

The Data Savvy Librarian

#DST4L



C. Erdmann, NCSU Libraries
@libcce | christopher_erdman@ncsu.edu

The Fall of the Library Fortress

@clauersen

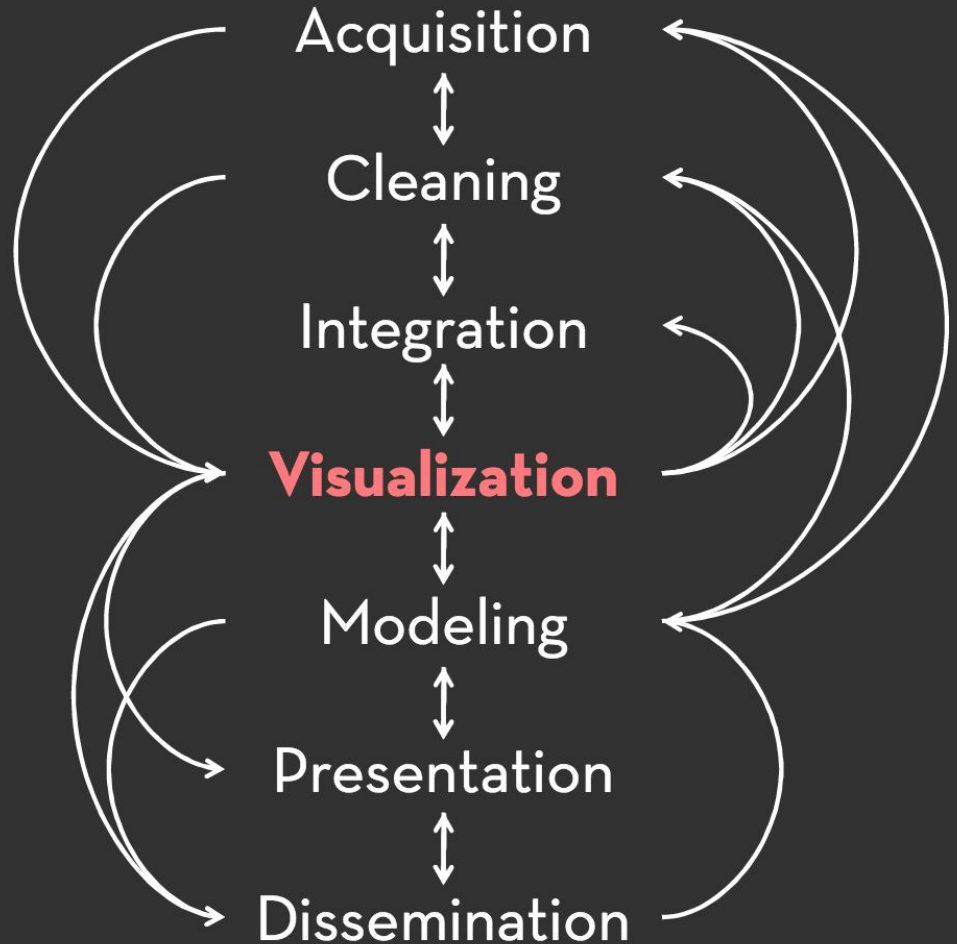




Meet Them at the Watering Hole | @snet_jklump

Experience the Research Data Lifecycle

Interactive Data Analysis
@jeffrey_heer



Hands on w/



Goals

- Allow librarians to experience and gain a better understanding of the research data lifecycle.
- Train librarians to have skills and practical knowledge to build and use data tools (OpenRefine).
- Forward a culture within libraries that enables data-centric approaches (change our mindset).
- Grow a community of data savvy librarians.

Empower librarians!

Following a Path & Providing Context



The Challenge Ahead



A Pythonic Solution

```
lockerrors_processing.py - F:\Portable Python 2.7.5.1\lockerrors_processing.py
File Edit Format Run Options Windows Help

import csv
import os
import re

bibtotal = 0
xmltext = []

os.chdir('I:\\Cataloging\\Database Management\\OCLC Error Reports\\Output\\')
for fn in os.listdir('.'):
    if fn.endswith('.udev'):
        f = open(fn, 'rU')
        lines = f.readlines()
        f.close()

        count = 0
        records = {}
        data = []
        reportdate = ''

        for line in lines:
            if len(line) == 2:
                if data[-1].rstrip().startswith('ERROR:'):
                    records[count] = data
                else:
                    count += 1
                    data = []
                    records[count] = data
            else:
                data.append(line)
                records[count] = data
                if data[-1].rstrip().startswith('RECORD REJECTED'):
                    count -= 1
                    data = records[count]
                    data.append(line)
                    records[count] = data
                if data[-1].rstrip().endswith('BATCHLOAD INPUT RECORD PREPARATION'):
                    reportdate = str(data)[10:21]

#Get only the Houghton (MH-H and MH-HT) records
hourecords = {}
count = 0
biblist = []

for k, v in records.items():
    errorcount = 0
    if any('aMH-H=' in v for v in v) or any('aMH-HT=' in v for v in v):
        count += 1
        linecount = 0
        houfields = {}
        houlines = []
```

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



Alpine Club. *Peaks, Passes, and Glaciers: Being Excursions by Members of the Alpine Club*. London: Longman, Green, Longman, and Roberts, 1862. <http://books.google.com/books?id=FXgSAAAAAYAAJ>.

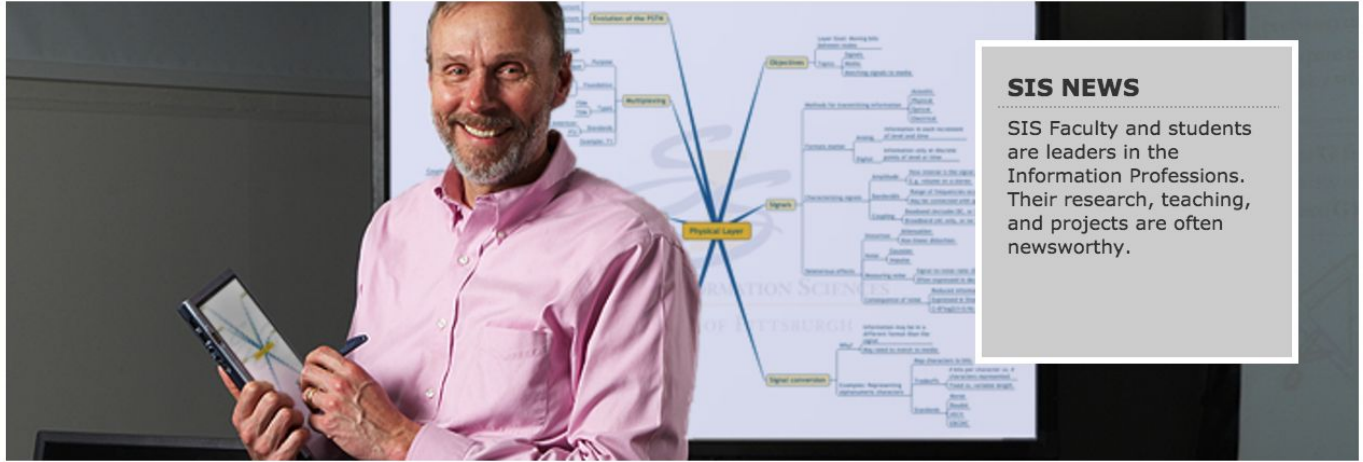
@vmdbks

Library Connect

Partnering with the Library Community

Data Scientist Training for Librarians: A course and a community

By Jeremy Guillette and James Damon, Harvard University | Nov 13, 2014



SIS NEWS

SIS Faculty and students are leaders in the Information Professions. Their research, teaching, and projects are often newsworthy.

- [ABOUT](#)
- [ADMISSIONS](#)
- [ACADEMICS](#)
- [PEOPLE](#)
- [RESEARCH](#)
- [ALUMNI](#)
- [CONTACT US](#)

SIS faculty awarded more than \$200,000 from the Institute of Museum and Library Services

09/15/2016

NCSU Libraries part of Data Science grant from IMLS

September 20, 2016

Share [f](#) [t](#) [e](#)



Image courtesy of Mikael K. Elbæk

Chris Erdmann, Chief Strategist for Research Collaboration at the NCSU Libraries, is part of a multidisciplinary team that was awarded a \$97,911 grant from the [Institute of Museum and Library Services](#) (IMLS) through their Laura Bush 21st Century Librarian Program. Headed up by researchers at the University of Pittsburgh, "[The Data Scientist as the 21st Century Librarian?](#)" will provide recommendations about the data-science skills that librarians need, and build model data science training for research librarians.

Erdmann's participation adds to the NCSU Libraries' portfolio of leadership efforts to train librarians in partnering with researchers across the lifecycle of their work through such in-house programs as the [Data and Visualization Institute for Librarians](#).

OTHER NEWS

[Pentair Foundation supports STEM programming, instruction at the NCSU Libraries](#)

A \$35,000 grant funds a new fellows program and sponsors the "Making Space" events series. [→](#)

[Do you know your animals?](#)

eMammal Lite, a Citizen Science app developed by the NCSU Libraries, needs your help identifying animals [→](#)

[Cha-ching for the NCSU Libraries](#)

Student "Penny Wars" campaign raises some coin—literally—for the Libraries. [→](#)

[all news](#) [→](#)

UPCOMING EVENTS

- | | |
|------------|---|
| DEC | Read Smart Book Discussion - Inside the O'Briens by Lisa Genova |
| 6 | |



DST4L

as a Service