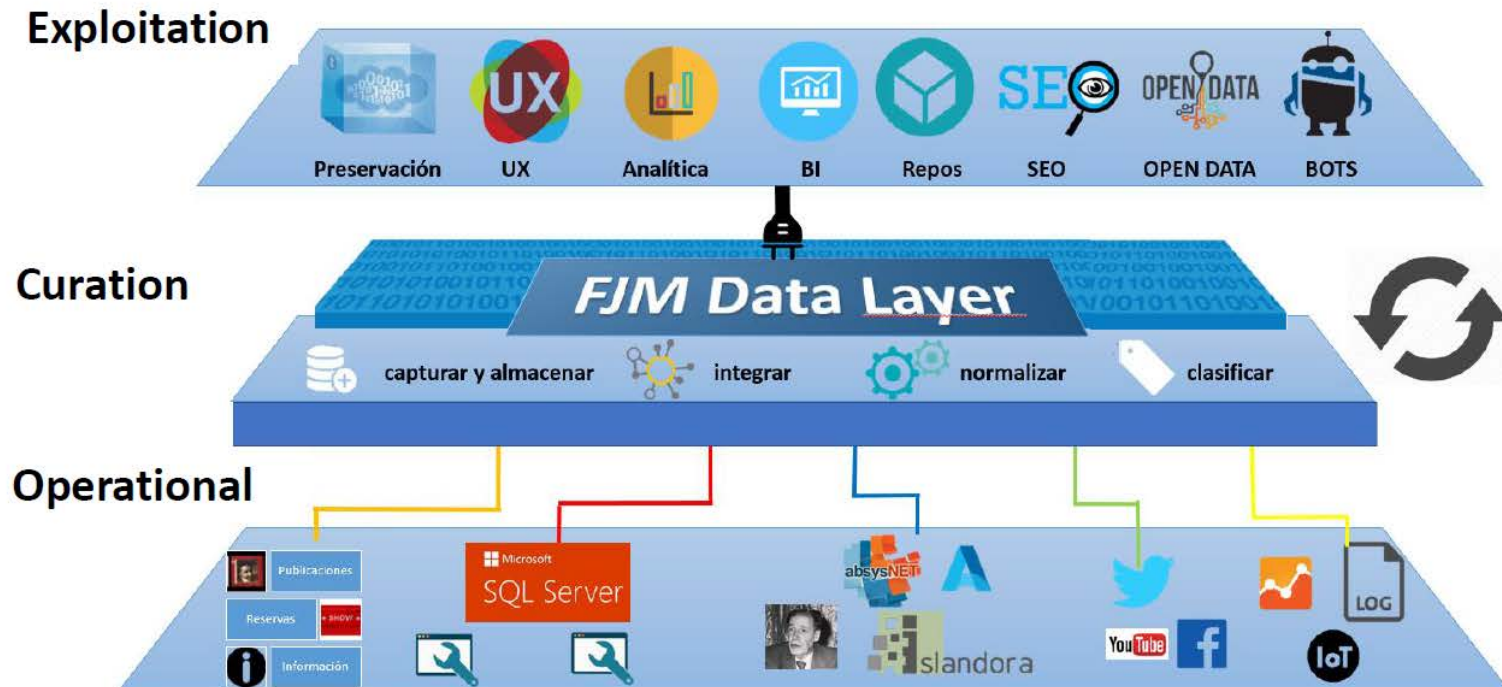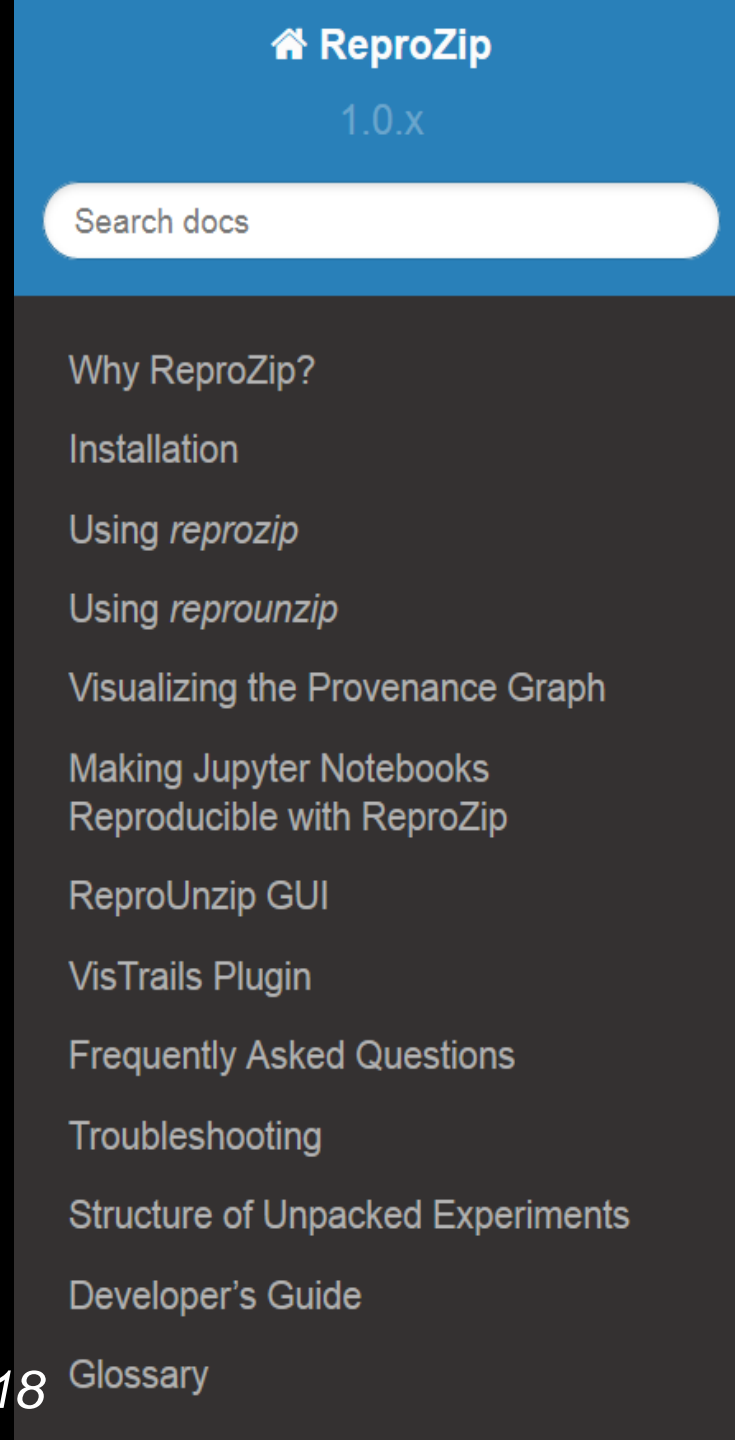
Data-savvy roles

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

Next Gen data stewardship: FAIReR data roles?

- *Reproducibility Librarian*
- *Research Data Auditor*
- *Data Verification/Validation Officer*

Data Librarian

Lyon Gottingen/CODATA Symposium March 2018

ReproZip
1.0.x

Search docs

Why ReproZip?

Installation

Using *reprozip*

Using *reprounzip*

Visualizing the Provenance Graph

Making Jupyter Notebooks Reproducible with ReproZip

ReproUnzip GUI

VisTrails Plugin

Frequently Asked Questions

Troubleshooting

Structure of Unpacked Experiments

Developer's Guide

Glossary

RDM & Reproducibility Librarian @ NYU (Vicky Steeves!)

Next Gen data science roles?

- Data pool / data lake as analysis substrate
- Predictive models and simulations
- Cognitive computing applications
- Growth of Decision Science

- *Cognitive Data Analyst*
- *Forecast Modeller*
- *Decision Engineer*

Lyon CODATA Conference October 2017
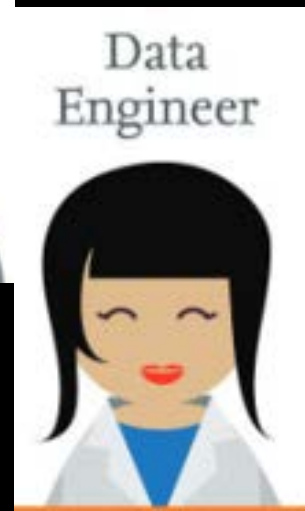
Library Data Analyst (faculty)

- Advertised March 2017
- Virginia Tech
- Promote value of library data
- Collects, analyzes, reports, documents….
- Expertise in SPSS, Stata, R, NVivo etc
- Business Intelligence Systems
- Liaison role with new Office of Academic Decision Support

Repositioning LIS education curriculum

MLIS Redesign: Stewardship and Science

"we need research-heavy librarians"

"we are looking for librarians with skills to support not only information discovery, but also analysis and creation -- in other words those who can provide expertise to students and faculty across the full research lifecycle."

LIS Training opportunities…..

…over to Chris