Working with a Data Librarian
What’s a Data Librarian?

Research skills

Technical skills

Deep knowledge of data landscape

Image via xkcd

HEY, LOOK, WE HAVE A BUNCH OF DATA! I'M GONNA ANALYZE IT.
NO, YOU FOOL! THAT WILL ONLY CREATE MORE DATA!
Current Work

• Supporting research
• Teaching
• Advising on data policy and infrastructure
  • *For example*: NIH Data Management and Sharing Policy
• Conducting research

Image by Terry Dagradi and courtesy of Cushing/Whitney Medical Library
Clinician-Librarian Collaboration

Images via The Noun Project. L to R credit: Wilson Joseph, Saeful Muslim, and Miroslav KURDOV.
Example of original, collaborative research

Original Investigation  | Geriatrics

January 6, 2023

Association of Nursing Home Exposure to Hurricane-Related Inundation With Emergency Preparedness

Natalia Festa, MD, MBA1,2; Kaitlin F. Throgmorton, MLIS3; Nora Heaphy, BA4; Maureen Canavan, PhD, MPH5; Thomas M. Gill, MD1

Author Affiliations  | Article Information


Key Points

**Question**  Are nursing homes exposed to potential hurricane-related inundation more likely to meet Centers for Medicare & Medicaid Services criteria for adequate emergency preparedness?

**Findings**  In this cross-sectional study of 5914 nursing homes, a higher prevalence of emergency preparedness deficiencies among nursing homes exposed to hurricane-related inundation in the Mid-Atlantic region was observed. Exposure status remained positively associated with the presence and number of emergency preparedness deficiencies after adjustment for facility characteristics, with the converse for facilities within the Western Gulf Coast.

**Meaning**  These findings suggest opportunities to reduce regional heterogeneity and improve the alignment of nursing home emergency preparedness with surrounding environmental risks.
Challenges Along the Way

• Access
• Management
• Processing
• Methodology
Approaches

• Interpreting and assessing data retrieval methods, documentation and quality

• Tracking data sourcing and citing data appropriately
Approaches

• Making data dictionaries and missingness tables
• Shifting from tools like Stata to Python
• Converting hard-to-document Excel processes to replicable code
• Documenting work in multiple modalities with Jupyter Notebooks
Approaches

• Reviewing data literature and methodology
• Implementing data processing methodologies, such as:
  • tidy data\(^4\)
  • split-apply-combine\(^5,6\)

Image via Analytics Vidhya
What we’ve learned from each other

Clinician perspective:
• The utility of publicly available data to answer questions with clinical and health services relevance
• Public data can present many challenges to cleaning and merging across multiple data sources
  • Unstructured data
  • Data with inconsistent identifiers across years and data sources (healthcare facility identification numbers, for example)
  • Merging data with inconsistent or error-laden identifiers presents many challenges
What we’ve learned from each other

Clinician perspective:
• Many data sources can be more amenable to cleaning and organizing in languages other than those in which clinician researchers are trained
• Accurately logging procedures used in data pre-processing is essential to replicability
• Data Librarians have rich expertise in addressing each of the challenges above
What we’ve learned from each other

Librarian perspective:
• Growing clinician ambition to take on complex, multi-disciplinary data projects
• More support needed, especially in early clinician-researcher career, for data training
• Opportunity to:
  • Fully participate in research data lifecycle
  • See the power of open data in action
How to succeed as a clinician-librarian team

• Communicate early and often
• Grow project portfolio over time
• Teach and trade skills
• Seek expertise elsewhere when needed
Advice to clinicians on involving librarians

• Learn what data support is offered by your library
• Learn and practice principles that your data librarian encourages as foundational to data science integrity and reproducibility
• Establish your team and expectations regarding roles and responsibilities during early stages of the project
• Ensure that there are opportunities for learning and professional development for all team members
Summing Up

DISCOVER  data support services
USE       open data
TAKE      a Team Science approach
Resources

• Data management planning guidance for NIH grants and beyond:
  • Ten simple rules for maximizing the recommendations of the NIH data management and sharing plan | PLOS Computational Biology
  • The FAIR Guiding Principles for scientific data management and stewardship | Scientific Data
  • Ten simple rules for the care and feeding of scientific data | PLOS Computational Biology

• Excellent data organization and cleaning principles:
  • Data organization in spreadsheets | The American Statistician

• For those interested in Python:
  • Python for Non-Programmers | python.org
References


3. LTCFocus Public Use Data sponsored by the National Institute on Aging (P01 AG027296) through a cooperative agreement with the Brown University School of Public Health. Available at www.ltcfocus.org. https://doi.org/10.26300/h9a2-2c26

