Federal Reserve Bank of Kansas City

Starting with "the end" in mind:

Building a map of data services at the Federal Reserve Bank of Kansas City

Christina M. Kulp, MLIS

Courtney R. Butler, MLS

The views expressed in this presentation are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

CADRE Center for the Advancement of Data and Research in Economics

Federal Reserve Bank of Kansas City

Why the Project?

- Data Management is a hot topic, but….

  What does it really mean?

- Improve Library services

  Research is part of the Bank's Mission!

- Lots of ideas / initiatives…something is still missing!

  Our Users!

Why the Library?

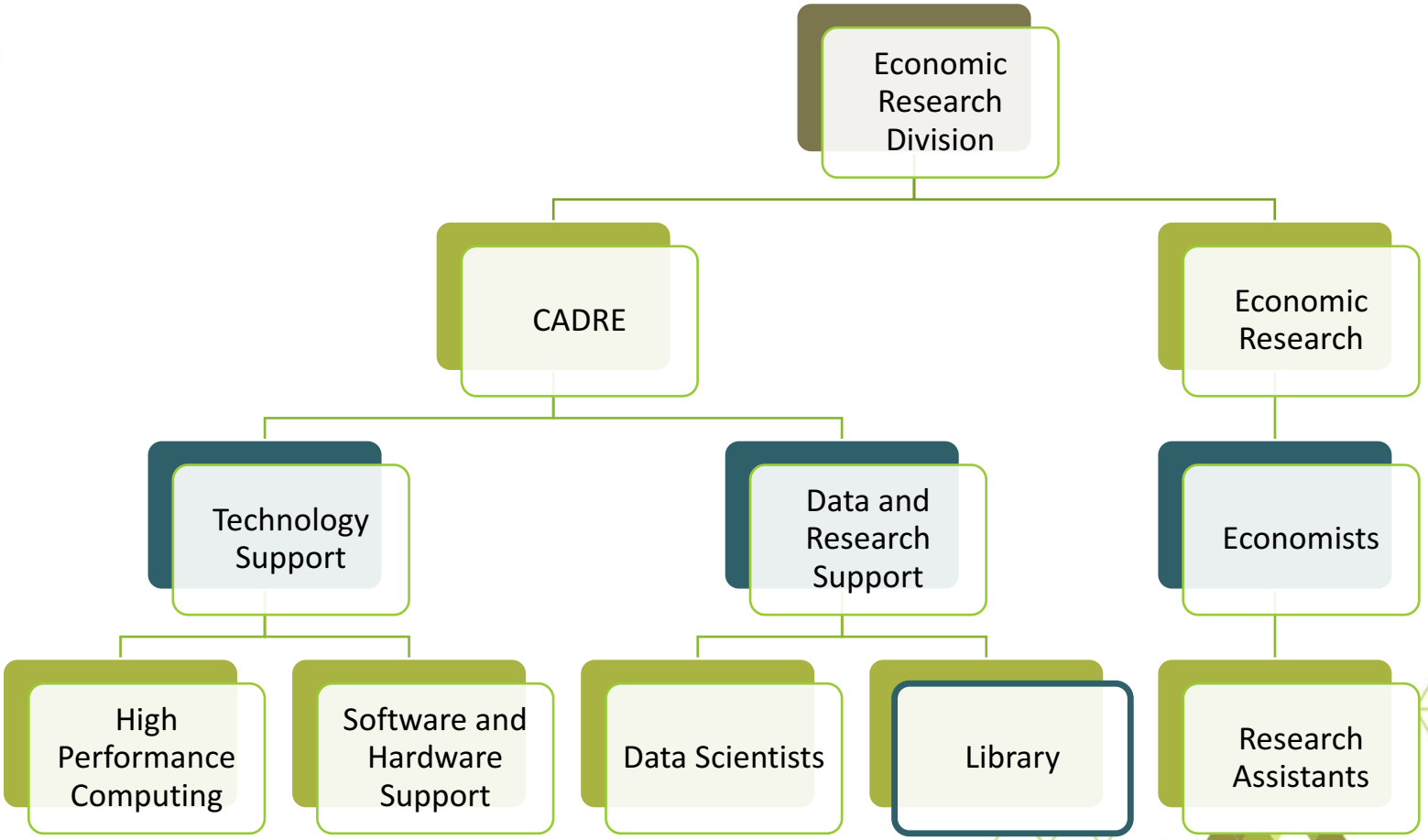
We book end the process….

- Data Services Librarian
  - Help find, trial, and purchase data at the start
- Data Curation Librarian
  - Preservation, metadata, and archive at the end

Build connections in between…

- Our duties and skills positions us between the data user and technical support
  - Cohesive narrative
- Get involved early to support better outcomes
  - Encourage best practices

CADRE
Center for the Advancement of Data and Research in Economics



Inspiration

Build a Village

"The Data Management Village: Collaboration among Research Support Providers in the Large Academic Environment"

- By Alicia H. Mohr, Lisa R. Johnson, and Thomas A. Lindsay
- In *Databrarianship: The Academic Data Librarian in Theory and Practice, 2016.*

Using a Life Cycle

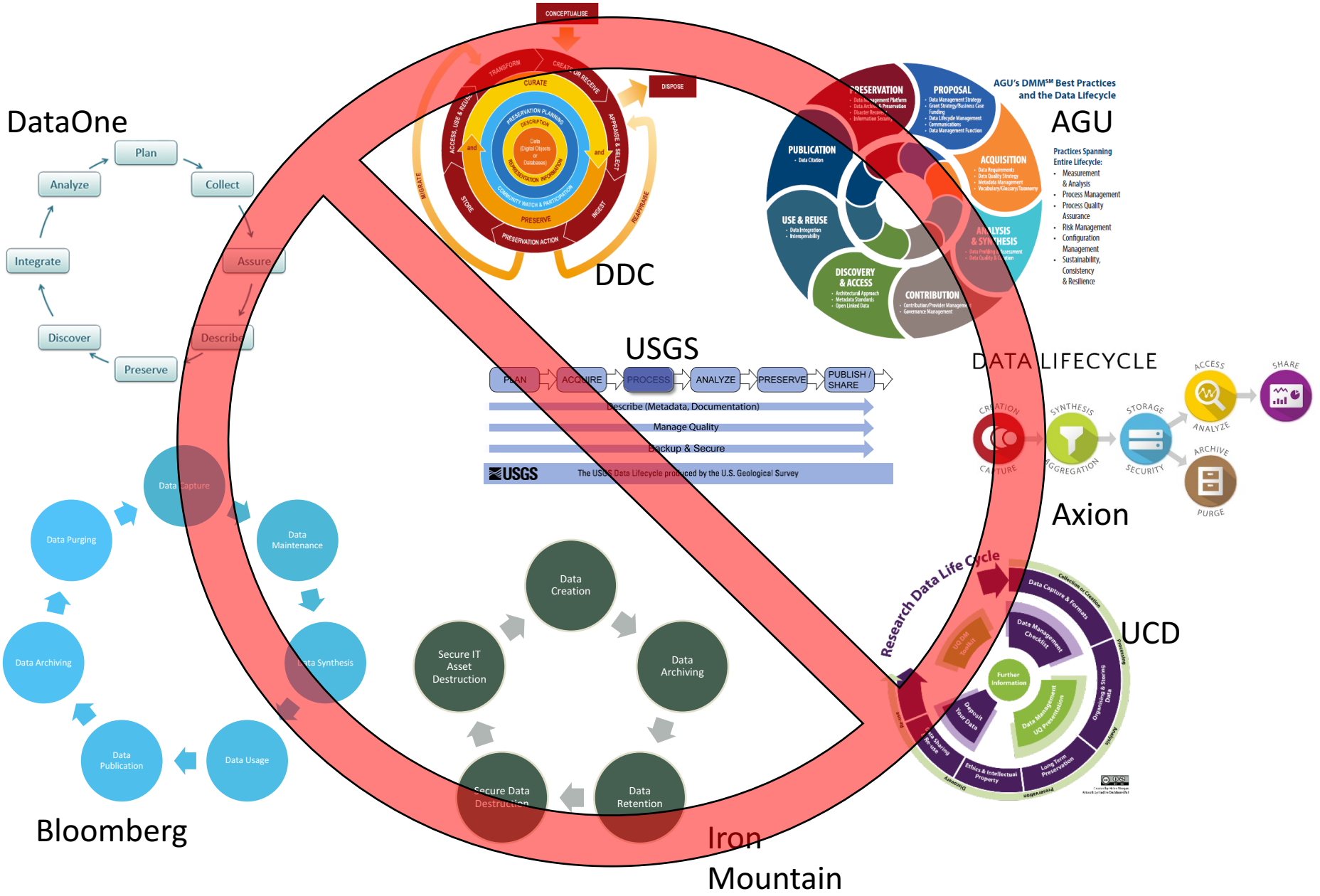
"The use of Life Cycle Models in Developing and Supporting Data Services"

- By Jake Carlson
- In *Research Data Management: Practical Strategies for Information Professionals, 2014.*

Mapping Process

- Find the right data life cycle
- Text analysis for data management concepts
- Build our own life cycle
- Identify our stakeholders
- Cross reference concepts (tasks) and stakeholders
- Test our assumptions and fill in blind spots with interview

Lesson One: Build your own life cycle.



Text Analysis

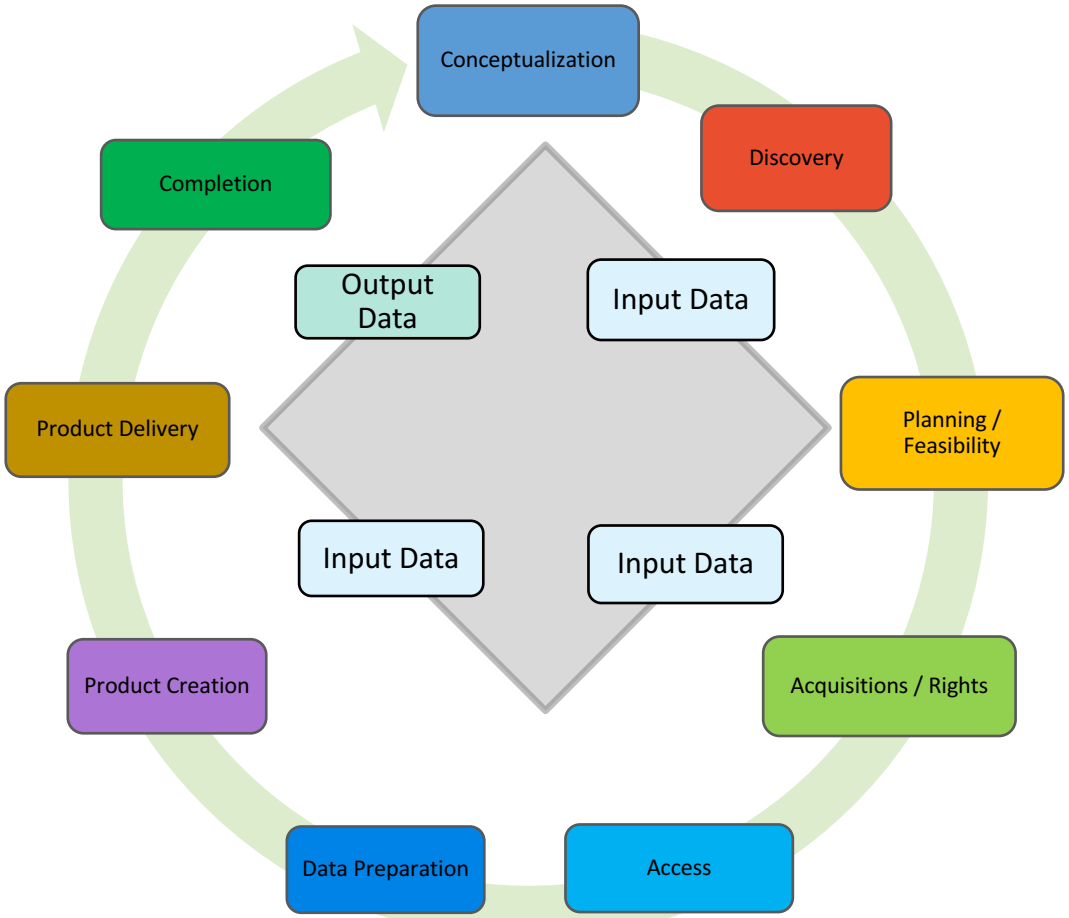
Conceptualization	Discovery	Planning	Acquisition (Rights)	Access
Proposal	Data identification	Data types	Acquisition - licensing, contracting	Data storage
Grantwriting	Trial/demo	Platforms	Classification - security	Backup/recovery
Ethics	Collection	Hardware	Classification - public v proprietary	Governance
Budget	Creation	Software	Intellectural property, copyright	Security
	Appraisal	Roles/Responsibilities	Authorized users	Receive/Ingest
	Select	Operational needs	Post-termination rights	Protect
	Evaluate	Data size		Authentication
		Duration		Assessment
		Metadata standards		Metrics
		Media		Usage
		Sharing/Repository		Compliance
		Security		Description
		Risk management		Metadata
		DMP		Data capture/incoming formats
		Recommendations (reproducibility)		

- Literature Review
- Collected key words
- Grouped into categories
- Our life cycle stages

Data Preparation	Creation	Delivery (Product)	Completion
Cleanse	Analysis	Deploy	**A (Preservation)**
Data Model	Aggregation	Work product	Preservation
Process	Synthesis	Share	Archive
Organize	Reproducible file creation	Publish	Security
Relate	Coding	Intellectual property	Integrity
Integrate	Evaluation of data	Citation	Ingest
Enrich	Data Model	Copyright	Migrate
Resolve	Evaluation of methods (internal, peer, vendor review)	Present	Curate
Optimize	Interpretation	Data release (public, upon request)	Deposit
Documentation	Stat modeling	DOI/unique ID	Metadata
Provenance	Hypothesis	Embargo	
Anonymize	Quality (missing value, outliers, estimates)		**B (Deletion)**
Normalize	Assurance		Dispose
Transform	Documentation		Purge
	Version control		
	Visualization		
	Data dictionary		
	Codebook		
	ReadMe		

Custom Life cycle

1. **Conceptualization** – Project Lead receives permission to pursue project and identifies basic data wants or needs.
2. **Discovery** – Project Lead matches potential data products to those wants / needs.
3. **Planning / Feasibility** – Coordination between stakeholders to determine if resources can support data needs of the project.
4. **Acquisitions / Rights** – System / Bank obtains legal right to utilize data for the project, e.g., signed contract.
5. **Access** – Process of downloading, organizing, storing, and governing the data for the project team to see and use.
6. **Data Preparation** – Process of cleaning, migrating, consolidating, and/or aggregating data in order to begin product creation.
7. **Product Creation** – Process of using the data to create the end product for the project, e.g., a publication or data set to share.
8. **Product Delivery** – Process of making the end product(s) available to its intended audience(s).
9. **Completion** – The project team hands off data products to support staff for operational tasks, e.g., preservation or purging.



Note: Not always a chronological cycle. Steps may be skipped, repeated, or no longer relevant for established work flows.

Lesson Two: Whose life cycle is this?

Stakeholders

Conceptualization	Discovery	Planning	Acquisition (Rights)	Access
Researcher/Analyst	Researcher/Analyst	Researcher/Analyst	Researcher/Analyst	Researcher/Analyst
External/Guest Researcher	External/Guest Researcher	External/Guest Researcher	Management	External/Guest Researcher
RA's	RA's	Research Assistant/Data Analyst	Administration	Research Assistant/Data Analyst
Management	Management	Management	Library	Management
Library	Library	Library	Procurement	Tech Support
Data Scientists	Legal	Tech Support	Legal	Library
	Procurement	Legal	Tech Support	Audit
	Tech Support	Data Scientist	Info Security	Info Security
			Audit	Data Manager

Data Preparation	Creation	Delivery (Product)	Completion
Researcher/Analyst	Researcher/Analyst	Researcher/Analyst	Researcher/Analyst
External/Guest Researcher	External/Guest Researcher	External/Guest Researcher	External/Guest Researcher
Research Assistant/Data Analyst	Research Assistant/Data Analyst	Research Assistant/Data Analyst	Research Assistant/Data Analyst
Tech Support	Management	Management	Archivist
Data Scientist	Tech Support	Library	Tech Support
Library	Data Scientist	Editor	Audit
	Editor	Admin Support	Library
	Library	Public Affairs	Data Manager
	Administration		Admin Support
	Legal		
	Data Manager		
	Admin Support		

- Who does Data Management?
- What stages are they involved in?

Cross-Referencing

Conceptualization	Discovery	Planning	Acquisition (Rights)	Access
Proposal	Data identification	Data types	Acquisition - licensing, contracting	Data storage
Grantwriting	Trial/demo	Platforms	Classification - security	Backup/recovery
Ethics	Collectio	Hardware	Classification - public v proprietary	Governance
Budget	Creation	Software	Intellectural property, copyright	Security
	Appraisa	Roles/Responsibilities	Authorized users	Receive/Ingest
	Select	Operational needs	Post-termination rights	Protect
	Evaluate	Data size		Authentication
		Duration		Assessment
		Metadata standards		Metrics
		Media		Usage
		Sharing/Repository		Compliance
		Security		Description
		Risk management		Metadata
		DMP		Data capture/incoming formats
		Recommendations (reproducibility)		

2

3

Conceptualization	Discovery	Planning	Acquisition (Rights)	Access
Researcher/Analyst	Researcher/Analyst	Researcher/Analyst	Researcher/Analyst	Researcher/Analyst
External/Guest Researcher	External/Guest Researcher	External/Guest Researcher	Management	External/Guest Researcher
RA's	RA's	Research Assistant/Data Analyst	Administration	Research Assistant/Data Analyst
Management	Management	Management	Library	Management
Library	Library	Library	Procurement	Tech Support
Data Scientists	Legal	Tech Support	Legal	Library
	Procurement	Legal	Tech Support	Audit
	Tech Support	Data Scientist	Info Security	Info Security
			Audit	Data Manager

1

Library
Do I understand what the research team is looking for?
What options already exist? (We have something or I know of something, including System)
Help with search options (terms, methods, locations, strategy, lit review)
Does the trial have a cost?
Does the trial have terms or conditions? (EULA, etc.)
Manage vendor relations
Is the trial available on-site or virtual?
Could this data be used beyond this project? Who should be involved?
How long does it take to get access to the trial? How long can we have access?
What is the approval process? (Procurement, Management, Legal, InfoSec, Research Automation)
Who do I need to work with to initiate the trial? (Vendor, Research Automation)
Should or can I find a subject matter expert to help?
Have we worked with this vendor before? Have we trialed or demo'd the data before?
Solicit feedback from end users
Do I need to get a quote on cost for the product?
How can I help with the evaluation process?
Who else offers this data? (besides the primary vendor)
Identify any potential barriers, restrictions, issues (e.g., requires FOIA, international vendor)
Prep for "planning" step
What criteria are the researchers using for evaluation?
Are there any security or privacy concerns for the trial data?
Do we have the resources we need to execute the trial?
Document for audit, compliance
Does the vendor understand our needs or mission?
Document all solutions investigated for use during acquisition

1. Stakeholder
2. Relevant Task
3. Concerns

Lesson Three: A life cycle is only part of the picture.

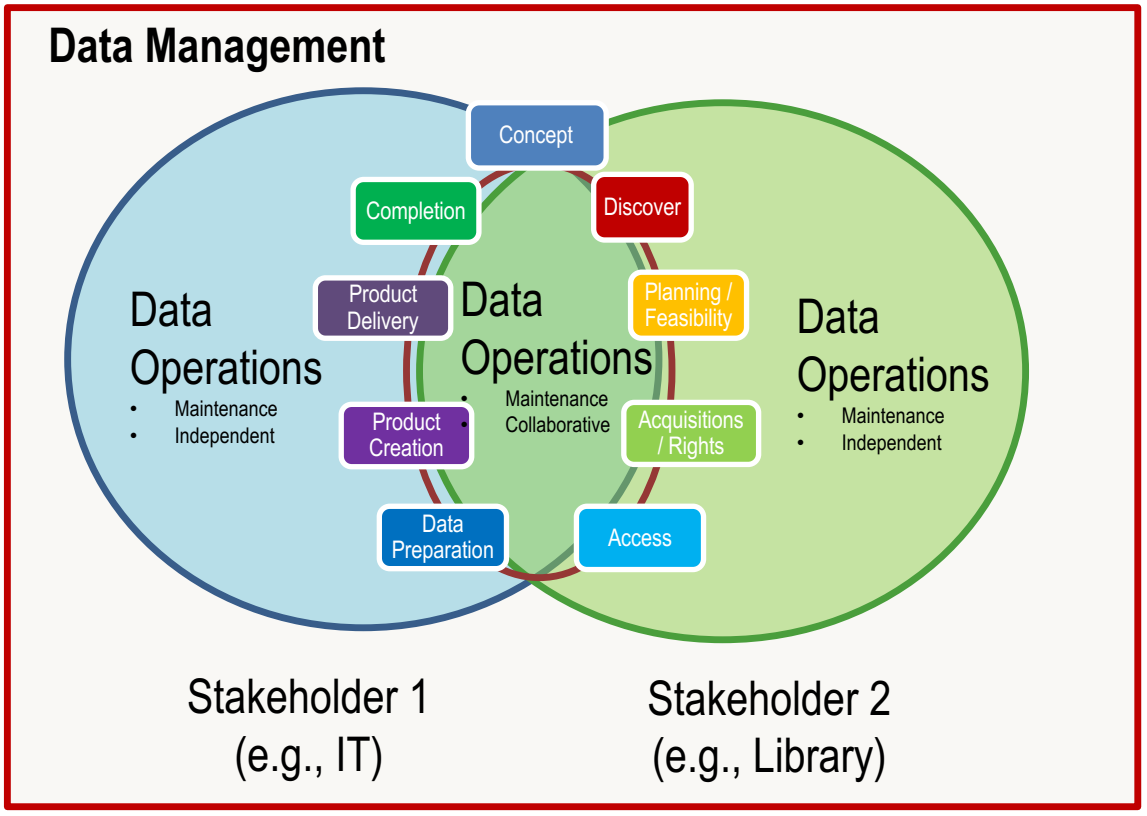
But what about…?

Recognized many tasks

- Belonged in many or all categories
  - Contract management
  - Data storage / file management
- Belonged in no categories
  - Building a cluster or VM environment
  - Maintaining a library catalog of datasets

Led to creation of “Operations” category

The Bigger Picture….



A life cycle implies a timeline with distinct stages – a project with a beginning and end

Data user life cycles focus on project-specific activities or tasks

Data operations involve both independent and collaborative tasks that occur behind the scenes

Data management encompasses both project and operational tasks

Good data management coordinates all tasks. Waiting until a project-oriented need arises is often too late for an effective response.

Lesson Four: Our map needs to connect all these pieces.

Interviews

Purpose of interviews:

- Test our assumptions
- Identify our blind spots
- Build awareness with stakeholders

Know your audience:

- Let them define data management
- We tailored our questions and slides
  - Data users – focus on life cycle
  - Technical support – try to get at operational tasks

Things We've Learned

Insights

- Each stakeholder does view data management differently.
- Our own definition is changing – are librarians more data curators?

Actions

- Our Economists want to know more about discipline best practices.
- Legal needs more context for contract reviews.

Next Steps

- Keep interviewing stakeholders.
- Share what we have learned.
- Strengthen the Library's role as:
  - Collaborators
  - Educators
  - Support at both ends of the pipeline
  - Build connections in between

Conclusion

- Build your own tools – like a life cycle.
- Talk to as many types of stakeholders as possible.
- Keep an open mind to identify blind spots.

Questions?