Ontology of Innovation Cases
Approach and Process via Conceptualisation to FAIR Domain Ontologies

Emanuele Ghedini – UNIBO
Ontology As Language

\[ i\hbar \frac{d}{dt} |\Psi(t)\rangle = \hat{H} |\Psi(t)\rangle \]

PHYSICS = MATH = APPLIED SCIENCES KNOWLEDGE AND DATA

Ontology
Ontology As Language

**Universal Language**
(i.e. transcend linguistic differences)

**Hard to learn**
(i.e. requires adequate training)

**Able to Perform Logical Reasoning**
(i.e. infer new knowledge from existing one)

**Ontology**

Provides Models of the World
(i.e. let us play with abstractions instead of material things)

When It Works, It Can Be Hidden
(i.e. things using it may confine its complexity behind the user interface layer)

\[ i\hbar \frac{d}{dt} |\Psi(t)\rangle = \hat{H} |\Psi(t)\rangle \]

**Math**
Ontology As Language

**Ontologies** can be the key for a quantum leap in **Applied Sciences** and **Industry** data (e.g. pervasive digitalization, knowledge sharing, Industry 4.0 to 5.0)

**Math** has been the key for the highest achievements in **Physics** (e.g. Newton, Maxwell, Einstein, Schrodinger, Standard Model)
Talking to Machines

Machines understand the logical language of ontologies (e.g. FOL, OWL-2), and already can be used to:

- **document data** (e.g. dcterms, DCAT)
- infer **new knowledge** (e.g. resolvers)
- support **AI** (e.g. ontology assisted AI)
Physics uses Quantities (numbers) to model the world
Ontologies use Concepts to build a (onto)-logical representation of the world
Jumping from Ontology to World requires Interpretation

Several conceptualizations exist for the same things, so that almost each human being is going to provide a different definition for a single term.

(physics is not affected since it works with quantities, and solve the issue with well defined measurement practices)
Everything we create to satisfy our needs begins with a **human meaningful input** and a **human meaningful output** (e.g. food, sound, product, action, picture, dataset).

E.g. Von Neumann architecture without I/O is a useless machine!

An ontology provides concepts to **semantically categorise such input and outputs**, according to a particular perspective, so that a user can understand what type of entities a particular tool may provide.

E.g. ChatGPT: good for assemble realistic language structures, but no idea about the meaning
Talking to Machines

Ontology Entities

- **class**
- **relation**
- **individual**

Example of axioms:
```lotus
ClassAssertion(:a :Car)
ClassAssertion(:r :Color)
ObjectPropertyAssertion(:hasColor :a :r )
```

Real-world objects
No specific ontological commitment about the meaning of ‘real’ and ‘object’ in OWL 2. Relying on common sense.
Talking to Machines

A great boost towards the use of formal ontologies in practice came in the ‘90 from the Semantic Web extension of the World Wide Web, thanks to the W3C standardization activities.

The objective Semantic Web is to make Internet data machine-readable.

Ontologies (in particular formal ontologies) play an important role in the Semantic Web and are placed in the higher levels of the architecture of languages.

However, the scope and ambition of such ontologies differ substantially from the ones coming from the philosophical community.

But an important concept as been introduces: the need for practical applications!
Talking to Machines

An OWL 2 Ontology is formally expressed in a persistent form by axioms declarations following a specific syntax (e.g. ASCII file with Turtle syntax)

```
@prefix : http://www.semanticweb.org/example/ontologies/example# .
@prefix a : http://www.w3.org/2002/07/owl# .
@prefix rdfs : <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl : <http://www.w3.org/2003/01/geo.owl#> .
@prefix owlxml : <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs2 : <http://www.w3.org/2003/01/geo/rdfs2#> .

http://www.semanticweb.org/example/ontologies/example.owl

---

### http://www.semanticweb.org/example/ontologies/example\ontology
+ hasColor rdf:type owl:ObjectProperty ;
  rdfs:subPropertyOf rdfs:ObjectProperty ;
  rdfs:range owl:Color .

### http://www.semanticweb.org/example/ontologies/example\Car
+ rdf:type owl:Class ;
  rdfs:subClassOf [ rdf:type owl:Restriction ;
    owl:onProperty hasColor ;
    owl:someValuesFrom owl:Color ] .

### http://www.semanticweb.org/example/ontologies/example\Flower
+ rdf:type owl:Class ;
  rdfs:subClassOf [ rdf:type owl:Restriction ;
    owl:onProperty hasColor ;
    owl:someValuesFrom owl:Color ] .

### http://www.semanticweb.org/example/ontologies/example\Red
+ rdf:type owl:Class ;
  rdfs:subClassOf owl:Color .

### http://www.semanticweb.org/example/ontologies/example\Yellow
+ rdf:type owl:Class ;
  rdfs:subClassOf owl:Color .

### http://www.semanticweb.org/example/ontologies/example\Green
+ rdf:type owl:Class ;
  rdfs:subClassOf owl:Color .

### http://www.semanticweb.org/example/ontologies/example\Blue
+ rdf:type owl:Class ;
  rdfs:subClassOf owl:Color .

---

```

---

11
Facilitate Users’ Life

TRANSLATORS

INDUSTRIAL USERS

Human Friendly Template

Graphical Conceptualisation

ONTIOLOGISTS

Formal Ontology
Terminology I (how we understand each others...)

**Innovation Challenge**: the high-level innovation objective addressed by specific cases expressed in natural language (e.g. master the process parameters, agile response to product, market and regulatory requirements)

**Innovation Case (or Application)**: a specific class of cases (e.g. design, manufacturing, sales) for which the challenge applies, involving processes and material objects and expressed in some human readable form (e.g. diagrams, flow charts).

**User Case**: a specific implementation of an innovation case for the user (e.g. a step in the manufacturing of a specific product, a manufacturing line of a company).

**Workflow**: data workflow expressed using MODA (or MODA-like) graphs representing the combination of methods for the prediction of properties of our objects/processes of interests (the totality or part of the user case).
Terminology II (how we understand each others...)

**Knowledge Source**: everything that can be used to provide new knowledge from existing one. It needs an INPUT (e.g. case set up for simulation, query for a database) and provide an OUTPUT.

**Knowledge Generator**: a knowledge source that generate new non-existing knowledge (e.g. a model that provides calculated properties for an object, the INPUT is the setup case and the OUTPUT are the properties)

**Knowledge Repository**: a knowledge source that provides already-existing knowledge (e.g. a database of properties, the INPUT is the name of the substance and OUTPUT is the desired property)
An innovation case is formalised as an ontological entities workflow with the purpose of identifying all the entities that play a role in the case to be referenced by properties and models.

A simplified object/process approach is chosen to facilitate understanding by users without experience in ontologies.
Innovation Case

List of all properties we need to deal with, for each innovation case.

Please consider that in the EMMO everything we know about an entity must come as a property, which is a symbolic representation of the object itself defined for a specific class of interpreters within a well defined observation process.

<table>
<thead>
<tr>
<th>Property Names</th>
<th>Property Description (semantics)</th>
<th>Data Description (symbolic)</th>
<th>Potential Interpreters</th>
<th>Source Data File Format</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Data Type</td>
<td>Data Range</td>
<td>Units</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measured by</td>
<td>Modelled by</td>
<td>Assigned by</td>
</tr>
</tbody>
</table>
User Cases

• **User Cases** are subsets of a specific **Innovation Case**, expressing the different user approaches to address the Innovation Challenge.

• Different User Cases for the same real world innovation case can be proposed, each one focusing on different set of properties or level of details of the objects, according to specific design objectives.

• User Cases are represented like the Innovation Case, by selecting objects (or macro objects), processes and properties of interest.

• Properties of real world objects/processes must be defined as problem inputs or KPIs, for each User Case.
Workflows

Workflows are intended as **data workflow**.

Workflows describe how **properties** are obtained by means of **tools** (e.g. simulation software) or **methodologies** (e.g. measurement).

Workflows are made of a connected **sequence** of **knowledge sources** (e.g. models, databases, experiment).

A knowledge source **stands for** (i.e. substitute, act as, predict the outcome of) a real process.

Workflows must be represented using **MODA-like** graphs, and refers in each step the real world properties and real world processes of the user case.
Workflows

Build one or more potential data workflow connecting data representing properties of real world entities part of the user case.
Generic Example

<table>
<thead>
<tr>
<th>Object</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Powder</td>
<td>Micrometric aluminium particles</td>
</tr>
<tr>
<td>Precursor</td>
<td>Size controlled aluminium particles</td>
</tr>
<tr>
<td>Reactor Output</td>
<td>Aluminium nanoparticles aggregates coming from the reactor</td>
</tr>
<tr>
<td>Final Product</td>
<td>Sonicated aluminium nanoparticles</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Process</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sieving</td>
<td>The process of selecting particles according to...</td>
</tr>
<tr>
<td>Plasma Synthesis</td>
<td>Vaporisation of materials through a plasma reactor for ...</td>
</tr>
<tr>
<td>Sonication</td>
<td>Application of sound energy to break nanoparticles aggregates into smaller...</td>
</tr>
</tbody>
</table>

OntoTrans Open Workshop II – 7th September 2023
## Generic Example

<table>
<thead>
<tr>
<th>Entity</th>
<th>Property Names</th>
<th>Property Description (semantics)</th>
<th>Data Description (symbolic)</th>
<th>Potential Interpreters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Powder</td>
<td>Powder's mean diameter</td>
<td>The mean diameter of powder's particles.</td>
<td>Data Type: Real scalar; Data Range: Positive; Units: m; Reference Physical Quantity: Diameter; Measured by: -; Modeled by: -; Assigned by: Seller; Other: -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Powder's Price</td>
<td>The mass price of the powders.</td>
<td>Data Type: Real scalar; Data Range: Positive; Units: €/kg; Reference Physical Quantity: Price per kg; Measured by: -; Modeled by: -; Assigned by: Seller; Other: -</td>
<td></td>
</tr>
<tr>
<td>Reactor Output</td>
<td>Nanoparticles mean diameter</td>
<td>The mean diameter of nanoparticles collected after the synthesis.</td>
<td>Data Type: Real scalar; Data Range: Positive; Units: m; Reference Physical Quantity: Diameter; Measured by: Internal Lab; Modeled by: Synthesis Model; Assigned by: -; Other: -</td>
<td></td>
</tr>
<tr>
<td>Aluminium Nanoparticles</td>
<td>Aluminum nanoparticles mean diameter</td>
<td>The mean diameter of nanoparticles collected after the sonication.</td>
<td>Data Type: Real scalar; Data Range: Positive; Units: m; Reference Physical Quantity: Diameter; Measured by: Internal Lab; Modeled by: Sonication Model; Assigned by: -; Other: -</td>
<td></td>
</tr>
<tr>
<td>Plasma Synthesis</td>
<td>Treatment time</td>
<td>The process time.</td>
<td>Data Type: Real Scalar; Data Range: Positive; Units: s; Reference Physical Quantity: Time; Measured by: Device Timer; Modeled by: -; Assigned by: -; Other: -</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plasma Power</td>
<td>The power of the plasma torch in the reactor.</td>
<td>Data Type: Real Scalar; Data Range: Positive; Units: W; Reference Physical Quantity: Power; Measured by: Reactor instrumentation; Modeled by: Operator; Assigned by: -; Other: -</td>
<td></td>
</tr>
<tr>
<td>Sonication</td>
<td>Sonication power</td>
<td>The power of the sonicator.</td>
<td>Data Type: Real Scalar; Data Range: Positive; Units: W; Reference Physical Quantity: Power; Measured by: Sonicator instrumentation; Modeled by: Sonication ML model; Assigned by: Operator; Other: -</td>
<td></td>
</tr>
</tbody>
</table>
Generic Example

**Innovation Challenge:**
Produce Al nanoparticles of selected diameter

**KPI:**
Plasma Power (e.g., the focus is on minimising process power)
Representational Capabilities

This was an easy example of conceptualisation, based on object, processes, properties, and a simple workflow.

But if we want to represent complex materials, structured datasets, multi-step workflows, we need an ontology with an adequate representational power.
Elementary Multiperspective Material Ontology (EMMO)

Started by practitioners in Materials Science in 2018 in order to produce a framework consistent with scientific principles and methodologies.

Developed in a number of projects under the governance of the European Materials Modelling Council (https://emmc.eu/)

**Elementary:** it is a fundamental, top-level ontology, while at the same time assuming the existence of objects that cannot be divided further in space and time

**Multiperspective:** it enables scientists to describe the world by resorting to different views or perspectives, thus providing a high expressive power and versatility

All relationships among objects can be of three types:
- **Causal** (connectivity)
- **Mereological** (parthood)
- **Semiotic** (entities that refer to other entities according to an agent)
Radical Physicalism

We decided to completely embrace a physics-based description of the world, starting from the more fundamental and universally accepted knowledge about the nature, from Ancient Greek φύσις (physis): the Standard Model of Particle.

The SMP is the modern version of the ancient philosopher’s ontology: it tells you everything about the world, up to its fundamental beings, in Ancient Greek ὄντα (onta)
Radical Physicalism

This is the taxonomy of the **EMMO SMP classes**, covering all possible types of fundamental entities (**Quantum**).
Radical Physicalism

Then we had to deal with **physical phenomena** describing how particles interacts. We decided to build a causality theory in order to represent the **Quantum Field Theory** interactions, using **Feynman diagrams** representations.

It was amazing to see the theory grew almost naturally, and how easily the Feynman diagrams rules led to a **causal graph representation** of physical phenomena as **Direct Acyclic Graphs** with a **maximal graph dimension of 4**, totally inline with our **four-dimensional** approach.

![Diagram](image)

**Figure 1.** EMMO’s Core Naturalistic Commitments (the arrows stand for dC relations)
In order to go beyond the standard model, we decided to go for the fusion of mereology and causality, creating a new discipline called mereocausality, that has been formalised in First Order Logic.

The adoption of AGEM theory also led to the definition of universe: the smallest and the largest entities are described using these 13 axioms.

The EMMO mereocausality theory is fully capable of representing everything that exist... but is a bit too detailed for practical usage!!!
EMMO represent world entities as a **causal network** (a Direct Acyclic Graph) of **quantum entities** whose types and interactions are governed by the **Standard Model of Particles**.
The EMMO defines **spatial and temporal relations** according to their mereocausality structure. The **rules** derived from Feynman diagrams are paramount for the creation of such relations i.e., spacetime emerges thanks to them.

**Spatial and Temporal Relations**

**Graphical Mereological Relations**

Express the mereological relations between concepts through a 4D graphical representation using the following schemes representing how objects and processes can relate each other. Each case is labelled according to how the grey boxes relate to the white ones.

**Proper Overlap**

Two entities that share some of their parts, without being one part of the other. All these relations are subrelation of **isProperOverlapOf**.

**Graphical Causal Relations**

Express the causal relations between concepts through a 4D graphical representation using the following relations, expressed by **black** arrows.

**Temporal Causality**

Describes the causal relations between entities that unfold in time.

**Spatial Causality**

Describes the causal relations between entities that unfold in space (symmetric).
Interactions between quantum systems are described by the **Standard Model Theory (QFT)**, expressed usually through **Feynman diagrams**, enabling the highest possible ontological detail and objectivity in the representation of materials.
Interactions are represented as mediated by **real particles** (i.e., a causal chain) so that it is possible to apply **non-quantum** approaches (i.e., classical or relativistic).

A **classical field** is represented as a **stream of real particles**.
Quantum Fields

Two-way interactions are mediated by virtual particles (symmetric relations) and represents space-like interactions between entangled particles, connecting the entities used to build the Hamiltonian for a quantum mechanical representation. A quantum field is represented as a symmetric relation.
Representing the World

The EMMO needs to represent all the entities between SMPs up to the universe. How do we deal with e.g., composite particles, atoms, molecules, materials, phases, solids, mixtures?
Multiscale Representation

The EMMO uses mereocausality to classify entities according to their structures, distinguishing between concepts like physical phenomena (e.g., scattering) or objects (e.g., atom).

This provides a seamless multigranular representation of world entities: exactly what you would like a material ontology to provide!
From Elementaries to Materials

Why we need a so deep fundamental level of detail in the description of materials?
Because we want:
- to speak the language of material scientists (e.g. physics, chemistry)
- to enable unambiguous representation of world entities ready for been translated into one of the existing physics-mathematical models (e.g. quantum mechanics)
- to enable multi-scale representation of world entities
- to provide a strong foundation for material-based taxonomies
Physical Quantities

We wanted to build a representational system able to embed ISO 80000 International System of Quantities, the Vocabulary of Metrology, and the SI International System of Unit in a way that is clearer and more expressive than existing ontologies (e.g., QUDT).

To do so we separated the semantic from the syntactic part in the representation of physical quantities.
Upper Metrological Concepts
SI Units
International System of Quantities
Physical Quantities

0.5 km

DATA VALUE
PREFIX
UNIT SYMBOL
MEASUREMENT UNIT

SEMANTIC MEANING
(e.g., length)
The Full Picture

Generic ontological representation of a modelling task, related to properties, data and material type.
Knowledge Management

We can create a Knowledge Base based on EMMO, that is built upon existing Data Management solutions, implementing data federation in practice.

Ontologies are a great way to make your data (or better, your knowledge) FAIR.

Data Federation is possible i.e. data can be made directly reachable from the KB through mapping with the graph database vendor querying system.

Data are not directly reachable from the KB but provide users with information about their location and accessibility.
The acronym CIF is used both for the Crystallographic Information File and for the Crystallographic Information Framework, a broader system of exchange protocols based on data dictionaries and relational rules expressible in different machine-readable manifestations.

The EMMO CIF ontology maps all concepts of CIF-DDL and CIF-CORE and provides a multiperspective approach to represent CIF data and to connect them to real world materials.
Conclusions

• The use of ontologies as human-to-machine and machine-to-human interfaces requires an approach that facilitates the participation of industrial users in the conceptualisation step (i.e. OntoTrans template).

• Human Translators are needed to guide the industrial users and formalise the user case in ontological form.

• The EMMO is used to build ontology modules to represent all the steps of Innovation Case to design, execution and documentation.

• Ontological representation is used to populate a knowledge base (OntoKB) that can be navigated by agents to harvest existing data, or used as training set for AI approaches towards automated decision making or workflow design.
In Layman Terms...

The ontology is expected to **represent**, to **document** and to **connect** together:

- the **innovation** and **user case** (i.e. the things: object and processes)
- their **properties** (i.e. the data that describes the things)
- the **tools** used to generate (e.g. models, characterisation) or retrieve (e.g. databases) the required properties (i.e. the knowledge sources)
- the **data workflow** used to generate the desired properties (i.e. the KPIs)
The OntoTrans project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 862136.