Workshop:
Practice of CRM-Based Data Integration

George Bruseker, Achille Felicetti, Mark Fichtner, Franco Niccolluci

CIDOC 2017

Tblisi, Georgia

25/09/2017

George Bruseker

- Data Mapping
- Philosophy
- Data structure
- Ontology & Semantics
- Data Provenance

Achille Felicetti

- Semantics
- Archaeology
- CRMArchaeo
- Epigraphy
- NLP Tools

Mark Fichtner

- Data Integration
- Design
- Data Modelling
- RDF/OWL
- Drupal

Franco Niccolucci

- Digital Infrastructures
- DCH Policy and Strategy
- Archaeology
- CRMArchaeo and CRMba

Schedule

| Time Slot	Topic	Facilitator
10:00 – 10:20	Introduction to Data integration and Semantics	George Bruseker
10:20 – 11:30	CRM Basics Introduction and Exercises	George Bruseker
11:30 – 12:00	**COFFEE**	
12:00 – 13:00	CRMArchaeo Introduction and Exercises	Achille Felicetti & Franco Nicolucci
13:00 – 14:00	**LUNCH**	
14:00 – 15:30	Semantic Mapping Techniques + 3M Tool Intro	George Bruseker & Achille Felicetti
15:30 – 16:00	**COFFEE**	
16:00 – 16:45	Practical Exercises with 3M Mapping Tool	George Bruseker & Achille Felicetti
16:45 – 18:00	Demonstration of Data Use in Wisski and ResearchSpace Platforms	Achille Felicetti & Mark Fichtner

Introduction to Data Integration and Semantics
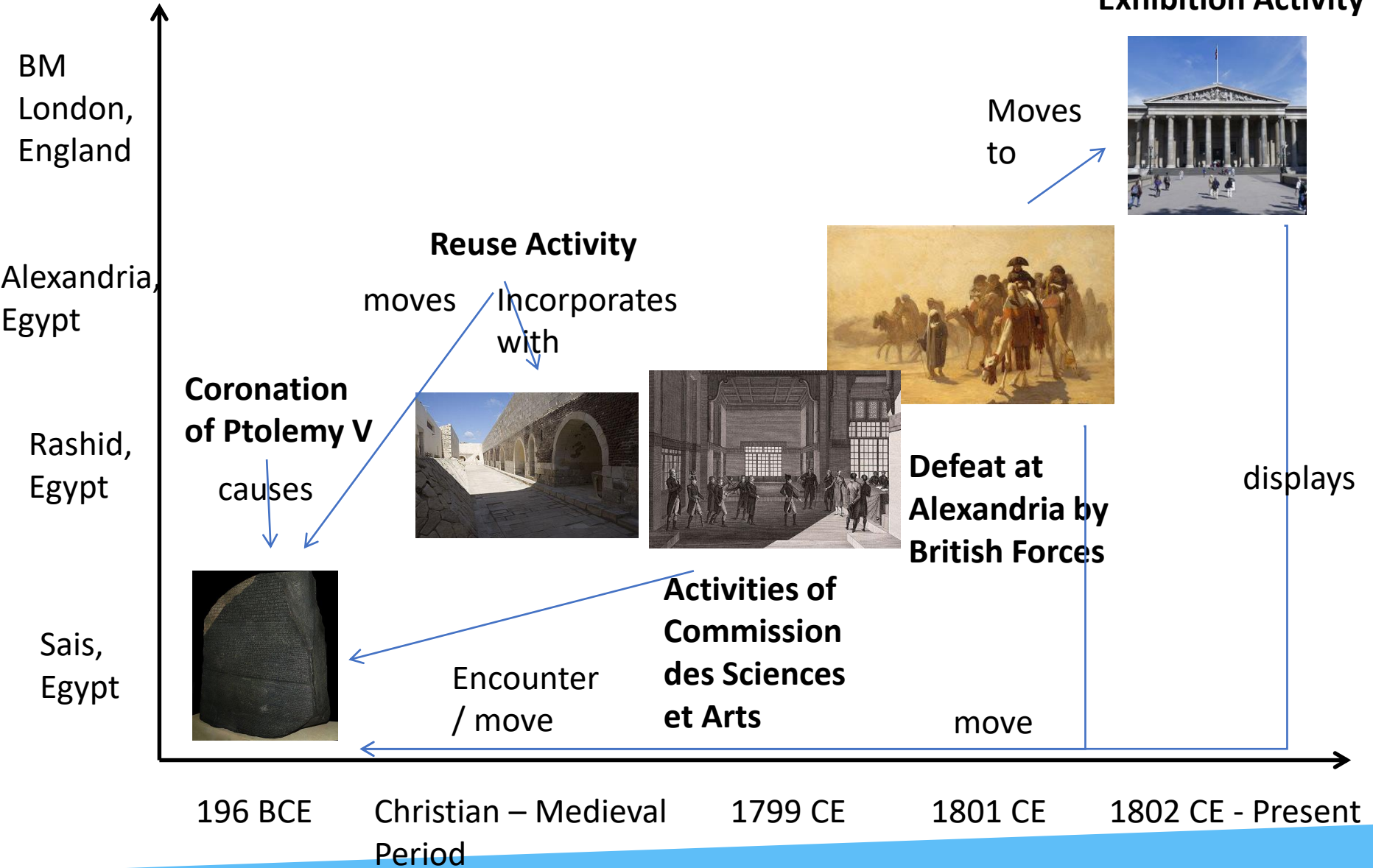
George Bruseker
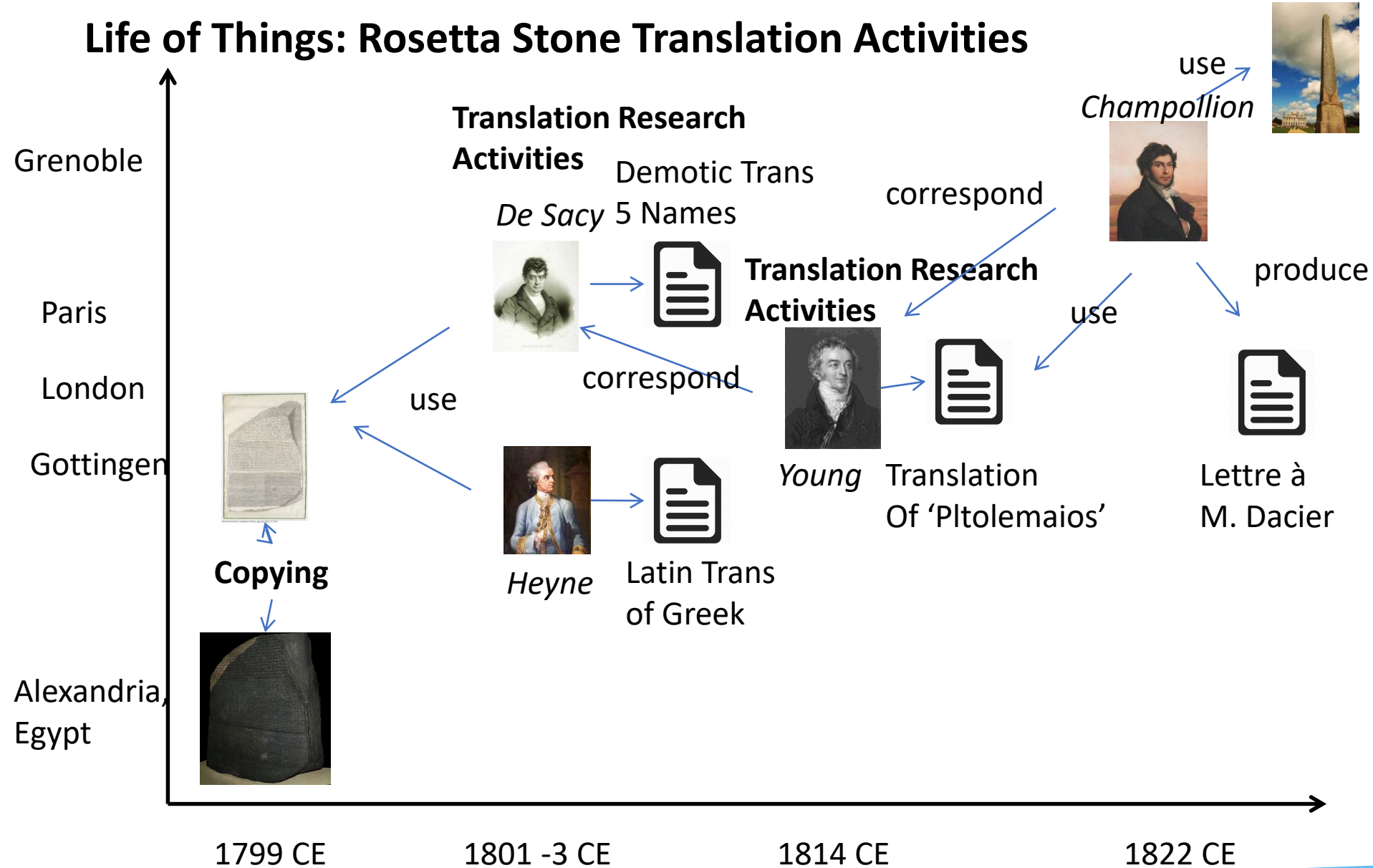
CIDOC 2017

Tblisi, Georgia

25/09/2017
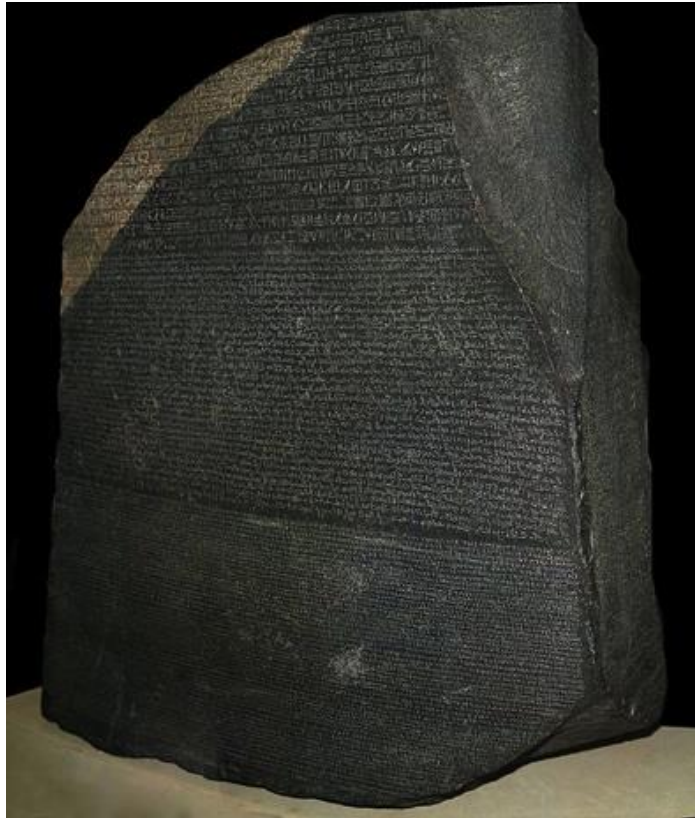
A Global Information
Riddle
Continued

Life of Things: Rosetta Stone Macro Movements

Exhibition Activity

BM London, England

Moves to

Reuse Activity

Alexandria, Egypt

moves   Incorporates with

Coronation of Ptolemy V

Rashid, Egypt

causes

Defeat at Alexandria by British Forces

displays

Sais, Egypt

Activities of Commission des Sciences et Arts

Encounter / move

move

| 196 BCE | Christian – Medieval Period | 1799 CE | 1801 CE | 1802 CE - Present |

Life of Things: Rosetta Stone Translation Activities

Intra-Institutional Information Diversity in Memory Institutions

Spectrum 4.0

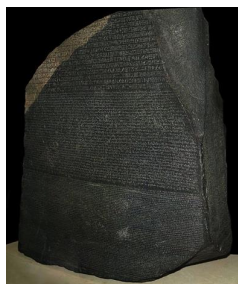**Registrar Team**

Exhibitions Team

Conservation Team

MARC 21

<ead>

LIbrary

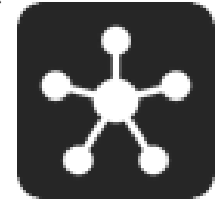**Curators & Researchers**

Archival Team

Inter-Institutional Information Universe

Projects

Partners

Aggregators

Museums, Library, Archives
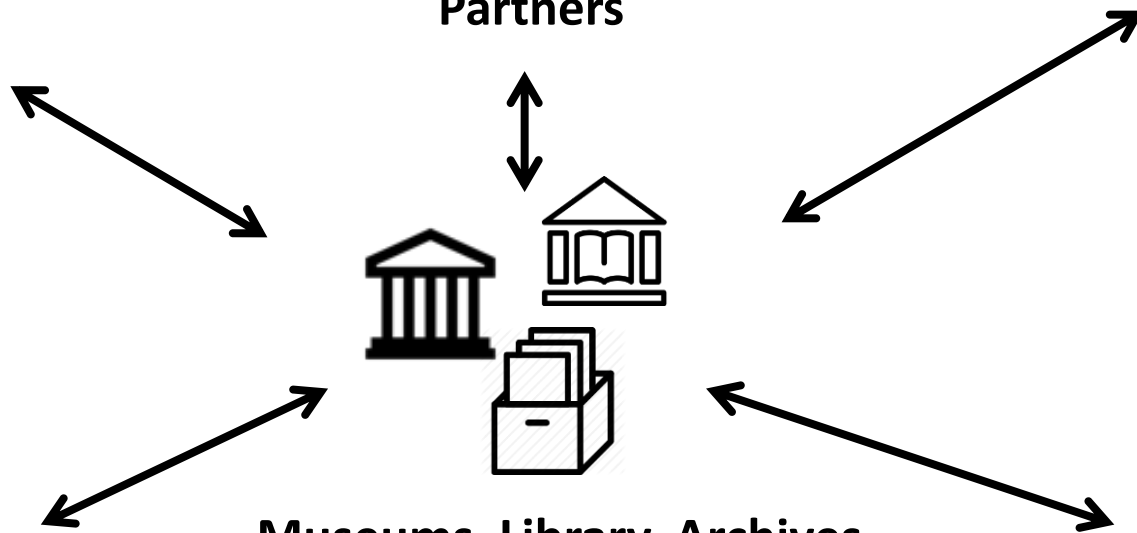
Businesses

Standards Institutions

CHRISTIE'S

europeana

Increased knowledge?
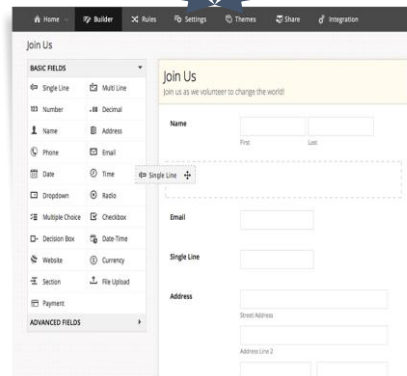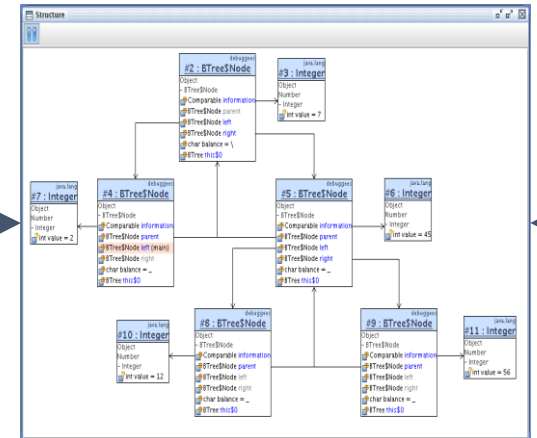
Specialists
(stuck?)
in their worlds

observe

publish

ground

input

Reach /
understand
/ critique

store

The CH Information Integration Challenge

- There is a common interest and a common empirical world under discussion

- Nevertheless, different factors intercede to create a disparate information landscape including:

  - scientific approach/discipline
  - tools
  - standards
  - Means

- Heterogeneity is not a a problem to be solved (in the sense of reduction); it is, rather, constitutional to our understanding.

- Still we need a means to present the big picture and to pass information from one specialist to another.
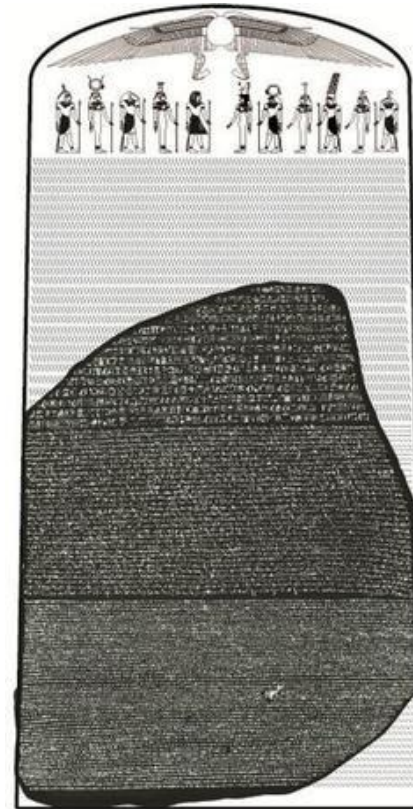
We need: a different approach

How to create successful integrations?

Schema Standard

- Useful in a limited context where agreement can be had amongst all actors or enforced

- Useful for more limited contexts, where domain is relatively standardized and closed.

Semantics-Ontology

- Useful when data to be covered is heterogenous

- Useful when teams are dispersed

- Useful when tools are necessarily diverse

- Can/should be combined with standards

What is an ontology? (What is it not?)
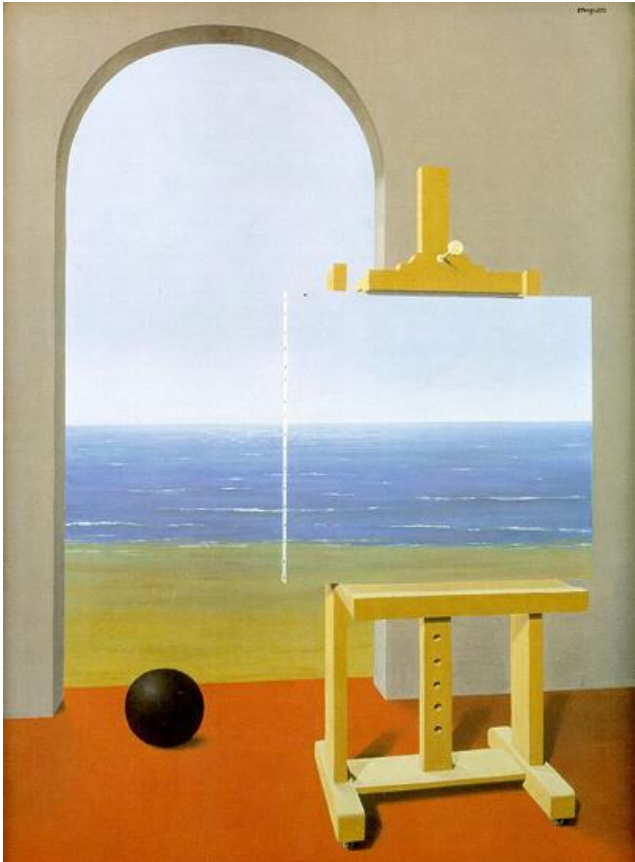
- A window / frame which allows expression of the world according to the kinds of statements used by a domain(s)

- Not a description of the world as such
(not Ontology)

- Plural not singular

- Heidegger's 'domain ontology'

What is an ontology? (What is it not?)

- Mix of computer science and philosophy
- Bridge between computer science world and data producers/researchers
- Formalization of knowledge domain
- Bound to a reality
- Creates a *lingua franca*
- Machine Processable, Human Readable

"A Universal Symbolism, very popular, very agreeable to the people...might be introduced if small figures were employed in the place of words, which would represent visible things by their lines, and the invisible, by the visible, which accompany them."
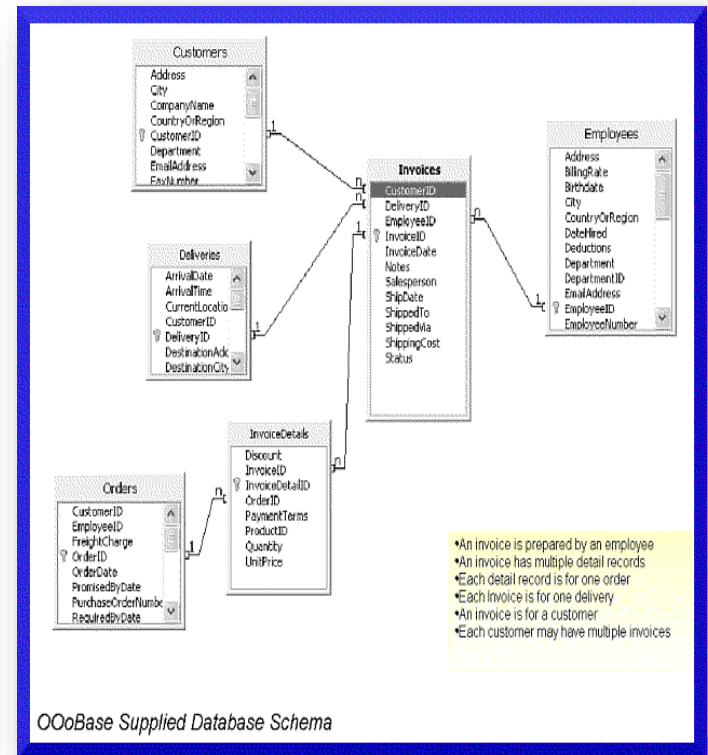
— Leibniz, 1679

Gottfried Wilhelm Leibniz.

What kind of data can a formal ontology produce?

Triples (Subject - Verb - Object)

VS.
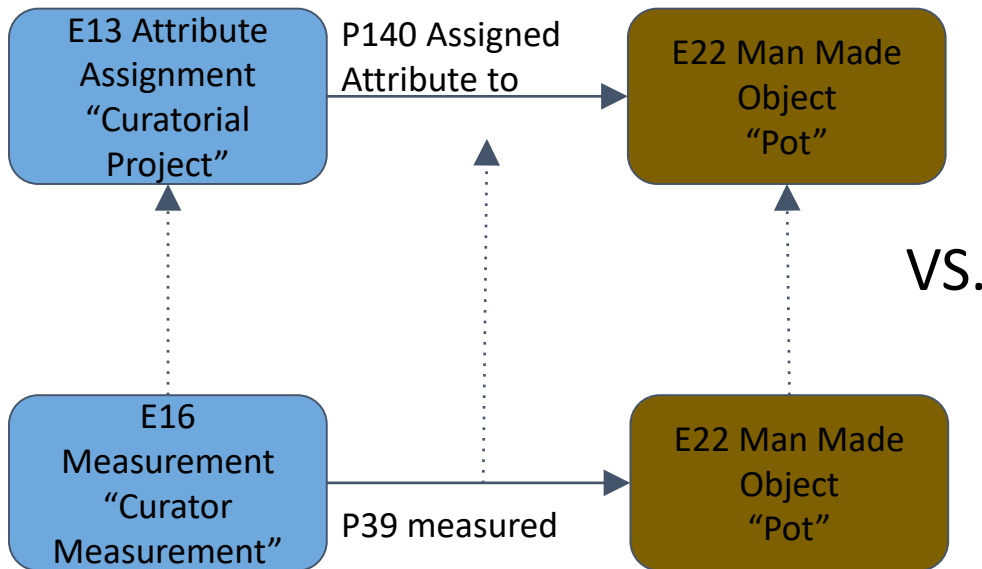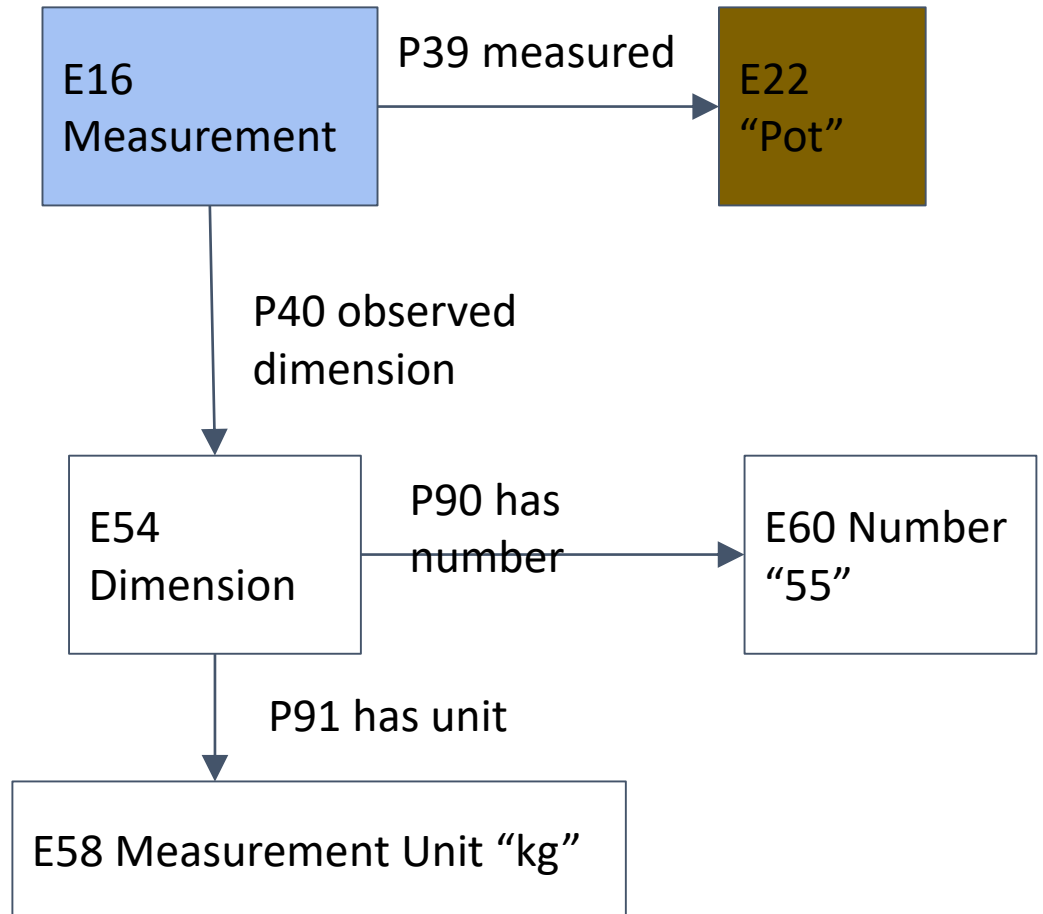
In a graph, supporting subsumption relation
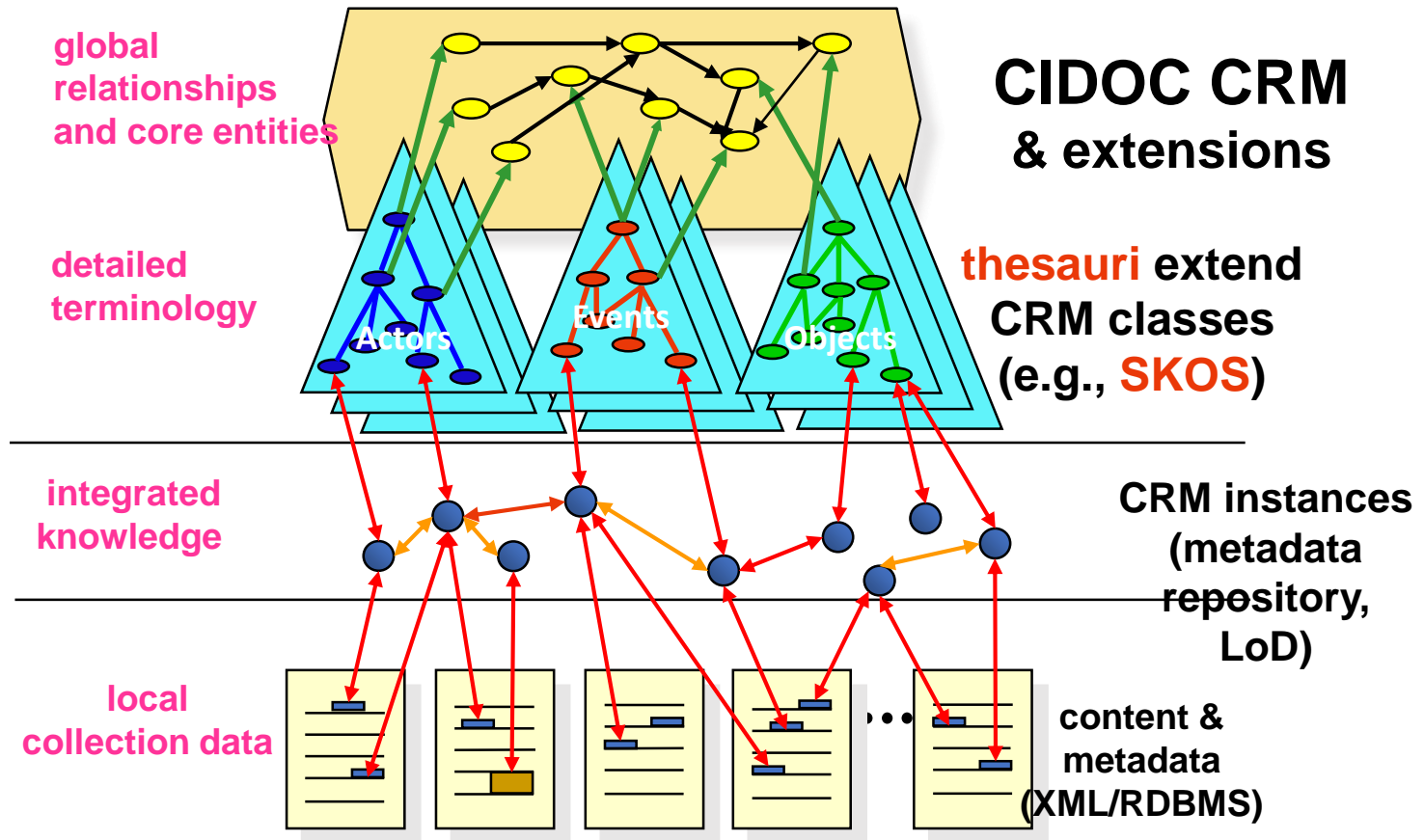
Traditional RDBMS

What kind of data can a formal ontology produce?

Object Table	
ID	101
Name	Pot
Weight (kg)	55

Rendering the implicit, explicit

E16 Measurement — P39 measured → E22 "Pot"

E16 Measurement — P40 observed dimension → E54 Dimension

E54 Dimension — P90 has number → E60 Number "55"

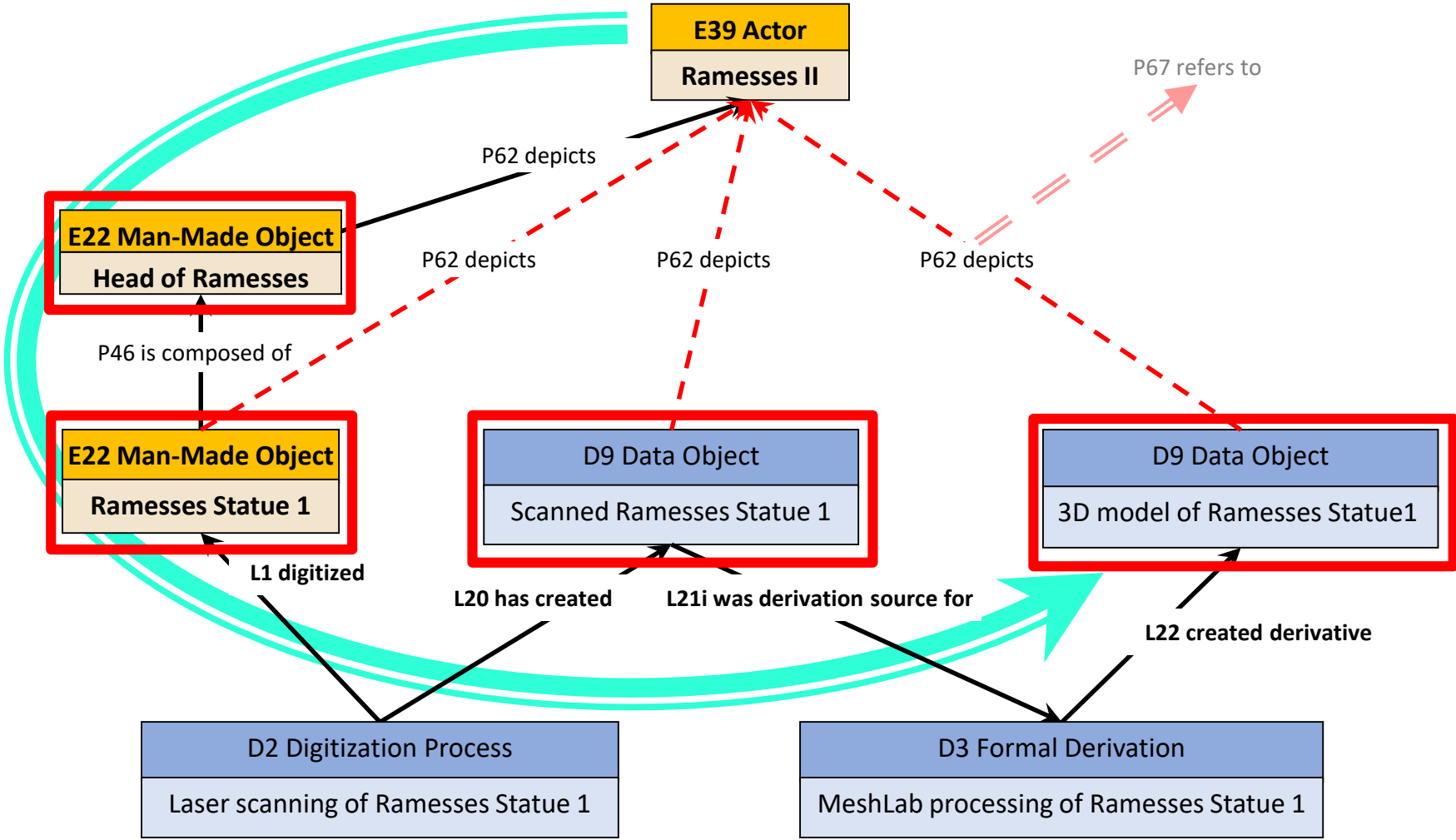E54 Dimension — P91 has unit → E58 Measurement Unit "kg"
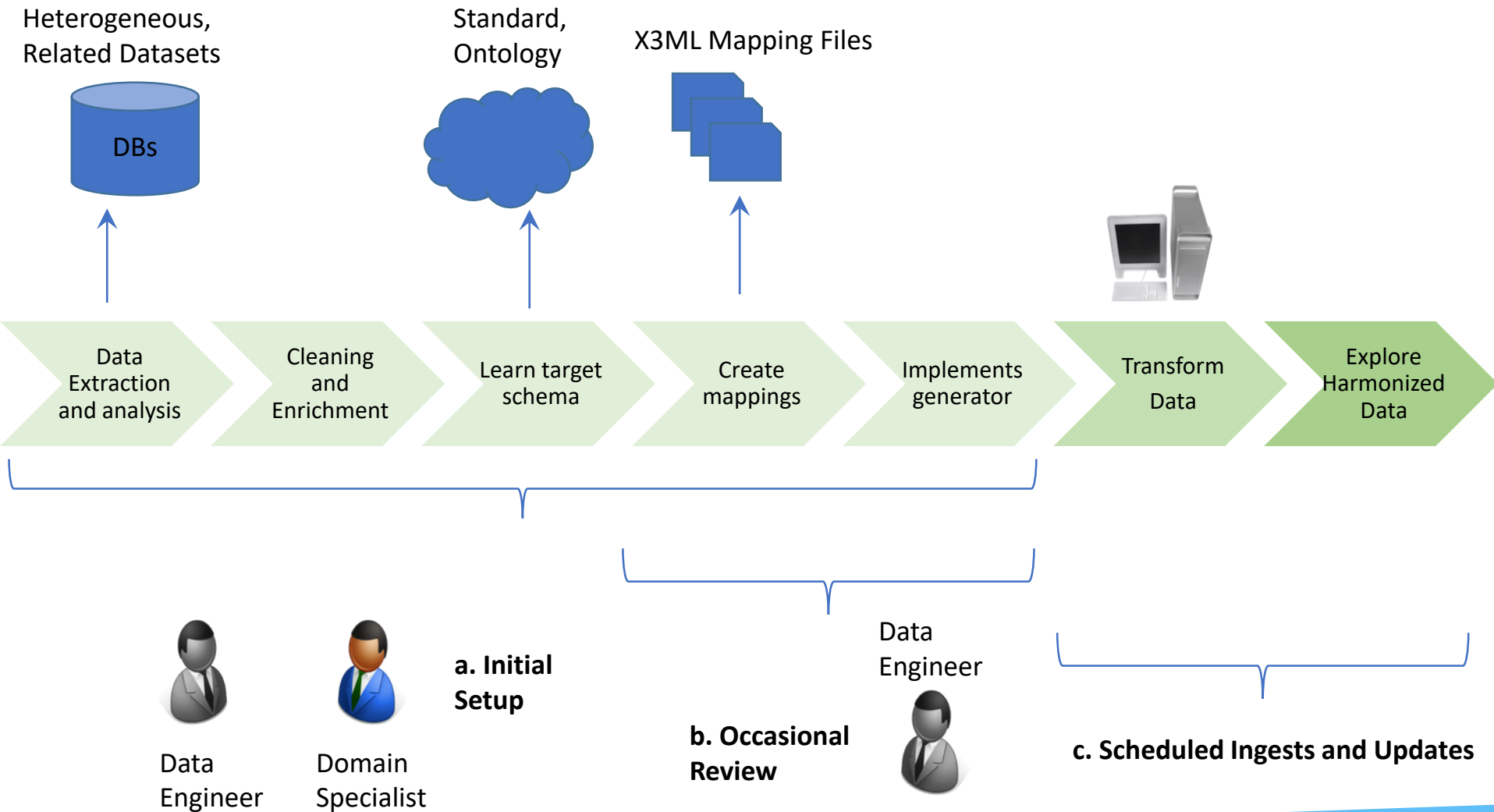
Integrated Data Picture

- Support interoperability of mutually relevant data sources
- enable sourced and verifiable facts from datasets
- foster referenceability and reusability of data
- foster structured argumentation on top of facts

Query: FIND things (?) that refer to Ramesses II

Workflow

Heterogeneous,
Related Datasets

Standard,
Ontology

X3ML Mapping Files

DBs

| Data Extraction and analysis | Cleaning and Enrichment | Learn target schema | Create mappings | Implements generator | Transform Data | Explore Harmonized Data |

Data Engineer   Domain Specialist

a. Initial Setup

b. Occasional Review

Data Engineer

c. Scheduled Ingests and Updates

Workshop Take-Aways

1) Understand the Process and the Goal of semantic data integration
2) Familiarization with CIDOC CRM and CRMarchaeo models
3) Understanding of basic mapping techniques
4) Familiarization with 3M Mapping Tool
5) Familiarization with ResearchSpace and Wisski Platform functionalities