An Interdisciplinary Approach

• A Unified Ontology-driven Modeling Methodology for databases, knowledge bases and OO-systems
  - Grounded in reality
  - Transparent to people
  - Rigorous
  - General

• Based on
  - Logic
  - Philosophy
  - Linguistics
The OntoClean Methodology

User Conceptualization

Conceptual Model

Ontology

Methodology

Minimal Top-Level Ontology

Ontology-Driven Modeling Principles

Useful Property Kinds

Formal Ontological Properties/Relations

The OntoClean Methodology

Formal Ontological Properties/Relations

Useful Property Kinds

Ontology-Driven Modeling Principles

Minimal Top-Level Ontology

User Conceptualization
Formal Ontology

• Theory of *formal distinctions and connections* within:
  - entities of the world, as we perceive it (*particulars*)
  - categories we use to talk about such entities (*universals*)

• Why *formal*?
  - Two meanings: *rigorous* and *general*
  - Formal logic: connections between truths - neutral wrt *truth*
  - Formal ontology: connections between things - neutral wrt *reality*

• **Goal**: characterizing particulars and universals by means of formal properties and relations.
The role of formal relations

• *Formal relations*
  - only depend on their relata
  - apply to multiple subdomains

• *Material relations*
  - Imply the existence of something besides the relata
  - Are specific of a few subdomains

• Role of formal relations
  - Formulate *general constraints* on the domain
  - Induce distinctions among entities that impose a general structure on the domain.
Formal Ontological Analysis

- Theory of Parts
- Theory of Wholes
- Theory of Essence and Identity
- Theory of Dependence
- Theory of Qualities
- Theory of Composition and Constitution
Mereology

• Primitive: *proper part-of* relation (PP)
  - asymmetric
  - transitive
  - \( Pxy =_{\text{def}} PPxy \lor x=y \)

• Axioms:

  **supplementation:** \( PPxy \rightarrow \exists z ( PPzy \land \neg z=x) \)

  **principle of sum:** \( \exists z ( PPxz \land PPyz \land \neg \exists w(PPwz \land \neg (Pwx \lor Pwy))) \)

  **extensionality:** \( x = y \leftrightarrow (Pwx \leftrightarrow Pwy) \)

*Excluded models:*
Extensionality and mereological invariance

• Extensionality: whenever the parts exist, \( x \) exists
  \( \textit{the whole is always the sum of its parts} \)

• Mereological invariance: \( x \) \textit{always keeps its parts}

• Examples of extensional entities:
  – Amounts of matter
  – Regions
  – Pluralities (pseudo-extensionality)

• Mereologically invariant (but non-extensional) entities:
  – A physical body (a lump of matter)
Identity and Unity

- Identity: is this my dog?
- Unity: is the collar part of my dog?
Unity

• A tentative formulation: $x$ is a whole under $\omega$ iff $\omega$ is an equivalence relation that binds together all the parts of $x$, such that

$$P(y,x) \rightarrow (P(z,x) \leftrightarrow \omega(y,z))$$

but not

$$\omega(y,z) \leftrightarrow \exists x (P(y,x) \land P(z,x))$$

• $P$ is the part-of relation
• $\omega$ can be seen as a generalized indirect connection
Unity Refined

\[ \delta_R(x) =_{df} R(x, x) \]

\[ \upsilon_R(x) =_{df} \sum_{\delta_R(x)} \land \forall y, z ((\delta_R(y) \land \delta_R(z) \land P(y, x) \land P(z, x)) \rightarrow R(y, z)) \]

\[ (x \text{ is unified by } R) \]

\[ \omega_R(x) =_{df} \text{Max}_{\upsilon_R}(x) \]

\[ (x \text{ is a whole under } R) \]

\[ \Sigma_\phi(x) =_{df} \forall y(P(y, x) \rightarrow \exists z(\phi(z) \land P(z, x) \land O(z, y))) \]

\[ (\text{sum of } \phi \text{s}) \]
Kinds of Whole

• Depending on the *nature of* $\omega$, we can distinguish:
  - *Topological wholes* (a piece of coal, a lump of coal)
  - *Morphological wholes* (a constellation)
  - *Functional wholes* (a hammer, a bikini)
  - *Social wholes* (a population)

* a whole can have *parts that are themselves wholes* (with a different $\omega$)
Parts vs. components

- A proper part is a *component* iff it is a whole
- We can have topological components, morphological components, functional components....
Unity and Plurality

• Ordinary objects: wholes or sums of wholes
  - Singular: no wholes as proper parts
  - Plural: sums of wholes with a common unity condition
    • Plural wholes (the sum is also a whole)
    • Collections (the sum is not a whole)
      - Members of collections are special parts!

• “Fiat” objects: everything else

• Role of topological wholes in perception (ordinary things)
Unity Disjointness Constraint

Properties with incompatible UCs are disjoint
Essence and Rigidity

• Certain entities have *essential* properties.
  - John must have a brain.
  - John must be a person.
• Certain properties are essential to *all* their instances (compare *being a person* with *having a brain*).
• These properties are rigid - if an entity is ever an instance of a rigid property, it must always be.
Permanent vs. Essential Properties

• Being *always* a student
• Being *necessarily* a student
• Quantifying on *possibilia*
Formal Rigidity

• $\phi$ is rigid (+R): $\forall x \diamond \phi(x) \rightarrow \Box \phi(x)$
  - e.g. Person, Apple

• $\phi$ is non-rigid (-R): $\exists x \diamond \phi(x) \land \neg \Box \phi(x)$
  - e.g. Red, Male

• $\phi$ is anti-rigid (~R): $\forall x \diamond \phi(x) \rightarrow \neg \Box \phi(x)$
  - e.g. Student, Agent
Rigidity Constraint

\[ +R \not\subset \sim R \]

- Why?
  - \[ \forall x \ P(x) \rightarrow Q(x) \]

Diagram:
- Q\(^-R\)
- P\(+R\)
- O10
Identity as a *primitive relation*

Identity is a *primitive equivalence relation* that satisfies Leibniz’s rule.

In general, *identity can’t be defined.* What we can have are just *informative constraints* (identity criteria)

...related to relevant *classes* of entities

...which are ultimately the result of our conceptualization of reality.
Synchronic Identity

• A plausible criterion:

  \[ x \text{ and } y \text{ are the same iff they have the same location} \]

• Exceptions:
  - Immaterial objects (holes...)
  - Events (singing while taking a shower...)
  - Constituting entities (the statue’s clay...)
Diachronic Identity

• It’s hard, but terribly useful for commonsense
• Requires some notion of persistence
• In addition, the *sameness* (or continuity) of certain properties is required
Identity is not similarity!

- At different spatio/temporal locations, two things may be identical but not similar, and vice versa.
- Identity and similarity are different concepts.
- Similarity may be seen as a generalization of geometrical congruence.
- Indiscernability of identicals (Leibniz’s law) vs. identity of indiscernibles.
Identity criteria

- **Classical formulation:**
  \[ \phi(x) \land \phi(y) \rightarrow (\rho(x,y) \leftrightarrow x = y) \]

- **Generalization:**
  \[ \phi(x,t) \land \phi(y,t') \rightarrow (\Gamma(x,y,t,t') \leftrightarrow x = y) \]
  (synchronic: \( t = t' \); diachronic: \( t \neq t' \))

- In most cases, \( \Gamma \) is based on the **sameness** of certain **characteristic features**:
  \[ \Gamma(x,y,t,t') = \forall z (\chi(x,z,t) \land \chi(y,z,t')) \]
Heuristics for Identity

• Finding necessary and sufficient ICs for a given property may be very hard.

• Heuristic 1: at least a sufficient IC.
• Heuristic 2: some essential parts or qualities
• Heuristic 3: some essential (non-rigid) properties
Sortals and other properties

• **Sortals** (*horse, triangle, amount of matter, person, student...*)
  - Carry identity
  - Usually correspond to *nouns*
  - High organizational utility
  - Main subclasses: *types* and *roles*

• **Non-sortals** (*red, big, decomposable, eatable, dependent, singular...*)
  - No identity
  - Usually correspond to *adjectives*
  - Span across different sortals
  - Limited organizational utility (but high semantic value)

• **Categories** (*universal, particular, event, substance...*)
  - No identity
  - Useful generalizations for sortals
  - Characterized by a set of (only necessary) formal properties
  - Good organizational utility
Local vs. Global ICs

- **Local IC:**
  \[ \phi(x, t) \land \phi(y, t') \rightarrow (\Gamma(x, y, t, t') \leftrightarrow x = y) \]

- **Global IC:**
  \[ \phi(x, t) \rightarrow (\Gamma(x, y, t, t') \leftrightarrow x = y) \]
Carrying vs. Supplying Identity

- **Supplying** (global) identity (+O)
  - Having an IC (or essential property) that doesn’t hold for a subsuming property
- **Carrying** identity (+I)
  - Not supplying identity, while being subsumed by a property that does.
- **Common sortal principle**: \(x = y \rightarrow\) there is a common sortal supplying their identity
- Theorem: only rigid properties supply global identity
Sortal specialization

- **Type** specialization (e.g. Living being $\rightarrow$ Person)
  - New features (especially essential properties) **affect identity**
  - Both necessary and sufficient ICs can be added while specializing types
    - Polygon: same edges, same angles
    - Triangle: two edges, one angle
    - Living being: same DNA, etc...?
    - Zebra: same stripes?

- **Role** specialization (e.g. Person $\rightarrow$ Student)
  - New features **don't affect identity**
Types and Roles

entity

life-form

person

African

European

* causal-agent

? 

* worker
Identity Disjointness Constraint

ICs impose *constraints* on sortals, making their ontological nature explicit:

Properties with incompatible ICs are *disjoint*

Examples:
- sets vs. ordered sets
- persons and passengers
- amounts of matter vs. assemblies
Inheriting identity constraints

If $P$ is rigid:

$x$ is the same $P$ as $y \iff x = y$

$\neg(-x$ is the same $P$ as $y$

- every $x$ is a $Q$

$\neg(-x$ is the same $Q$ as $y$

Physical Object

Person
Example - Identity

• Is time-interval a subclass of time-duration?
  - Initial answer: yes
• IC for time-duration
  - Same-length
• IC for time-interval
  - Same start & end
Example - Identity

- occurrent
  - time-duration
    - time-interval
      - One hour
        - 2-3 PM Tues.
        - 3-4 PM Weds.
Part, Constitution, and Identity

- **Structure** may change identity
- **Mereological extensionality** is lost
- **Constitution** links the two entities
- Constitution is asymmetric (implies *dependence*)
Ontological Levels and IC/UC Kinds

- Physical
  - Mereological (an amount of matter, a collection)
  - Topological (a piece of matter)
  - Morphological (a cubic block, a constellation)
- Functional (a tool, a biological organ)
- Biological (a human body)
- Cognitive/Intentional (a person, a robot)
- Social (a company)

√ Correspond to different kinds of IC/UC
√ All levels except the mereological one have non-extensional IC
√ A generic dependence relation links higher levels to their immediate inferior.
How ontological levels simplify taxonomies

- social-event
- mental-event
- physical-event
- communication-event
- perceptual-event
Why bother with this?

- **Formal ontological analysis** requires analyzing all properties according to their meta-properties - This is a **lot** of work!

- Why perform this analysis?
  - Makes **modeling assumptions** clear, which:
    - Helps resolving known conflicts
    - Helps recognizing unknown conflicts
  - Imposes **constraints** on standard modeling primitives
    (generalization, aggregation, association)
  - Elicits **natural distinctions**
  - ...results in more **reusable ontologies**
Resolving Ontological Conflicts

- Two well-known ontologies define:
  - \textit{Physical Object} is-a \textit{Amount of Matter} (WordNet)
  - \textit{Amount of Matter} is-a \textit{Physical Object} (Pangloss)

- \textbf{Amount of Matter}
  - unstructured / scattered “stuff”
  - Identity: mereologically extensional
  - Unity: intrinsically none (anti-unity)

- \textbf{Physical Object}
  - Isolated material body
  - Identity - three options:
    - None
    - Non-extensional
    - Extensional
  - Unity: Topological

\textbf{Conclusion:} the two concepts are \textit{disjoint}. Physical objects are \textit{constituted} by amounts of matter
Overloading Subsumption
Common modeling pitfalls

- Instantiation
- Constitution
- Composition
- Disjunction
- Polysemy
Taxonomic Constraints

- $+R \not\subset \sim R$
- $-I \not\subset +I$
- $-U \not\subset +U$
- $+U \not\subset \sim U$
- $-D \not\subset +D$

- Incompatible IC’s are disjoint
- Incompatible UC’s are disjoint

- Categories subsume everything

- Roles can’t subsume types
Instantiation (1)

Does this ontology mean that My ThinkPad is a ThinkPad Model?

ThinkPad Model

Ooops…

My ThinkPad (s# xx123)

T21

Question: What ThinkPad models do you sell?
Answer should NOT include My ThinkPad -- nor yours.
Instantiation (2)

- Notebook Computer
- ThinkPad Model
- T Series
- My ThinkPad (s# xx123)
- model
- T 21
Composition (1)

Computer

Disk Drive

Memory

Micro Drive

Question: What Computers do you sell?
Answer should NOT include Disk Drives or Memory.
Composition (2)

Computer

<table>
<thead>
<tr>
<th>part-of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk Drive</td>
</tr>
<tr>
<td>Memory</td>
</tr>
</tbody>
</table>

Micro Drive
Disjunction (1)

Computer has-part Computer Part

Disk Drive

Micro Drive

Memory

Camera-15 has-part Flashcard-110

Unintended model: flashcard-110 is a computer-part
Disjunction (2)

Computer \text{ has-part } \text{Disk Drive } \lor \text{Memory } \lor \ldots
Polysemy (1) (Mikrokosmos)

Physical Object  Abstract Entity

Book

.....

Question: How many books do you have on Hemingway?
Answer: 5,000
Polysemy (2)
(WordNet)

Physical Object

Book
Sense 1

.....

Abstract Entity

Book
Sense 2

Biography of Hemingway
Question: What types of matter will conduct electricity? Answer should NOT include computers.
Constitution (2)

Entity

Amount of Matter  Physical Object

Metal  Clay  Computer

consstituted
Property Dependence

• Does a property holding for $x$ depend on something else besides $x$?
  - $P(x) \rightarrow \exists y Q(y)$
  - $y$ should not be a part of $x$

• Example: Student/Teacher, customer/vendor
## Basic Property Kinds Table

<table>
<thead>
<tr>
<th>+</th>
<th>+</th>
<th>+</th>
<th>±</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>Quasi-type</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>Mixin</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>~</td>
<td>+</td>
<td>Mat. role</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>~</td>
<td>-</td>
<td>Phased sortal</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>+</td>
<td>±</td>
<td>Category</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>~</td>
<td>+</td>
<td>Formal role</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Attribution</td>
</tr>
</tbody>
</table>
A formal ontology of properties

Property

Sortal
+I

Non-sortal
-I

Role

Rigidity

Non-rigid
-R

Rigid
+R

Formal Role

Attribution

Category

Material role

Mixed

Phased sortal
-D +L

Type

Quasi-type

-R

-R+D

-I
The **Backbone Taxonomy**

Assumption: *no entity without identity*

- Since identity is supplied by types, every entity must instantiate a type
- The taxonomy of types spans the whole domain
- Together with categories, types form the **backbone taxonomy**, which represents the **invariant structure** of a domain (rigid properties spanning the whole domain)
Flowchart of backbone taxonomy:

- Non-sortals
  - Attributions
- Sortals
  - Mixins
- Backbone Taxonomy
  - Categories
  - Formal Roles
  - Top Types
    - Types & Quasi-Types
  - Material Roles
- Phased Sortals
Well-founded ontology design

An ontology-cleaning example
Dealing with *Ontological Relativism*

• Deciding about the meta-properties carried by a given property...

  Is up to **YOU**!

• But a *common agreement* must be achieved about the formal meaning (and practical utility) of meta-properties
Property Analysis
Entity, Location

• Entity
  - Everything is an entity
  - -I-U-D+R
  - Category

• Location
  - A generalized region of space.
  - +O: by its parts (mereologically extensional).
  - ~U: no way to isolate a location
  - -D+R
  - Type
Property Analysis
Amount of Matter, Red

- **Amount of Matter**
  - unstructured / scattered “stuff” as lumps of clay or some bricks
  - +O: mereologically extensional
  - ~U: intrinsically no unity
  - -D+R
  - Type

- **Red**
  - Really Red-thing, the set of all red-colored entities
  - -I-U-D-R
  - Formal Attribution
Property Analysis
Agent, Group

• Agent
  - An entity playing a part in some event
  - +I-U: no universal IC/UC
  - +D: on the event/action participating in
  - ~R: no instance is necessarily an agent
  - Formal role

• Group
  - An unstructured collection of wholes
  - +O: same-members
  - ~U: unstructured, no unity.
  - Type
Property Analysis
Physical Object, Living Being

• Physical Object
  - Isolated material objects.
  - +O: same spatial location (only synchronic, no common diachronic IC).
  - +U: Topological
  - -D+R
  - Type

• Living Being
  - +O: same-DNA (only nec.)
  - +U: biological unity
  - -D+R
  - Type
Property Analysis
Food, Animal

• Food
  - +I-O~U: amt. of matter
  - +D: something that eats it.
  - ~R: being food is not necessary...
  - Material Role

• Animal
  - +O: same-brain
  - +U: biological unity
  - -D+R
  - Type
Property Analysis
Legal Agent, Group of People

• Legal Agent
  - A legally recognized entity
  - +L: All legal systems have a defined IC, has-same-legal-ID
  - -U: no universal unity
  - +D: on the legal body that recognizes it
  - ~R: not necessary
  - Material Role

• Group of People
  - See Group
  - +I-O~U-D+R
  - Quasi-type
Property Analysis
Social Entity, Organization

- **Social Entity**
  - A group of people together for social reasons
  - -I: no universal IC
  - +U: social-connection
  - -D+R
  - category

- **Organization**
  - A group of people together, with roles that define some structure
  - +O: same-mission and way of operating
  - +U: functional
  - -D+R
  - Type
Property Analysis

Fruit

- Fruit
  - An individual fruit, such as an orange or banana
  - +O: same-plant, same-shape, etc. (only nec.)
  - +U: topological
  - -D+R
  - Type
Property Analysis
Apple, Red Apple

- Apple
  - +O: shape, color, skin pattern (only nec)
  - +U: topological
  - -D+R
  - Type

- Red-Apple
  - +I-O: from Apple
  - +U: from Apple
  - -D
  - ~R: no red apple is necessarily red
  - type-attribution mixin
**Property Analysis**

**Vertebrate, Person**

- **Vertebrate**
  - Really vertebrate-
    animal
  - A biological
classification that adds
new membership
criteria (has-backbone)
  - +I-O: from animal
  - +U: from animal
  - -D+R
  - quasi-type

- **Person**
  - +O: same-fingerprint
  - +U: from animal
  - -D+R
  - Type
Property Analysis
Butterfly, Caterpillar

• Butterfly
  - +L: same-wing-pattern
  - +U: biological
  - -D
  - ~R: the same entity can be something else (a caterpillar)
  - Phased sortal

• Caterpillar
  - +L: spots, legs, color
  - +U: biological
  - -D
  - ~R: caterpillars become butterflies and change their IC
  - Phased sortal
Property Analysis

Country

- **Country**
  - A place recognized by convention as autonomous
  - +L: government, sub-regions
  - +U: countries are countable (heuristic)
  - -D
  - ~R: some countries do not exist as countries any more (e.g. Prussia) but are still places
  - Phased sortal
Remove non-rigid properties

Entity
- I-U-D+R

Location
+ O-U-D+R

Amount of matter
+ O~U-D+R

Physical object
+ O+U-D+R

Living being
+ O+U-D+R

Group
+ O~U-D+R

Group of people
+ I-O~U-D+R

Social entity
- I+U-D+R

Agent
- I-U+D~R

Organization
+ O+U-D+R

Person
+ O+U-D+R

Location
+ O-U-D+R

Amount of matter
+ O~U-D+R

Physical object
+ O+U-D+R

Living being
+ O+U-D+R

Group
+ O~U-D+R

Group of people
+ I-O~U-D+R

Social entity
- I+U-D+R

Agent
- I-U+D~R

Organization
+ O+U-D+R

Person
+ O+U-D+R

Location
+ O-U-D+R

Amount of matter
+ O~U-D+R

Physical object
+ O+U-D+R

Living being
+ O+U-D+R

Group
+ O~U-D+R

Group of people
+ I-O~U-D+R

Social entity
- I+U-D+R

Agent
- I-U+D~R

Organization
+ O+U-D+R

Person
+ O+U-D+R
Analyze taxonomic links

- ~U can’t subsume +U
- Living being can change parts and remain the same, but amounts of matter can not (incompatible ICs)
- Living being is constituted of matter
Analyze taxonomic links

- ~U can’t subsume +U
- Living being can *change parts* and remain the same, but amounts of matter cannot (incompatible ICs)
- Living being is *constituted* of matter
Analyze taxonomic links

- \sim U$ can't subsume $+U$
- Physical objects can change parts and remain the same, but amounts of matter can not (incompatible ICs)
- Physical object is constituted of matter
Analyze taxonomic links

- ~U can’t subsume +U
- Physical objects can change parts and remain the same, but amounts of matter can not (incompatible ICs)
- Physical object is constituted of matter
Analyze taxonomic links

- **Meta-properties fine**
- **Identity-check** fails: when something stops being an animal, it does not stop being a physical object (when an animal dies, its body remains)
- **Constitution again**
Analyze taxonomic links

- **Meta-properties fine**
- **Identity-check** fails: when an entity stops being an animal, it does not stop being a physical object (when an animal dies, its body remains)
- **Constitution again**
Analyze taxonomic links

- ~U can’t subsume +U
- A group *can’t change parts* - it becomes a different group
- A social entity can change parts - it’s more than just a group (incompatible IC)
- Constitution again
80
ER2002

Entity \(-I-U-D+R\)

- Amount of matter
- Location \(+O-U-D+R\)

- Living being \(+O+U-D+R\)
  - Animal \(+O+U-D+R\)
  - Vertebrate \(+I-O+U-D+R\)
  - Social entity \(-I+U-D+R\)
  - Person \(+O+U-D+R\)
  - Group of people \(+I-O-U-D+R\)
  - Organization \(+O+U-D+R\)
  - Group \(+O-U-D+R\)

- Agent \(-I-U+D~R\)

- Analyze non-rigid properties

- \(\sim R\) can’t subsume \(+R\)
- Really want a type restriction: all agents are animals or social entities.
  - Subsumption is not disjunction!
• ~R can’t subsume +R
• Another disjunction: all legal agents are persons or organizations
Analyse non-rigid properties

- ~R can’t subsume +R
- Another disjunction: all legal agents are persons or organizations
Analyze non-rigid properties

• ~R can’t subsume +R
• Apple is not necessarily food. A poison-apple, e.g., is still an apple.
• ~U can’t subsume +U
• Caterpillars are wholes, food is stuff.
Analyze non-rigid properties

- \(\neg R\) can't subsume \(R\)

- Apple is not necessarily food. A poison-apple, e.g., is still an apple.

- \(\neg U\) can't subsume \(U\)

- Caterpillars are wholes, food is stuff.
• Identity check: a location can’t change parts...
• 2 senses of country: geographical region and political entity.
• Split the two senses into two concepts, both rigid, both types.
There is a relationship between the two, but not subsumption.
- Caterpillars and butterflies cannot be vertebrate
- There must be a rigid property that subsumes the two, supplying identity across temporary phases

Look for missing types
Look for missing types
Analyze Attributions

- No violations
- Attributions are discouraged, can be confusing.
- Often better to use attribute values (i.e. Apple Color red)
The backbone taxonomy

- Entity
  - Location
  - Amount of matter
  - Physical object
  - Living being
  - Social entity
  - Group
  - Group of people
  - Country
  - Geographical Region
  - Fruit
  - Apple
  - Lepidopteran
  - Vertebrate
  - Animal
  - Person
  - Organization
Use OntoClean for all your ontology cleaning needs!