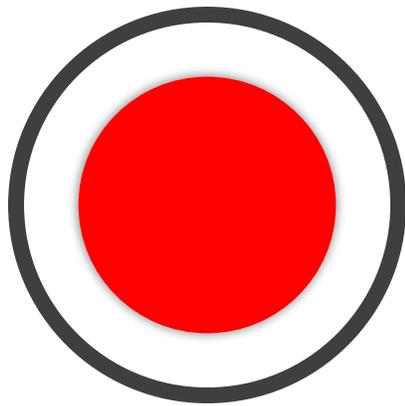# ASK ME (ALMOST) ANYTHING!
DATA MANAGEMENT AND QUALITY FOR NON-DATA-MANAGEMENT PROFESSIONALS

Michiko I. Wolcott, Principal Consultant

Msight Analytics

October 20, 2021

AGENDA

Introduction

Data Management: An Overview

What is "data quality"?

The "System" for Data Management

Open Q&A: Please use the chat panel to submit your questions

DATA MANAGEMENT PAIN POINTS

PAIN #1
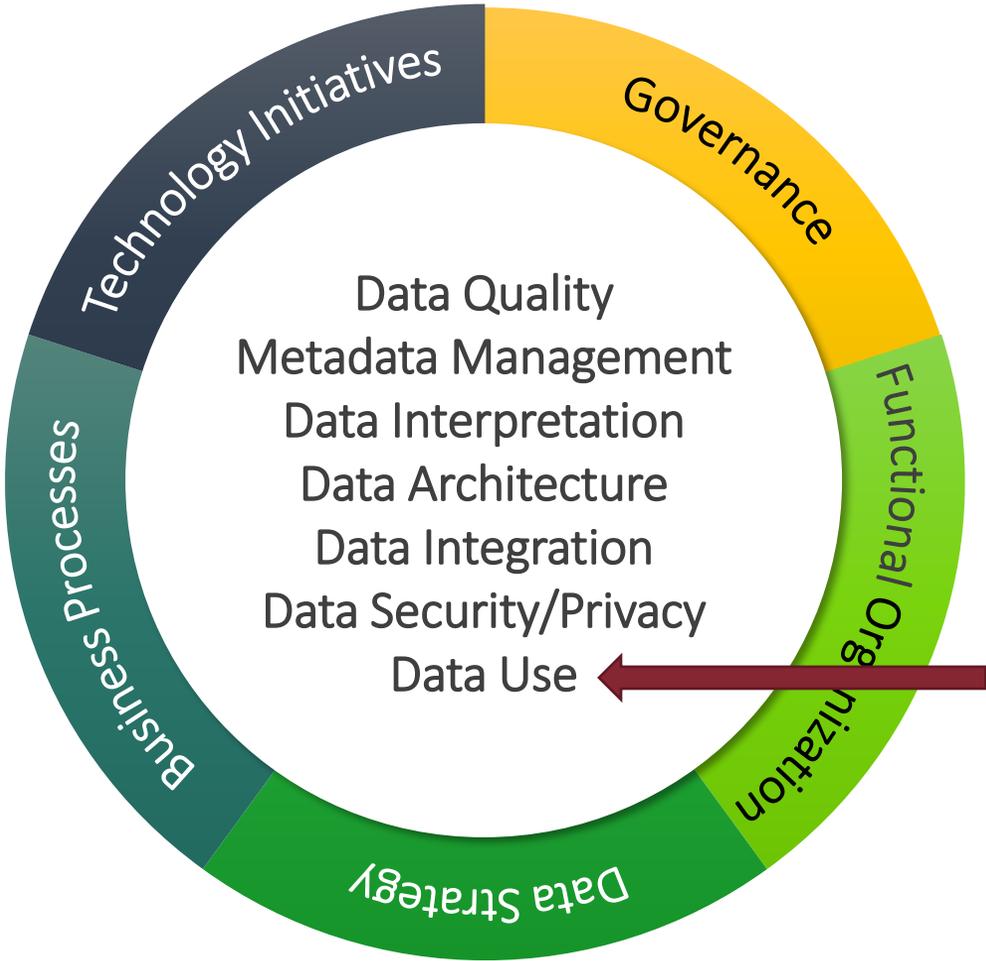
Lack of proper documentation about data

PAIN #2

Errors in the data discovered by users

PAIN #3

Data contents do not make sense to users.

THE WORLD OF DATA MANAGEMENT

Data Management: "The development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles."

(DAMA-DMBOK 2nd ed.)

STAGES OF THE INFORMATION LIFECYCLE

Data → GENERATE/ ACQUIRE → PREPARE → USE/ MAINTAIN → SUNSET

Analytics: Design → Develop/ Deploy → Use/ Maintain → Sunset

Insights: Develop → Use Maintain → Sunset

Msight
Analytics · Consulting

WHAT DO WE MEAN BY "QUALITY"?

"In technical usage, 'quality' can have two meanings:
1. the characteristics of a product or service that bear on its ability to **satisfy stated or implied needs**;
2. a product or service **free of deficiencies**."

Source: American Society for Quality, https://asq.org/quality-resources/quality-glossary/q

⬇

Compliance with some set of criteria, standards, or requirements

⬇

Noncompliance of standards or requirements = Deficiency, Defect

⬇

Perfect quality = Complete absence of defects

Data Cleansing ≠ Data Quality

Characteristic, state

Activity, task

DIMENSIONS OF DATA QUALITY

| COMPLETENESS | • How populated/complete is it? |
VALIDITY	• Are the values technically valid?
ACCURACY	• Is it free of errors?
CONSISTENCY	• Is it free of discrepancies?
INTEGRITY	• Is it coherent?
TIMELINESS	• Does it reflect the timing of interest (current and historical)?
UNIQUENESS	• Is it free of duplications and fragmentations?
REASONABILITY	• Does it reasonably reflect the intent and the reality?

EXTRACT

TRANSFORM

LOAD

ANALYZE!!
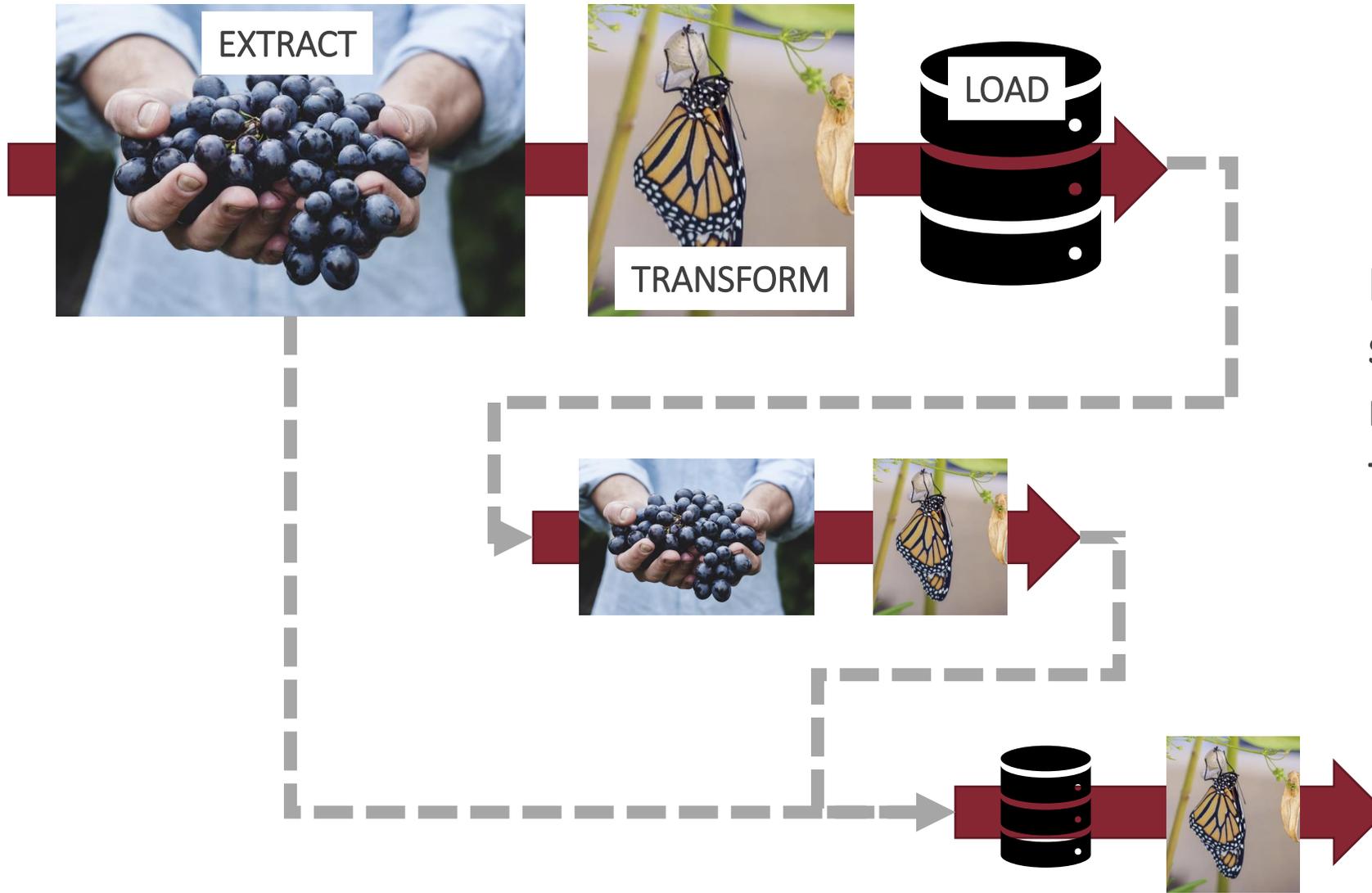
Each one of these steps presents multiple opportunities for data defects.

Msight
Analytics · Consulting

COMPLETENESS

VALIDITY

ACCURACY

CONSISTENCY

INTEGRITY

TIMELINESS

UNIQUENESS

REASONABILITY

Technology-led data quality: custody-centric (logic over content)
- Can address each dimension only partially.
- Reasonability is rarely addressed.

EXTRACT

TRANSFORM

LOAD

ANALYZE!!
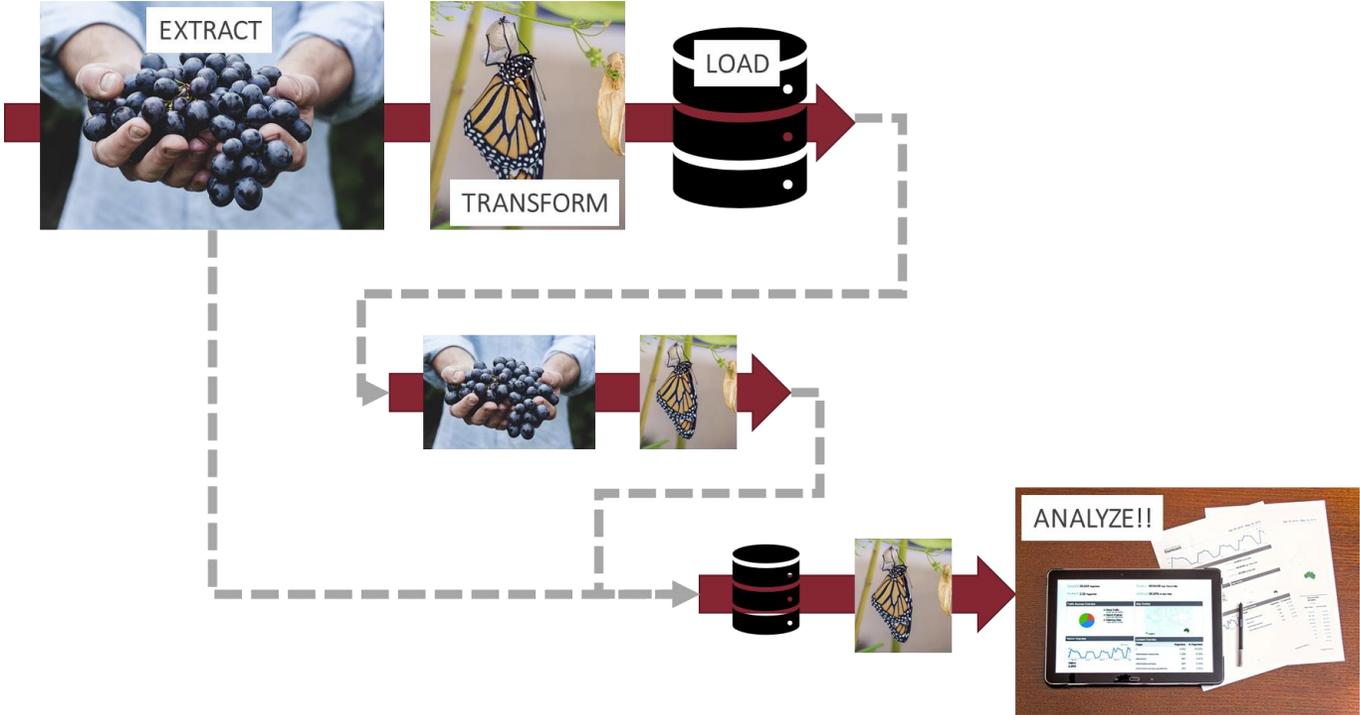
User-led (including analytics-led) data quality: reactive and incomplete
- Can address quality only when it gets to them.
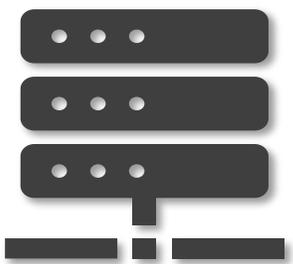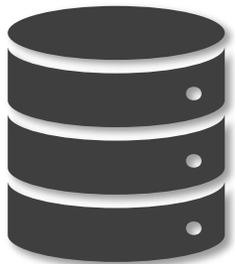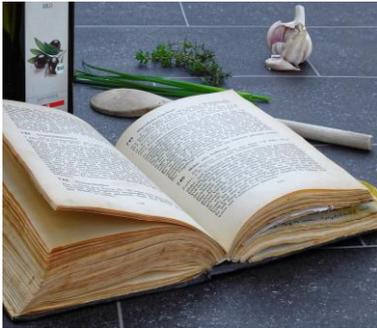- Data quality is a means to an end, not the end itself.

Msight
Analytics • Consulting

EACH FUNCTION BRINGS DIFFERENT VALUES TO THE CONSUMER.

FRAMEWORK FOR FUNCTIONAL STRUCTURE

Enterprise Leadership			Enterprise alignment
Data Owner	Analytics Owner	Technology Owner	Enterprise owners
Data Advisory	Analytics Advisory	Technology Advisory	Work group(s)
Data Functions	Analytics Functions	Technology Functions	Day-to-day execution
Business/Research Interest			Identification of business/research needs

Within Data Advisory, Analytics Advisory, and Technology Advisory: "Expertise" across top, "Business Areas" along side.

Left arrow (upward): Identify and escalate

Right arrow (downward): Decide and communicate

Msight
Analytics · Consulting

WHAT SHOULD YOU DO?

- Advocate for establishing a proper data function.

- Continue to be diligent about the data you do see.

  - Technology: Learn as much as you can about the data content and how that relates to reality from the users' perspective.

  - Analytics: Learn about data quality methodologies to help identify problems, albeit reactively.

    > At the minimum: Perform a data audit for every project data you receive, as soon as you receive it.

More at https://msightanalytics.com
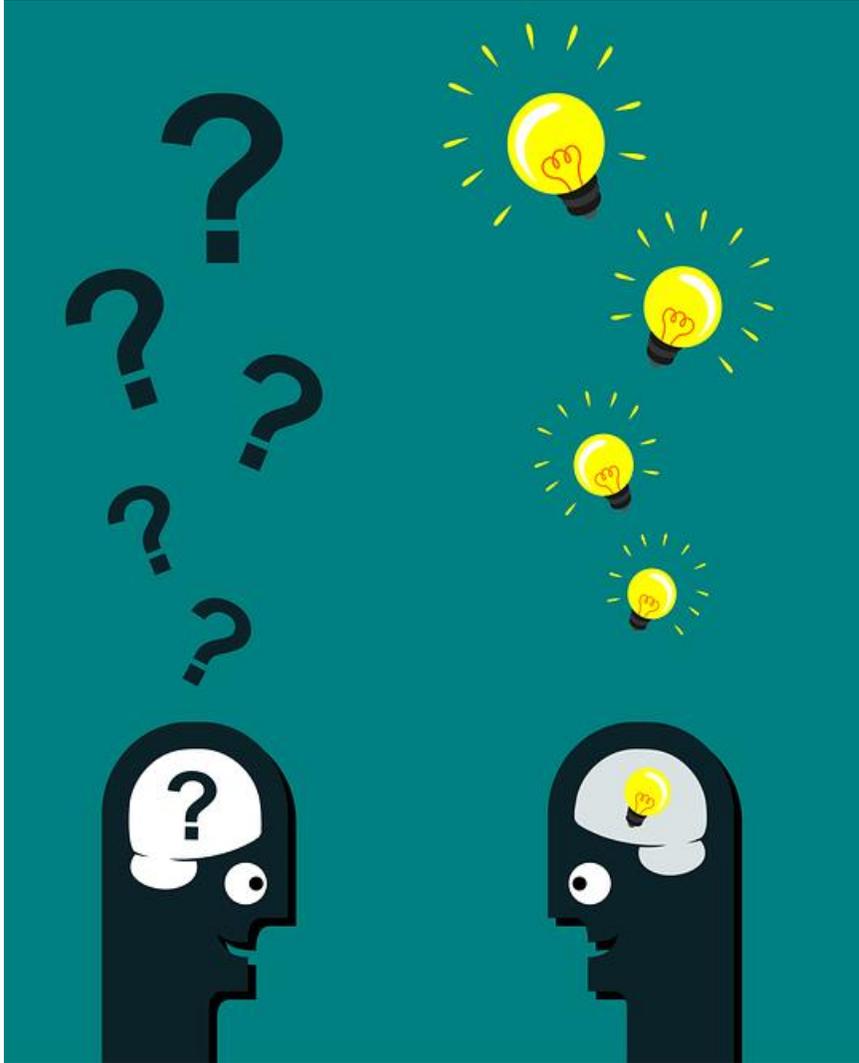
Michiko I. Wolcott, Principal Consultant
Msight Analytics
info@msightanalytics.com
🐦 @Michiko_Wolcott
in linkedin.com/company/msight-analytics
f @msightanalytics