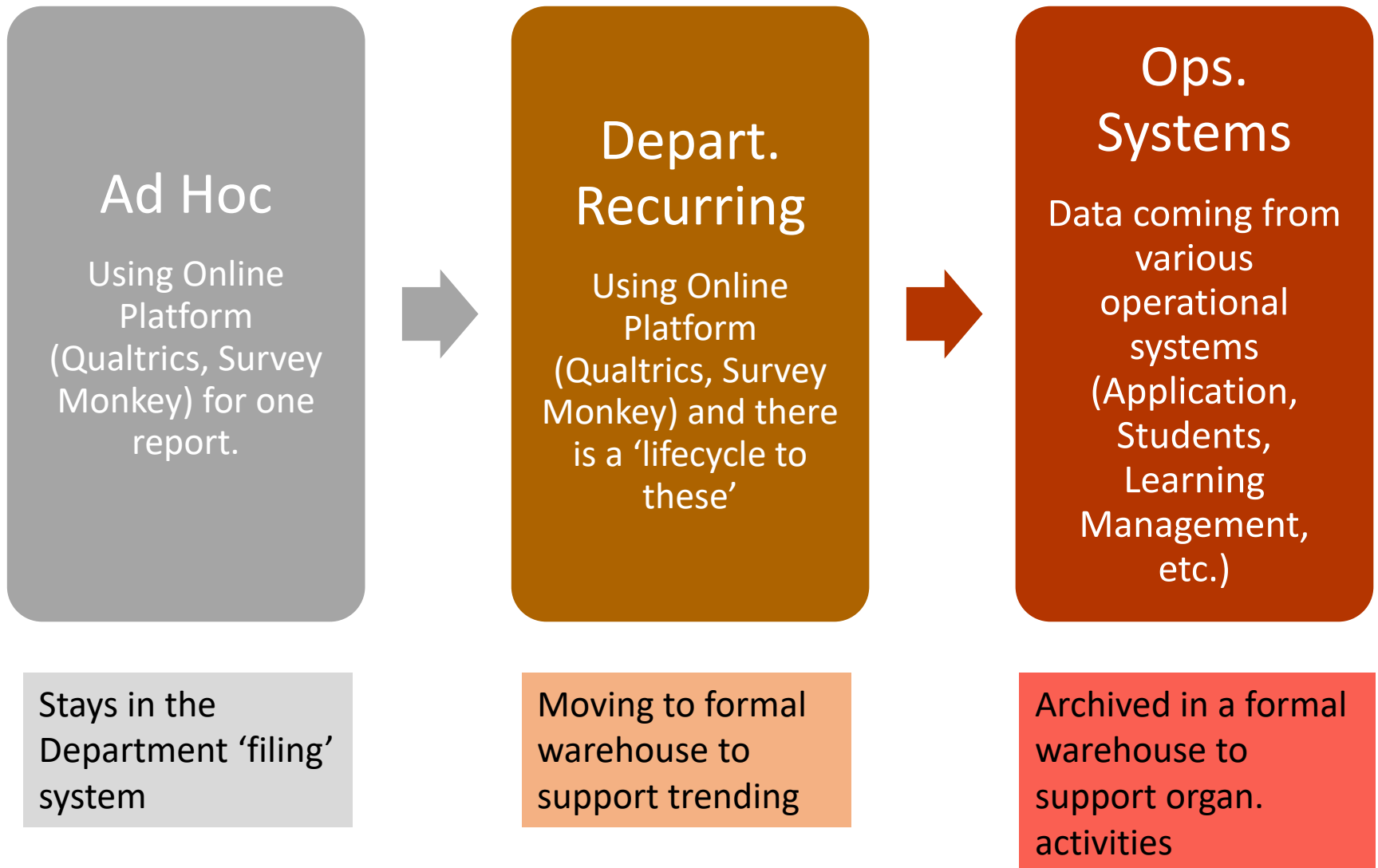


Session Outcomes

- Describe the basic components of data management and it can improve IR professional work
- Explain the value of business side documentation and its differences from IT documentation
- Develop an action plan to take back to their institution

Data Collection Activities in Modern Data Unit



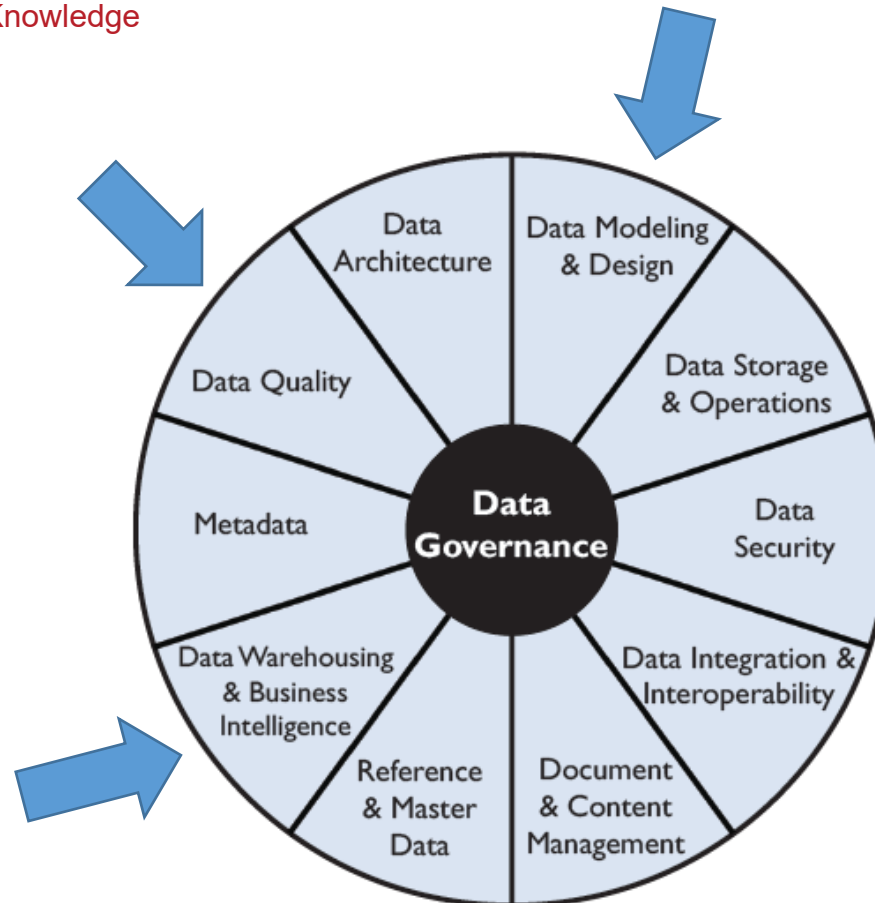
DAMA DMBOK

The Data Management Association's Guide to the Data Management Body of Knowledge

Data Management

is the development, execution, and supervision of plans, policies, programs and practices that deliver, control, protect, and enhance the value of data and information assets through their lifecycles (p. 17)

Data Governance



DAMA-DMBOK2 Data Management Framework

Copyright © 2017 by DAMA International

From Research Assistant to IR Data Steward

Research Assistant

- Dedicated project where the vision, direction, and management is supplied by the PI
 - Literature Review
 - Cleaning a data set
 - Developing & maintaining a code book
 - Using syntax with comments to track analysis decisions and results

Data Steward

- Describes and demonstrates responsibility for data and processes to ensure effective control, increase value, and use of data assets*
 - Data Quality including Metadata
 - From Source to Asset
 - Rules & Standards for Processing
 - SME and 'Data Trustees'

From Codebook to Documentation

Demonstrating Business Value

Code Book

- Extraction of the Variable description from a Statistical Processing Software
 - Variables: Names & Labels
 - Parameters: Type and Size
 - Question or Description
 - Data values

Data Steward Documentation

- Describes the context, process and value of the asset
 - Source to Repository
 - Quality
 - Lifecycle
 - Processing

Data Asset Documentation: IT & Business

Both are Needed but for different reasons

IT Side

- Emphasis is on the technical processes the data asset
 - Transactional System → Staging → Repository Lifecycle
 - Size, security classification
 - Change Log
 - Data values
 - Permissions for access

Business Side

- Emphasis is on Value & Use
 - Source → Repository Lifecycle
 - Quality using Statistical tests and identifying high value variables (100% fidelity?)
 - “Why” Changes to variables
 - Processing for Data Marts, Reports (Static & Dynamic)
 - Permissions for use

7 Key Components

Fundamentals

Documentation is about

- > 'who'
- > did 'what'
- > 'when' with a lot of 'how'
- > and a bit of 'where' and 'why'

Tools: Word or Excel

Data Documentation Form

Change Log

Update Date:	
Data Steward (Author):	
Short Description of Last Change	
'Impact' changes	

Data Asset Description & Lifecycle

Data Asset Name (formal and informal), Acronym	
Annual Data Modification Window:	
Technical Data Steward or DBA:	
Source Feed	
Repository Location	
Typical Processing Time from Source to Repository	

1. Summary Change Log

Update Date:	October 29, 2020
Data Steward (Author):	G. Garrison
Short Description of Last Change	We moved from APP system* the 2019-2020 application information. See the APP system document for operational opening/closing dates and processing routines. Everything ran smoothly after code improvements.
'Impact' changes	<p>APP system added a new admissions decision option, "Decline after Accept" (labeled '7') per policy about background checks. See Dean's Committee note for full description.</p> <p>VP of Ed Operations requested summary descriptive analysis to check on policy.</p>

2. Data Asset Description & Lifecycle

Data Asset Name (formal and informal), Acronym and description	Tables in DW1 are called APPSYM. There are 10 each year that are updated, added annual applicant records, admissions decisions, demographics. Reference table ID for department usual joins.
Annual Data Modification Window:	APPSYM officially closes on the last day a school enrolls students. Processing checks begin 4 weeks before to test and resolve any possible data migration issues in DWDEV.
Technical Data Steward or DBA:	D. Lei
Source Feed	APP system = Application System, local propriety software built by our IT software department.
Repository Location	DWDEV is used for staging for cleaning and generating calculations. DW1 final location
Variable creation done by calculations	Contact information (Address & Zip) converted to GIS vectors. Summary columns created for applications and admissions decisions.

3. Data Asset Value striving for 'single source of the truth'

<p>Privacy and Compliance Classification and Flags</p>	<p>Confidential (contains both personal and institutional level)</p> <ul style="list-style-type: none"> • Matriculation indicator is check with STUREC. STUREC is under FERPA restrictions. • 7% of the records are international and about ½ are under GDPR restrictions. Contact data is stored in the Data Hub under retention policy flags.
<p>Major Static Reports</p>	<p>Summary descriptive tables feed annual press release and website front page infographic; aggregated tables to 'Facts' website; IPEDS Admissions Survey</p>
<p>Major Dynamic Feeds (Dashboards)</p>	<p>Aggregated tables feed the Power BI Data Mart that then is embedded in 'Facts' website</p>
<p>Data Mart Description and Location</p>	<p>Using Power BI Cloud.</p>
<p>Core Variables</p>	<p>Apps, Applications, Reapps, Admissions Decisions, Demographics</p>
<p>Core Quality Assessment</p>	<p>Applicants, Applications, Reapplicants, = 100%, Admission Action: 5% missing; Demographics: R&E 17% missing, S/G 5% Missing</p>

4. Data Asset Details & Access

Repository Structure	RDBMS or Relational DB
Data Hub Exchanges	Applicant contact information, year of application
Extraction Tools (SQL, PSQL)	SQL for basic querying; R for stats work
MetaData Description	Load date, record id with embedded year
Reference Data Description	Department codes join to department names
Other organizational access	Marketing Depart. for profiles of future applicants

5. Business Glossary Terms

Example: In 'Admis_Decision' added another value (7) for "Decline after Accept" a departmental action after Background check review.

6. Data Quality Report

Itemized report of the of the key variables of value and assessed for 'Completeness' (no missing), 'Uniqueness' (no duplicates), and 'Validity' (no unexpected values)

7. Processing Routines & Code

To guide analysis and reports generation both internal and external

- Name of the Process/Routine:
 - Description:
 - How to Calculate (including caveats)
 - Code:

Developing your Action Plan

Rationale:

- Remember you are documenting institutional memory
- Documenting your data value
- Expanding your data culture

How to get started

- Build this into your routines as you work through the lifecycles of your data project.
- Value the documentation time and effort. Initiation is hard, maintenance is easier.
- Adjust the Log description
- Adapt to fit your data culture



Please use the WHOVA Mobile App to rate this session. CAIR uses the session evaluations to determine the winners of the Best New Presenter, and Best Presenter awards each year.

SCHOOL OF
EDUCATIONAL STUDIES
Claremont Graduate University

Questions and Comments

Thank You!

#datastrong
#CAIR2020
#IR

Gwen Garrison, PhD
gwen.garrison@cgu.edu