Towards the Ontological Analysis and Modularization of CRM (v.6.2.1)

Presentation based on joint paper with **Béatrice Markhoff** and **Perrine Pittet** to appear in Proceedings of FOIS XI 2020

Emilio M. Sanfilippo,
Laboratory for Applied Ontology ISTC-CNR
emilio.sanfilippo@cnr.it

48th CRM meeting

Introduction

Goal of our study:

- To support the use of CRM for **knowledge representation** and **data management tasks** in the DH community

  Examples (not limited to):

  - Data modeling
  - Data publishing, e.g., via Web platforms
  - Data sharing
  - Data integration
  - Exploitation of automated reasoning procedures

Analyzed the **logical** and **ontological** foundations of CRM

Introduction: Formalism

State of the art:

- **Mainly** driven by **Semantic Web** languages (RDF, RDFS, OWL2, SPARQL)

Limited expressivity (in comparison to, e.g., first-order logic) but

- Good computational properties
- Well-supported by technologies and programming libraries
- Integrable with, e.g., relational databases (via **OBDA**)
- At the heart of **LOD** and **FAIR** approaches

Introduction: Ontological foundations

Use of methodologies and theories well-known in AI and ontology engineering, mainly

- **OntoClean**:[1] e.g., rigidity vs. anti-rigidity (see also OntoUML)[2]
- **Formal ontology**: theories of objects, events, qualities, dependence, constitution, parthood, etc. broadly used in, e.g., foundational ontologies like DOLCE[3] and UFO[4]

---

[1] Guarino, N., & Welty, C. A. (2004). An overview of OntoClean. In Handbook on ontologies (pp. 151-171). Springer, Berlin, Heidelberg.

[2] Guizzardi, G. (2005). Ontological foundations for structural conceptual models.

[3] Borgo, S., & Masolo, C. (2009). Foundational choices in DOLCE. In Handbook on ontologies (pp. 361-381). Springer, Berlin, Heidelberg.

[4] Guizzardi, G., Wagner, G., Almeida, J. P. A., & Guizzardi, R. S. (2015). Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story. Applied ontology, 10(3-4), 259-271.

Introduction: Ontological foundations

Focus on:

- Use of *quantifications* (i.e., cardinalities on relations)
- Use of relations with disjunctions (e.g., *P53 has former or current location*)
- E92 Spacetime Volume and E93 Presence
- E4 Period
- E72 Legal Object
- E54 Dimension
- *Ongoing work* on conceptual objects

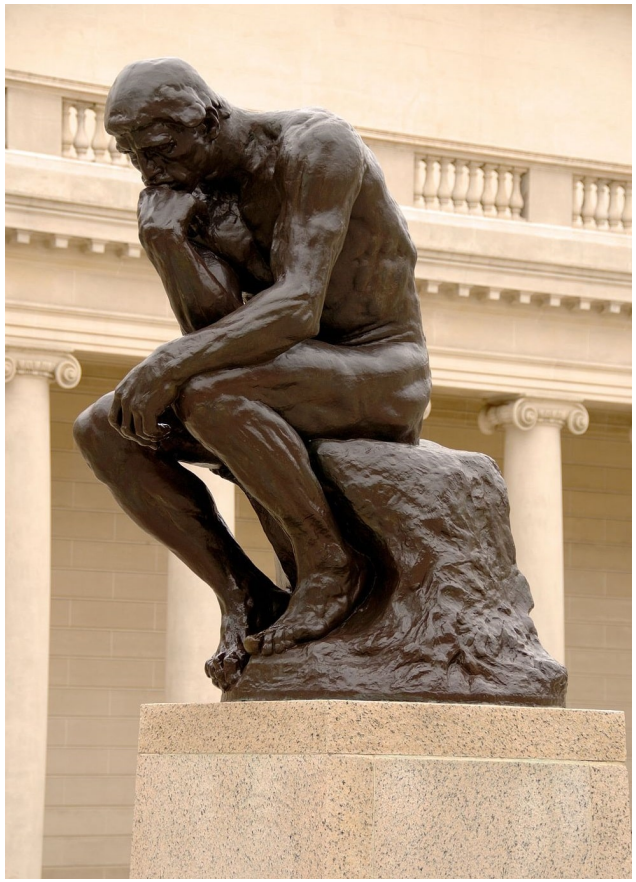
Introduction: Contributions

The presented study is – **hopefully** – a contribution for

1. The use of CRM as a **modular** Semantic Web, **OWL** ontology
2. The robustness of the ontological foundations of CRM

Part I: Ontological Analysis

Disclaimer: (Perhaps) Limited understanding of CRM



E92 Spacetime volume

Some key points:

- "comprises 4 dimensional point sets (volumes) in physical spacetime [...]. They may derive their identity from being the extent of a material phenomenon [...]" [CRM, p.41]
- **Example:** the Battle of Trafalgar, the shooting of Nelson during the Battle of Trafalgar, etc.
- From a **modeling stance**:
  - E92 **subsumes** *E18 Physical Thing* (persistent item, v6.2.1, now removed); *E4 Period* (temporal entity); and *E93 Presence*, **meaning that**
    - Physical things, periods, and presences **are** 4 dimensional point sets!

E92 Spacetime volume: Analysis

For example, *E18 Physical Thing* is a

- Persistent item, therefore an **endurant** [CRM, p.35], AND
- Spacetime volume, therefore something that is separable into temporal parts as a **perdurant**
  - E.g., one can – via *P166 was a presence of* – model the *presence of* a physical thing

This has been changed in CRM 7.0 (June 2020)!

Three- and Four-dimensionalism

Three- (3D) and **four**- dimensionalism:

- 3D: objects (i) can have only spatial parts; (ii) endure through time, i.e., they can be present at different times
  - Ontology of **endurants**: My desk can be divided in various spatial parts, e.g., its engineering components (e.g., 1 top, 4 legs, 8 screws, etc.)
- 4D: (i) objects can have both spatial and *temporal parts*; (ii) at each instant of time t at which an object o is present, what is present is a temporal part of o existing at and only at t
  - Ontology of **perdurants**: My-desk-at-t, My-desk-at-t' etc.

Commonly seen as **alternative** ontological positions[5]

[5] Wahlberg, T. H. (2014). The endurance/perdurance controversy is no storm in a teacup. Axiomathes, 24(4), 463-482.

E92 Spacetime volume and E93 Presence: Analysis

Example of E93 [n CRM 7.0]:

- The Roman Empire on 19 August AD 14

What is a geopolitical entity in the scope of CRM?

- **IF** it is a **period** (as CRM seems to assume), it is fine
- **IF** it is a **persistent item** (a complex social organization indeed), the case above seems misleading (i.e., a persistent item would be a space-time volume)

E72 Legal object

"[...] material or immaterial items to which instances of *E30 Right*, such as the right of ownership or use, can be applied" [CRM, p.33]

- High-level class in CRM, it subsumes *E18 Physical Thing* and *E90 Symbolic Object* (and all their subclasses)
- **For example**, persons, features, man-made objects, information objects, etc. they are legal objects

Rigidity vs Anti-rigidity (OntoClean)

Property (e.g., *being a person*, *being a student*):[6]

- **Rigidity**: a property is rigid when it necessarily holds for all its instances
  - If John is a **person**, he is necessarily as such whenever he exists, i.e., he can not stop being a person while remaining the same entity;

- **Anti-rigidity**: a property is anti-rigid when it does not necessarily hold for all its instances
  - John is a **student** but he is **not** necessarily as such. If John stops being a student, he still keeps his identity as a person (i.e., there is nothing 'fundamental' changing in his identity)

[6] Guarino, N., & Welty, C. A. (2004). An overview of OntoClean. In Handbook on ontologies (pp. 151-171). Springer, Berlin, Heidelberg.

Rigidity vs Anti-rigidity (con't)

Restriction:

- Classes referring to **anti-rigid properties** can **not** subsume classes referring to **rigid properties**

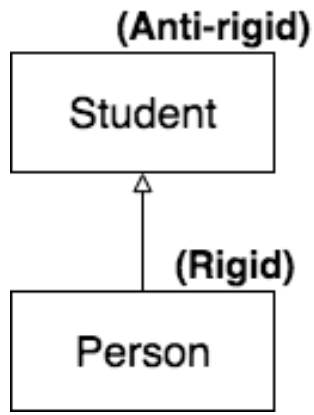


Figure: **NOT** allowed if Student is *anti-rigid* and *Person* is rigid

E72 Legal object: Analysis

E72 Legal Object seems to model anti-rigid properties:

- Legal properties that entities do not necessarily satisfy but that they can acquire within socio-legal contexts

E18 Physical Thing seems to model rigid properties:

- An entity can not stop being an instance of E18 while preserving its identity

If this consideration is correct:

- E72 can **not** subsume E18
  - E.g., human beings are **not** necessarily legal objects

E72 Legal object: Analysis (con't)

A proposal:

- Reconsider the position of E72 in the CRM taxonomy
- Introduce a modeling approach that makes explicit the representation of legal objects as properties that are only contextually satisfied (**social roles**)[7]
  - A human being *counts as* a legal object when this-and-that

[7] Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gangemi, A., & Guarino, N. (2004). Social Roles and their Descriptions. In KR (pp. 267-277).

E4 Period

Two main assumptions in CRM (see [CRM p.3])

1. **Granularity:** "there are no assumptions about the scale of the associated phenomena" (atomicity vs. complexity)
2. **Ontological nature:** "sets of coherent phenomena **or** cultural manifestations occurring in time and space"

I will focus on (2)

E4 Period: Analysis

E4 Period has two different meanings, i.e., it classifies temporal entities satisfying different identity/unity criteria:

1. "Sets of coherent phenomena": e.g., John walking from office to train station, John and Marry getting married, the birth of John and Mary's baby, etc.
   - Temporal entities satisfying identity/unity criteria that are not necessarily culturally based

2. "Cultural manifestations occurring in time and space": e.g., Middle Age, Italian Renaissance, Jurassic, etc.
   - Temporal entities necessarily satisfying cultural identity/unity criteria
   - Their identity/unity depend on a community of agents ascribing them a certain cultural value

E4 Period: Analysis (con't)

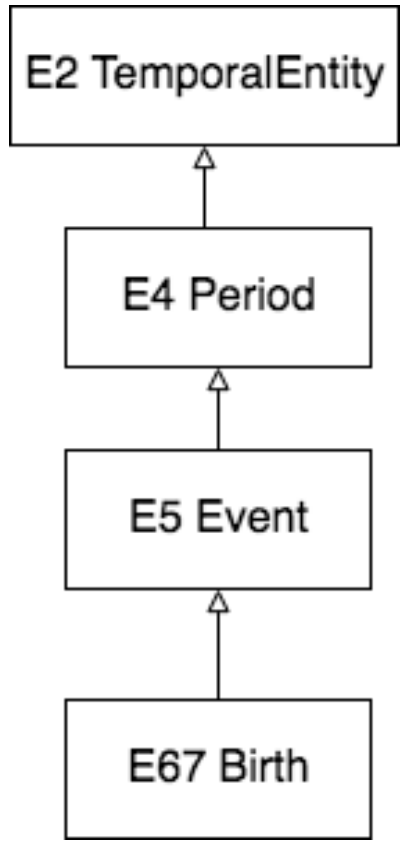


Figure: How do you differentiate between an birth event with cultural value (e.g., Leonardo Da Vinci's birth) and a birth event without such a value?

E4 Period: Analysis (con't)

A proposal:

- Reconsider the relation between *E5 Event* and *E4 Period*
- Introduce a modeling approach that can explicitly capture the **ascription of cultural value** to some temporal entities (persistent items, too?)

E54 Dimension

"**Quantifiable** properties that can be measured by some **calibrated means** and can be approximated by values, i.e. points or regions in a mathematical or conceptual space, such as natural or real numbers, RGB values etc" [CRM p.26]

Modeling pattern:

- *P90 has value E60 Number*
- *P91 has unit E58 Measurement Unit*

E54 Dimension: Analysis

What about **qualitative** values? For example,

- My chair's color is scarlet (red, blue, etc.)
- My chair's weight is heavy (light, etc.)

These could be useful to document entities with cultural heritage value, e.g., stained glasses, furniture, etc

Qualities and quality spaces

Foundational ontologies like DOLCE and UFO

- **Qualities:** individual specifically dependent entities like the individual color (weight, height, etc.) of my chair
- **Quality kinds:** disjoint classes of resembling qualities (color-qualities, weight-qualities, length-qualities, height-qualities, etc.)
- **Quality spaces:**[8] provide (topological, mereological, metric, etc.) structures to organize **qualities values**. For example,
  - In a quality space for colors, scarlet is a subregion of red
  - In a (different) quality space for colors #19D538 is a subregion of #00FF00 (green)
  - In a quality space for weights measured in kilos, 8kg is less than 8,5kg
  - In a (different) quality space for weights, *light* is disjoint with *heavy*

[8] Similar to **conceptual spaces** in the sense of Gärdenfors, P. (2004). Conceptual spaces: The geometry of thought. MIT press.

E54 Dimension: Analysis (con't)

Proposal:

1. To enlarge to scope of *E54 Dimension* to cover various kinds of qualities including those that are **not** necessarily measurable by calibrated means
2. Therefore, to explicitly cover the modeling of **qualitative values**, e.g., via an approach like the one previously discussed[9]

[9] see also: Masolo, C., & Borgo, S. (2005). Qualities in formal ontology. In Foundational Aspects of Ontologies (FOnt 2005) Workshop at KI (pp. 2-16).

Part II: Ontology modularization

Basic ideas:[10]

- It can be interpreted as decomposing potentially large and monolithic ontologies into (a set of) smaller and interlinked components (modules)
- Module M is an ontology existing in a set of modules such that, when combined, make up a larger ontology
- There is no universal way to modularize an ontology

[10] Khan, Z. C., & Keet, C. M. (2015). An empirically-based framework for ontology modularisation. Applied Ontology, 10(3-4), 171-195.

CRM modularization

Goal:

- Selective use, development, and maintenance
- Formal representation in **OWL** (based on Erlangen release[11])

Example:

- Modeling of the mereological/topological structure of a man-made object
- Modeling of a social group (e.g., a group of artists)

without linking to temporal information

[11] `https://github.com/erlangen-crm/ecrm`

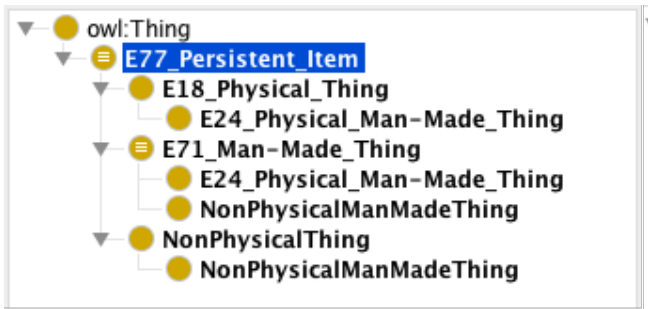
CRM modularization: Overview

The library of modules includes **18 modules** (preliminary work):[12]

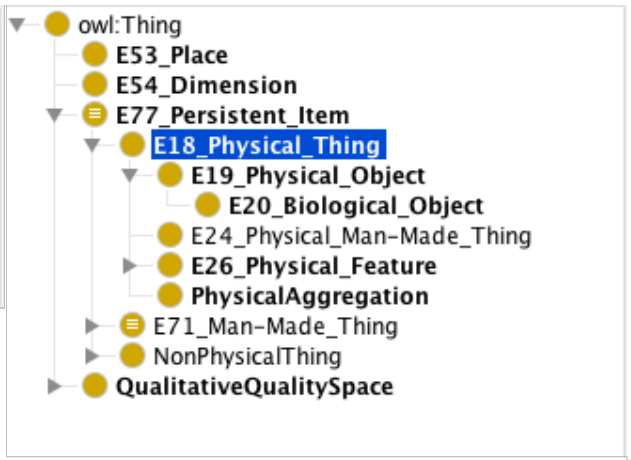
- 6 modules for persistent items
- 8 modules for temporal entities
- dimension-module (covers qualitative values)
- place-module
- top-module: the highest classes of the ontology
- whole: union of all modules (whole CRM ontology)

[12] https://github.com/emiliosanfilippo/cidoc-modularization

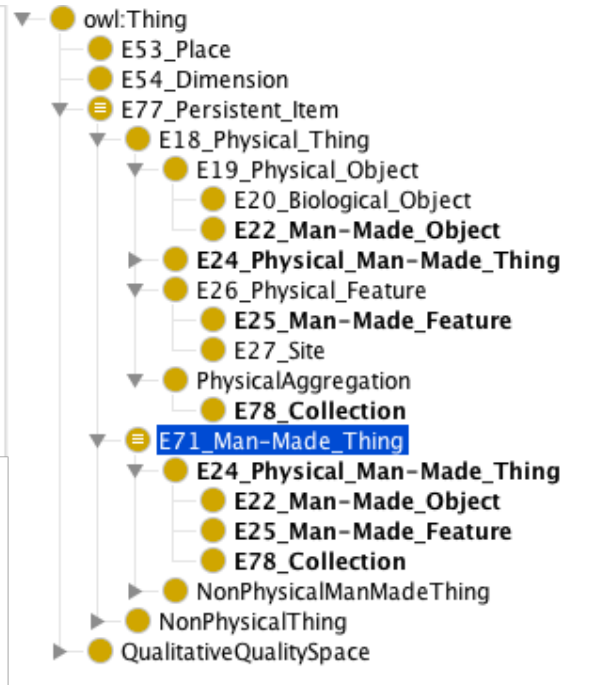
CRM modularization: persistent items (hints)



Persistent item top



Physical Things



Artefacts

CRM modularization: Example

Assume you need to represent the **physical structure** of a man-made object (e.g., a car with cultural heritage value), dimensions included

- `cidoc:artefact-module` is all you need

If you need to add **temporal information** about the production event, including information about the **creator**, you need to import also

- `cidoc:actor-module`
- `cidoc:modification-activity-module`

Remarks: Relations with disjunctions

Re-engineering of relations using disjunctions, e.g.,

- *P53 has former or current location*: two different meanings (former vs. current location)
  - It subsumes *P55 has current location*: what about *has former location*?
  - **Proposal:**
    - Either use P53 (and all similar relations) as a general modeling relation subsuming 2 relations (has former location and has current location) OR
    - Avoid relations with disjunctions

Remarks: Cardinality restrictions

Cardinality restrictions (*quantifications*), e.g,:

- All physical things (E18) *consists of* (P45) material (E57)
- E18 subsumes *E26 Physical Feature*
- E26 covers things like scratches and holes [see CRM, p.15] which are commonly understood as **immaterial entities** in formal ontology[13]

IF holes and similar features are immaterial entities in CRM, the cardinality of P45 likely needs revision

[13] See Casati, R., & Varzi, A. C. (1994). Holes and other superficialities, MIT

Remarks: Shortcuts

For data modeling purposes with SW technologies/languages, introduce **shortcuts** possibly by reusing existing LOD vocabularies

- object o **created** in date d
  - E.g., **Dublin Core**: http://purl.org/dc/terms/created – **shortcut for**

Def: $created(o, d) \equiv$
$PhysicalManMadeThing(o) \wedge Date(d) \wedge \exists e, t(Production(e) \wedge$
$hasProduced(e, o) \wedge hasTimeSpan(e, t) \wedge identifiedBy(t, d))$

Conclusions

Ontological analysis

- It could be useful to compare CRM with existing modeling theories used in ontology engineering

Formalization

- Considering the massive use of SW technologies and languages, a stable OWL version of CRM is a desiderata

Modular structure

- **Advantage:** selective exploitation and maintenance
- (Possible) **Disadvantage:** architecture runs the risk of becoming complex; it may require more cognitive effort to be learnt especially by novel users

Thank you very much!

Emilio M. Sanfilippo
Laboratory for Applied Ontology ISTC-CNR
emilio.sanfilippo@cnr.it

Paper presented at FOIS 2020 about CIDOC-CRM, please check the conference website or send me an email