June 7 , 2016

WISCONSIN CHAPTER

Integrating the Business Glossary into the Data Management Lifecycle

You Say Info, I say Data
… We can't call the whole thing off!

The search for the elusive Critical Data Element

Don Soulsby
Sandhill Consultants

SANDHILL

Who we are:

Sandhill Consultants

Is a group of federated people, partners, products and processes that help our clients:

Build comprehensive

Data Architecture

Resulting from a persistent

Data Management Process

Founded on a robust

Data Governance Practice

Produces **Trusted, Governed Data**

Don Soulsby

- Vice President Architecture Strategies, Sandhill Consultants
- CMMI – Enterprise Data Management Expert
- Enterprise Architecture, Business Intelligence, Metadata Consultant
- Data/Process modeller
- Educator & Industry Speaker

Meta an abstract, high-level analysis

Wright person who creates, builds, or repairs something specified

SANDHILL

The Elusive

Critical Data Element

YOU SAY INFO, I SAY DATA
PART 1

SANDHILL

2105 - Trust & Traceability

FROM FIN TO FORK

1 CODE THE CATCH
Fish harvesters identify their catch with a unique code using a tag or label.

2 UPLOAD INFO
Info about when, where and how your seafood was caught is uploaded to ThisFish.info.

3 HANDLE + SHIP
Other folks in the supply chain can also upload info about its handling and shipping.

4 TRACE TO DISCOVER
Trace your seafood's code at ThisFish.info to discover its story and connect to your harvester.

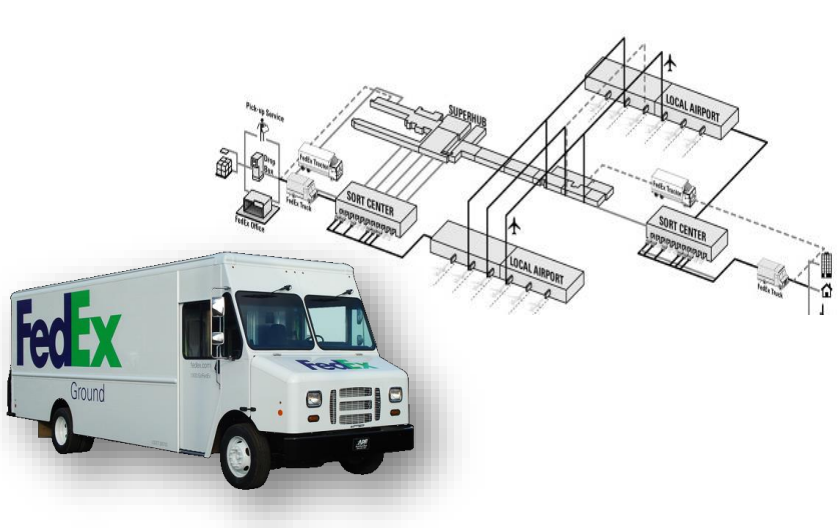
SANDHILL

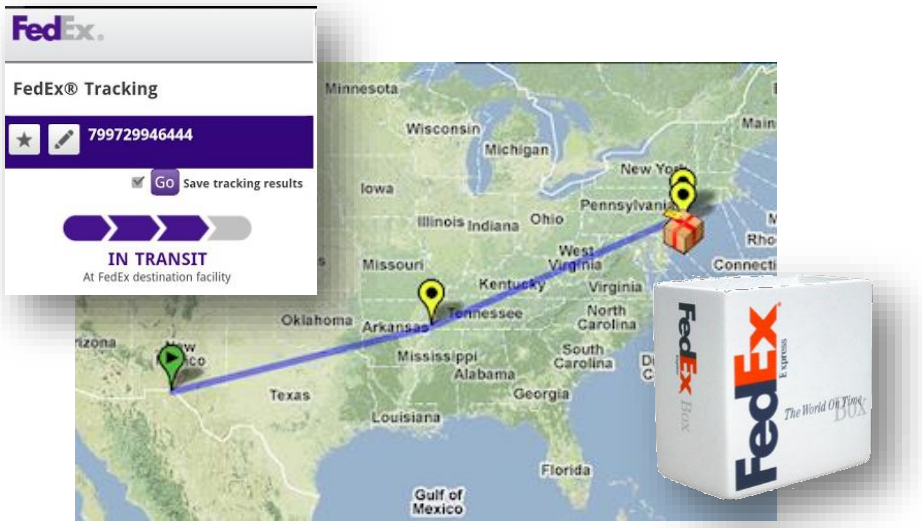
Help Data Architecture not go…

Frank
Lloyd
Wrong

SANDHILL

Operational vs Customer Service Data





"Information about the package
is as important as the package itself"

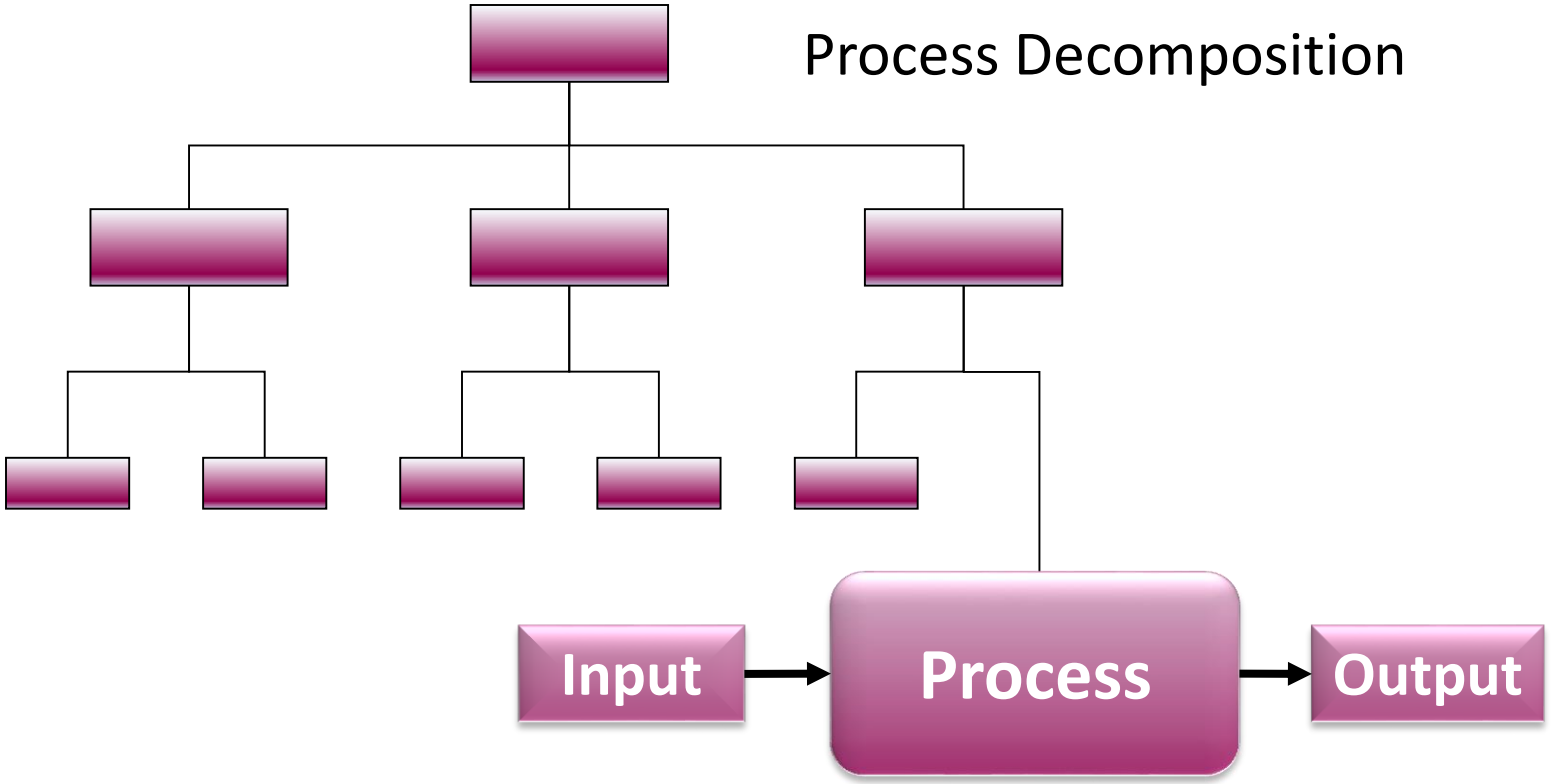
Frederick W. Smith, CEO Fedex,

Industrial Age

Information Age
Integrated Age
Interoperable Age

SANDHILL

Process Oriented Systems Engineering



Data derived from Process

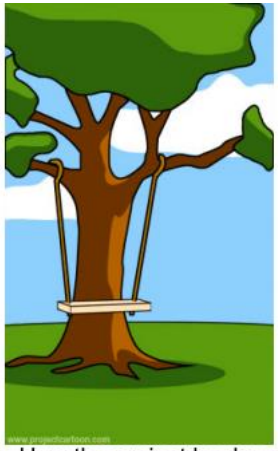
SANDHILL

Given an Understanding of the Requirements

SANDHILL

Understanding Requirements

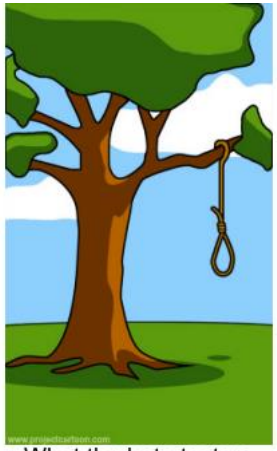
How the customer explained it



How the project leader understood it

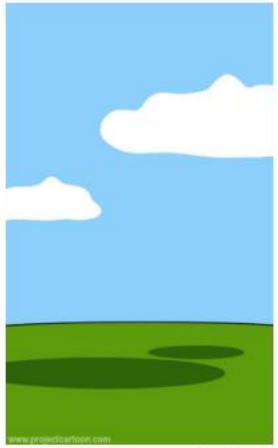
How the analyst designed it

How the programmer wrote it

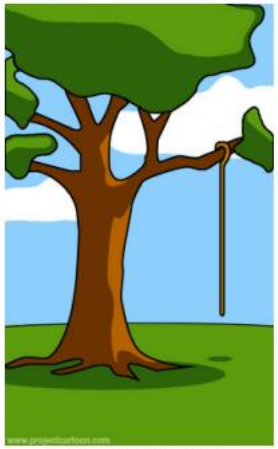


What the beta testers received

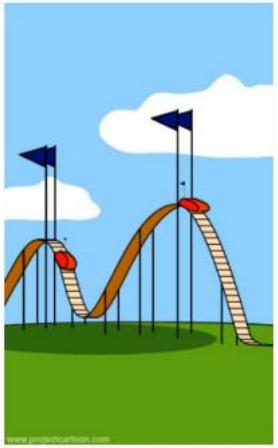
How the business consultant described it



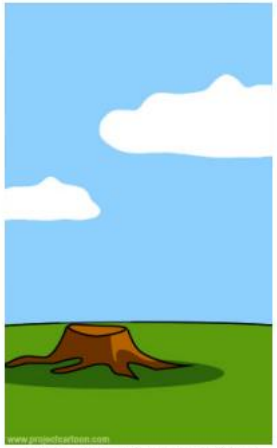
How the project was documented



What operations installed



How the customer was billed



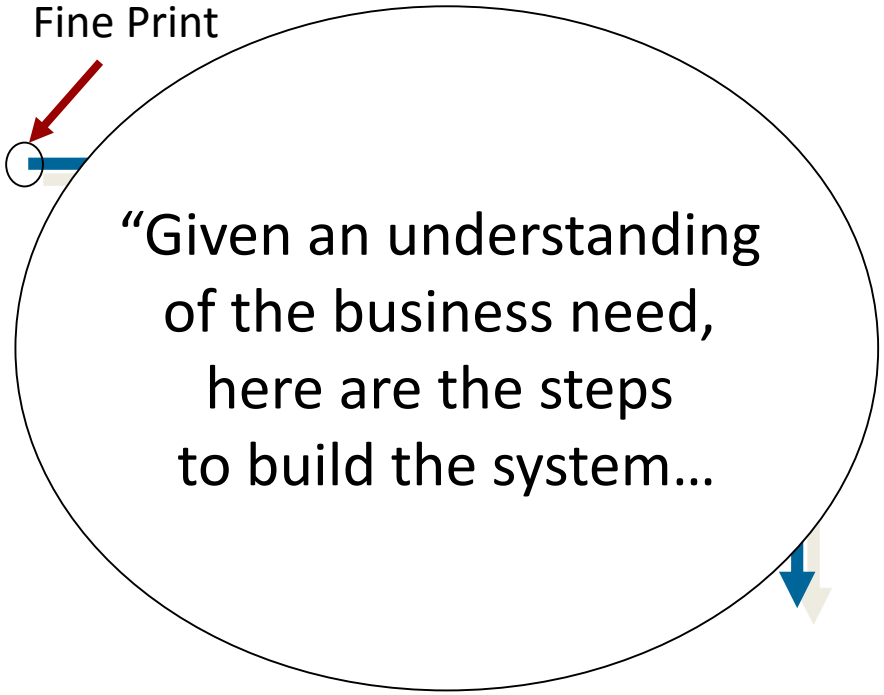
How it was supported

What marketing advertised

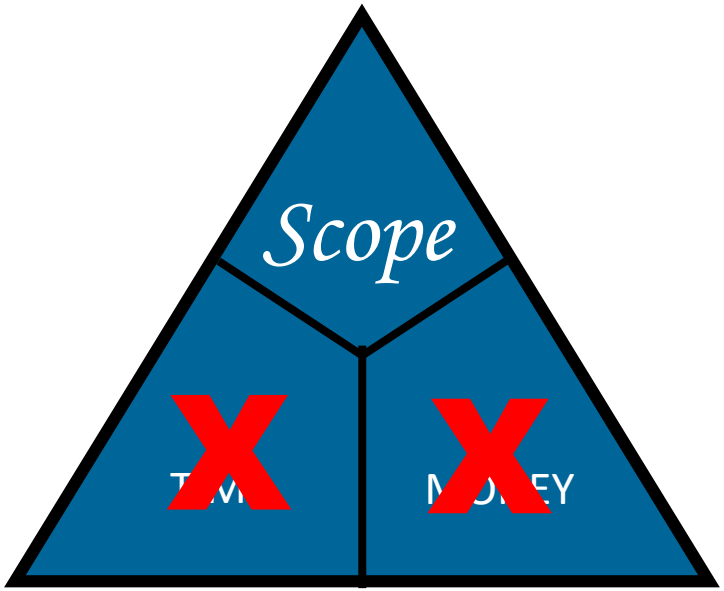
What the customer really needed

SANDHILL

Lack of Integration - Interoperability



Waterfall Approach



SANDHILL

Biggest problem

FROG

We do not

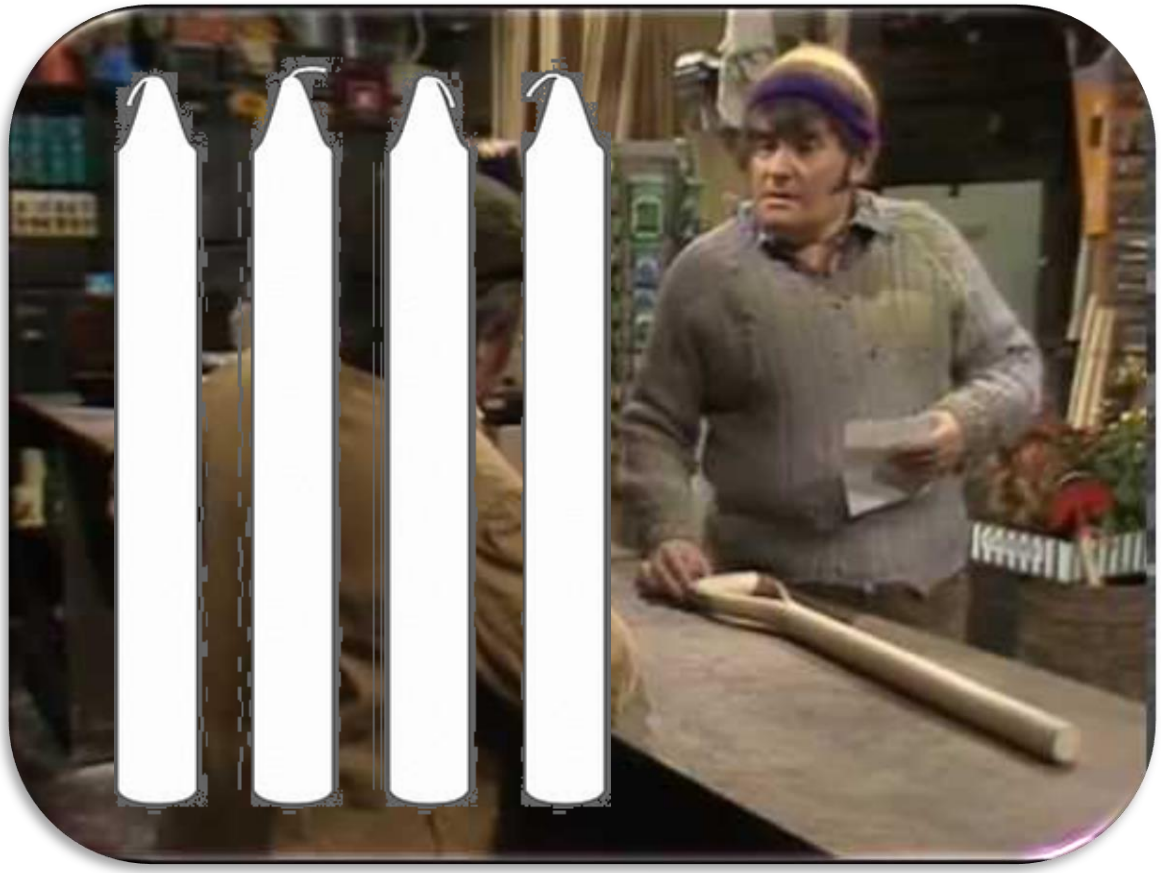
Communicate well!

- No Definitions
- Too many definitions
- We think we know…
- No Context !

SANDHILL

Fork Handles

https://www.youtube.com/watch?v=gi_6SaqVQSw

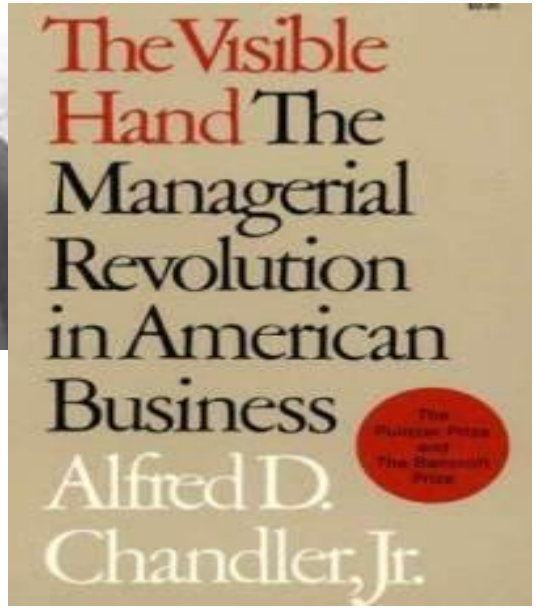


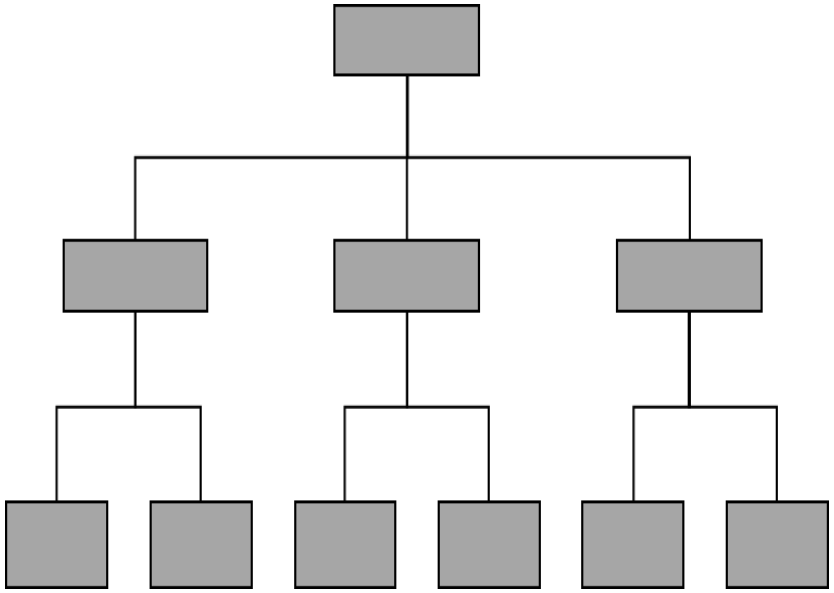
"…Two great countries separated by a common language"
G.B.Shaw

SANDHILL

Structure follows Strategy

Alfred D. Chandler, Jr.





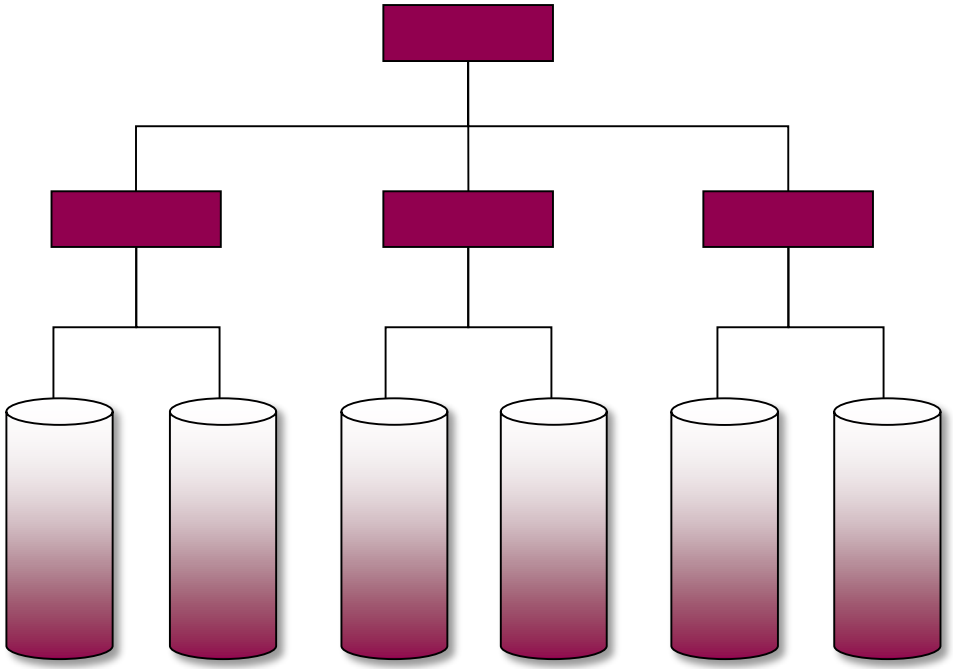
SANDHILL

Form Follows Function

Louis Sullivan
‘Form (ever) follows function"

The principle is that the shape of a building or object should be primarily based upon its intended function or purpose.

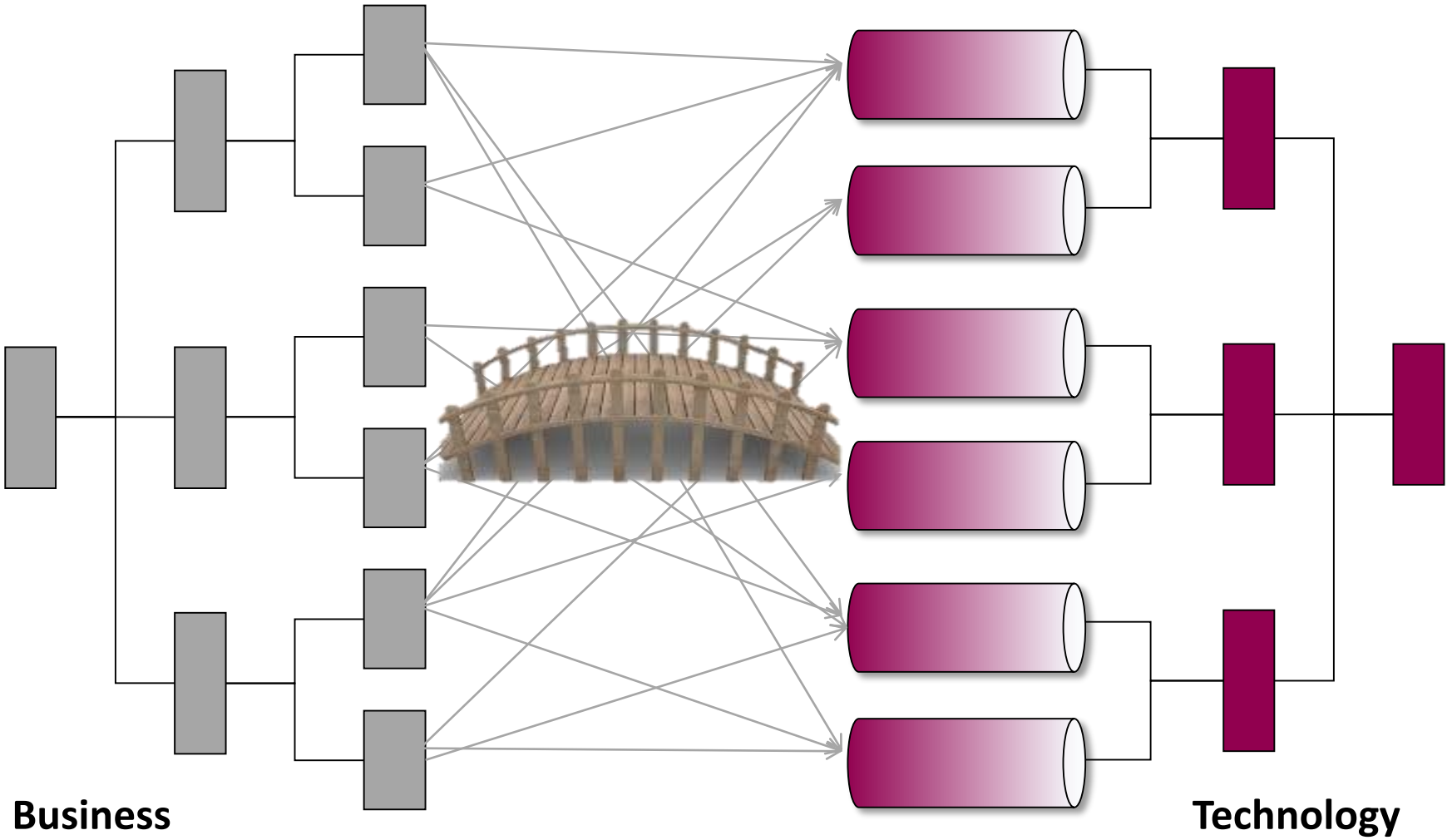
Process Decomposition
Service Oriented Architecture



Building “ING” data

SANDHILL

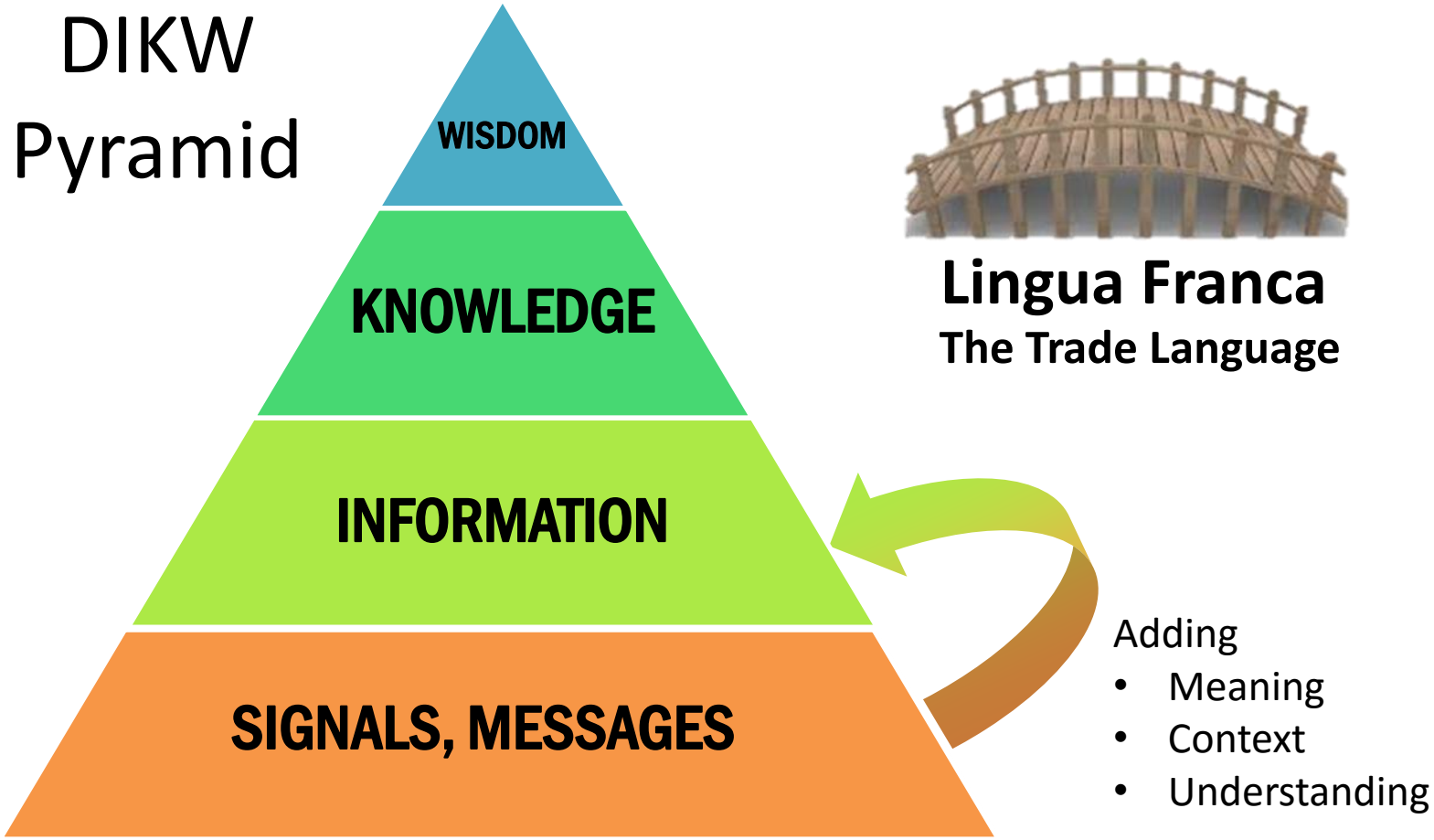
Data Process Challenge



SANDHILL

The Business Need

DIKW Pyramid



Lingua Franca
The Trade Language

Adding
- Meaning
- Context
- Understanding

1955

SANDHILL

Data Management

- Guiding principle of Data Management is that data is an enterprise asset.
- First step is to have an Inventory of Assets
- Not all assets have the same value and require the same level of management.
- Need to classify assets to be able monetize them.

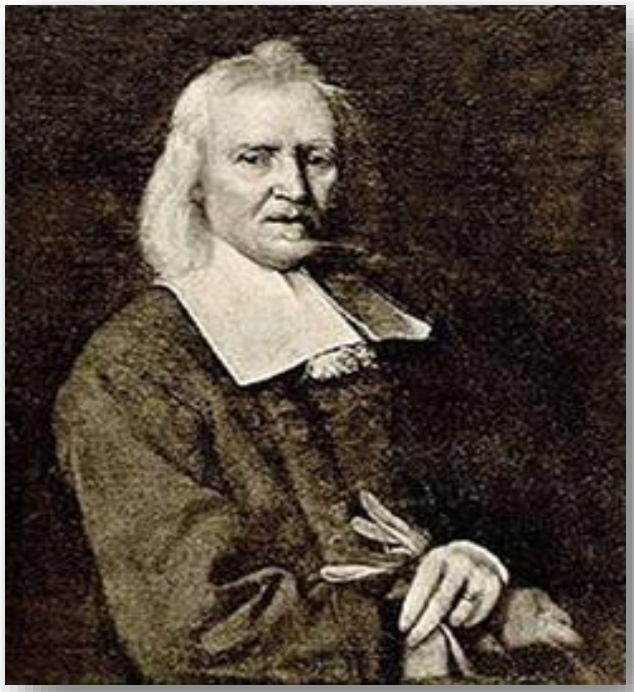
SANDHILL

Data Quality Ownership

This is a story about four People named **EVERYBODY, SOMEBODY, ANYBODY** and **NOBODY**. There was an important job to be done and **EVERBODY** was sure that **SOMEBODY** would do it. **ANYBODY** could have done it, but **NOBODY** did. **SOMEBODY** got angry about that, because it was **EVERBODY's** job. **EVERBODY** thought **ANYBODY** Could do it, but **NOBODY** realized that **EVERYBODY** wouldn't do it. It ended up that **EVERBODY** blamed **SOMEBODY** when **NOBODY** did what **ANBODY** could have.

SANDHILL

Major Challenges – Data Assets



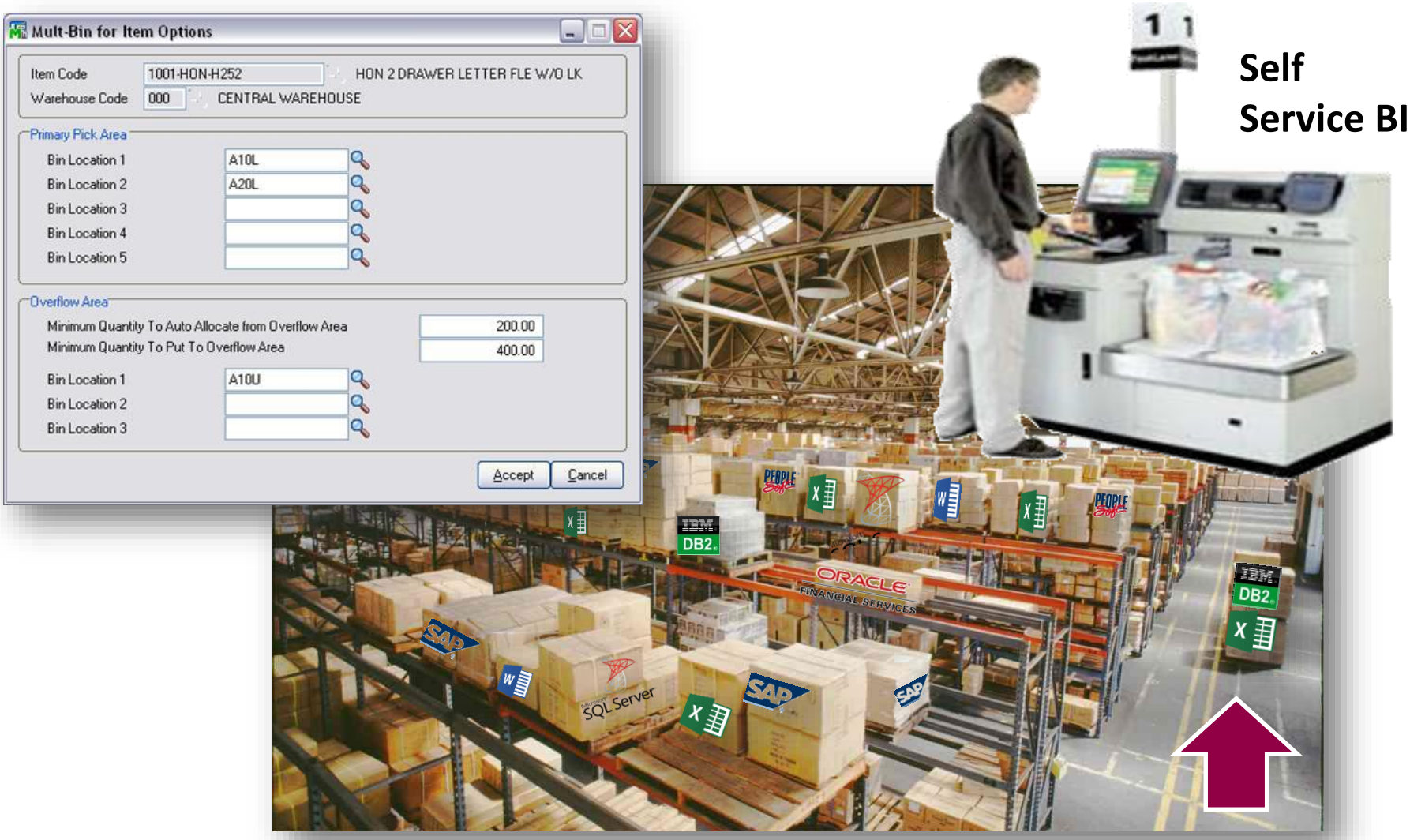
That which is Everybody's business is Nobody's business

Izaak Walton (17th Century)

- Lack of Senior Executive support, do not see data assets on the books.
- Data in many distributed systems with no common data model and multiple repositories of metadata with limited search capabilities.
- Difficulty in getting funding centralized and strategically planned.
- IT project culture, limited support for cross project data programs – master data.

SANDHILL

Data Asset Inventory



SANDHILL

Classification

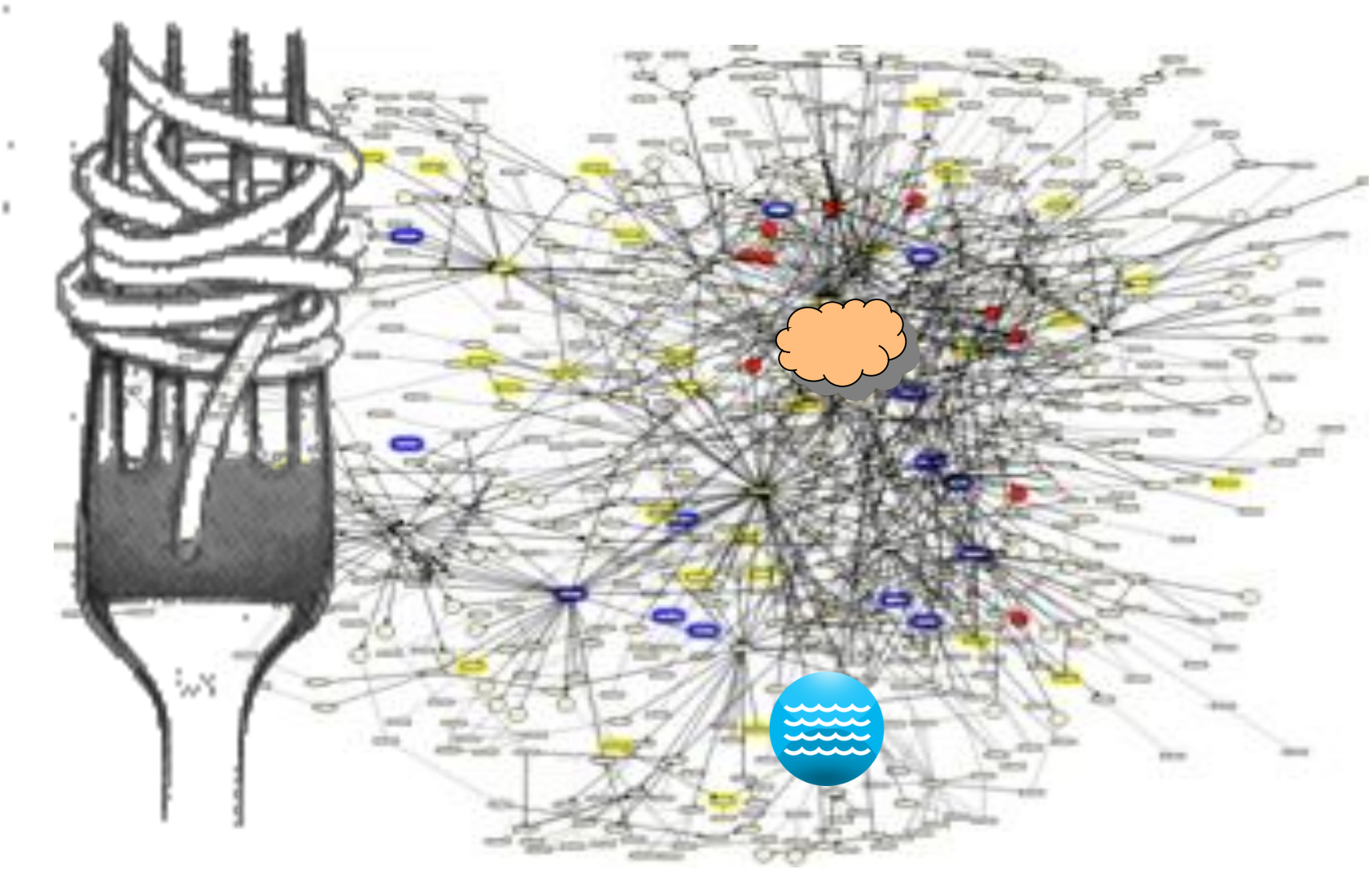
UCA
Uniform Classification of Accounts

Railway Assets

Track and Roadway	
Account Number	Account Description
101	Land
102	Grading
103	Rail
105	Ties
106	Paved Concrete Trackbed (PACT System)
107	Other Track Material
109	Ballast
111	Track Laying and Surfacing
113	Used Track Material in Store
115	Bridges
117	Culverts
119	Tunnels
121	Fences, Snow Sheds and Rock Sheds
123	Public Improvements
125	Other Right-of-Way Property

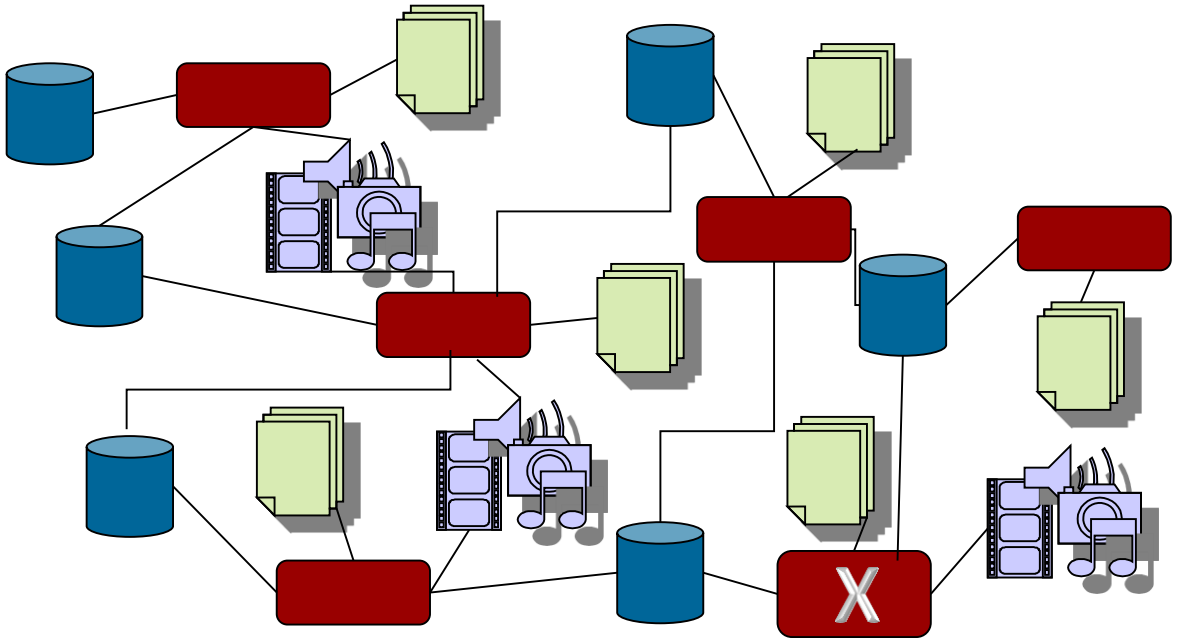
SANDHILL

Data Management Environment



SANDHILL

The Management Problem?

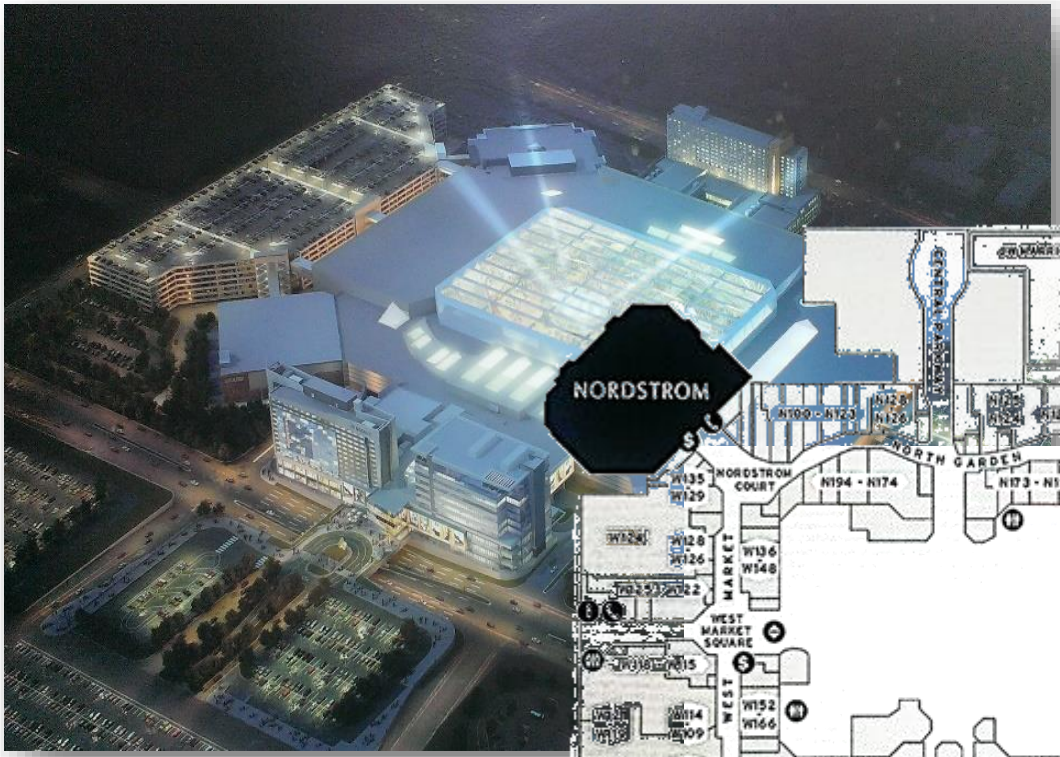


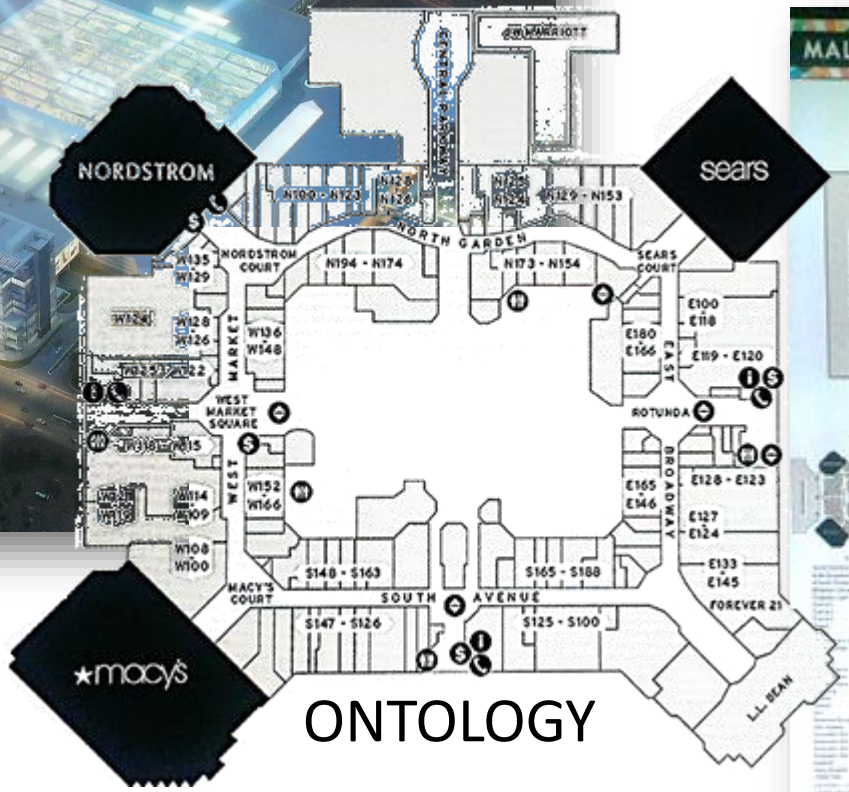
You cannot manage what you cannot measure,
You cannot measure what you cannot define.

You have a ***Finding*** problem!

SANDHILL

Data Governance Mall



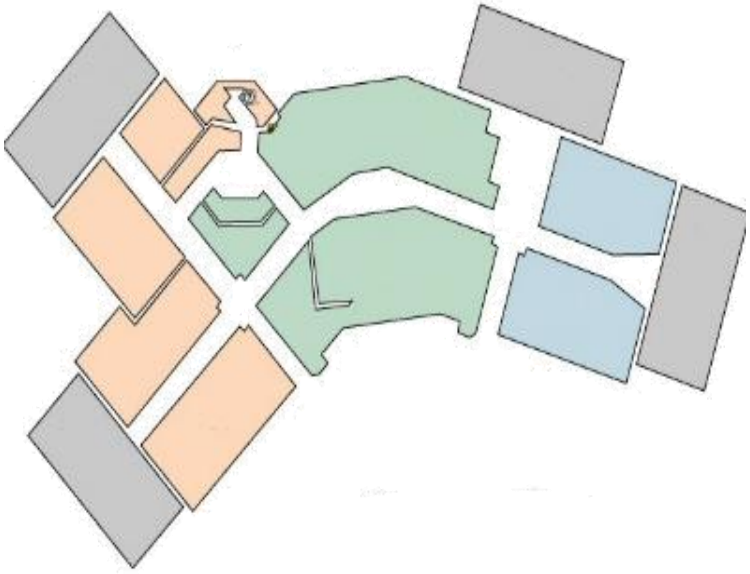


ONTOLOGY

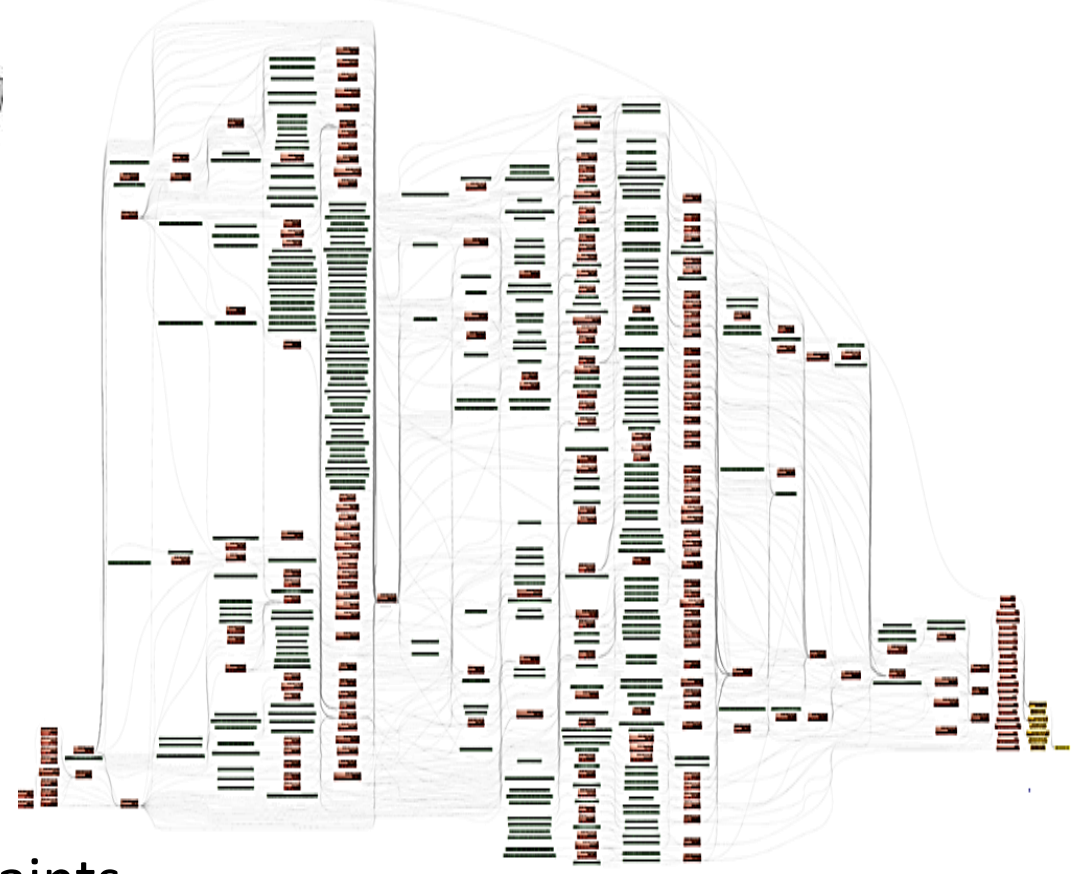
TAXONOMY

SANDHILL

The Data Governance Mall



The Mall Ontology



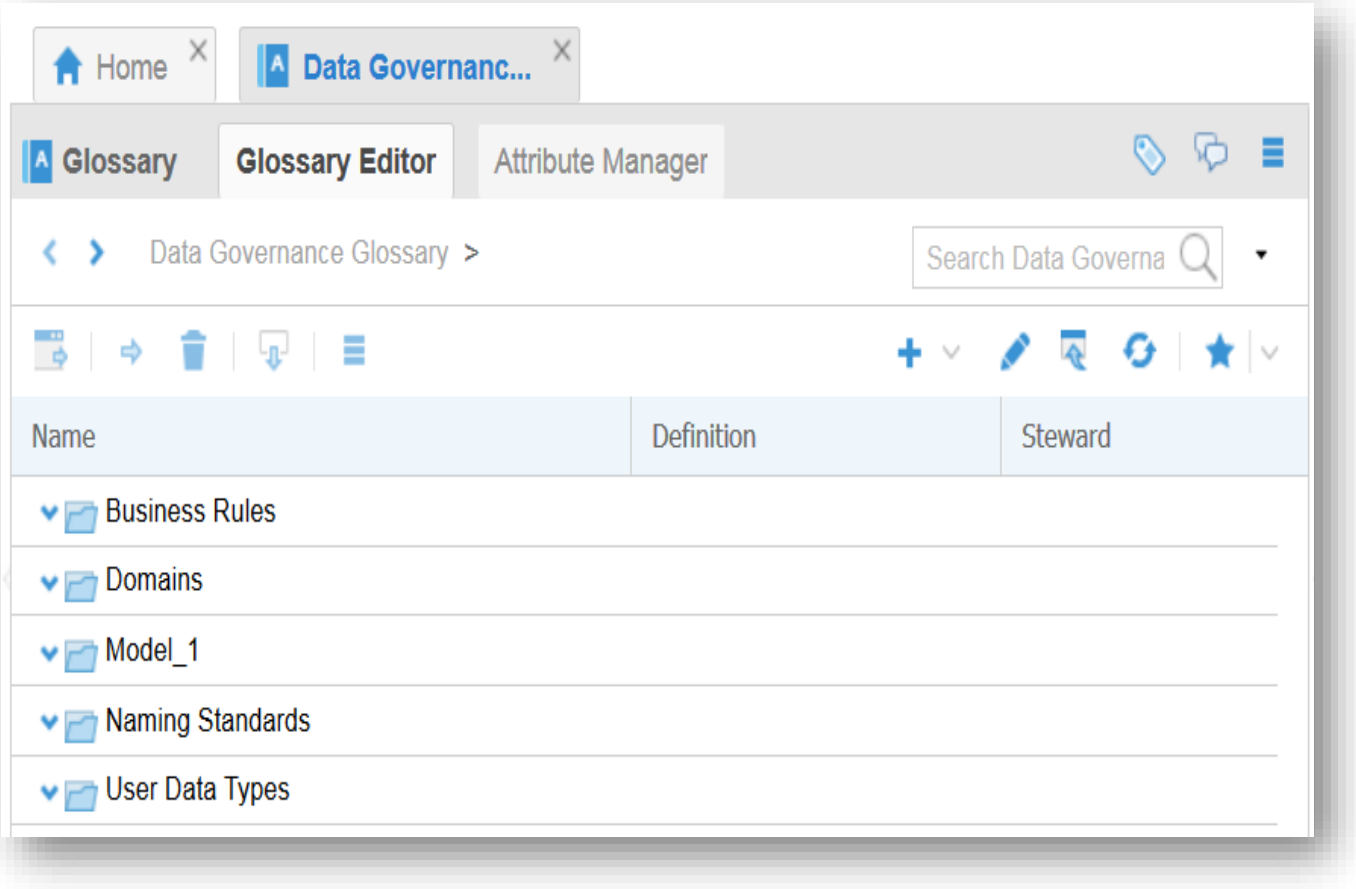
Objectives:

- Context
- Understandability
- Searchability
- Usability
- Hide technological constraints

SANDHILL

The Business Glossary

Mall Taxonomy



SANDHILL

Taxonomy and Ontology

Taxonomy

- The science or the technique used to make a classification.
- A classification; especially, a ***classification in a hierarchical system.***
- The science of ***finding***, describing, classifying and naming organisms.

Ontology

- (logic) A logical system involving theory of classes, developed by Stanislaw Lesniewski (1886-1939).
- (computer science, information science) A structure of concepts or ***entities*** within a ***domain***, organized by ***relationships***; a system model.

SANDHILL

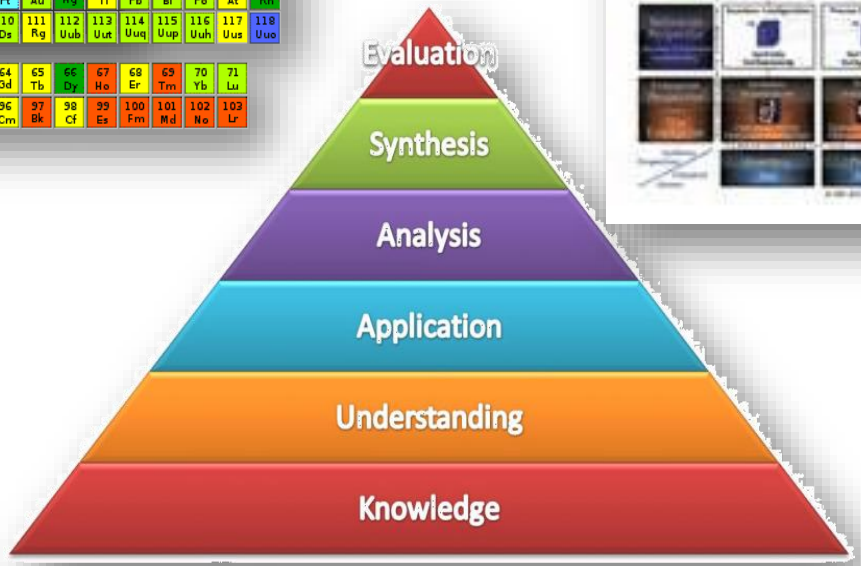
Taxonomies

- Accounting
- Automotive Parts and Equipment
- Banking
- Building and Construction Management
- Finance and Investment
- Food and Beverage
- Food Service Equipment and Supplies
- General Business
- Higher Education
- Human Resources
- Industrial Equipment and Supplies
- Information Technology
- Insurance
- K12
- Legal
- Local Government
- Manufacturing
- Medical Conditions and Specialties
- Medical Equipment and Supplies
- Mining
- News
- Non-Profit
- Oil and Gas
- Personal Care Products
- Product and Service
- Project Management
- Real Estate
- Records Retention
- Sales and Marketing
- Scientific and Technical Equipment
- Sensory
- Sentiment
- Skills
- Wastewater Utility
- Water Utility

wand®
precision classification for search

SANDHILL

Ontologies



Bloom's Learning Domains

SANDHILL

Enterprise Ontology - Metadata

Business Metadata

- Semantic definitions and business context for enterprise information systems.

Technical Metadata

- Logical and physical design and blueprint information for enterprise information assets.

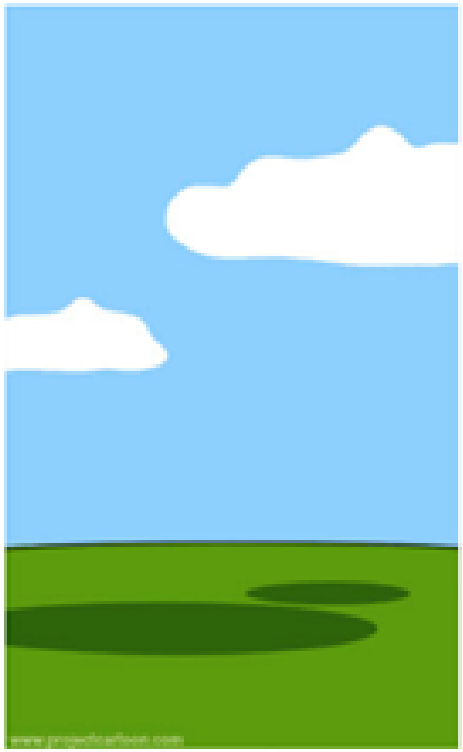
Operational Metadata

- Information describing the activity of information systems in production.

DATA IN CONTEXT

SANDHILL

Getting on the list of 10!

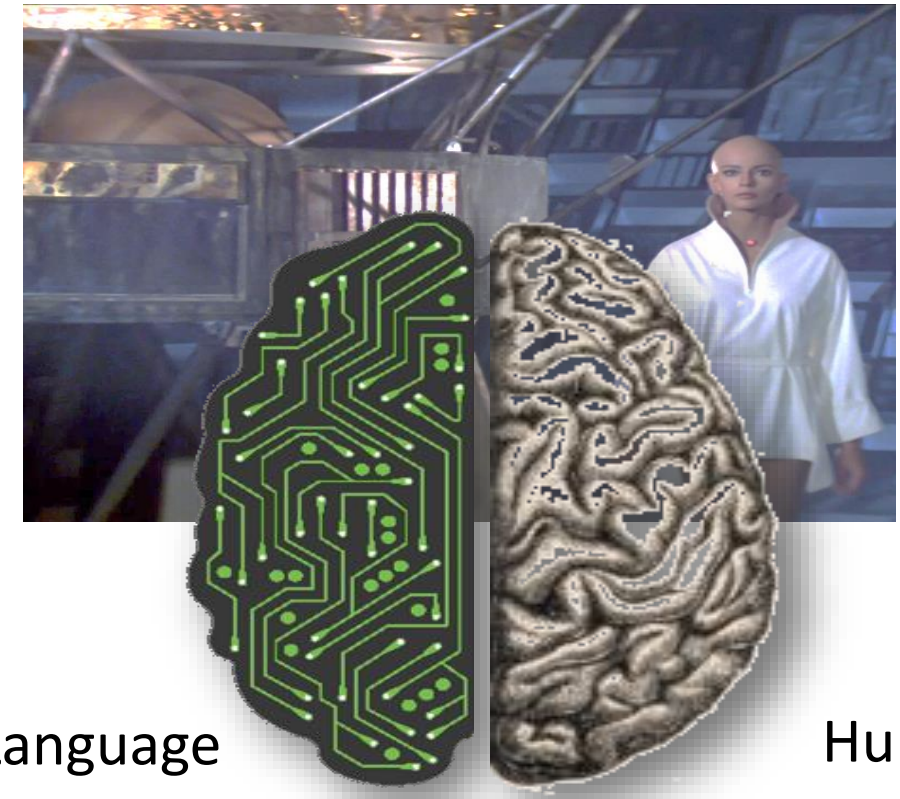


How the project was documented

1. …
2. …
.
.
9. …
10. …
11. **METADATA**
12. **DOCUMENTATION**

SANDHILL

The Corporate Brain Challenge



SOFTWARE
Programming Language

WETWARE
Human Language

DATABASE

(code)
Technical Metadata

TAXONOMY

(text)
Business Metadata

SANDHILL

Business Metadata

Business Metadata refers to facts about the business. Typically, they are not stored in conventional system.

Business Metadata is…

- Contextual information that business users access directly
- Is expressed in a language business people can understand

SANDHILL

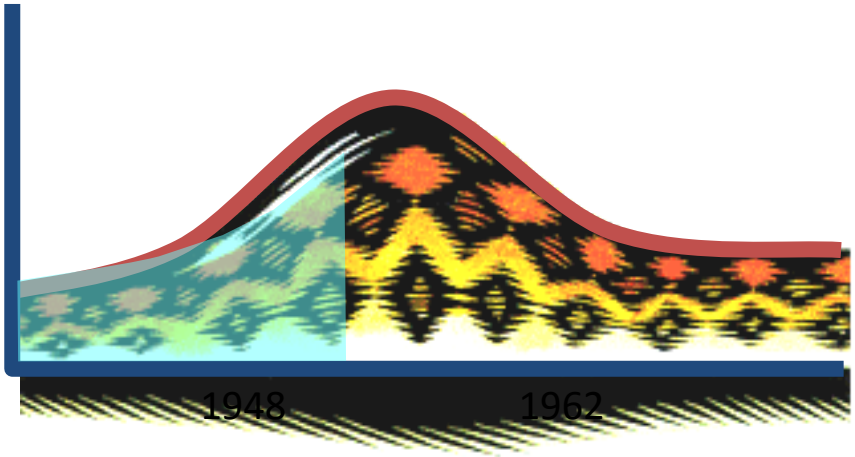
Putting Data Together

- Compliance and Regulations
- Globalization, Mergers and Acquisitions
- Metrics and Analytics – Data Warehouse
- Data Quality Issues – Big Data
- Customer 360^{O} - CDI/PIM – Master Data
- Enterprise Integration Software - ERP
- Graying of the workforce

SANDHILL

Bye Bye Boomers

The Pig and the Python



"The Pig and the Python: How to Prosper from the Aging Baby Boom"
by David Cork, Susan Lightstone

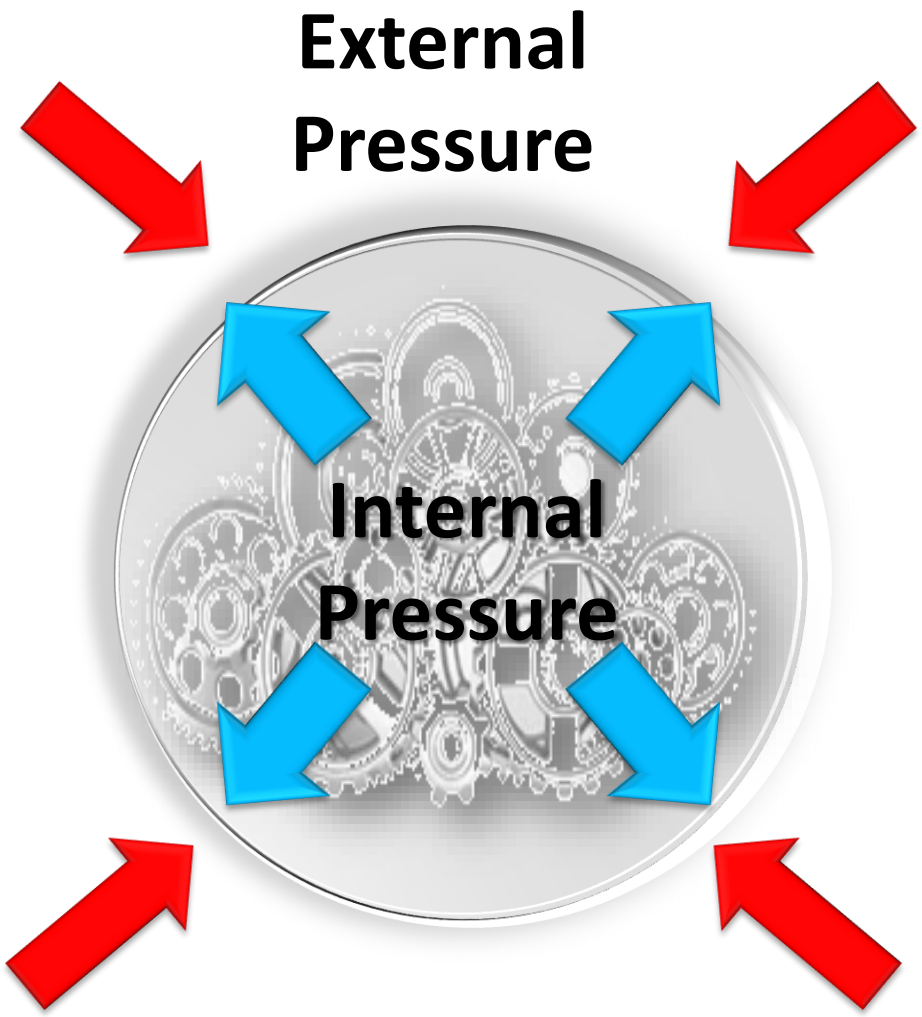
The Magic

Age + Years of Service

Courtesy of 2008 – "200.5 E"

SANDHILL

The Only Thing That Is Constant Is Change*



SANDHILL

* Heraclitus – 500 BCE

Metadata Driven Change Management

Human Oriented

REACTIVE

Business Glossary

Business Domain

PROACIVE

Conceptual Data Model

Forward Engineering

Logical Data Model

Middle Out

Metadata Repository

PROACTIVE

Physical Data Model

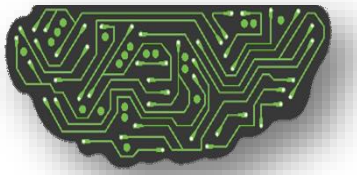
Reverse Engineering

REACTIVE

Technology Domain

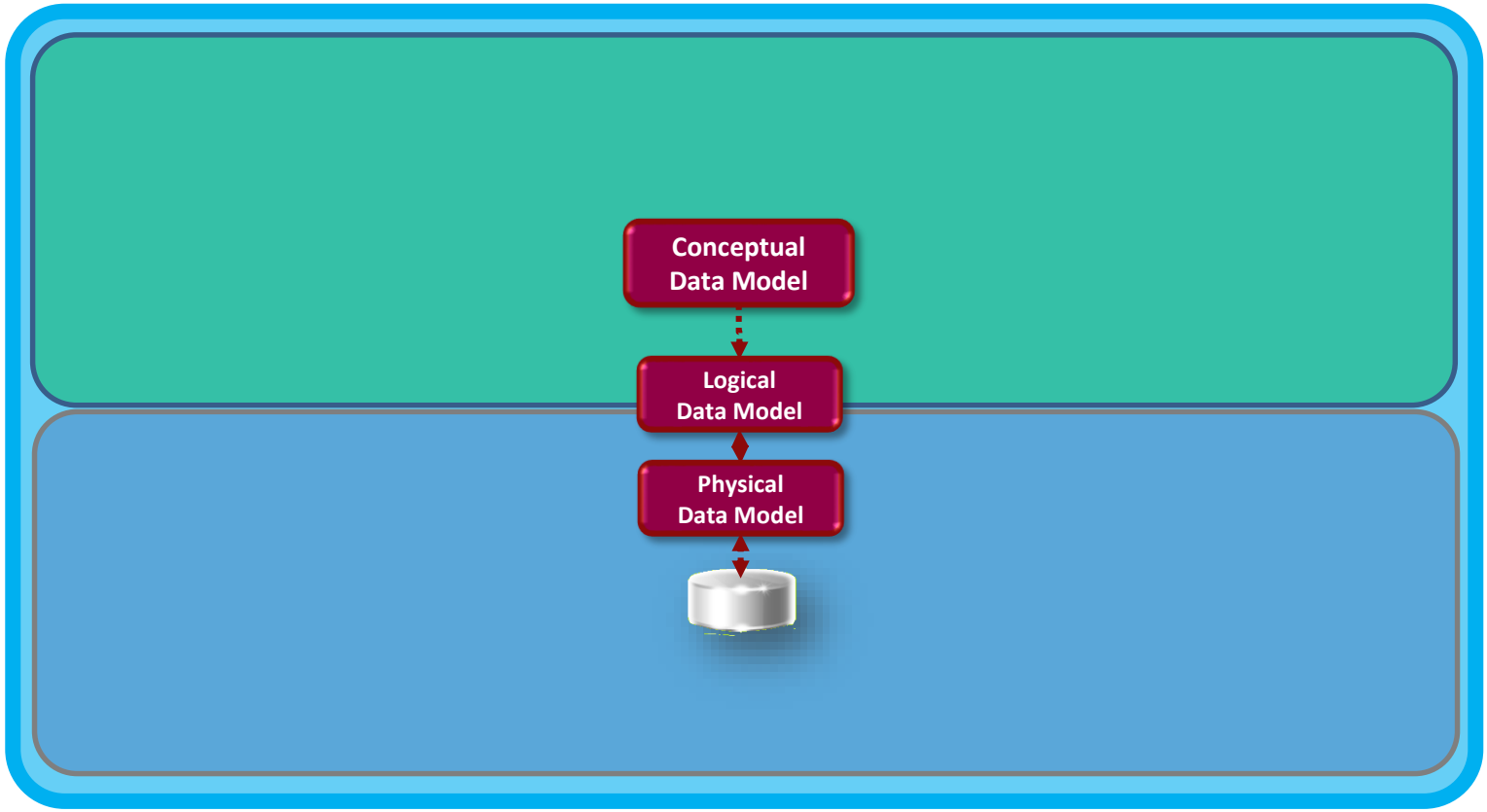
DATABASE

Computer Oriented



SANDHILL

Application Development



COMPANY

Business Domain

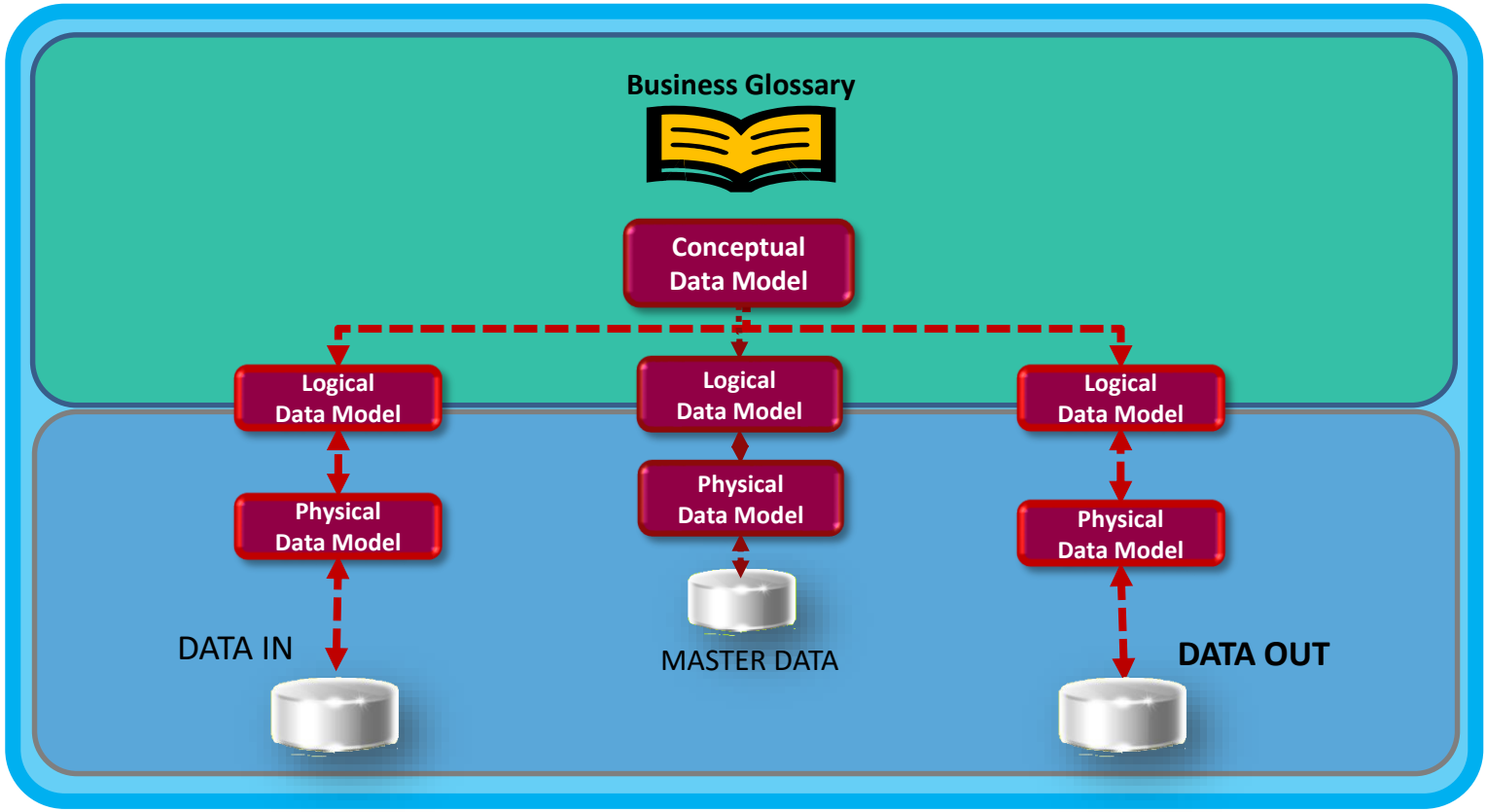
IT Domain

SANDHILL

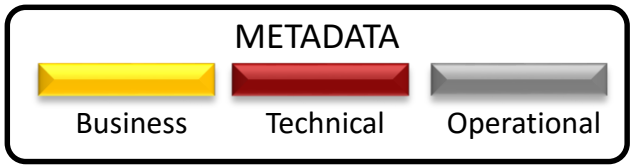
METADATA

Business | Technical | Operational

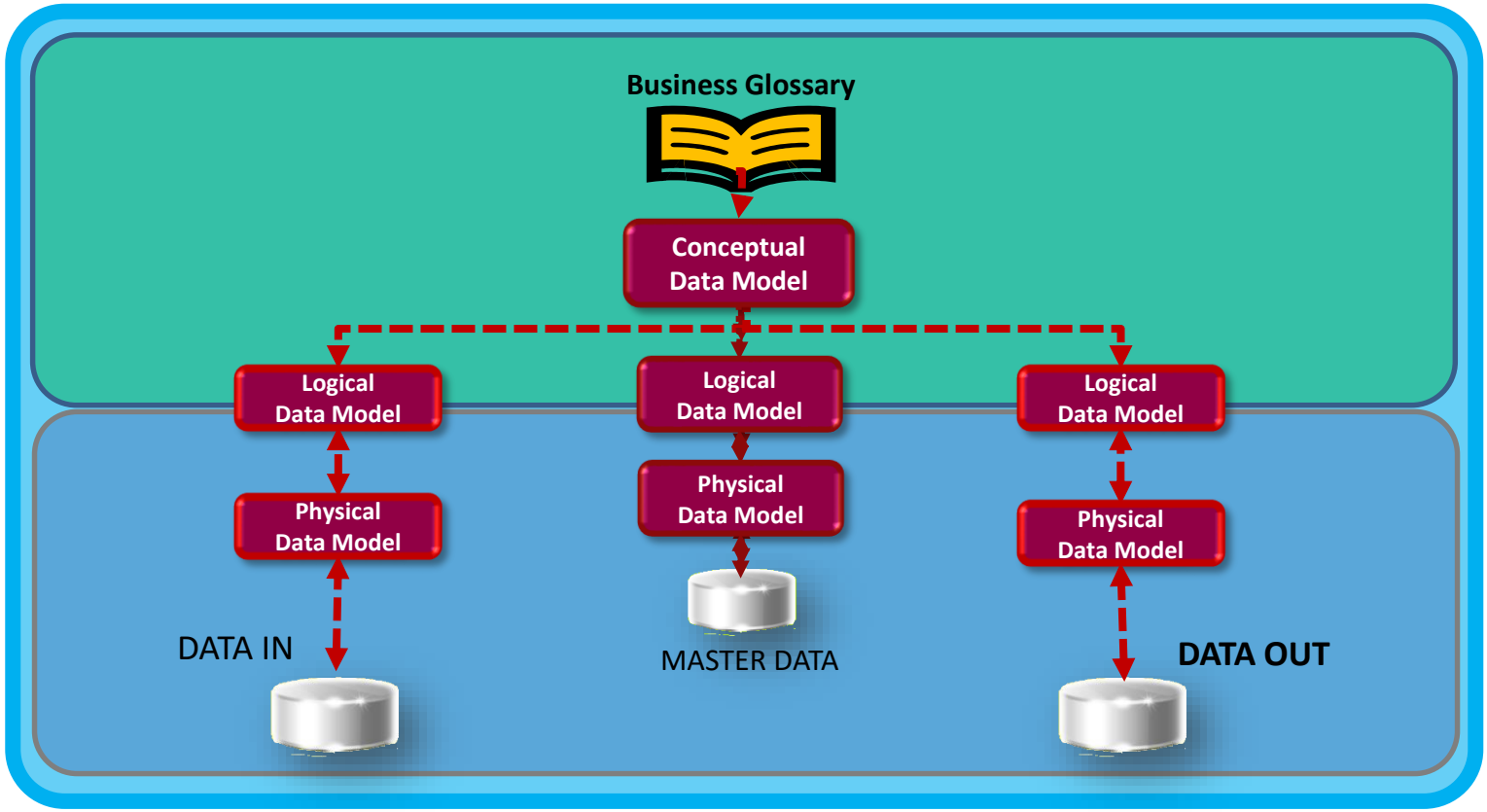
Application Development



SANDHILL



Semantic Traceability

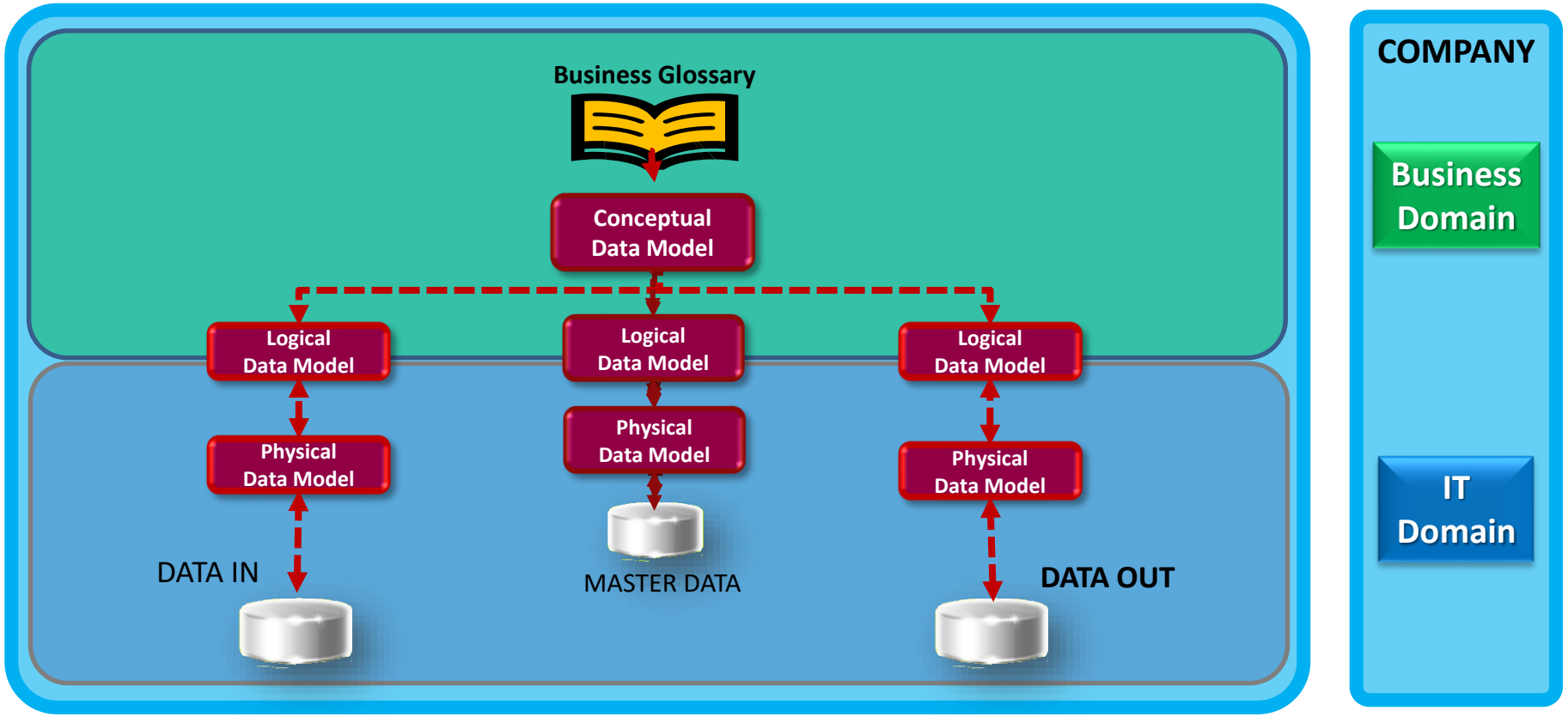


SANDHILL

METADATA

Business | Technical | Operational

Impact Analysis

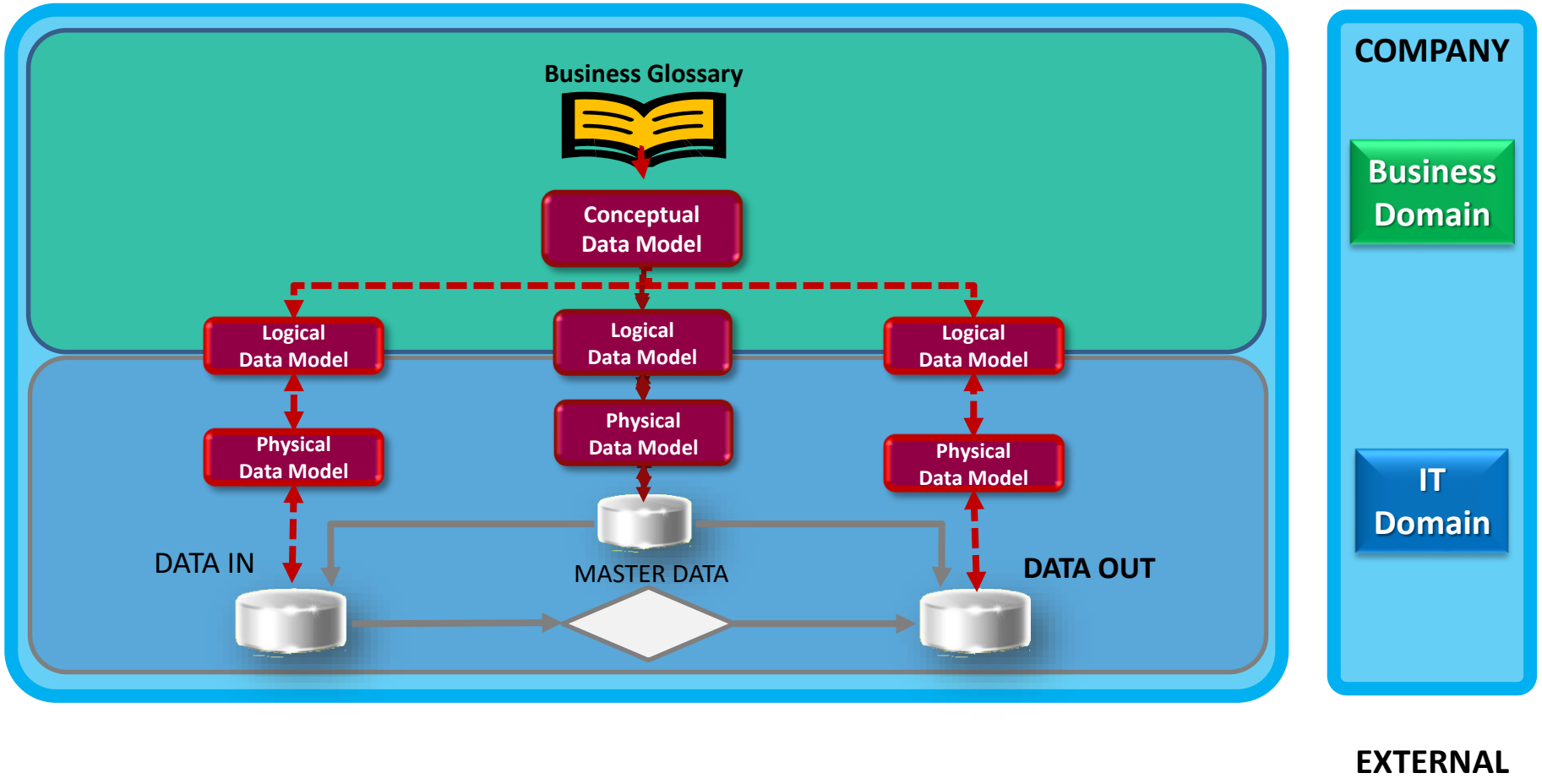


SANDHILL

METADATA

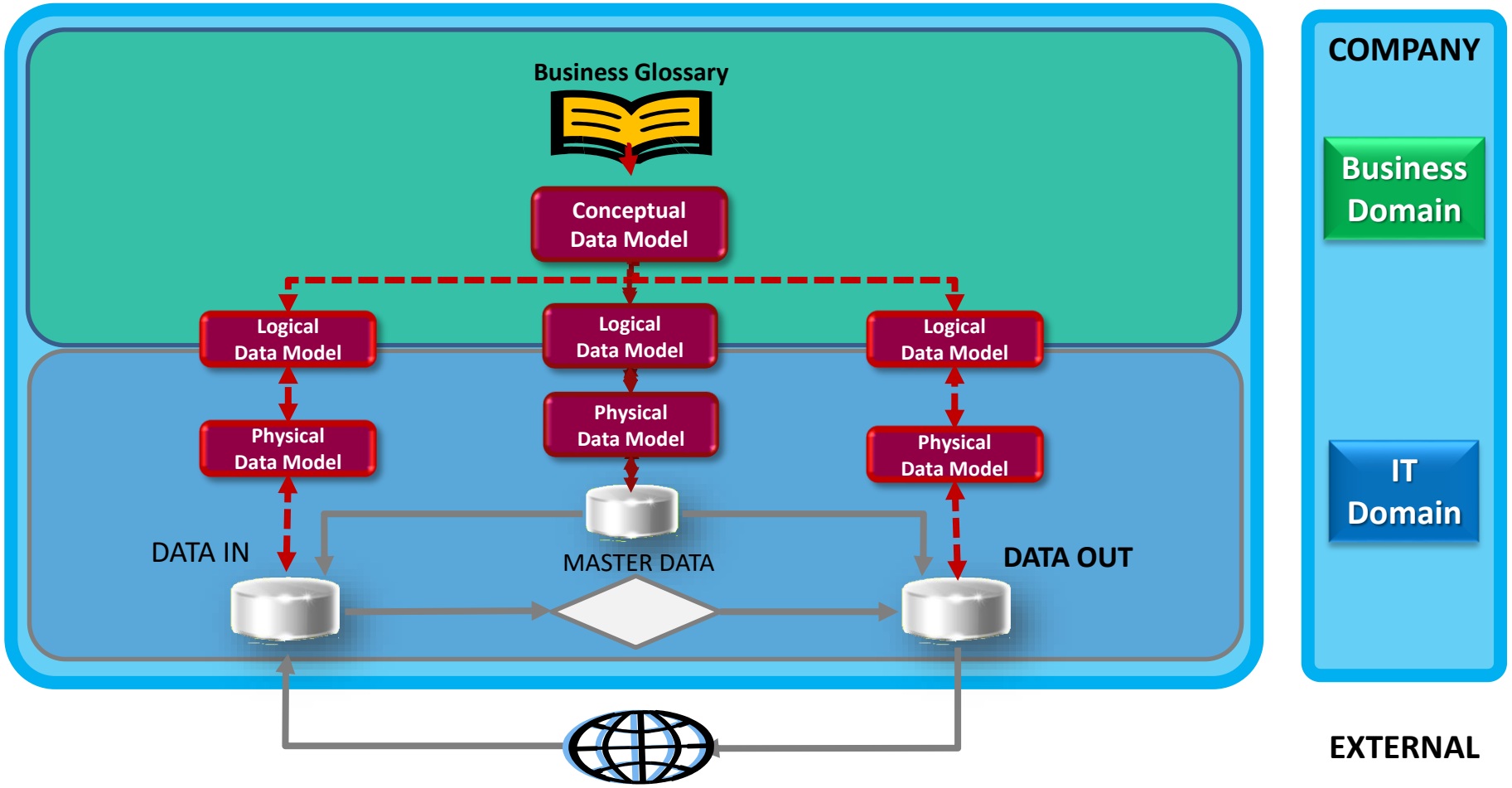
Business Technical Operational

Operational Data Traceability

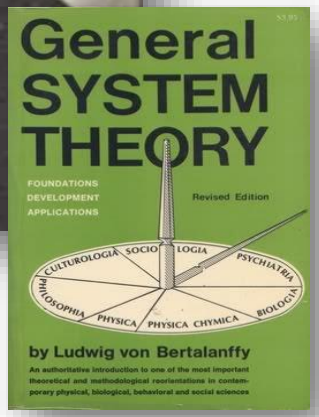


SANDHILL

External Data Traceability

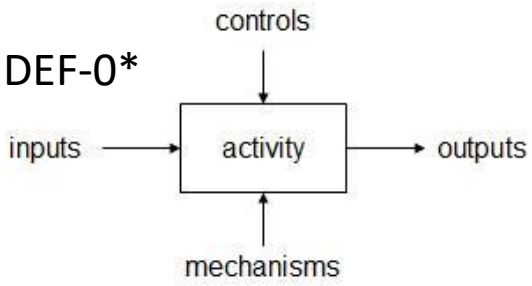


General Systems Theory

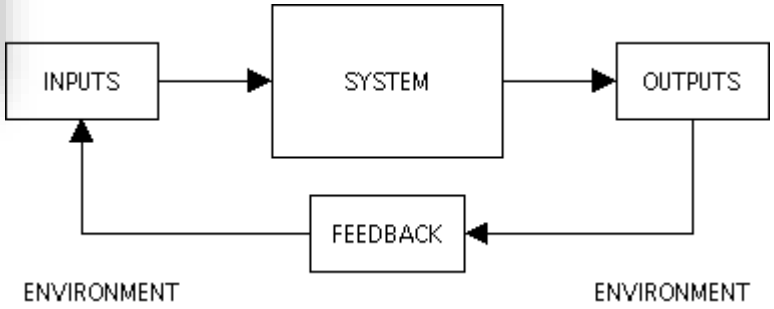


Karl Ludwig von Bertalanffy was an Austrian-born biologist known as one of the founders of general systems theory (GST).

IDEF-0*

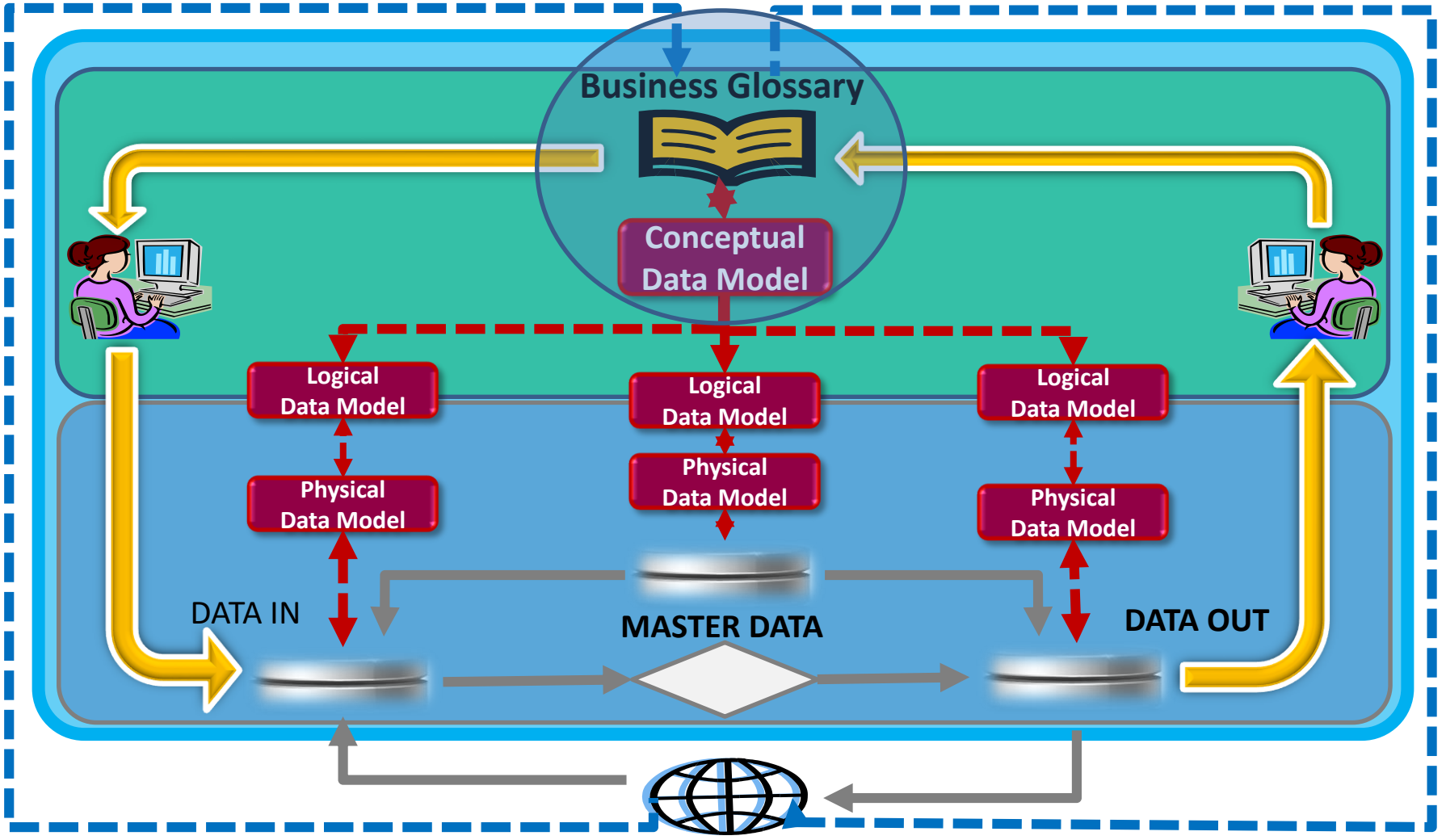


*ICAM Definition, Integration DEFinition
Integrated Computer-Aided Manufacturing



SANDHILL

Enterprise Feedback



SANDHILL

What is a Critical Data Element ?

Competing with High Quality Data: Concepts, Tools, and Techniques for Building a Successful Approach to Data Quality

by Rajesh Jugulum

July 2014

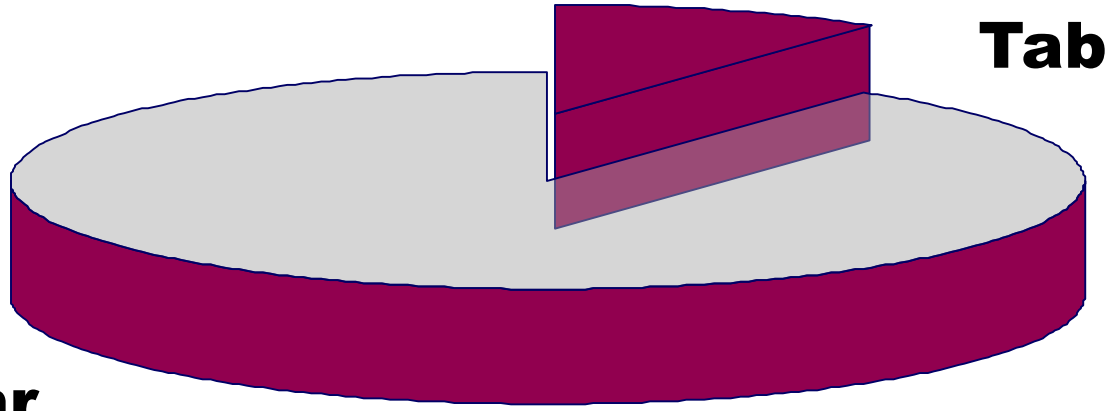
- Critical data elements (CDEs) are defined as “the data that is critical to success” in a specific business area (line of business, shared service, or group function), or “the data required to get the job done.”

SANDHILL

Data Megatypes*

Tabular Data (Structured):

Formatted data typically held as rows and columns in a table or database



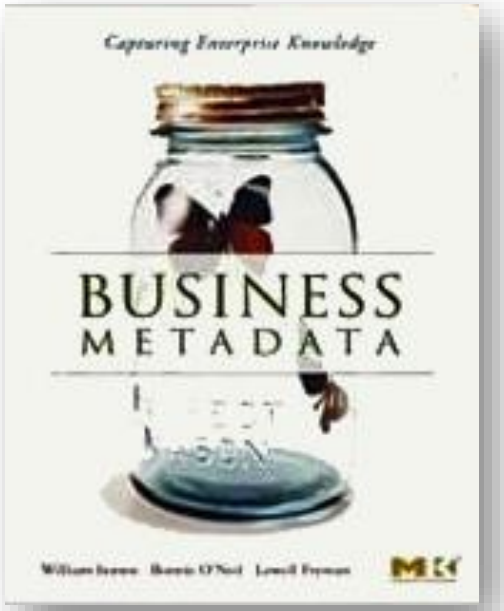
Non-Tabular Data (Unstructured):

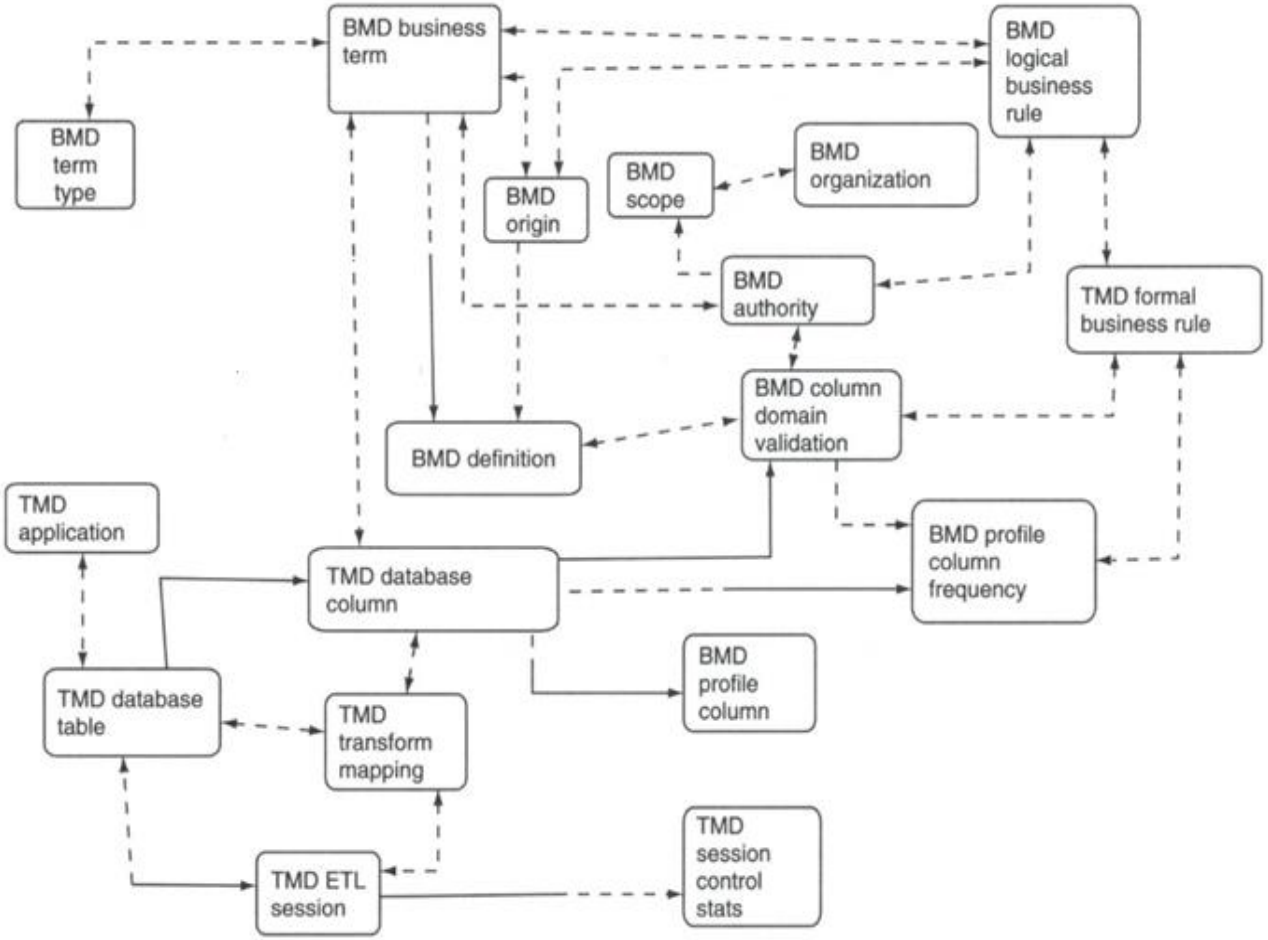
Data which has free form such as text, images, multimedia

* Michael H. Brackett

SANDHILL

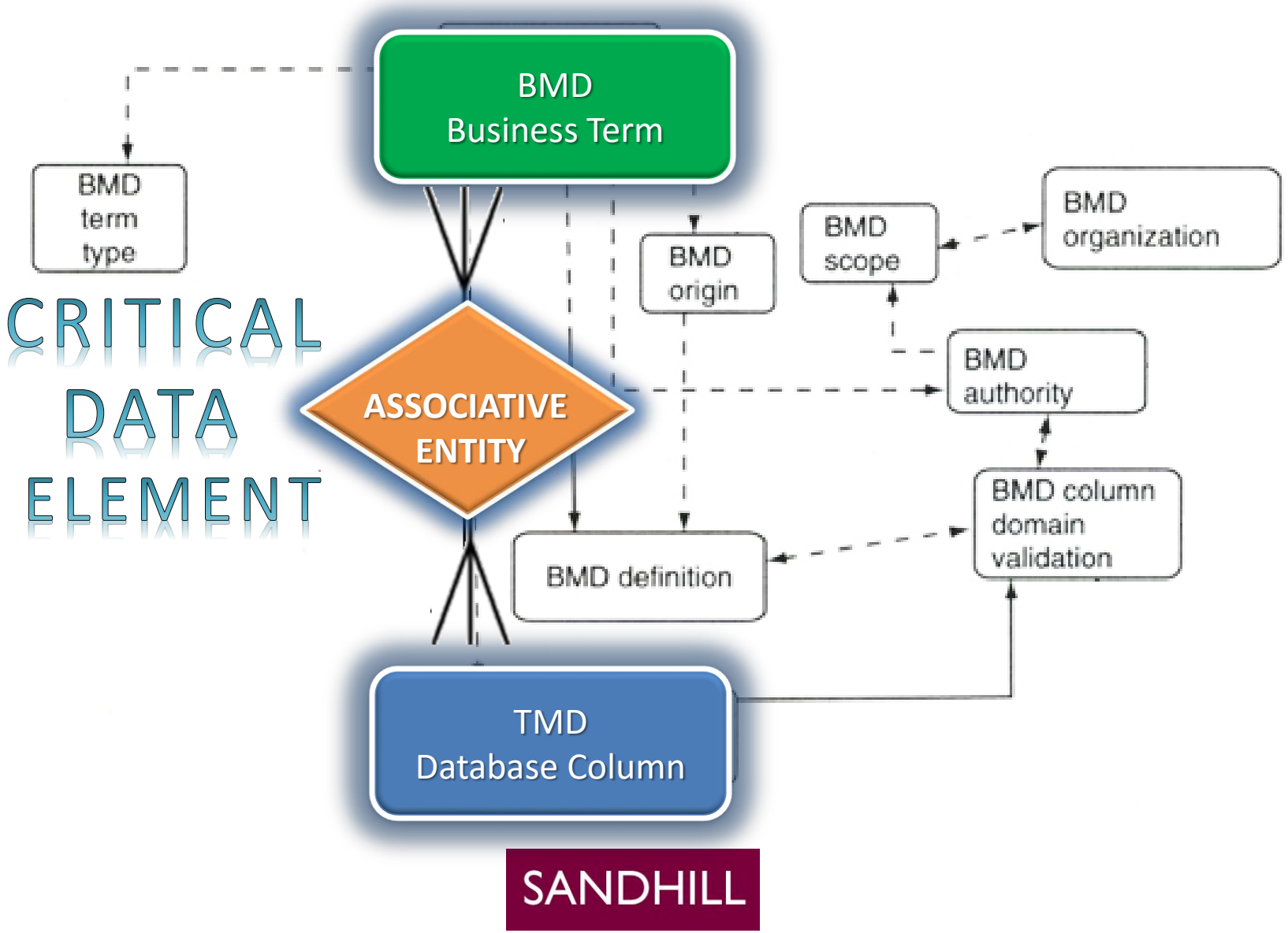
Business Metadata – Inmon, O'Neil, Fryman





SANDHILL

Business & Technical Metadata

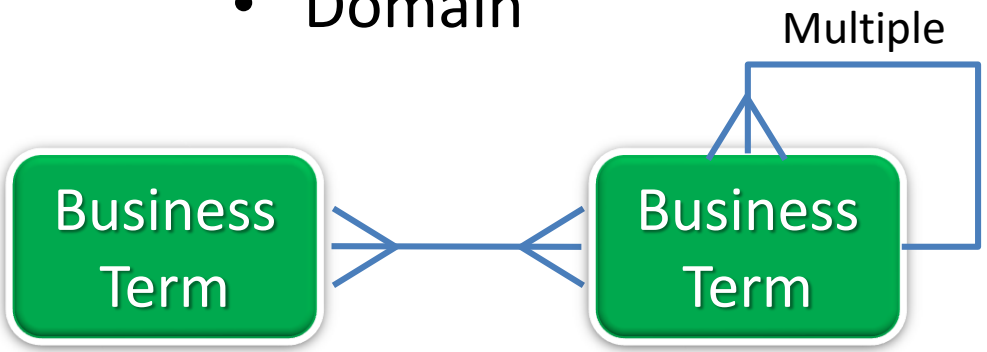


SANDHILL

Critical Data Element

- Business Term
  - Name
  - Description
  - Workflow Status
  - Abbreviation
  - Acronym
  - Synonym
- Term Type
  - Topic
  - Term
  - Qualified Term

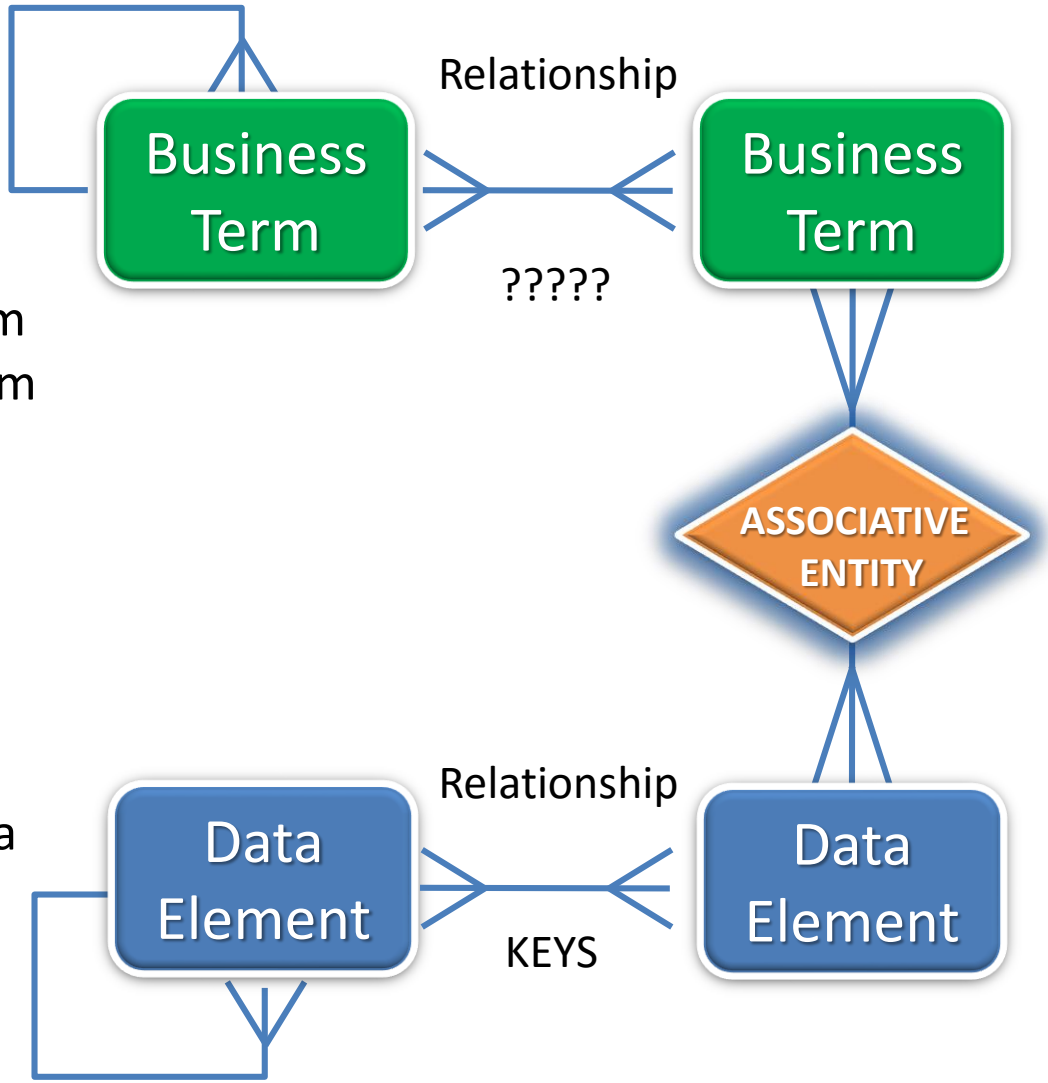
- Origin
- Business Rule
- Authority
- Scope
- Organization
- Domain



SANDHILL

Glossary is not a Dictionary

- Business Concept or Category
- Business Term
- Qualified Term



- Subject Area
- Entity
- Attribute

It is a

- Dictionary
- Thesaurus
- Model

SANDHILL

Semantic Relationships…

- is defined for
- is defined for a
- is due to
- is held by
- is issued by
- is organization unit of
- is paid via
- is rated upon
- is recorded for a
- is responsibility of
- mandates
- may be applied to
- executes
- has gender of
- has related
- identifies location of
- identifies marital status of
- identifies shipment terms associated with
- is consolidated upon
- is of status type
- is result of
- is segment of
- is subdivision of
- is used for
- rates
- applies
- belongs to
- collateralizes
- converts
- focuses upon
- has name preference of
- has preference
- has relationship with
- holds title of
- identifies language processing associated with
- involves
- is affiliated with
- is charged against
- is consolidated in
- is employed by
- is generated by
- is included in
- is measured by
- is placed by
- is prepared upon completion of
- is priced upon
- is related to and based upon planned projections of
- is reported by
- is shipment method of
- is stored as
- measures audience duplication with
- participates in
- performs
- receives
- refers to employment position
- responds to
- adjusts
- characterizes
- has accounting period of
- has credit status of
- has identifying NAICS
- has identifying SIC
- has legal form identified by
- has marital status of
- has payment status of
- has placed
- has processing status of
- has relationship defined by
- holds
- identifies credit status of
- identifies interests of

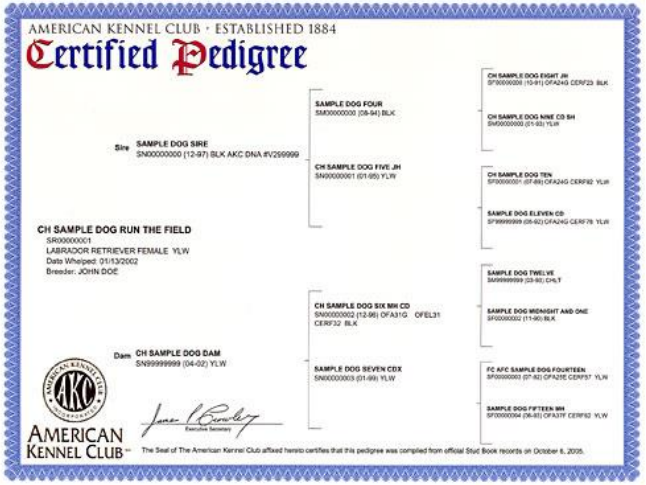
SANDHILL

Data Model Design Layers

Logical Model	**Physical Model**
Entity	Table
Attribute	Column
Key	Index
Relationship	Constraint

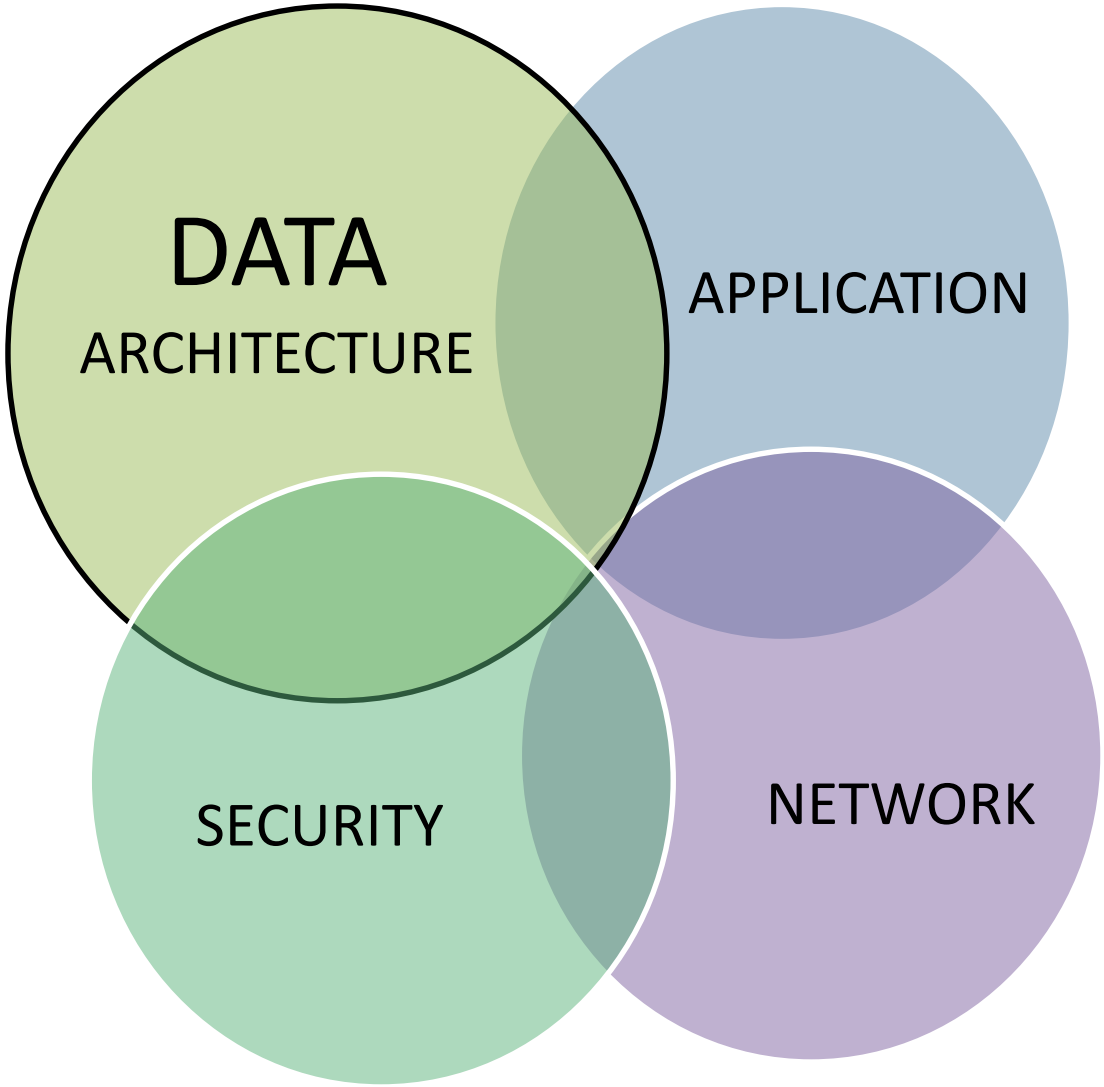
SANDHILL

Critical Data Element - Risk



SANDHILL

Data Architecture in EA



SANDHILL

Metadata Integration

The integration of business and technical metadata has become the holy grail of Data Governance.

"But just like the holy grail, it seems to be a never-ending quest fueled by the belief that technology is going to solve all our problems."

Rick Sherman
DM Review Magazine

SANDHILL

Tibetan Proverb

水头 不清水尾 浑

“ If upstream is dirty,
downstream will be muddy”

Thank You

[shuǐtóu] [bù] [qīngshuǐ] [wěi] [hún]

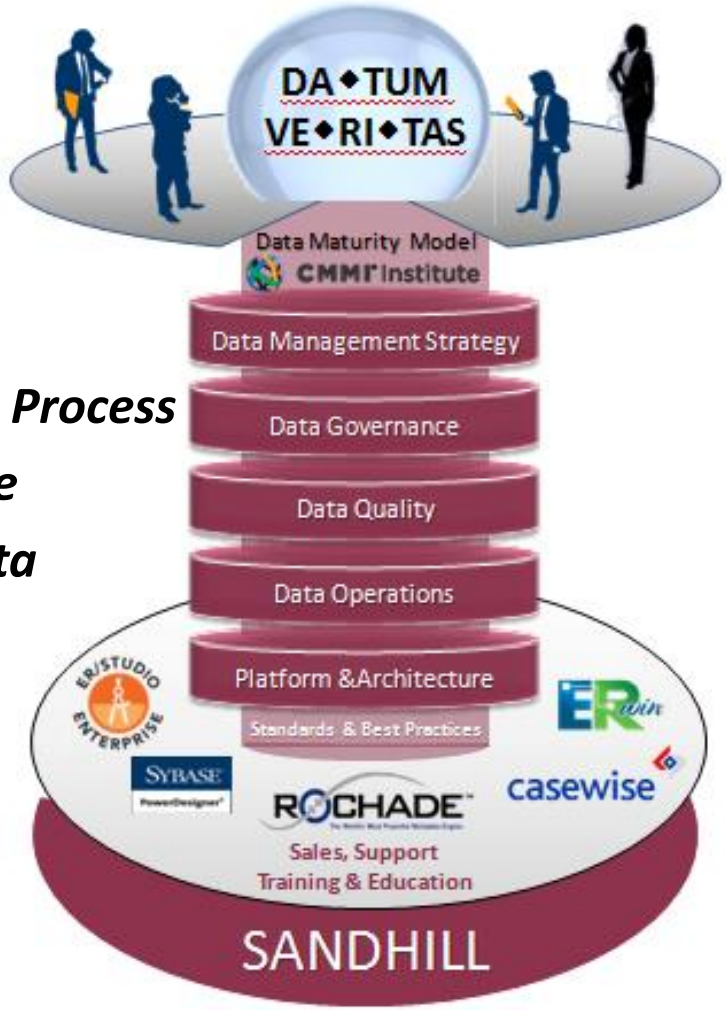
SANDHILL

Sandhill Consultants:

Sandhill Consultants

Is a group of people, products and processes that help our clients:

- Build comprehensive ***Data Architectures***
- Resulting from a persistent ***Data Management Process***
- Founded on a robust ***Data Governance Practice***
- With Da'tum Ver'it'as – ***Trusted, Verifiable Data***



SANDHILL