The California State University
WORKING FOR CALIFORNIA

California State University

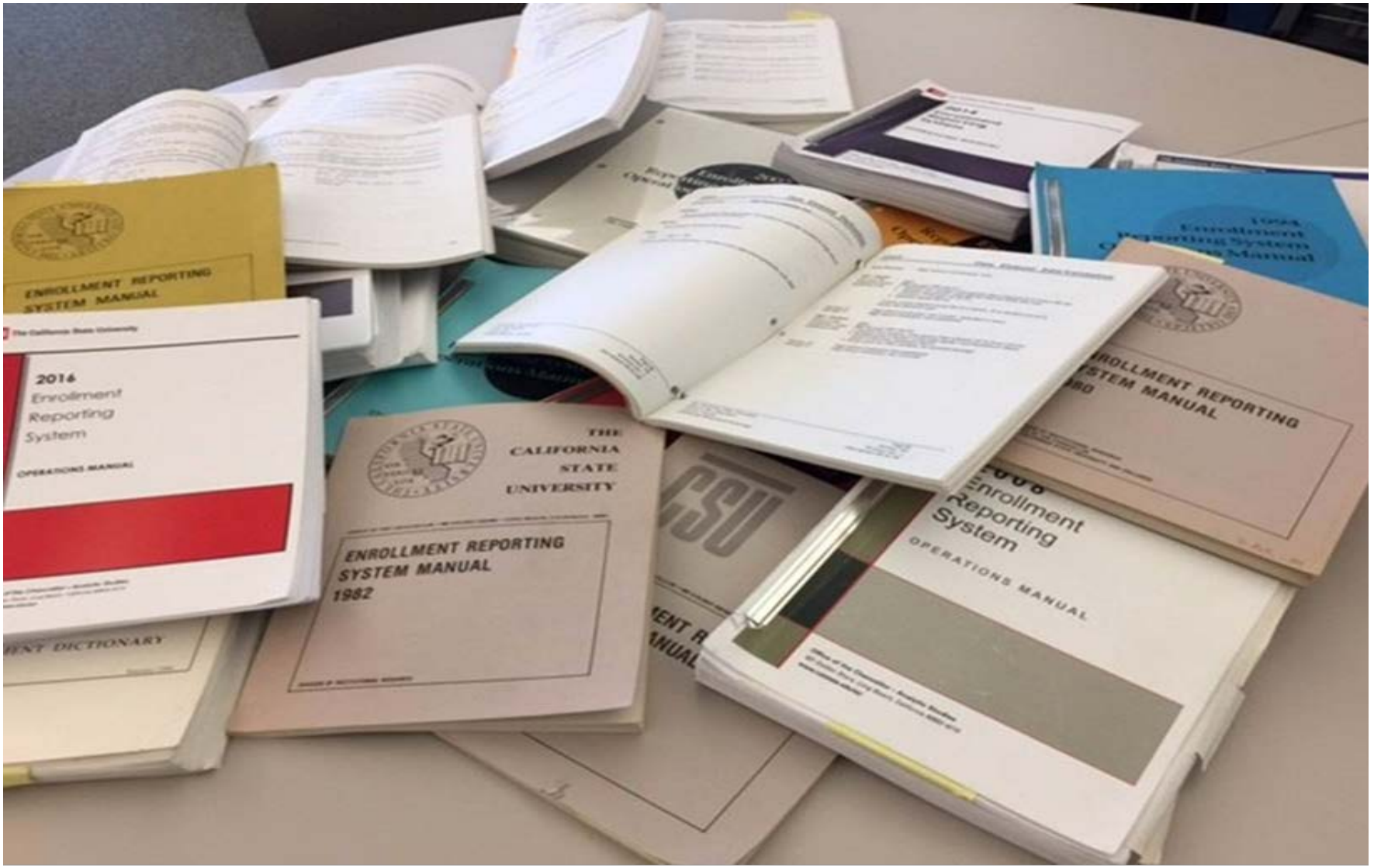
Best practices for developing and managing a data dictionary

Monica Malhotra, Deputy Director (mmalhotra@calstate.edu)
Lisa Limbeek, Web Coordinator (llimbeek@calstate.edu)


Applying for a job at IKEA
Make a chair and take a seat.
CANARY PETE

CSU The California State University
WORKING FOR CALIFORNIA

Hard Copies of Data Dictionaries from the 1973 onwards



What is Data Dictionary?

- Provides the names, definitions, and attributes of data elements within a data system.
- A tool for recording and processing information (metadata) about the data that an organization uses.
- It provides collective understanding of how data will be used and expressed within a given context.

CSU The California State University
WORKING FOR CALIFORNIA

Element Name	CITIZENSHIP CODE

Data Element Definition
A code that indicates whether a student is or is not a citizen of the United States, and the type of visa for noncitizens. This code does not denote fee status.
Structure
One (1) alphanumeric character.
Values

Value	Meaning
Y	United States citizen
I	Non-U.S. citizen, immigrant (applied for and received Form I-551 "green card")
F	Non-U.S. citizen, F visa (student)

Well Designed Data Dictionary Improves the Data Quality

- Improves documentation and control
- Ensures data integrity
- Eliminates redundancy
- Increases consistency
- Easier data analysis
- Better means of estimating the effect of change
- Allows effective communication

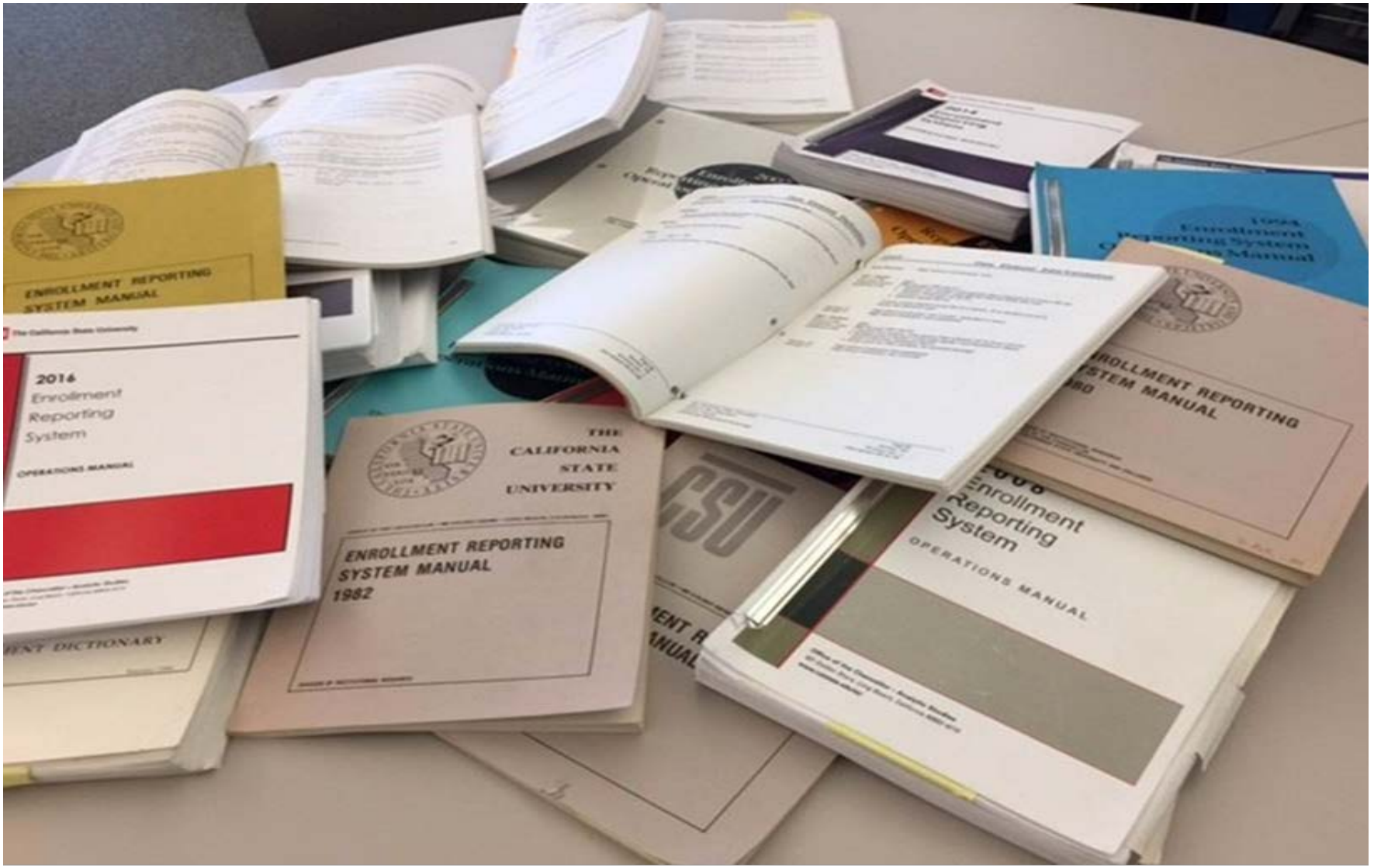
Designing a Data dictionary: Key Components

- Traceability: Knowledge of every collection that particular data appears
- Archive functions: Access to old definitions and evolution of changes.
- Intentional thoughtful design: Interface and design that is based on the needs of the end user.
- Data quality checks: automatic, regular checks of data accuracy and completeness built into the maintenance processes for the dictionary.


REALITY
CHECK
AHEAD

CSU The California State University
WORKING FOR CALIFORNIA

Hard Copies of Data Dictionaries from the 1973 onwards



Data Element Business Rules (Could include any of the material below):

- **Data Element coding (allowed values) and intra-element validation details or reference to other documents:** Explanation of coding (code tables, etc.) and validation rules.
- **Related data elements:** List of closely related data element names when the relationship

- **Validity dates for the data element definition:** Validity dates, start and possible end dates for when the data element is or was used. There may be several time periods when the data element has been used.
- **History references:** Date when the data element was defined in present form, references to superseded data elements, etc.

- Short application domain definitions and references to other documents needed to understand the meaning and use of the data element.
- **Source of the data in the data element:** Short description of where the data is coming from. Includes rules used in calculations producing the data element value.
- **Validity dates for the data element definition:** Validity dates, start and possible end dates for when the data element is or was used. There may be several time periods when the data element has been used.
- **History references:** Date when the data element was defined in present form, references

Basis of Admission

- Fall 1973 - Spring 2006,

"P" = Applicants who were eligible for admission as First Time Freshmen and who have less than 56 units of college credit, 2.0 GPA, eligible under "grandfather" clause 40804.

- Fall 2012

"P" = Admission as an Upper Division Transfer, Regular Admission, Applicants who have completed 60 units of transferable college credit via Transfer Associate of Arts/Sciences Degree – SB 1440.

- Fall 1974- Spring 2006,

"L" = First Time Freshmen <u>Alternate/Other admission</u> - Students admitted in pilot programs under section 41250.

- Fall 1995 – Spring 1999

"L" = First Time Freshmen <u>Regular Admission</u> - Students admitted in pilot programs under section 41250 or who met the UC eligibility requirement.

- Fall 1999 onwards

"L" = First Time Freshmen <u>Alternate/Other Admission</u> - Students admitted in pilot programs under section 41250.

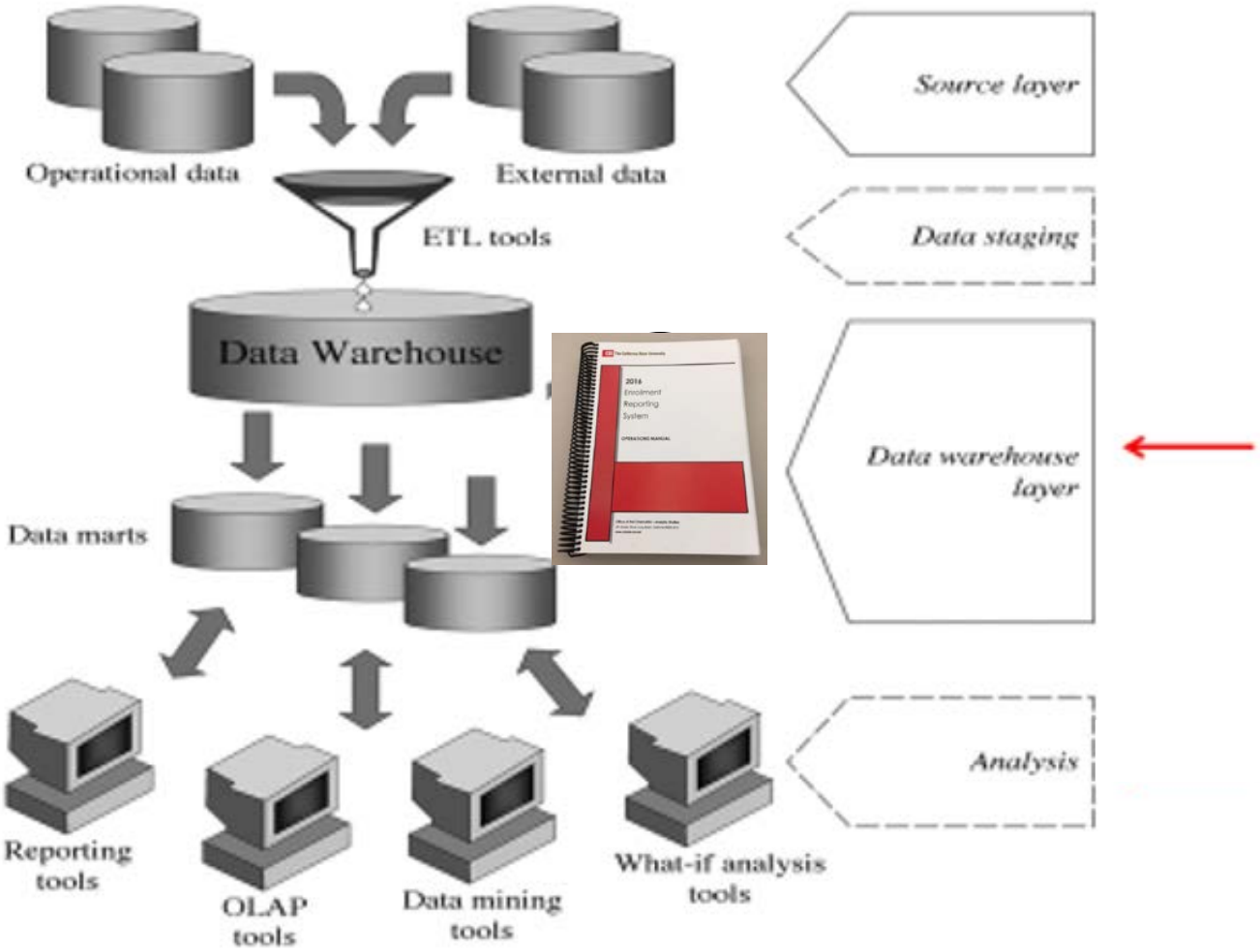
Current Structure of data dictionary
Stand alone - Not integrated with relational system

Identified Two Fundamental Steps

Manually checked, validated and documented all historical changes over time since 1973. Web based interface for updating and maintaining data dictionary.

https://asd.calstate.edu/ded/

Second step: Integrate the data dictionary to the Centralized Data Warehouse



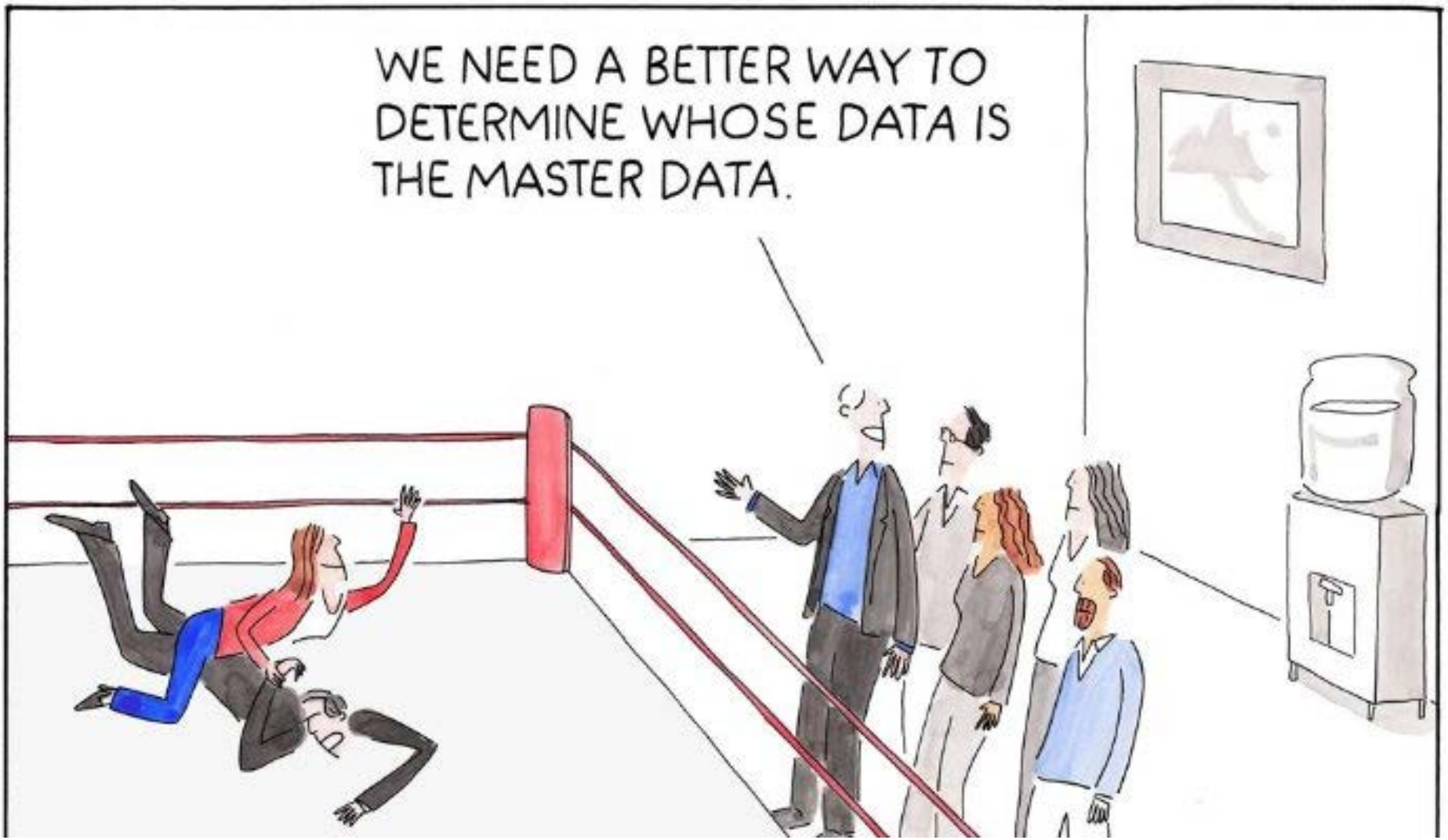
What is Metadata?

“Game of Thrones”

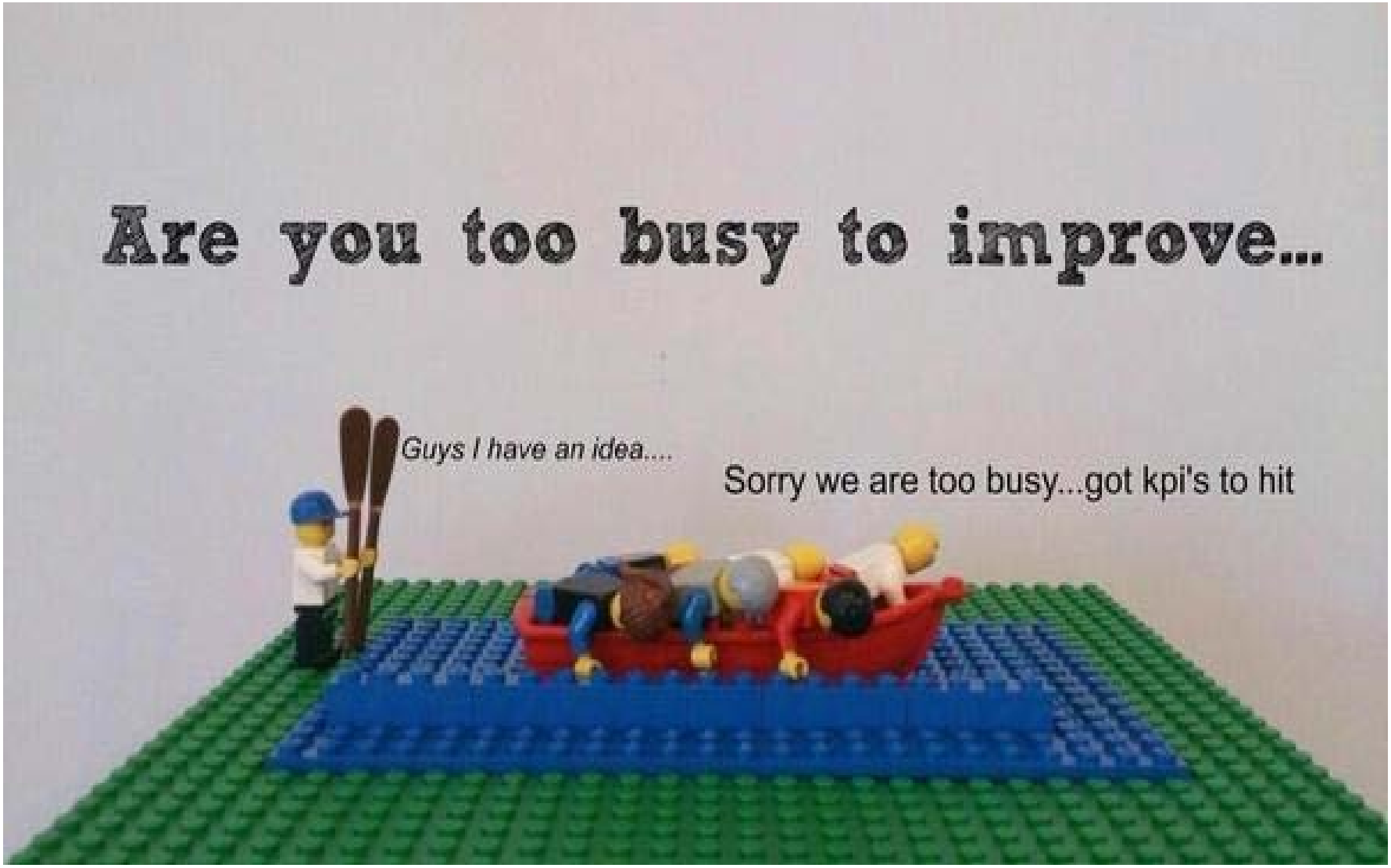
- Season: 1
- Episode Number: 7
- Episode Name: “You Win or You Die”
- Original Air Date: May 22, 2011
- Runtime: 58 minutes
- Director: Daniel Minahan
- Episode description: ”Ned confronts Cersei about her secrets;Joe takes his Night’s Watch vows; Drogo promises to lead the Dorthraki to King’s Landing.”

Recommended Practices

- Data governance process should be in place
- Keep dictionary current and relevant
- Single Entity for reference
- Dictionary should not be a Static or a stand-alone tool
- Clearly defined ownership of collections, elements and due dates
- Collaborative and interactive effort between IT and Business
- DED is a report from the metadata
- See the data dictionary as a process, not a product


WE NEED A BETTER WAY TO DETERMINE WHOSE DATA IS THE MASTER DATA.


IT-BUSINESS
Collaboration


Are you too busy to improve...
Guys I have an idea....
Sorry we are too busy...got kpi's to hit

CSU The California State University

Monica Malhotra, Deputy Director (mmalhotra@calstate.edu)
Lisa Limbeek, Web Coordinator (llimbeek@calstate.edu)