ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

Ontology Best Practice

Arkopaul Sarkar(ENIT)

NI4OS
National Initiatives for Open Science in Europe

OntoComm ons “Ontology-driven data documentation for Industry Commons” has received funding from the European Union’s Horizon Programme call H2020 -NMBP-TO-IND-2020-singlestage, Grant Agreement number 958371

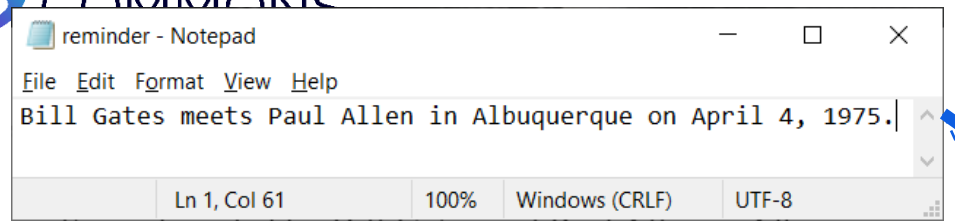
ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

"PEOPLE CANNOT SHARE KNOWLEDGE IF THEY DON'T SPEAK A COMMON LANGUAGE."
(THOMAS DAVENPORT, 1997)

- Common symbol (syntax)
- Common meaning (semantic)
- Common intention (semiotic)

Indus valley seals, 2500–2400 B.C.E

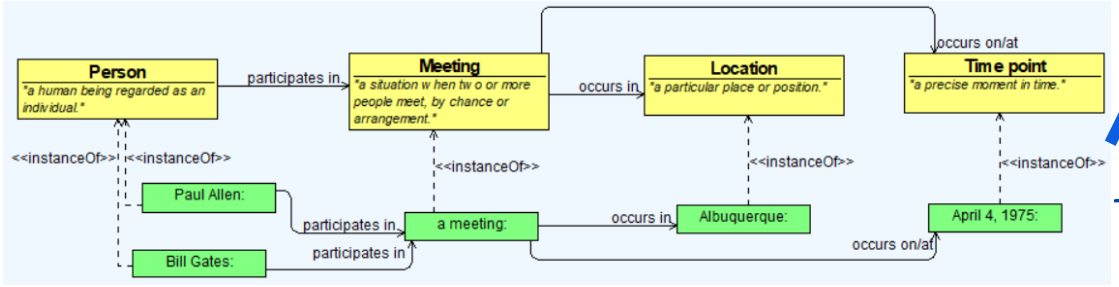
ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS



Free text, no syntax!

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<note>
  <to>Biil Gates</to>
  <from>Paul Allen</from>
  <heading>reminder</heading>
  <body>meets in Albuquerque on April 4, 1975</body>
</note>
```

Structured information, syntax but no semantic!



Knowledge graph, both syntax and semantic!

In simple terms, ontology uses a set of formalized objects, relations, concepts, and properties to help both *human* and software agents to interpret data.

What is an ontology?

Corpus Aristotelicum

A philosophical study belonging to the branch Metaphysics.

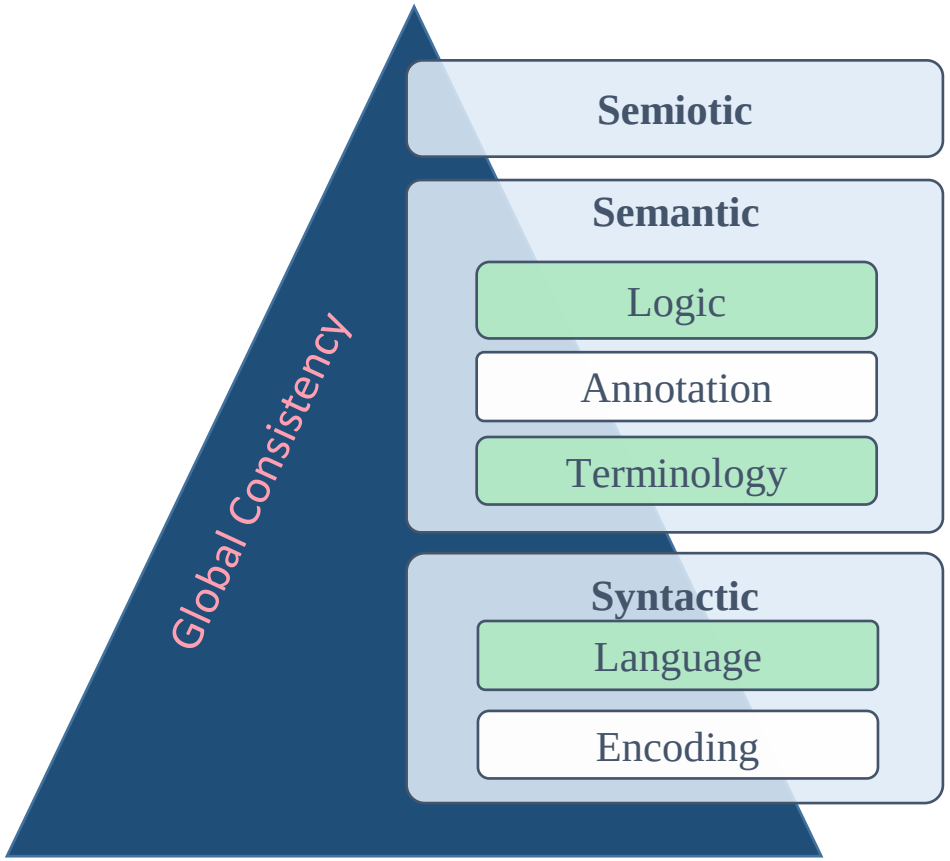
Ontology is the philosophical study of the nature of being, existence, or reality, as well as the basic categories of being and their relations…

Computer science considers a more pragmatic definition. An ontology is a…

Formal, explicit specification of a shared conceptualization. (Studer, 98)

- Machine readable
- Concepts, properties, functions, axioms are explicitly defined
- Consensual knowledge
- Abstract model of some domain

Concerns of ontology development



Interpretation, Context

Formal system of mathematics to validate the ontology model and perform reasoning.

Labels, identifiers, and other description associated with terms and ontology file.

Classes and relations

Formal languages for construction (OWL, RDF, CL, KIF etc.)

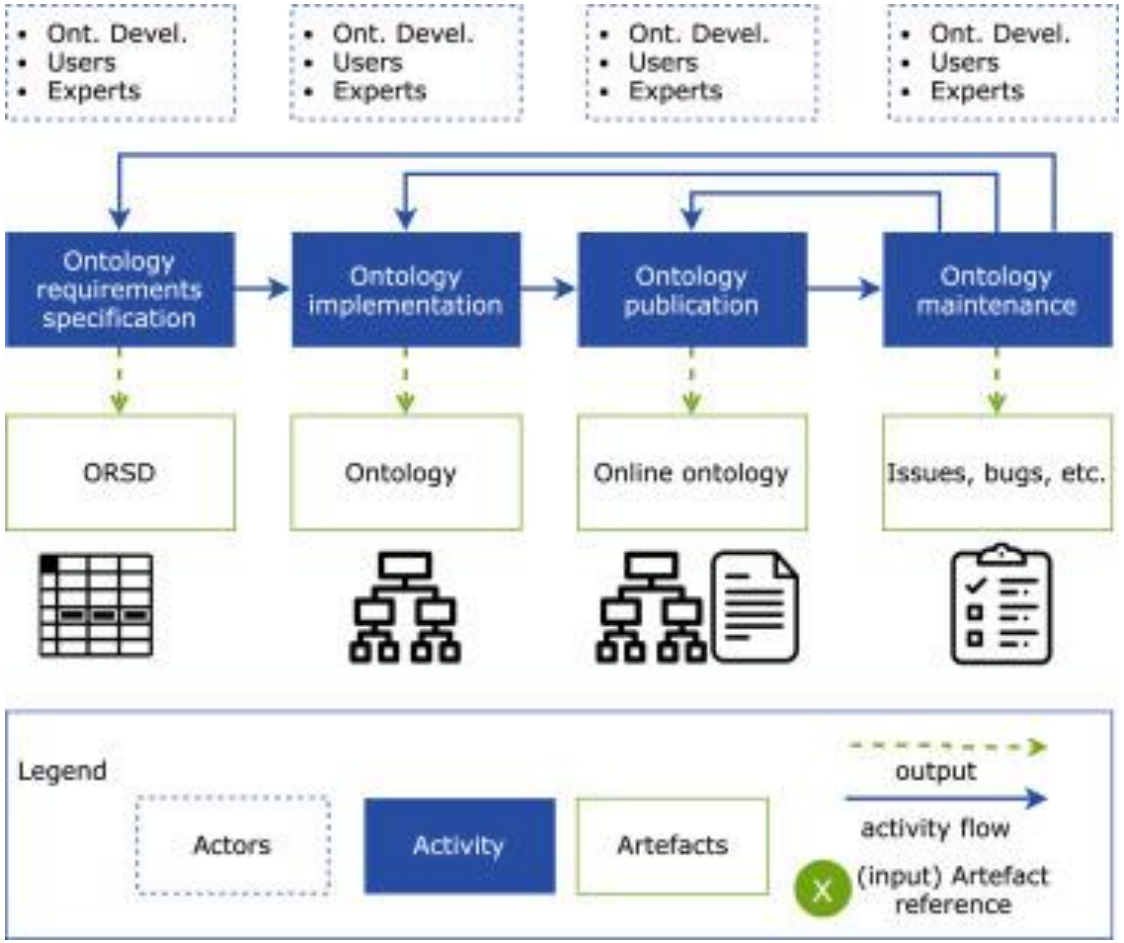
Format of writing/storing (serialization) ontology in a file (XML, Turtle, Manchester etc.)

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

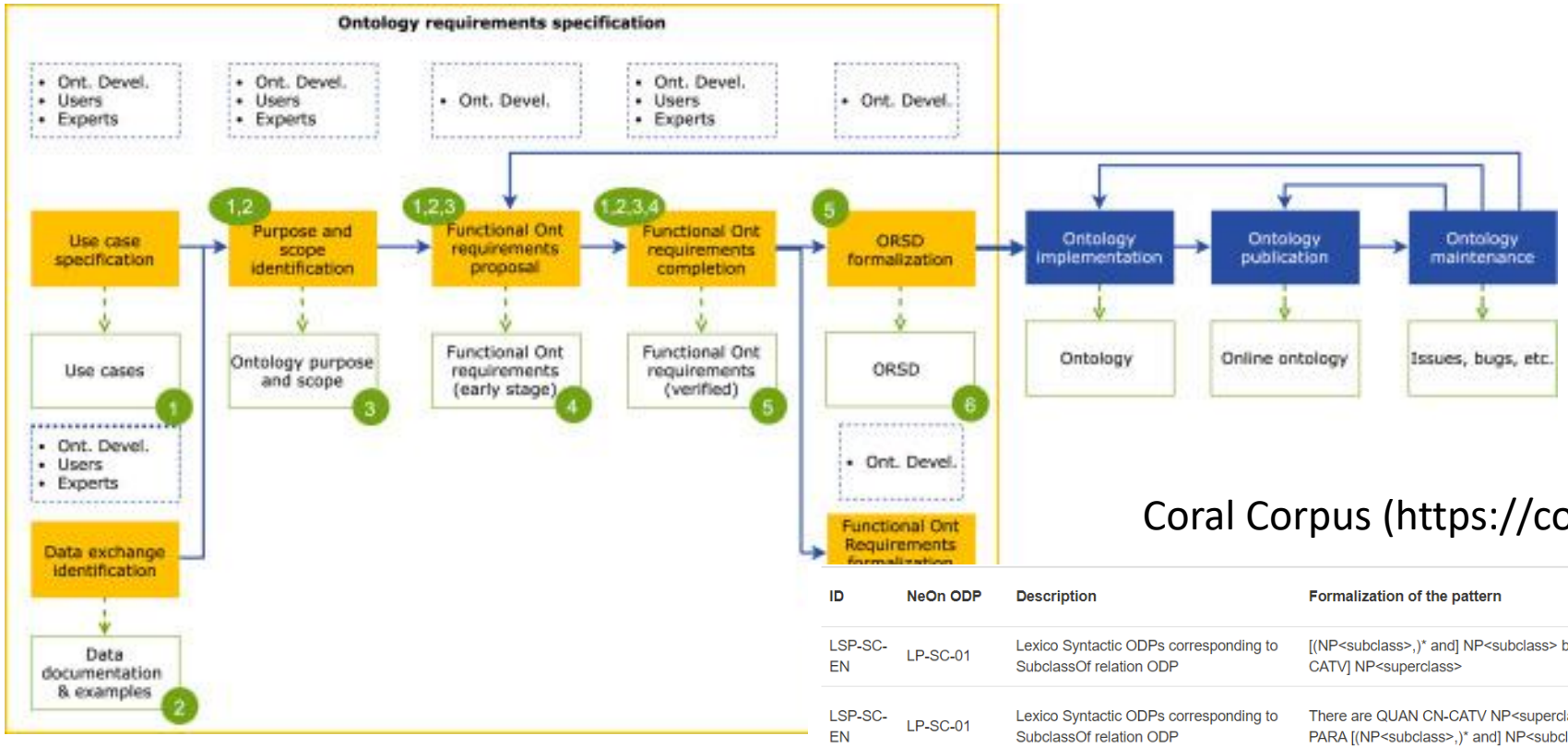
Ontology Development Methodology

Methods	Author	Life-cycle	Requirement Gathering Method	Source of Terms/relationship	Direction of scope
TOVE	Grüninger & Fox, '95	Yes	Competency Question	Competency Question	Middle-Out
EO	Uschold & King, '95	Yes	Specification	Domain knowledge	Middle-Out
Methontology	Fernández et al.	Yes	Competency Question, Expert Knowledge, Text	Specification, Expert Knowledge	Middle-Out
KBSI IDEF5		Yes	completion criteria	Data analysis, Expert Knowledge	NA
KAKTUS	Benaras et al. '94	No	Specification of application(s)	Pre-existing ontologies	Top-Down
Plinius	Mars et al. '94	No	completeness criteria	NL text, Formal Theory	Bottom-up
ONION	Steve & Gangemi '96	No	completeness criteria	Pre-existing ontologies	Aggregation
SENSUS	Swartout et al. '97	No		NL text	Bottom-up

Linked Open Terms (LOT) methodology



Requirement Engineering



Coral Corpus (https://coralcorpus.linkeddata.es/)

ID	NeOn ODP	Description	Formalization of the pattern	Example of usage	OWL constructs	DL Expresivity
LSP-SC-EN	LP-SC-01	Lexico Syntactic ODPs corresponding to SubclassOf relation ODP	[(NP<subclass>,)* and] NP<subclass> be [CN-CATV] NP<superclass>	An orphan drug is a type of drug / One of the users is the manager	subclassOf, Class (Thing, Nothing)	AL
LSP-SC-EN	LP-SC-01	Lexico Syntactic ODPs corresponding to SubclassOf relation ODP	There are QUAN CN-CATV NP<superclass> PARA [(NP<subclass>,)* and] NP<subclass>	There are several kinds of memory: fast, expensive, short term memory and long-term memory	subclassOf, Class (Thing, Nothing)	AL
LSP-SC-EN	LP-SC-01	Lexico Syntactic ODPs corresponding to SubclassOf relation ODP	[A(n) | QUAN] example of | CN-CATV NP<superclass> be |include [PARA] [(NP<subclass>,)* and] NP<subclass>	Types of criteria for assessing applications are: quality, safety and efficacy	subclassOf, Class (Thing, Nothing)	AL
LSP-SC-EN	LP-SC-01	Lexico Syntactic ODPs corresponding to SubclassOf relation ODP	NP<superclass> be | CATV [PARA] [(NP<resource1>,)* and | or], NP<subclass>	The devices can be classified into categories: FunctionRelated, EneryRelated and BuildingRelated	rdfs:subClassOf Class (Thing, Nothing)	AL
LSP-MI-EN	LP-MI-01	Lexico Syntactic ODPs corresponding to Multiple Inheritance ODP	NP<subclass> [can]be NP<superclass> and NP<superclass>	Amphibians are water-living and land-living animals	subclassOf, Class (Thing, Nothing)	AL

Competency questions

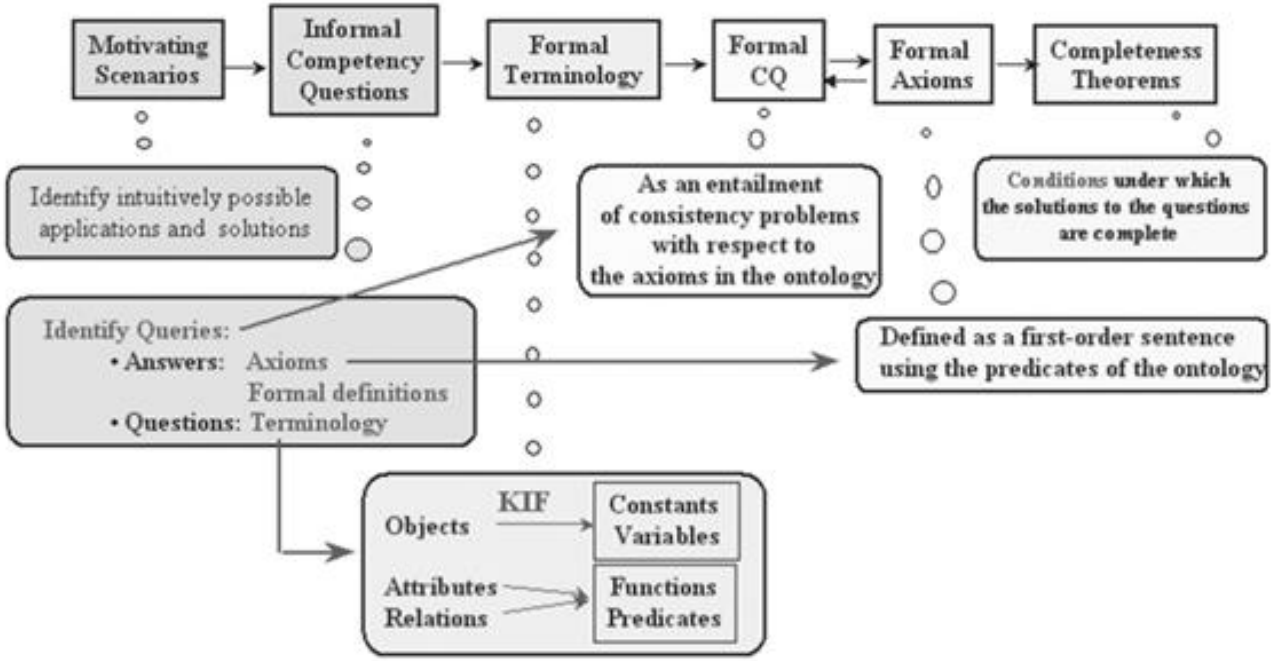
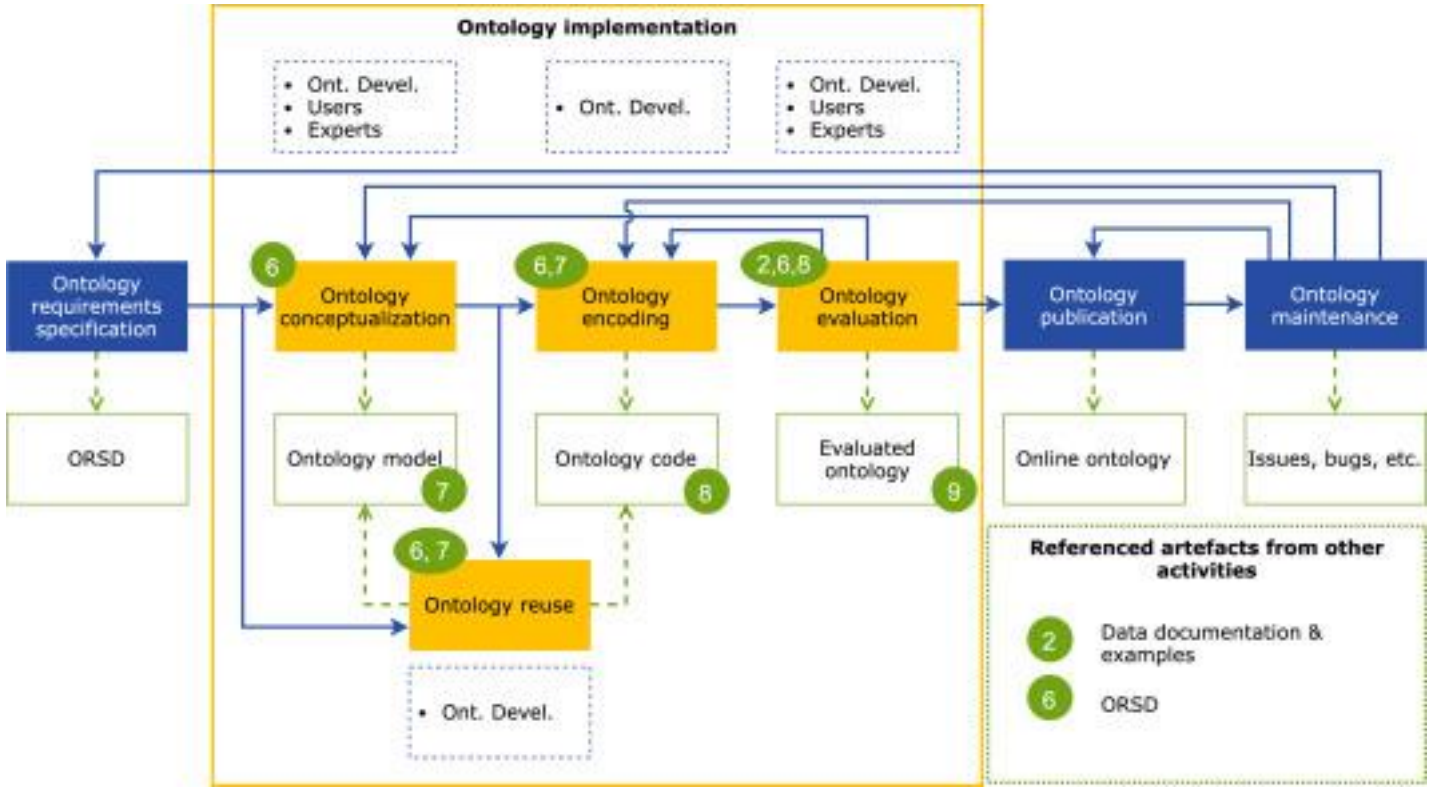


Figure 3 Main steps of Grüninger and Fox's Methodology

```
Alice lectures MATH456
MATH456 is a ScienceCourse
Bob teaches MIE123
MIE123 is an EngineeringCourse
```

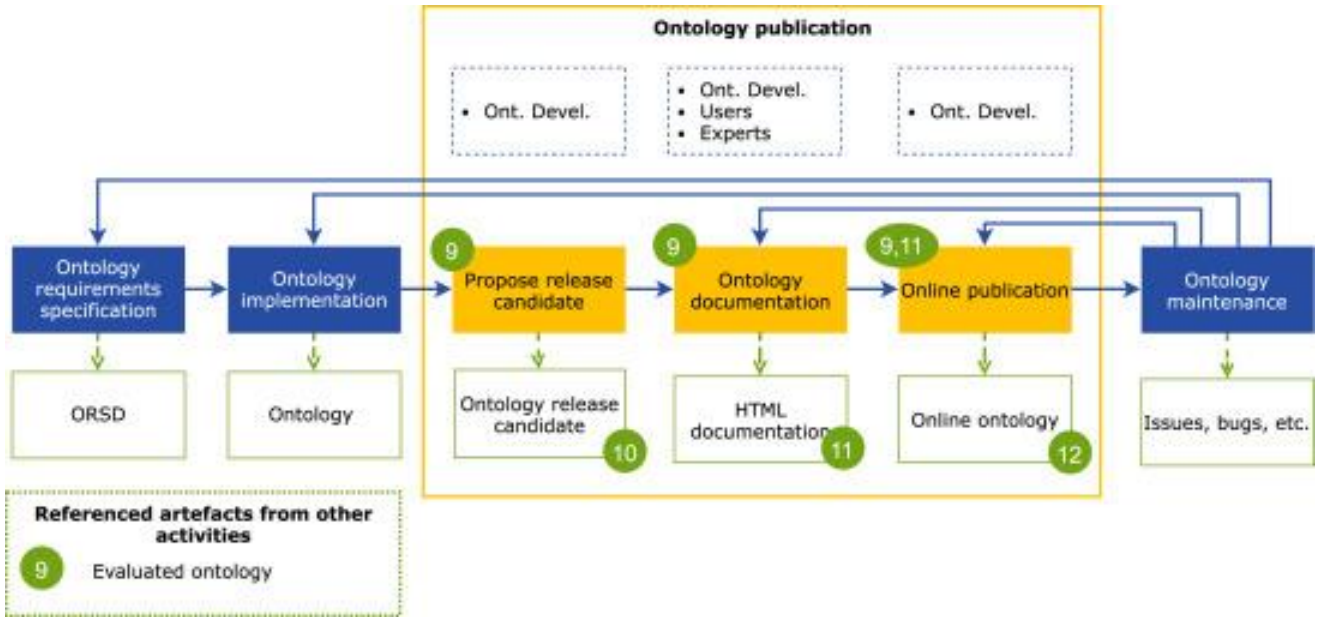
- *Is Alice a professor in the Math Department?*
- *Is Bob enrolled in any courses?*

Ontology Implementation



- Conceptualization
  - Mind map
  - Characterisation
  - Natural language definition
  - Domain knowledge, standards
- Ontology reuse
  - Upper-level ontology (Top, Mid)
  - Ontology Design Pattern
  - Hard and soft reuse
- Ontology encoding
  - Ontology development 101 (Noy and McGuiness)
  - Syntax and Semantics
  - Documentation (FAIR concerns)
- Ontology Evaluation
  - Technical quality (clarity, consistency, encoding bias)
  - Applicability (precision, recall, fitness, relevance, completeness)
  - Usability (adabtibility, expandability)

Ontology Publication

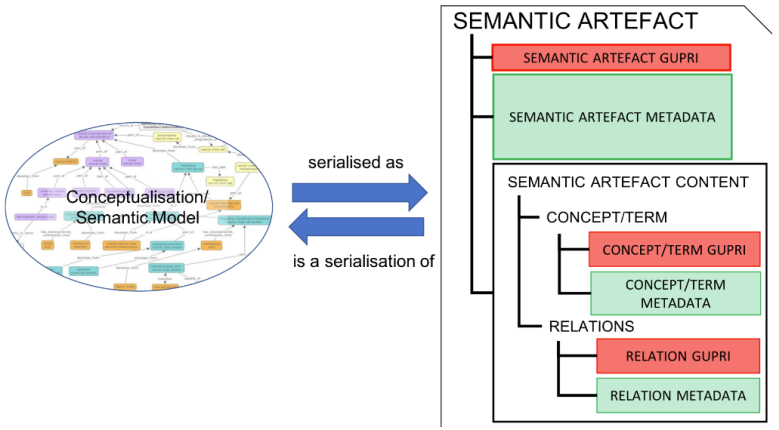


- Release candidate
  - Review and issue management
  - CI / CD
- Ontology Documentation
  - Artifact level and Content level
  - FAIR metadata
  - Visualisation
  - Guide
- Ontology publication
  - Ontology repository (IndustryPortal, MatPortal, BioPortal, OLS, OntoHub)
  - W3C submission
  - FAIR repositories

Ontology and FAIRness and FAIR Ontology

- Ontologies help in interoperability, but they are not interoperable themselves
  - also, not findable
  - also, not accessible
  - also, not reusable

  Research is underway and literatures on this topic is already available.



- Ontology is the “I” of the FAIR
  - the “I” (interoperability) of FAIR is only possible with the support of information structures that are ontologically consistent and that make explicit the ontological commitments that they inevitably make.
  - Domain and application ontology need to adopt Top and Mid level ontology
    - Alignment (“Not ontologies but Ontology”)
    - Semantic congruency

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

FAIR Recommendation

P-Rec. 1	Globally Unique, Persistent and Resolvable Identifiers must be used for Semantic Artefacts, their content (terms/ concepts/ classes and relations), and their versions.
P-Rec. 2	Globally Unique, Persistent and Resolvable Identifiers must be used for Semantic Artefact Metadata Records. Metadata and data must be published separately, even if it is managed jointly.
P-Rec. 3	A common minimum metadata schema must be used to describe semantic artefacts and their content.
P-Rec. 8	Human and machine-readable persistence policies for semantic artefacts metadata and data must be defined.
P-Rec. 9	Semantic artefacts must be made available as a minimum portfolio of common serialisation formats
P-Rec. 14	Standard vocabularies should be used to describe semantic artefacts
P-Rec. 15	Provenance information regarding the reuse of components from third-party semantic artefacts should be made explicit.

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

FAIR Recommendation (cont.)

#	
P-Rec. 16	The semantic artefact must be clearly licensed for use by machines and human
P-Rec. 17	Provenance must be clear for both humans and machine
P-Rec. 4	Semantic Artefact and its content should be published in an appropriate semantic repository
P-Rec. 5	Semantic repositories must offer a common API to access Semantic Artefacts and their content in various serialisations for both use/ reuse and indexation by any search engines
P-Rec. 6	Build semantic artefact search engines that operate across different semantic repositorie
P-Rec. 7	Repositories should offer a secure protocol and user access control functionalities
P-Rec. 10	Foundational Ontologies may be used to align semantic artefacts
P-Rec. 11	Semantic mappings between the different elements of semantic artefacts should be serialised in machine-readable format
P-Rec. 12	Crosswalks, mappings and bridging between semantic artefacts should be documented, published and curated

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

Thanks

Questions?

FOLLOW US ON

Contact

www.ontocommons.eu

Arkopaul Sarkar, asarkar@enit.fr

OntoComm ons “Ontology-driven data documentation for Industry Commons” has received funding from the European Union’s Horizon Programme call H2020 -NMBP-TO-IND-2020-singlestage, Grant Agreement number 958371