Formal Ontologies:

Seeking Theoretical Foundations

Martin Doerr

in collaboration with Maria Daskalaki

Center for Cultural Informatics,
Institute of Computer Science
Foundation for Research and Technology - Hellas

Heraklion, Crete, September 24, 2015

Overview

- Our Background
- Engineering Goals
- Formal Ontologies in Computer Science today
- Our Approach
- Provisional Foundational Positions & Open Problems
- Conclusion

Foundations of Ontologies

Our Background

The Centre for Cultural Informatics at ICS_FORTH

- Deals with information systems for cultural-historical knowledge and science data since 1992
- Applications in various museums, archives, research infrastructures
- Focus on information modelling and information integration by learning and interpreting the epistemological process of various disciplines:
  - *museum disciplines, archaeology, art conservation, archiving, library science, clinical studies, analytical sciences, geology, biodiversity*
  - *critical-empirical investigation of data created by domain experts and their explanation of data fields and data examples*
  - *master degree postgraduate courses in archaeology and computing*
- Our basic research: core ontology for empirical sciences and realistic argumentation processes / support of knowledge revision

Foundations of Ontologies

Engineering Goals

Main Results:

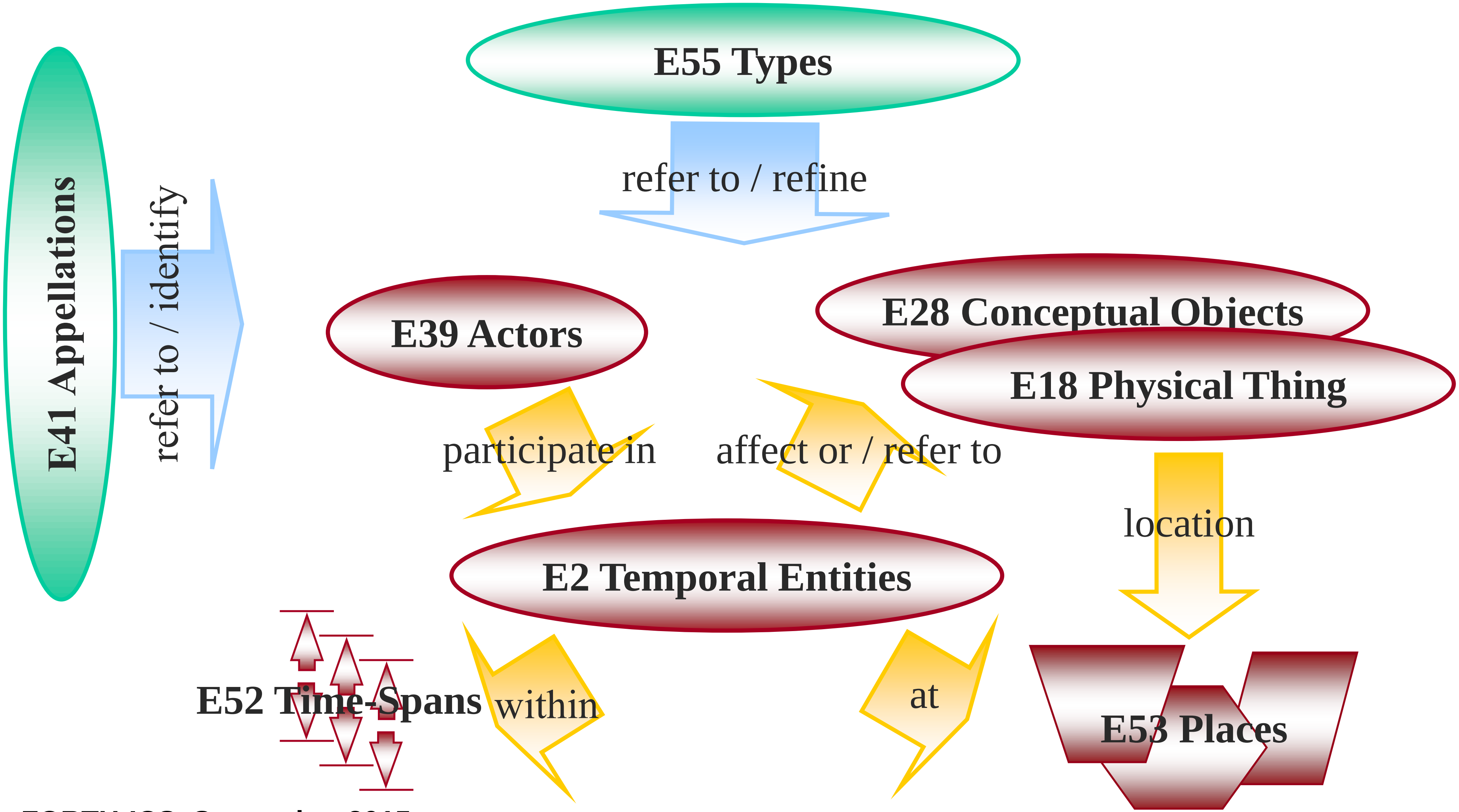
- CIDOC CRM (ISO21127), a formal ontology for global cultural-historical data integration, continuously being extended
- increasingly taken up in European funded research infrastructures and by private clients for globally aggregating large amounts of facts (e.g., British Museum, Getty Research Institute, Germanic National Museum)

Engineering Problem

- Hundreds(?) of domain experts have to define a formal translation of their data structures to the global ontology ("mapping”).
- Hundreds of experts have to learn the CRM, learn why a CRM concept is a good match, when a new concept has to be added, and what makes a good new concept for information integration.
- Teaching philosophical choices as practical guidelines ….have we understood the choices?

Foundations of Ontologies

Top-level classes useful for integration



Class example: E2 Temporal Entity

E2 Temporal Entity

Scope Note:

This class comprises all phenomena, such as the instances of E4 Periods, E5 Events and states, which happen over a limited extent in time.

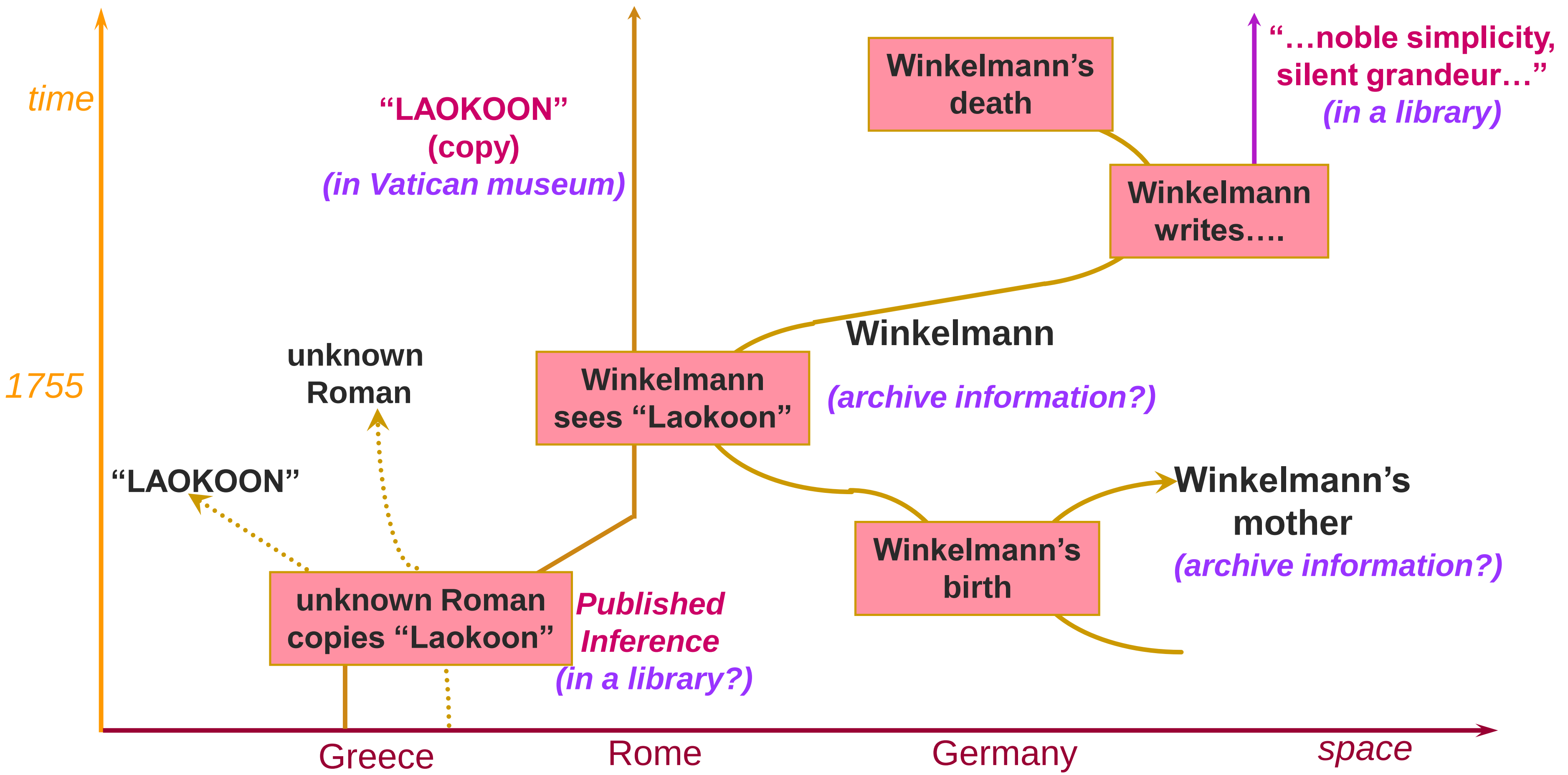
In some contexts, these are also called perdurants. This class is disjoint from E77 Persistent Item. This is an abstract class and has no direct instances. E2 Temporal Entity is specialized into E4 Period, which applies to a particular geographic area (defined with a greater or lesser degree of precision), and E3 Condition State, which applies to instances of E18 Physical Thing.

Temporal Entity-Main Properties

- E2 Temporal Entity
  - Properties: P4 has time-span (is time-span of): E52 Time-Span
- E4 Period *(IsA E2)*
  - Properties: P7 took place at (witnessed): E53 Place
    P9 consists of (forms part of): E4 Period
    P10 falls within (contains): E4 Period
- E5 Event *(IsA E4)*
  - Properties: P12 occurred in the presence of (was present at): E77 Persistent Item
    P11 had participant (participated in): E39 Actor
- E7 Activity *(IsA E5)*
  - Properties: P14 carried out by (performed): E39 Actor
    P20 had specific purpose (was purpose of): E5 Event
    P21 had general purpose (was purpose of): E55 Type
    P16 used specific object (was used for): E70 Thing
    P125 used object of type (was type of object used in) E55 Type

Foundations of Ontologies

History as a Connected Event Graph



Foundations of Ontologies

Formal Ontologies in Computer Science

Formal ontologies were introduced in computer science in the 1990'ies

- As a means overcome idiosyncratic database design preventing data exchange and to integrate data of the same domain under a common schema
- …by formulating the shared "domain conceptualization" of experts as a logical theory (conceptual model, knowledge representation schema, "T-Box") using a "vocabulary" of entity and relation concepts to describe "possible states of affairs" in a "domain".
- The "domain" is taken as a set of already identified, well-distinguished objects. Expert conceptualizations are taken as "intuitions", logical constraints on relations between objects should yield "unambiguous concepts" and unambiguous communication.
  - *"mother" = "human" & "female" & "has_child"*
  - *For instance, B.Smith defines "has part" by transitivity and extensionality laws.*

The Formal Ontologies Crisis

Around 2000, computer science recognized the need of methodologies to create good ontologies

- Many processes were described that should effectively produce good ontologies, top-down or iterative. They turned out to be "intuitive".
- Objective quality metrics were sought.
- Computer science could not find engineering methods acceptable to the discipline's self-understanding. The experts' understanding of their concepts is taken as "**black box**" (A.Gangemi 2006)
- Literature on good practice of ontology engineering disappeared after 2006

Current domain convictions:

- Everybody may have his own ontology (we are back in the 1980'ies).
- Now any aggregate of universals (schema, terms) may be called "ontology"

how people understand each other remains a great riddle…

Foundations of Ontologies

Our Approach

In contrast, we follow an empirical method of verifying the adequacy of the ontology:

- We observe how experts relate their concepts to real situations.
- We ask experts
  - *how instances of their concepts "behave" (e.g., can two museum objects become one? Can you take a photo of a "place"? After 100 years, is it the same place?)*
  - *when they are ambiguous (e.g., is WWII an event or a period?)*
  - *what are the exceptions (e.g., birthplace of baby born in airplane)*
- Comparing reality and expert answers we propose to the expert modifications of their concepts that are logically consistent and consistent with their observations and reasoning and the precision they expect up to which the ontology can characterize and predict reality *(e.g., begin of existence)*
- We test how widely these models are adequate across disciplines.

We model a part of reality that can be enclosed in discrete human concepts.

I present here

- an (intuitive) proposal of philosophical positions and questions
- with direct bearing to justify our (intuitive) methods
- and to produce teachable quality criteria for our “formal” ontologies

My questions:

- Are these positions known in philosophy?
- Are they questioned, and if yes, why?
- Can we reject the questioning with objective arguments?
- Can we improve and justify positions relevant for our task?
- Can we formulate these positions in a way that people can learn effective ontology engineering from us?
-

Foundations of Ontologies

Position 1: Knowledge and Information

Information is arrangements of symbols encoding propositions from a sender to a target audience.

Knowledge in the sense ("I know that X") is "justified belief". It is propositions with symbols that are resolved to reality and a trust value.

The way of justification is a question of scientific method.

Only humans can resolve symbols to real things (so far). They can "know a thing" (second sense of "knowing"). Therefore only humans can know.

Information can induce knowledge in humans by relating the known to the unknown. Knowledge of humans can be tested by asking questions and/or by provoking reactions.

Position 2: Reality

Reality acts and reacts independently from the intention of observer. The observer is part of the reality and perceives by interacting.

There is only one reality. It is that which provides the potential to make observations by different individuals comparable.

Reality is ultimately interconnected. Nothing exists in isolation. There are no discrete objects with precise boundaries in the mathematical sense.

Some phenomena exhibit distinctness. Their discreteness is relative to the persistence of a combination of properties within some limits, e.g., my body alive and limited in time.

By virtue of being one reality, different definitions of distinct/discrete things imply objective relations between things such defined.

Foundations of Ontologies

Is this one tree or two?

Position 3: Knowledge and Reality

Knowledge in the sense ("I know that X") is justified belief. Justification describes the way how knowledge relates to reality.

Some concepts can be seen as recipes to turn sensory impressions into discrete propositions.

Reality has the potential to reveal the same or comparable factual/categorical phenomena to multiple people by observation, i.e., independent behavior from the observers' intentions.

The expectation of truth in our knowledge comes from the trust in justification and independent observation. It is a likelihood of being true, and a method to learn better.

Truth becomes a continued process of trying to "know better".

(likelihood is not a concession of ignorance!)

Position 4: Relativity of Concepts

A concept must be useful for something, it must have a function in a discourse, pursuit and/or survival.

A concept has an "intension", a sort of recipe to determine instances

From instances of a concept we expect a "behavior", i.e., potential or necessary properties (a person has a friend, a mountain is in my way, a car may kill me, a house protects from rain).

The function of the concept is the ability to conclude from intension to potential.

A concept is "good", when it constrains well potential properties of interest to a subset of reality.

Position 4: Relativity of Concepts

Therefore, concepts depend on intended use/function.

Examples:

- **Human**: Tut Ankh Amun does not exist any more, he died in 1323BC. Tut Ankh Amun is in the Cairo museum. (continuity of life function versus body substance. Proof of cause of death on the mummy).
- **Bottle**: When does a bottle for urine sample stop to exist, and when a bottle of sample urine? (six different urine samples Armstrong provided during the 1999 Tour tested positive for the performance-enhancing drug EPO when examined in 2004 by a French lab fine-tuning EPO testing.)

This is neither subjective nor arbitrary. It is functional.

Therefore, we do not ask domain experts, "What is a bottle?" But "what do you do with these bottles".

Position 5: “Anatomy of a Concept”

It appears to us that a concept (of real phenomena?) is determined by

- substance
- identity conditions (same, one or two?),
- conditions for coming into/ ending existence,
- unity conditions (when is something a part of an instance)?
- Then the expected potential is a question of experience. The definition will be modified, until experience provides the best approximation of a potentiality of interest.

Is this correct? If yes, how do we define these conditions, do they form categories of their own?

Once properties/relations justify categories, they must be the real primitives of ontology, prior to categories. Then, what is their nature?

Foundations of Ontologies

Conclusions

Information Technology needs ontology as a means to communicate and manage knowledge at a global scale.

Computer Science has no epistemology for the adequacy of an ontology.

We propose an approach combining methods from empirical sciences with effective philosophical principles.

Only if such principles can effectively be justified against competing ones and result in a teachable method, science and scholarship has a chance to carry their notion of truth into "information age".