Ontolog: Database And Ontology Mini-Series

Data and process revisited: ontology driving a paradigm shift in the development of business application systems

Chris Partridge
BORO Centre – Chief Ontologist

Caveat – area of interest

- Large Operational Business Systems, e.g.
  - ERP Systems
  - CRM Systems
  - FX Settlement Systems
  - Trading Accounting Systems
  - Retail POS Systems
  - Telecom Billing Systems
- Semantic Web Applications
  - "Collective Knowledge" Systems
  - Social networking – Facebook
  - Wikis
  - Bio-medical dictionaries
- Inference
  - Logics – first o[illegible] description

Thesis

1. There is a common interest in 'what exists' (= ontology) in philosophy and business systems
2. The business systems community works within a paradigm for this that is not adequate for the development of modern large complex systems
3. A review of where ontology fits into the development process leads to a proposal for revising this paradigm
4. The revised paradigm leads to a reassessment of one of today's key IT problems – the re-development of legacy systems

Ontology: Philosophy and Business Systems

- Goal:
  - Establish that there is a common interest in what exists (= ontology).
  - Start with a brief history of ontology

A brief history of Ontology

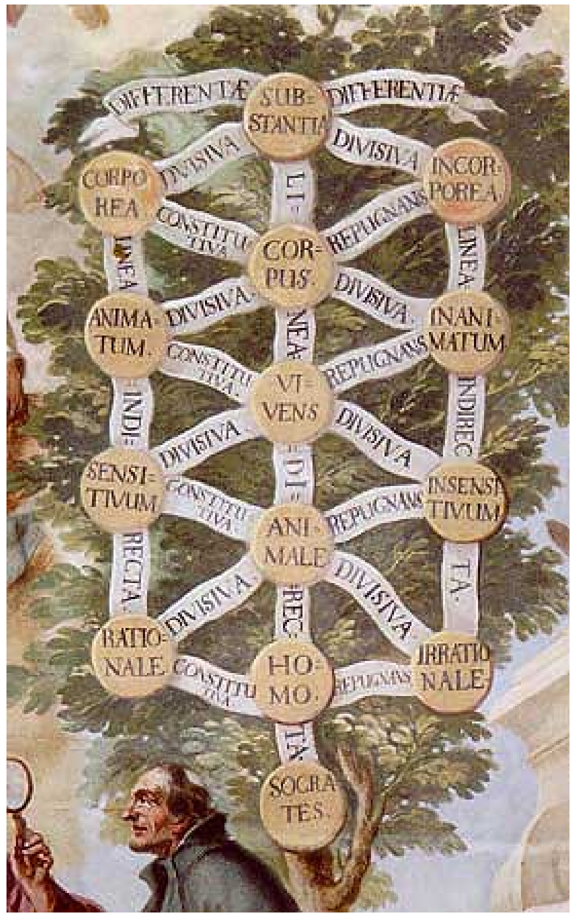
- History of the word
  - The word ontology is from the Greek ὄν, genitive ὄντος: of being (part. of εἶναι: to be) and -λογία: science, study, theory.
  - While the etymology is Greek, the oldest extant record of the word itself is the Latin form ontologia, which appeared in 1661, in the work *Ogdoas Scholastica* by Jacob Lorhard (Lorhardus) and in 1631 in the *Lexicon Philosophicum* by Rudolph Göckel (Goclenius). By this stage, it is regarded as forming the basic subject matter of metaphysics.

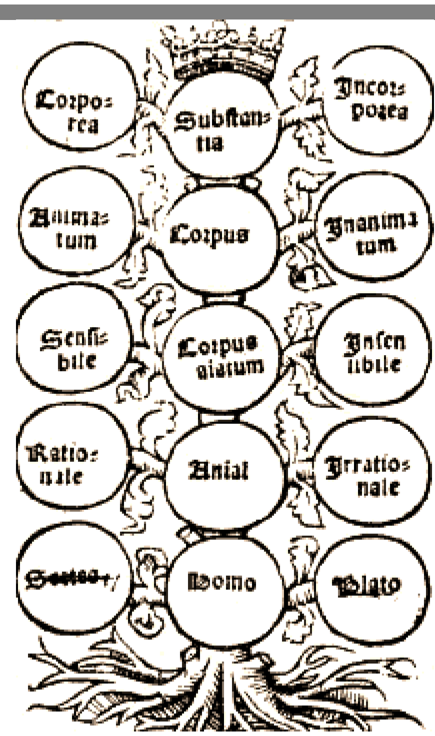
Ontology: The practice

- Origins
  - Ontology as a mode of analysis is generally thought to have originated in early Greece and occupied most famously Aristotle. , who created the first system of ontology in the form of an ontology of substances – represented pictorially in the tree of Porphyry.

- 20th Century views
  - Quine claimed that the question ontology asks can be stated in three words ‘What is there?’ – and the answer in one ‘everything’. Not only that, “everyone will accept this answer as true” though “there remains room for disagreement over cases.” (“On What There Is”).
  - Jonathon Lowe has a more technical definition - an ontology is "the set of things whose existence is acknowledged by a particular theory or system of thought." (E. J. Lowe, The Oxford Companion to Philosophy)

Tree of Porphyry





Business Systems Ontology

1. Regard business systems as ‘theories’ of their business domains (Naur 1985)
2. Recasting the philosophical description in these terms thus:
   - IS ontology: The set of things whose existence is acknowledged by a particular business system.
     - A common way of characterizing this ‘acknowledgment’ relationship is as one of ‘ontic commitment’. (Quine 1969).

Naur, P., "Programming as Theory Building," Microprocessing and Microprogramming, 15, (1985), 254-261.

Quine, W. V., Ontological relativity, and other essays, Columbia University Press, New York, 1969.

Ontology versus Ontological model

- An ontology is a set of objects whose existence is acknowledged by a particular business system.
- However, the relationship between these objects and the system may be quite convoluted
- So we, if we want to see the ontology, we need a model of it.
  - For this to be a good representation of the ontology is needs to directly reflect it.
  - There obvious method is a simple semantics where each object in the ontology has a direct relationship with the corresponding representation in the model. This is a well-established technique.
  - One way to think of an ontological model is that the representations are names of the objects in the ontology – from a Fregean perspective as reference and no sense (from a Millian perspective as denotation without connotation). In (Marcus, R.B., Modalities : philosophical essays. 1993), explicitly following in the footsteps of Mill and Russell this is called ‘tagging’. (I call this strong reference in my book *Business Objects: Re - Engineering for re - use* (1996).)
- There is a loose way of speaking where the ontological model is called an ontology. This can lead to use-mention confusion. However, so long as the context makes clear what it being referred to, there is no issue.

Business Systems - 20th Century views

- Connection with Ontology and ‘Reality’ recognised from the start
  - The issue is ontology, or the question of what exists. (Mealy 1967. p. 525)
  - For some time now my work has concerned the representation of information in computers. The work has involved such things as file organizations, indexes, hierarchical structures, network structures, relational models, and so on. After a while it dawned on me that these are all just maps, being poor artificial approximations of some real underlying terrain. (William Kent 1978, “Data And Reality: Basic Assumptions in Data Processing Reconsidered”)

- Resulting view codified / standardised in the early 80s
  - ANSI-SPARC - Griethuysen, J.v. ISO/TC97/SC5/WG3-N695 - Concepts and Terminology for the Conceptual Schema and the Information Base., ANSI, New York, NY, 1982.

- Hirschheim, R.A., Klein, H.K. and Lyytinen, K. Information Systems Development and Data Modeling: Conceptual and Philosophical Foundations. Cambridge University Press, Cambridge, 1995.

George Mealy: in more detail

- Mealy distinguishes three distinct realms of interest:

  the real world itself,
  ideas about it existing in the minds of men, and
  symbols on paper or some other storage medium.

  The latter realms are, in some sense, held to be models of the former. Thus, we might say that data are fragments of a theory of the real world, and data processing juggles representations of these fragments of theory. No one ever saw or pointed at the integer we call "five" – it is theoretical – but we have all seen various representations of it, such as:

  V $(101)_2$ 5_8 *5* *0.5E01*

  and we recognize them all as denoting the same thing, with perhaps different flavours. …The issue is ontology, or the question of what exists. (Mealy 1967. p. 525)

- Mealy, G. H. 1967 "Another Look at Data," Proceedings of the Fall Joint Computer Conference, November 14–16, Anaheim, California

- In a footnote Mealy refers to Quine's essay "On What There Is".

Bill Kent: in more detail

- A message to mapmakers: highways are not painted red, rivers don't have county lines running down the middle, and you can't see contour lines on a mountain.
- For some time now my work has concerned the representation of information in computers. The work has involved such things as file organizations, indexes, hierarchical structures, network structures, relational models, and so on. After a while it dawned on me that these are all just maps, being poor artificial approximations of some real underlying terrain.
- These structures give us useful ways to deal with information, but they don't always fit naturally, and sometimes not at all. Like different kinds of maps, each kind of structure has its strengths and weaknesses, serving different purposes, and appealing to different people in different situations. Data structures are artificial formalisms. They differ from information in the same sense that grammars don't describe the language we really use, and formal logical systems don't describe the way we think. "The map is not the territory" [Hayakawa].

Current Business Systems Paradigm

- Goal:
  - Establish the characteristics of this paradigm – particularly in relation to the data-process distinction.
  - Start to establish the issues this raises.

Look at the situation in Kuhnian terms

- Look at the situation in Kuhnian terms.
  - Fixing the paradigm: Happened in the late 70s, early 80s.
  - Absorbed into the community’s unspoken assumptions: No longer considered a suitable / fruitful topic for analysis.
- First, expose this paradigm
  - For an exposition, we need to return to the documents produced at the time it was under discussion.
    - ANSI-SPARC - Griethuysen, J.v. ISO/TC97/SC5/WG3-N695 - Concepts and Terminology for the Conceptual Schema and the Information Base
  - Also need to make currently implicit assumptions explicit.
    - In particular, the data-process distinction.
- Then motivate a challenge to it.

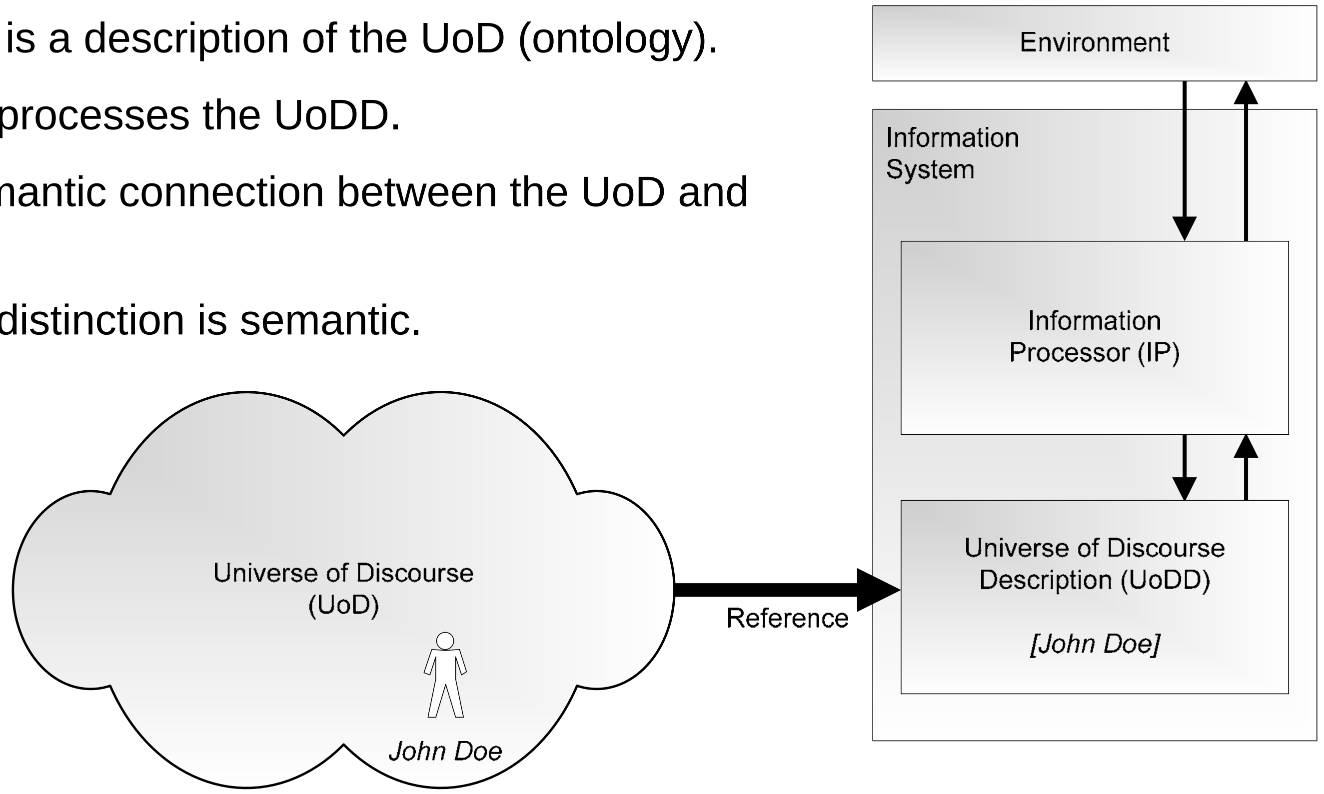
The Information System's relation to its Universe of Discourse (ontology)

UoDD (data) is a description of the UoD (ontology).

IP (process) processes the UoDD.

No direct semantic connection between the UoD and IP.

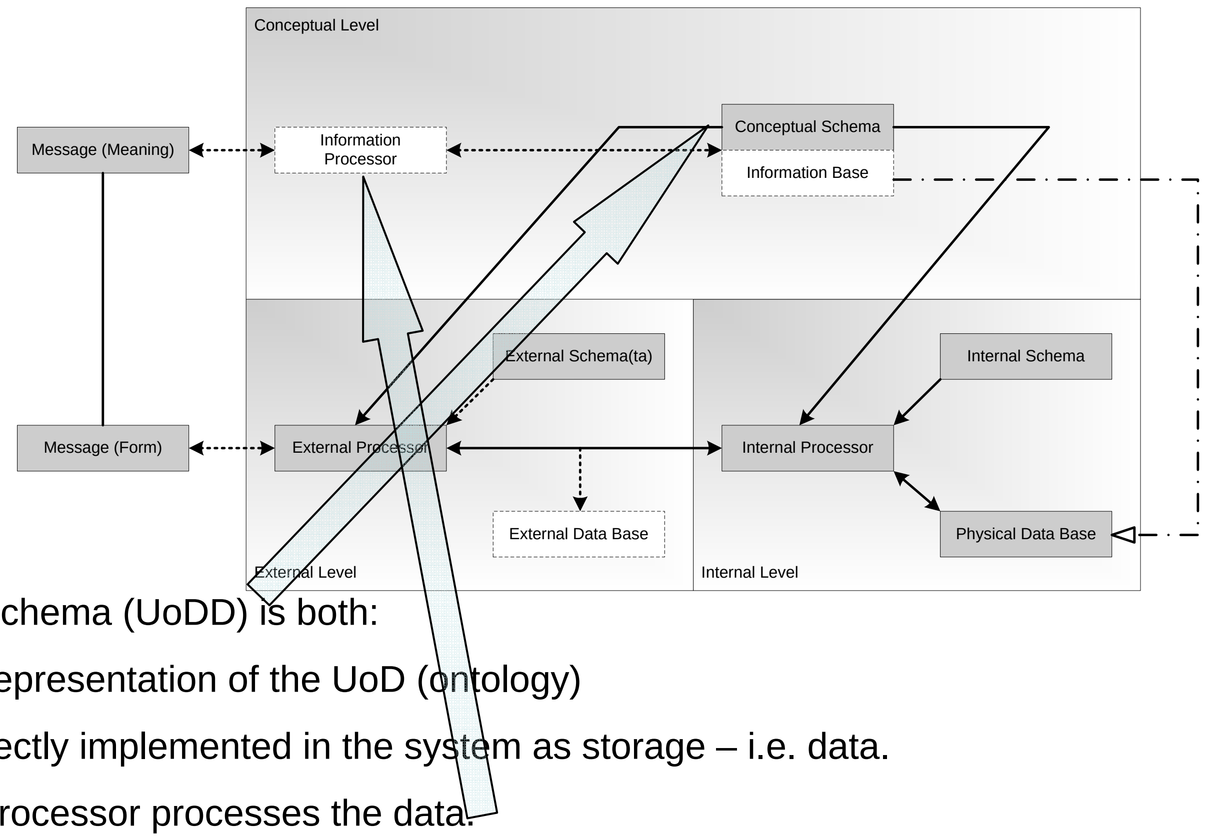
=> IP/UoDD distinction is semantic.



Reference: Griethuysen, J van, "ISO/TC97/SC5/WG3-N695 - Concepts and Terminology for the Conceptual Schema and the Information Base.," ISO/TC97/SC5/WG3-N695, ANSI, 1982.

Equivalence of IP/UoDD and Data (storage) / Process

For example: ANSI/SPARC: Three level architecture



Conceptual Schema (UoDD) is both:

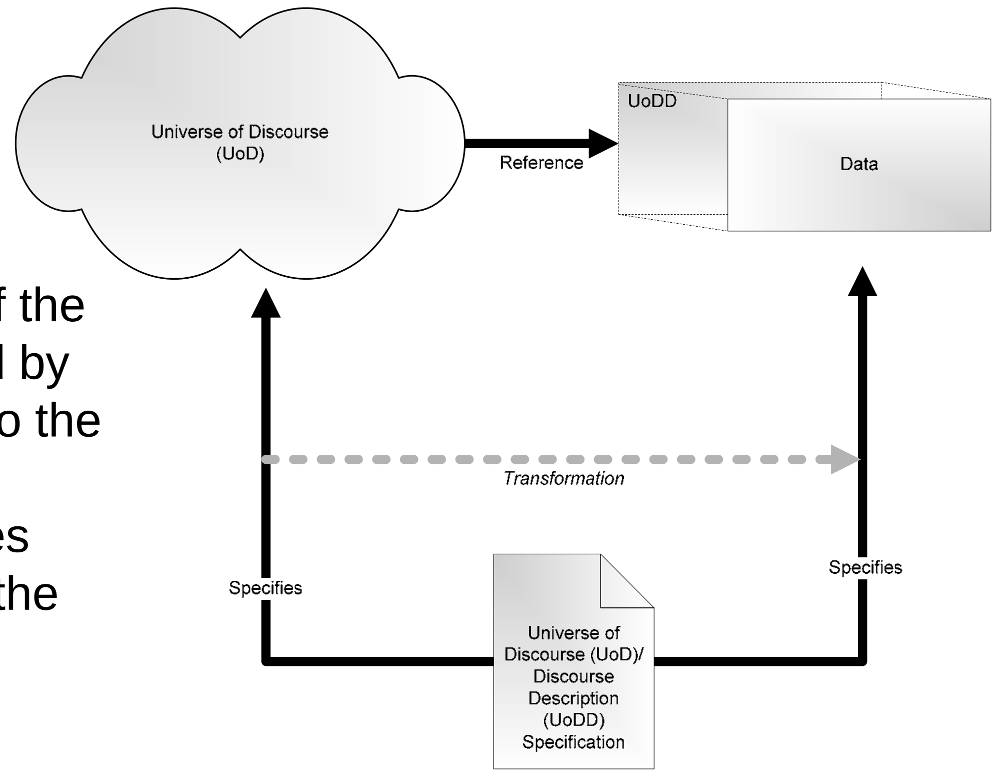
1) A representation of the UoD (ontology)
2) Directly implemented in the system as storage – i.e. data.

Information Processor processes the data.

Distinction reflected in the development process

For example: shift from UoD (ontology) to UoDD specification

The specification of the UoD is transformed by re-interpretation into the specification of the UoDD (data), it does not directly impact the IP (process).



From OMG MDA's perspective

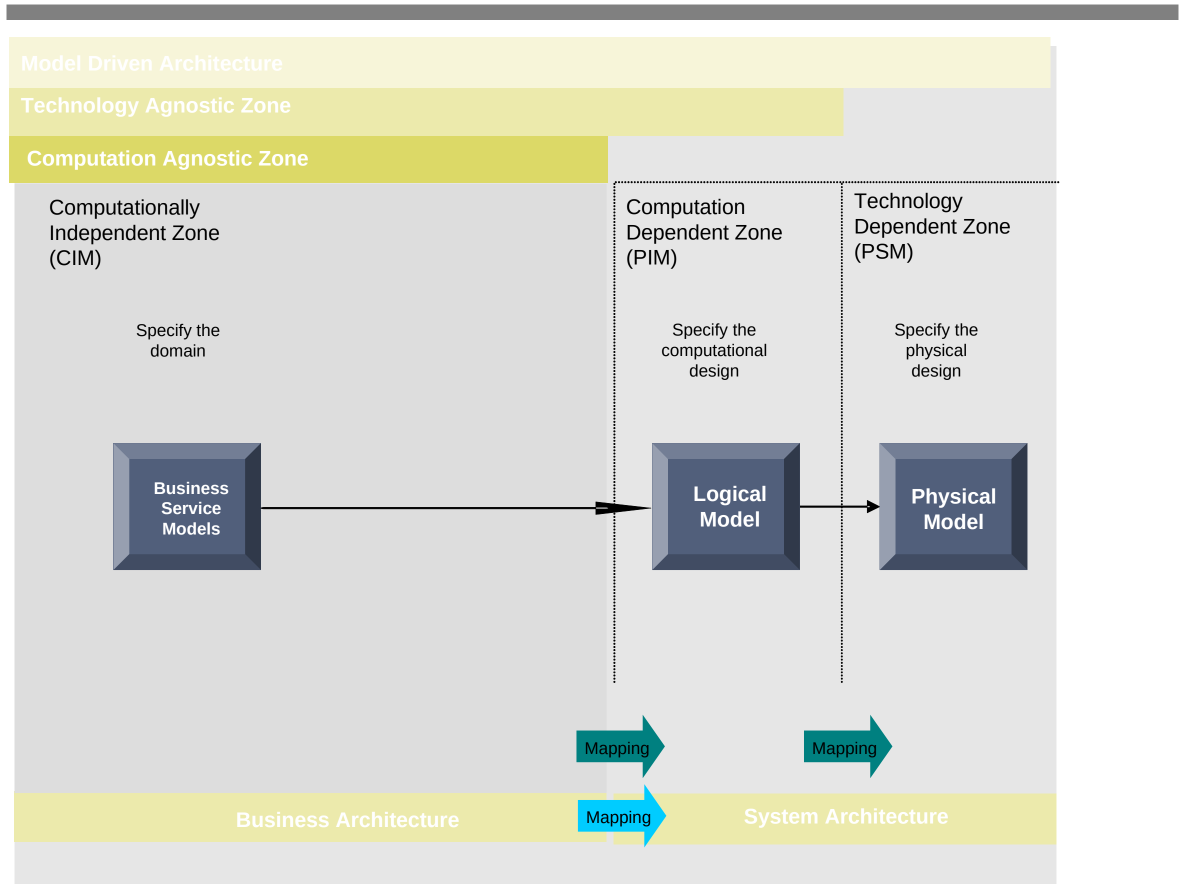
Process

Specify the domain.

Map this directly onto the logical data store.

Contrast this with the often indirect mapping into the physical data design.

First mapping is assumed to be trivial, second mapping is not. It is much studied



Challenge: What?

- Assumption that the mapping from UoD (ontology) to UoDD (data) is trivial as they are isomorphic. Hence the UoD and IP (process) are completely segregated.
- Not clear in what sense this is meant to be true:
  - Analytic – they are logically isomorphic.
  - Descriptive – they just are isomorphic.
  - Normative – they should be isomorphic.

Challenge: How?

- Standard philosophical technique of a thought-experiment counter-example.
- Experiment - Stages:
  - Examine ontic commitment of two ‘legacy’ systems with the same UoD – hence same ontology.
  - Show that elements of the UoD’s ontology are implemented as data in one system and process in the other.

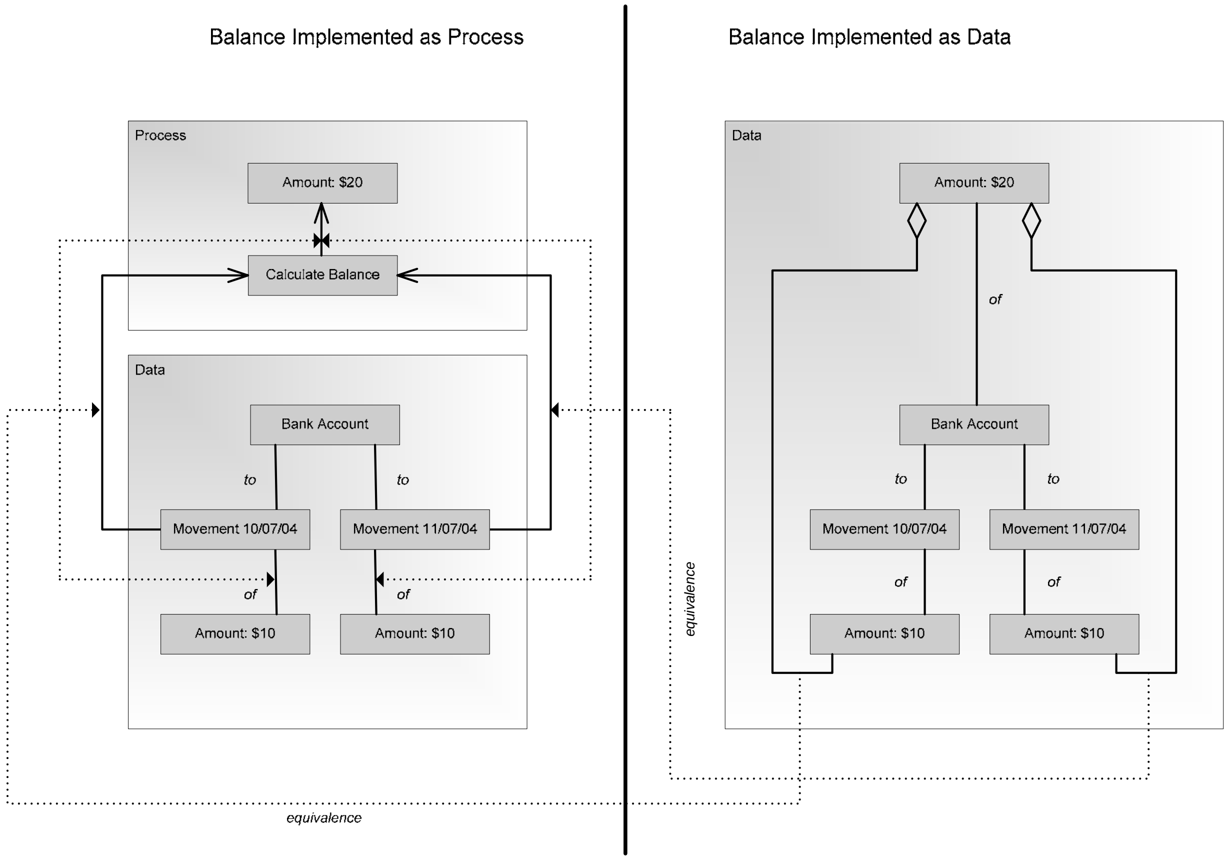
The systems' ontic commitment

- Consider two ISs that include this in their domain (UOD).
  - Two credit transactions of £10 which go to make up a balance of £20.
- Examining their ontological commitment (within a foundational ontology) we find a commitment to these objects and their relationships:
  - £10 Amount x 2
  - £10 Movement x 2
  - Bank Account
  - £20 Amount (balance).
- Issue is the ontic commitment to the balance – and so the relationship between the movements and the balance.

Balance implemented once as data and once as process

The first system records the two transactions and also records the balance and the fact that the transactions go to make up the balance.

The second system only records the two transactions and has a process that calculates the balance.



Conclusion

- As current paradigm insists on a trivial translation, then one MUST start with a system model.
- If one moves to the new paradigm, then:
  - One starts with the UoD (ontology)
  - One can and should map this onto both the data and the IP (process).
    - The mapping from UoD to system involves a decision on what should be data and what should be process.
    - The UoD is not isomorphic to the data. The mapping from UoD to system is not trivial.

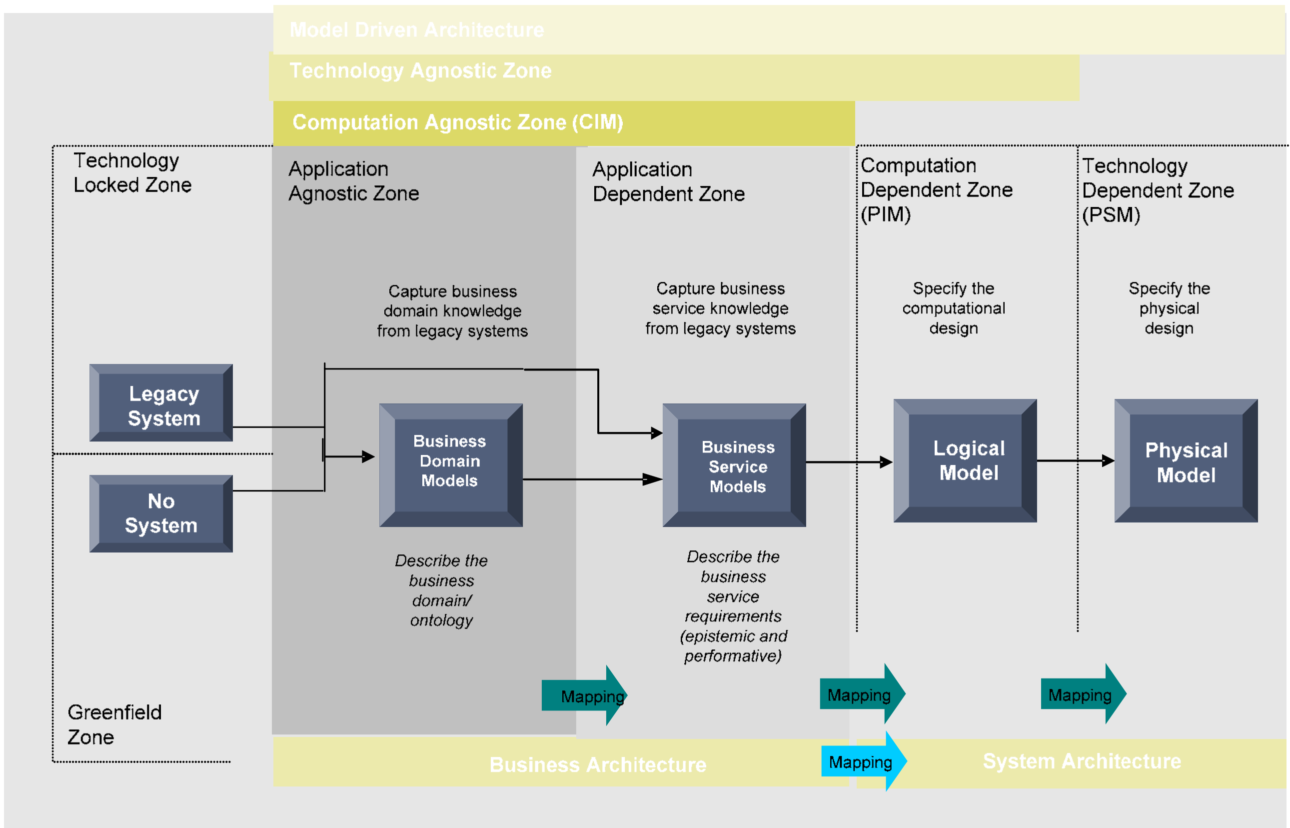
Implications for the development process

- This suggests radical changes in the early stages of the development process.
  - Currently designers construct a business model with their data/process decisions already embedded in it
  - This new perspective suggests that they should start with an ontology (with no data-process distinction) and then explicitly make their data-process decisions
- In this new development process
  - Ontology (the ontological model) has a clear role.
  - There is a non-trivial process of mapping the UoD/ontology into BOTH data and process

Epistemological layer

- In addition, this translation has a further layer.
- There are epistemological concerns that need to be reflected in the can and should map onto both the data and the IP (process).
  - This implies the need for both an ontology model and an epistemology model.
- We need to develop methods (and methodologies) for making these translations.

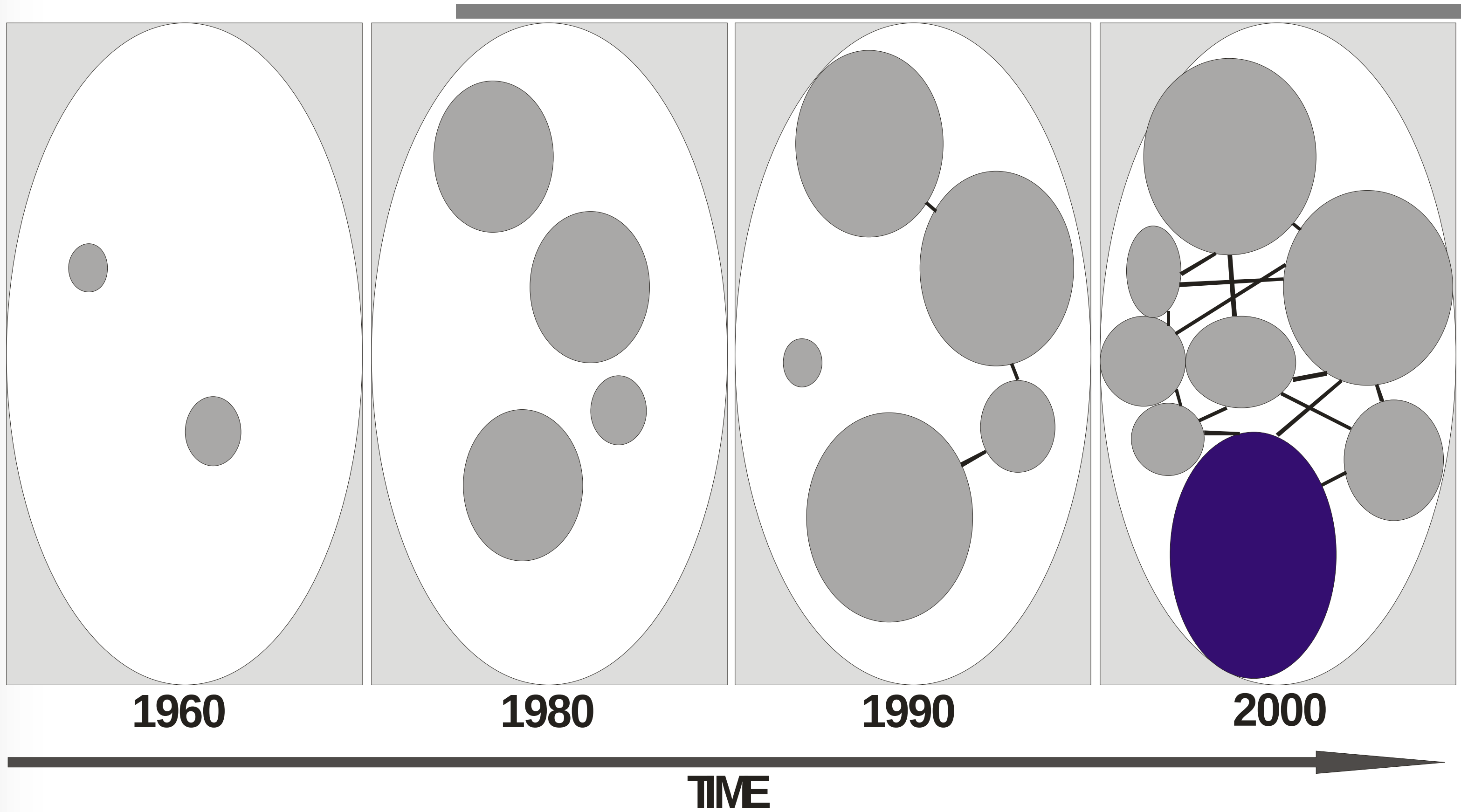
Visualising the revised development process



Legacy systems' ontologies

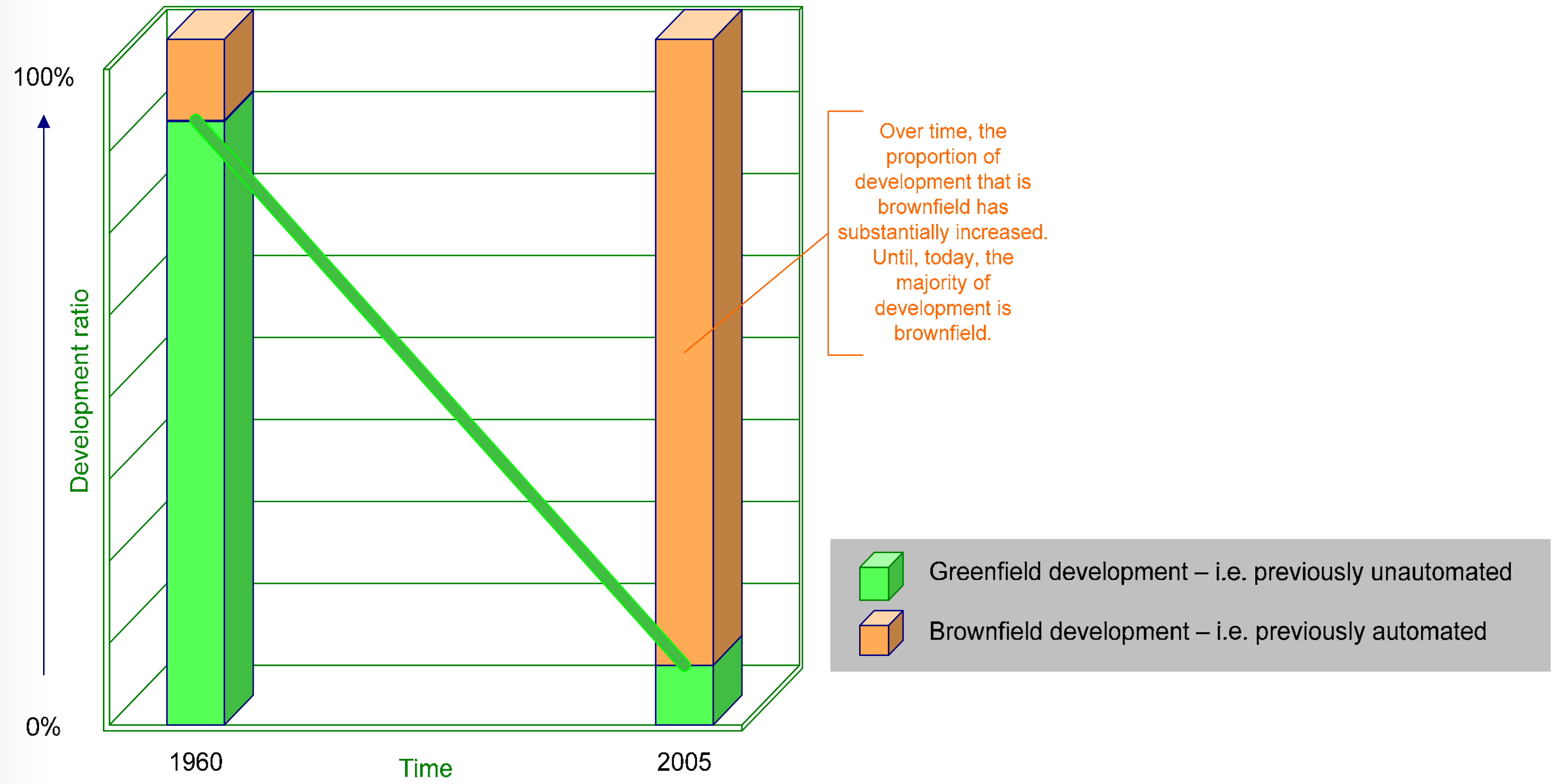
- Goal:
  - Identify ontology's roles in managing legacy systems

Background: Expanding Islands of Automation



Emerging automation: occupying the automatable space

Shift from Greenfield to Brownfield development



Legacy re-engineering approaches

- Virtually non existent.
- Methodologies not following the shift from green field to brownfield development
- Most large corporations stuck with a significant problem

Typical Legacy System

- System is:
  - Big - it spans the range of the business,
  - Old – originally developed in the early 70s-80’s,
  - Functionally rich – reflecting the enormous investment of knowledge and experience embedded in it.
- But it is showing its age - it is :
  - Inflexible and lacking important functionality.
  - Costly to maintain and enhance.
  - Not really a suitable basis for future developments.

Old paradigm and legacy systems approaches

- Approaches hampered by old paradigm's assumption that forces development to start with a system model.
- Legacy system based upon old technology, and old design approaches. I.e. there system model is significantly out of date.
- No clear route to salvaging enormous investment. No clear route to replacing the system.

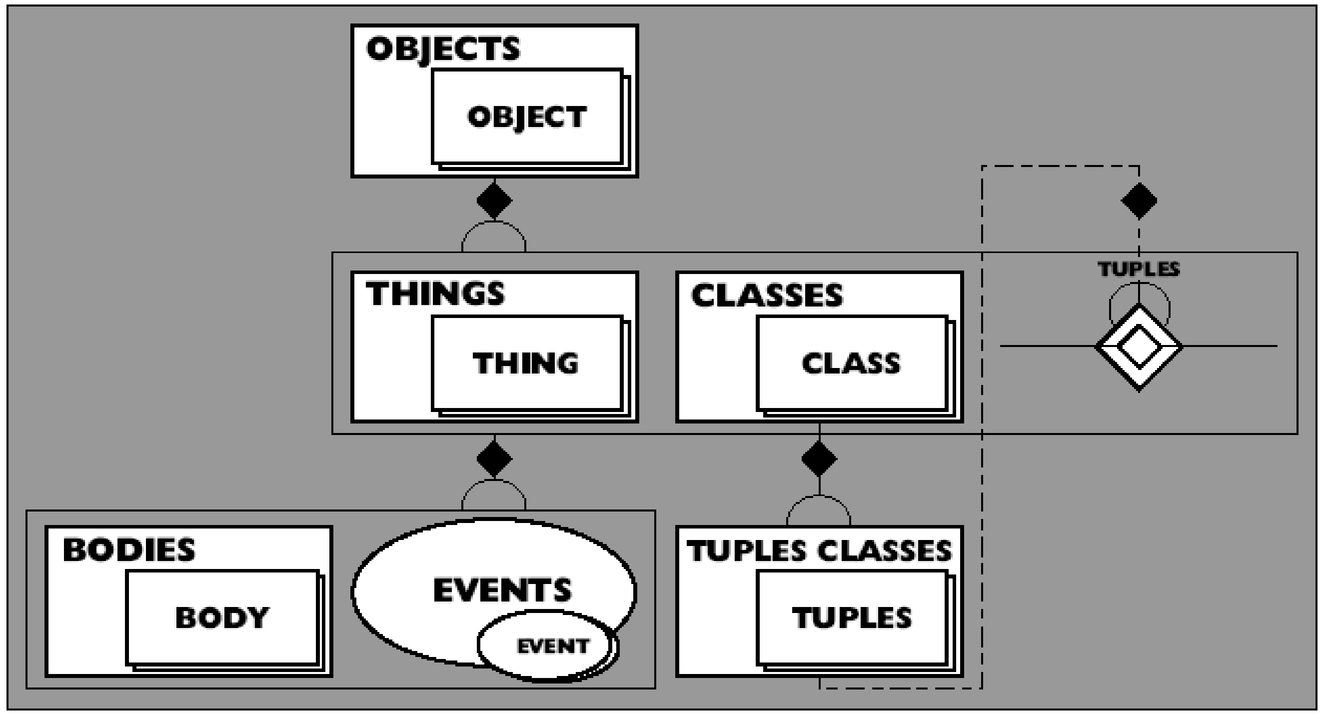
New paradigm and legacy systems approaches

- New paradigm suggests starting with an ontology.
- This captures the investment in understanding the business, without being tainted by the system design decisions.

Features of the ontology to support the analysis

- Need a framework
  - Need a clear top ontology
  - Need a clear decision on a number of metaphysical choices

Example: Top ontology



Top level needs to be comprehensive, to cater for all the types of object in the legacy systems.

Criteria of Identity

- One nice feature of some ontologies, is that each ontological category has a criteria of identity
  - Things = 4D extension,
  - Classes = instances,
  - tuples = places

Metaphysical choices

- A list of some choices
  - Universals and particulars
  - Perdurantism versus endurantism
  - Presentism versus eternalism
  - Absolute versus relative space, time and space-time
  - Modally extended versus unextended individuals
  - Materialism and non-materialism
  - Extensionalism versus non-extensionalism – I – Universals
  - Extensionalism versus non-extensionalism – II – Particulars
  - Topology of time – branching or linear.

See Partridge, C. (2002). LADSEB-CNR - Technical report 06/02 - Note: A Couple of Meta-Ontological Choices for Ontological Architectures.

My choices

- Perdurantism
- Eternalism
- Relative space-time
- Modally unextended individuals
- Materialism
- Extensionalism – I – Universals
- Extensionalism – II – Individuals
- Topology of time –linear.

Making metaphysical choices

- They are not independent, so need to be made consistently.
- Choices need to reflect the business system ontology's (engineering) goals.

Summary

1. There is a common interest in ‘what exists’ (= ontology) in philosophy and business systems
2. The business systems community’s paradigm, while playing lip service to UoD/Ontology works against it
3. Propose starting the development process with an ontology
4. This suggests a new approach to the re-development of legacy systems