

University of North Carolina at Chapel Hill, Health Sciences Library

Libraries as Indispensable Partners and Connectors in Data Science



Nandita Mani, PhD, MLIS
Associate University Librarian
for Health Sciences & Director,
UNC Health Sciences Library



Fei Yu, PhD, MPS
Health Informatics Librarian,
UNC Health Sciences
Library & Assistant Professor,
SILS



Michelle Cawley, MLS
Head of Clinical, Academic,
and Research Engagement
UNC Health Sciences Library



Presentation Outline

- The value libraries bring to the Data Science ecosystem
- Opportunities for Collaboration between libraries and iSchools
- Pedagogical Considerations when teaching Data Science

Slide 2 Notes: Presentation Outline

Libraries play an integral role as learning partners and connectors throughout the data science lifecycle in curriculum and research. We will discuss the UNC library system-wide data science initiative around the following issues (1) how to scaffold the integration of data science competencies through layers of education and help solve real-world problems; (2) how to harness the expertise from interdisciplinary experts, and incorporate pedagogical approaches that facilitate problem-based learning; (3) how to bridge the gaps in skill acquisition, application of data science tools, and allocation of resources by playing an intermediary role. The discussion will be framed around social sciences and humanities, digital humanities, and a special emphasis on biomedical and health sciences. Examples will include the research collaborations that the UNC Health Sciences Library established with several campus units, including North Carolina Translational and Clinical Science Institute, UNC School of Medicine, School of Pharmacy, and other entities on campus that host large volumes of data.

What is Data Science?

- The work of transforming data into information and then transforming information to knowledge
- The process of using data to find solutions or predict outcomes

Slide 3 Notes: What is Data Science?

For the purposes of this discussion, these are the definitions we are using as it relates to Data Science.

Areas of Library Expertise

- **Tools & Programming**

R Studio | Python | Tableau |
Reference Managers | Domain-
Specific Tools and Programming

- **Data Sourcing & Acquisition**

Data Catalog | Navigating Public &
Proprietary Options | Data
Management | Data Licensing

- **Data Creation**

Digital Humanities

- **Data Cleaning & Preparation**

Visualization & Other Data Presentation | Data
Preservation & Archiving | Data Management

- **Analysis & Tool Testing**

GIS | Impact Measurement & Visualization
Text Analysis | Network Analysis | Inference
Statistics

- **Research Methods**

Systematic Review | Bibliometrics | Semantic
Analysis

Slide 4 Notes: Areas of Library Expertise

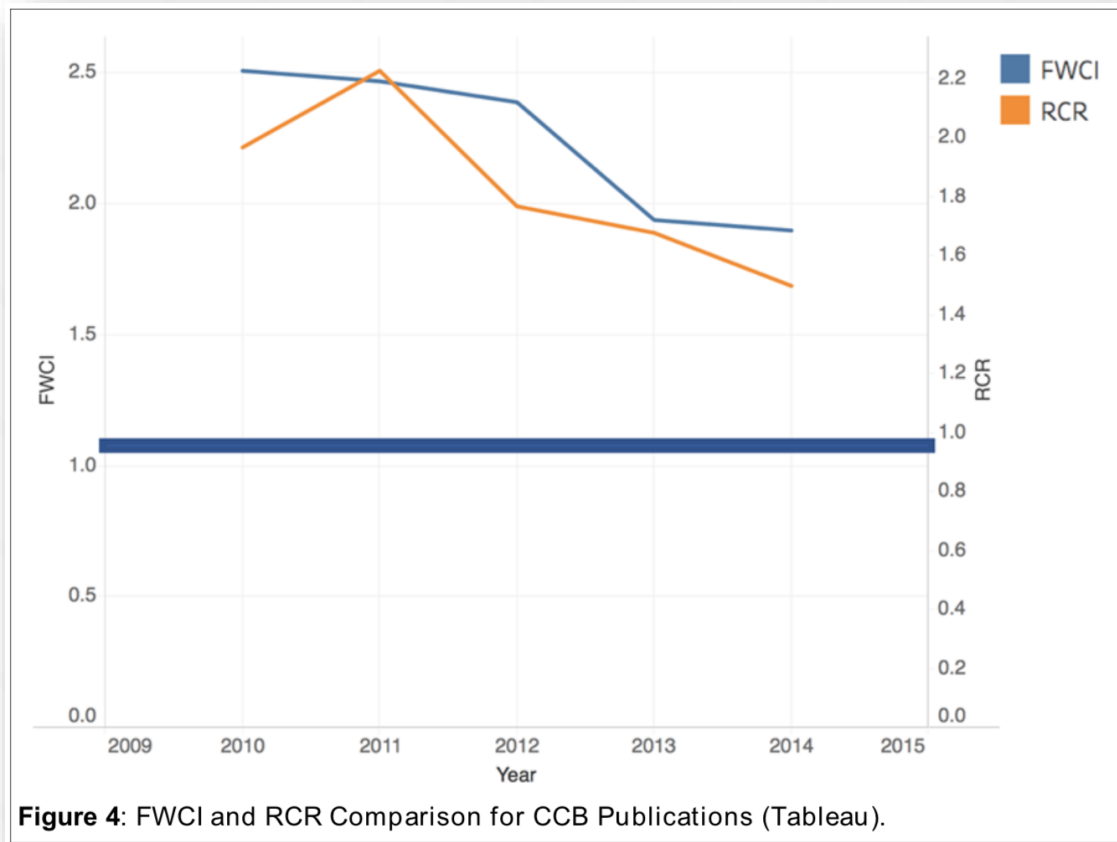
Librarians and archivists have expertise in a range of areas related to data science. Here are some examples of expertise held by library staff that can be integrated into the larger data science context.

Slide 5 Notes: Library Partnerships – Digital Humanities

Examples that showcase how libraries engage in various aspects related to data science:

- [#SilentSam](#) is a twitter data curation project at the University of North Carolina at Chapel Hill (UNC-CH) Libraries in wake of the recent protests regarding the Confederate Monument on the UNC-CH Campus. The University Libraries used a python tool, twarc, to harvest tweet data for specific hashtags searches (#silencesam and #silentsam). In addition, libraries provide instructions on how to use the data set and [finding aids](#).
- [On the Books: Jim Crow and Algorithms of Resistance](#) is a text mining and machine learning project of the UNC-CH Libraries in collaboration with experts in African American history, information science, legal information, and digital research. Funded by the Andrew W. Mellon Foundation, this project aims to identify racist language in legal documents and help expose the wide-range effects of Jim/Jane Crow on the American South.

Partnerships: CCB Program at UNC Lineberger Cancer Center



Librarians compared the **citation impact** of Cancer Cell Biology (CCB) Program at UNC Lineberger Cancer Center with the average research papers funded by NIH using Relative Citation Ratio (RCR) and the world average publication in the same field and published in the same time using Field Weighted Citation Impact (FWCI).

Source: Yu, F., & Hayes, E. B. (2018). Applying data analytics and visualization to assessing the research impact of the Cancer Cell Biology (CCB) Program at the University of North Carolina at Chapel Hill. *Journal of eScience Librarianship*, 7 (1), 1-20. Doi: <http://dx.doi.org/10.7191/jeslib.2018.1123>

Slide 6 Notes: Partnerships – CCB Program at UNC Lineberger Cancer Center

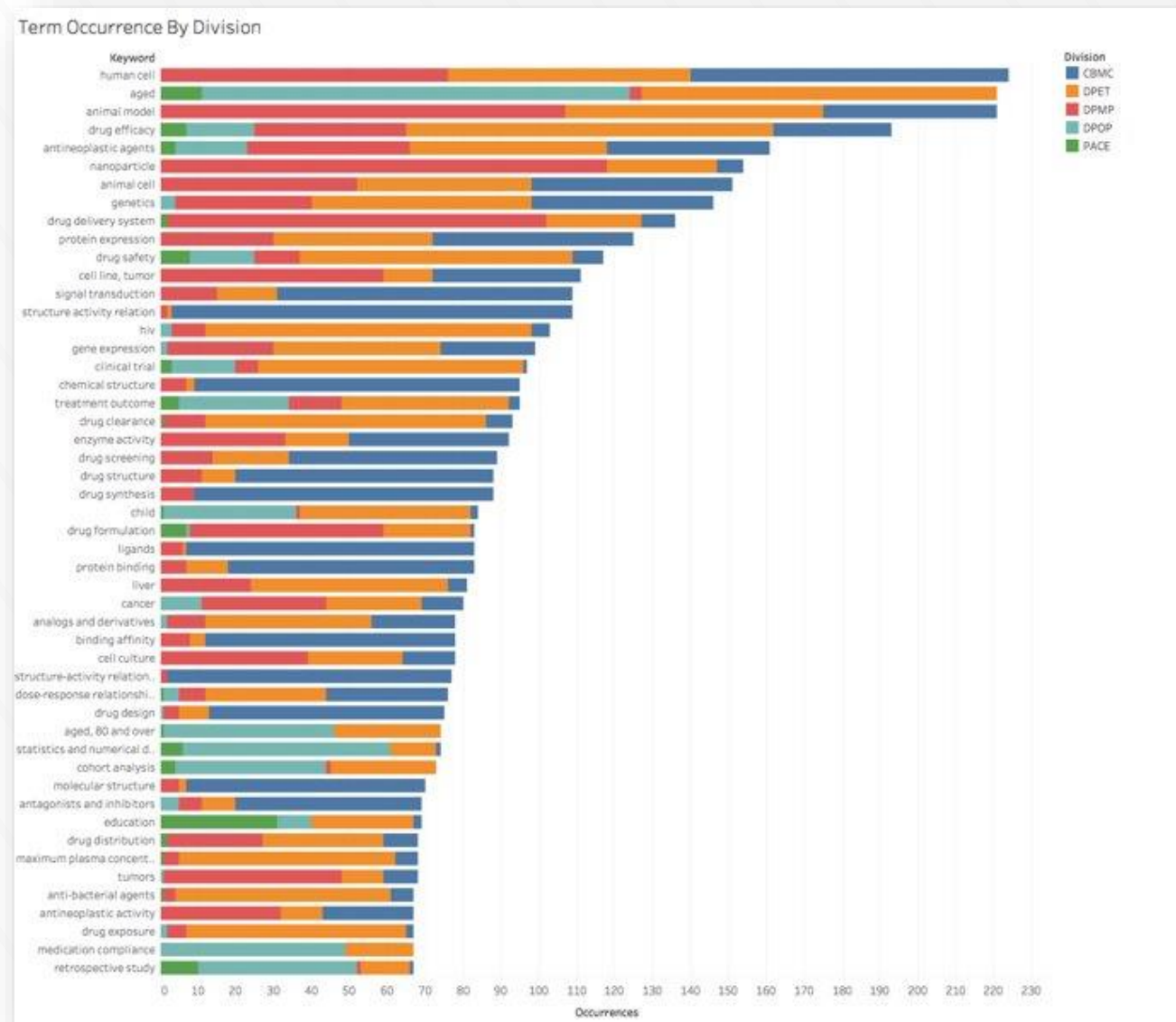
Lineberger research impact using bibliometrics: The UNC HSL conducted a research impact assessment project for the Cancer Cell Biology (CCB) Program in the Institution's cancer center through bibliometric data analysis and visualization. The analysis not only revealed the significant scientific output and citation impact of the CCB program in the examined time periods, but also provided quick insights to the administrators in terms of identified patterns, trends, and gaps for their supported research activities.

Slide 7 Notes: Partnerships – North Carolina Translational and Clinical Sciences Institute (TraCS)

We applied text-mining techniques to NC TraCS-supported publications and constructed a research topic network using key-terms extracted from titles and abstracts. The four clusters that these key terms formed within the network demonstrated that NC TraCS-supported research has focused on all phases of translational research, including but not limited to lab discovery, clinical research, physician care, and population science. HSL has been helping TraCS validate their mission and measure their supported translational science activities at an institution.

Partnerships: UNC Eshelman School of Pharmacy

Librarians visualized overlapping areas of curricula for the five divisions at the UNC Eshelman School of Pharmacy in support of **curriculum transformation.**



Slide 8 Notes: Partnerships – UNC Eshelman School of Pharmacy

Library participation on our Eshelman School of Pharmacy Graduate School Visioning Committee afforded the opportunity to bring to bear library expertise in wrangling data in the form of bibliometrics, which were then used to highlight research collaboration amongst faculty in various divisions, and research topics explored over a set period of time to then ascertain where opportunities may exist for furthering curricular threads.

Opportunities for Collaboration between Libraries and iSchools

Explore Data Science Questions Relevant to Libraries, for example:

- Where do UNC-affiliated authors publish? How likely are they to publish in open access journals? Where are they likely to publish in the next 5 years?
- What are the efficiencies and usability of library technologies around using AI and machine learning tools and approaches?
- What is the efficacy of information retrieval using AI approaches?
- How can we apply predictive analytics to grants, patents, and literature data?

Engage Librarians as Professors of Practice

- Establish *Professor of Practice* program where senior level experts bring practical application to instruction at iSchools.
- Integrate librarians as adjunct instructors in SILS courses; data science offers the opportunity for further integration.
- Explore opportunities for MOUs and joint appointments between libraries and iSchools.
- Develop research methods and resources (e.g., [CALA Program](#), RAs, work-study and field experience).

Slide 9 Notes: Opportunities for Collaboration Between Libraries and iSchools

Engaging librarians as Professors of Practice can be encouraged by integrating expertise in research methods and resources, teaching in SILS curricula, and by providing the opportunity for joint appointments or MOUs. Facilitating the establishment of a library associate program (e.g., UNC's CALA Program <https://library.unc.edu/carolina-academic-library-associates/>) and providing opportunities for Research Assistants (RAs) and field experience students to work in libraries helps increase awareness of how library theory and practice are integrated in library systems and how areas such as data science can be explored by working in a library environment.

Opportunities for Collaboration between Libraries and iSchools

- UNC HSL and UNC School of Information and Library Sciences (SILS) are currently exploring an opportunity around partnering to:
 - Develop an open-access, web-based user interface for machine learning tools.
 - Build capacity among librarians for applying AI-based approaches to bibliographic data including through a three-day curriculum.

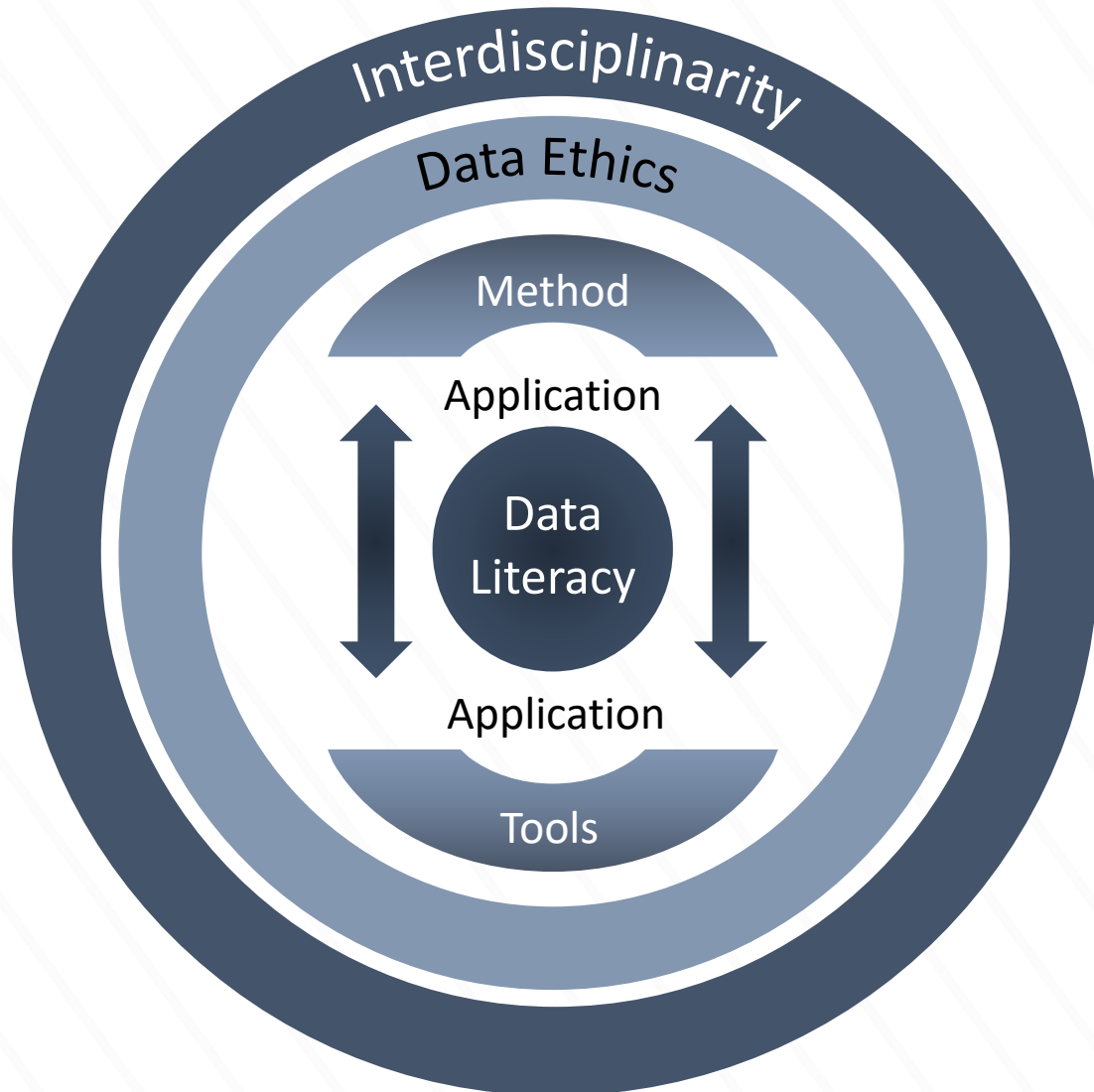
U.S. Association for Research Libraries (ARL) 2019-2021 Priorities:

- Facilitate a culture of innovation within research libraries.
- Provide data and analytics on research library practices, effectiveness, and impact.

Slide 10 Notes: Opportunities for Collaboration Between Libraries and iSchools

Building connections between libraries and iSchools can help to bridge gaps in skill acquisition and provide opportunities to apply data science tools to real world problems.

Curricular Model for Data Science



Begin with **foundational components** of data literacy including, understanding jargon, data life cycle, [FAIR Data Principles](#), Open Science, and interpreting data among other topics.

As learners move beyond fundamentals, the next phase will focus on specific **tools** (e.g., R, Python, Tableau, statistical software) and **methods** (e.g., data mining, machine learning, bibliometric analysis, systematic review) to use in answering complex questions.

As students engage in this process it is equally important to provide opportunities for **application** of these competencies in context (i.e., towards answering a particular question or addressing a specific problem).

Ensuring that **ethical implications** are considered from an interdisciplinary lens (e.g., algorithmic bias, Personally Identifiable Information) will help learners have a more holistic perspective of how data science skills and information can be integrated.

When approaching a question or problem it is optimal to convene an **interdisciplinary** team including subject matter experts as well as individuals with data science expertise.

Slide 11 Notes: Curricular Model for Data Science

- When considering a data science curricular structure, ensuring **foundational data literacy elements** are at the crux of the curriculum is essential.
- Exposing learners to **tools and methods** by which data science can be explored and providing opportunities to apply these skills will help ensure concepts are understood in-context.
- **Data ethics** and seeking to investigate data science questions from an **interdisciplinary lens** will enable a more holistic understanding of the question at hand and will help elucidate areas where further investigation or inquiry may be necessary. Taking a holistic view of teaching data science from an interdisciplinary lens is critical for students to understand how various information elements and discipline-based knowledge come together to answer questions and solve problems.
 - Can be modified by discipline: Identify stakeholders to understand the context of the problem you are trying to address; maximize instructional time by encouraging students to participate in library-led workshops or training. This can then reduce the amount of time spent in-class on specific tools so that face to face time can be spent applying concepts to specific cases/problems.