Oral Presentation

An Information Artifact Ontology Perspective on Data Collections and Associated Representational Artifacts

August 27, 2012 – Palazzo dei Congressi, Pisa, Italy

Werner CEUSTERS, MD
Center of Excellence in Bioinformatics and Life Sciences, Ontology Research Group
and
Department of Psychiatry

University at Buffalo, NY, USA

http://www.org.buffalo.edu/RTU
Introduction: data generation and use

- Observation & measurement
- Data organization
- Model development
- Further R&D (instrument and study optimization)
- Application

Δ = outcome

Generic beliefs

Add, verify, use
A crucial distinction: data and what they are about

First-Order Reality

observation & measurement

\[ \Delta = \text{outcome} \]

Representation

data organization

Further R&D
(instrument and study optimization)

Model development

Generic beliefs

Application
Data must be unambiguous and faithful to reality …
... even when reality changes

- Are differences in data about the same entities in reality at different points in time due to:
  - changes in first-order reality?
  - changes in our understanding of reality?
  - inaccurate observations?
  - registration mistakes?

Methods to achieve faithfulness and data clarity

<table>
<thead>
<tr>
<th>Sources</th>
<th>Data generation</th>
<th>Data organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection sheets</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Instruction manuals</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Interpretation criteria</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Diagnostic criteria</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Assessment instruments</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Terminologies</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Data validation procedures</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Data dictionaries</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ontologies</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

If not used for data collection and organization, these sources can be used post hoc to document, and perhaps increase, the level of data clarity and faithfulness in and comparability of existing data collections.
OPMQoL: Project goals

• to obtain better insight into:
  – the complexity of pain disorders, pain types as well as pain-related disablement and
  – its association with mental health and quality of life,
• to develop an ontology for this subdomain incorporating a broad array of measures consistent with a biopsychosocial perspective regarding pain,
• to integrate five existing datasets that broadly encompass the major types of pain in the oral and associated regions.
Considered datasets

- ‘US Dataset’ (724 patients) resulted from the NIH funded RDC/TMD Validation Project,
- ‘Hadassah Dataset’ (306 patients) from the Orofacial Pain Clinic at the Faculty of Dentistry, Hadassah,
- ‘German Dataset’ (416 patients) of patients seeking treatment for orofacial pain at the Department of Prosthodontics and Materials Sciences, University of Leipzig,
- ‘Swedish Dataset’ of 46 consecutive Atypical Odontalgia (AO) patients recruited from 4 orofacial pain clinics in Sweden as well as data about age- and gender-matched control patients,
- ‘UK Dataset’ (168 patients) of facial pain of non dental origin present for a minimum of three months.
Most important challenge

• The data sets cover more or less the same domain, but …

• the data within each data set are collected independently from each other, with distinct, partially overlapping data collection and data organization tools.
Vision: linking data using distinct assessment instruments
This talk …

• … covers just the first step in this endeavor.

• Goals:
  – to obtain a clear understanding of how the various information sources made available to the project relate to each other,
  – find ways for how this understanding can contribute to further advancing our insight in how information in general precisely relates to that what it is information about.
Challenges

1. Alignment of
   – the terminological perspective according to which the assessment instruments and data collections are designed on the one hand,
   – with the (realism-based) ontological perspective on the other hand;

2. Carry out this alignment in line with the principles of Ontological Realism.

Additional materials used

• Two relevant initiatives follow Ontological Realism:
    • describes the boundaries of each, but does not give a unifying perspective
    • describes information artifacts but is vague about what information is.
Additional materials used (2)

- Two papers covering “representational units”
Methodology

• manual analysis of the OPMQoL source materials in function of
  – the Information Artifact Ontology,
  – the Klein/Smith paper,

• definition of the most generic types of compositional elements of the sources and the sources as a whole,

• classification of these types in the taxonomy of the IAO and further clarification of the relationships amongst them in a UML-diagram,

• where deemed required, RUs were added to the IAO and modifications to existing IAO definitions proposed.
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Content Entity (ICE)</td>
<td>an entity that is generically dependent on some artifact and stands in relation of aboutness to some portion of reality [4]</td>
</tr>
<tr>
<td>Representational Artifact (RA)</td>
<td>an ICE which is believed to represent a portion of reality external to the representation (modified from [5])</td>
</tr>
<tr>
<td>Representational Unit (RU)</td>
<td>a RA which according to the structural conventions it is designed, is not built out of any other RAs</td>
</tr>
<tr>
<td>Denotator</td>
<td>a RU which denotes directly an entity without providing a description [6]</td>
</tr>
<tr>
<td>Term</td>
<td>a RU which is a general expression in some natural language used to refer to portions of reality (modified from [5])</td>
</tr>
<tr>
<td>Composite Representation</td>
<td>a RA built out of constituent sub-representations as its parts (modified from [5])</td>
</tr>
<tr>
<td>Data Collection</td>
<td>a composite representation built out of measurement data</td>
</tr>
<tr>
<td>Data Dictionary</td>
<td>a composite representation describing, inter alia, what data items in a data collection are about, including a data format specification</td>
</tr>
<tr>
<td>Terminology</td>
<td>a RA consisting of terms (modified from [5])</td>
</tr>
<tr>
<td>Ontology</td>
<td>a RA comprising a taxonomy as proper part, whose RUs are intended to designate some combination of universals, defined classes, and certain relations between them [3]</td>
</tr>
<tr>
<td>Realism-based Ontology</td>
<td>an ontology built out of RUs which are intended to be exclusively about universals and certain relations between them, intended to mimic the structure of reality, and which correspond to that part of the content of a scientific theory that is captured by its constituent general terms and their interrelations [3]</td>
</tr>
</tbody>
</table>
Results (2): relationships

Terminology component

- **term**
  - part-of 0..*
  - has-part 0..*
  - 1..* expressed-by
  - 1 means
  - broader 0..*
  - narrower 0..*
  - 0..1 expresses
  - corresponds-to 0..*

- **concept**
  - 1..* expressed-by
  - 1..* narrower

- **terminology**
  - 1

Data component

- **data dictionary**
  - 1 uses
  - 1..* used-for

- **data collection**
  - Uses
  - 1..* used-for

- **measurement datum**
  - 1

- **data collection ontology**
  - 0..1 uses
  - 1..* used-for

- **assessment instrument ontology**
  - 0..1 uses
  - 1..* used-for

- **assessment instrument**
  - 0..1 uses

Ontology component

- **representational artifact**
  - 1
  - 1..* expressed-by

- **bridging axiom**
  - 1 uses
  - 1..* used-for

- **ontology**
  - 1
  - 1..* used-for

- **denotator**
  - 1
  - 1 denotes

Reference ontology

- **ontology**
  - 1
  - 1..* used-for
Key terms

- **concept**: meaning of a term as agreed upon by a group of responsible persons,
- **entity**: anything which is either a universal or an instance of a universal,
- **portion of reality**: any entity or configuration of entities standing in some relation to each other.
Discussion

• Most significant change proposals for IAO:
  – addition of *Representational Unit* (RU) and *Representational Artifact* (RA),
  – *aboutness* targets *portions of reality* rather than only *entities*,
  – distinction between 'just' *being about* a portion of reality and *representing* a portion of reality.
  • False or misleading information is still *about* something, but does not *represent* that something,
  • Avoids underspecification.
Discussion (2)

Bridging axioms are able to glue terminologies and ontologies together without resorting to solutions incompatible with Ontological Realism.
Discussion (3)

- Compare with an OR-incompatible proposal:
Conclusion

• It is possible to use terminologies – even concept-based terminologies – and realism-based ontologies in one framework that remains compatible with Ontological Realism.

• Future work is required:
  – more precise treatment of aboutness, representation and denotation
  – relationship with thoughts and mental representations.
Acknowledgements

• National Institute of Dental and Craniofacial Research
  – The work described is funded in part by grant 1R01DE021917-01A1 from the National Institute of Dental and Craniofacial Research (NIDCR). The content of this presentation is solely the responsibility of the author and does not necessarily represent the official views of the NIDCR or the National Institutes of Health.

• Alan Ruttenberg (IAO custodian)