Building Data Science Skills

Enhancing Core Capabilities and Understanding Unique Value

The views expressed here are those of the author and do not necessarily reflect the official views of the Federal Reserve Bank of Kansas City or the Federal Reserve System

Courtney R. Butler
Senior Data Librarian

My Background

- MLS with specialization in Archives and Records Management
- Work history: government archives, public library special collections, academic library reference desk, hospital clinical librarian
- Current position:
 - Support both research and operational staff
 - Responsibilities have included data acquisition, research reproducibility, data curation, data preservation, and data management

Expectations

- Increasing overlap between data science and data librarianship
- Highly valuable
 - Expert coder
 - Advanced degree in related subject area
- Actual requirements should depend on the role and needs of the organization
 - Be realistic
 - Place value on librarian skillset

Consider...

- You do not have to “do” the thing to support the thing
- Librarians are lifelong learners
- Data science is not a whole different animal
- Each organization’s patrons have unique needs

Illustrations

- Speak the Language
- Understand Needs
- Provide Resources
- Foster Connections
- Advocate for Good Practices

Speak the Language

- Essential concepts and terminology
 - Basic coding and troubleshooting
 - Common coding languages and packages
 - Common file formats
 - Common methodologies

Learn the Language

- Papers
- Google
- Trainings and conferences
- Books and standards
- Researchers!

Provide Resources

- Trainings
 - Internal curriculum
 - MOOCs
 - Vendor visits
- Finding data
 - Subject guides
 - Data catalogs
 - Reference interviews

Understand Needs

- Data Licensing

- Usage

- Research vs operational
 - Methods
 - Machine learning
 - Web scraping
 - NLP
 - Publication/sharing

- Storage and Access

- Viewing vs downloading
 - HPC
 - API vs large scale data feeds

Foster Connections

- Support communication across stakeholders:
 - IT
 - Legal
 - Procurement
 - Other researchers
 - Information security
 - Records/Archives
 - Audit
 - Data providers

Advocate for Good Practices

- Reproducibility and Data Management
 - Documentation
 - Metadata
 - Code commenting
 - Open source formats
 - Versioning
 - Storage and backups
 - File naming and organization
 - Containerization
 - Preservation

Takeaways

- Maintain a growth mindset
- Emphasize strengths and unique value
- Build partnerships
- Connect with library community

Thank you!

courtney.butler@kc.frb.org