Stefan Schulz
Medical University of Graz (Austria)

purl.org/steschu

Realism vs. Pragmatism

Ontologies as sustainable knowledge representation artefacts

Workshop on knowledge management and the future of our society

Trondheim, Norway,
September 8th, 2014

The Landscape of Representation

meaning of domain terms

Universal properties of domain entities

Contingent characteristics of entities

"methanal" is a synonym of "formaldehyde"

"all cell membranes contain lipids"

"Ebola infections are rare"

"all fetuses were embryos"

"adult humans have typically 32 teeth"

"cell division" is broader than "mitosis"

"the surgical removal of a gallbladder is named "cholecystectomy"

" Lmn-2 interacts with Elf-2 "

"fungi are not plants"

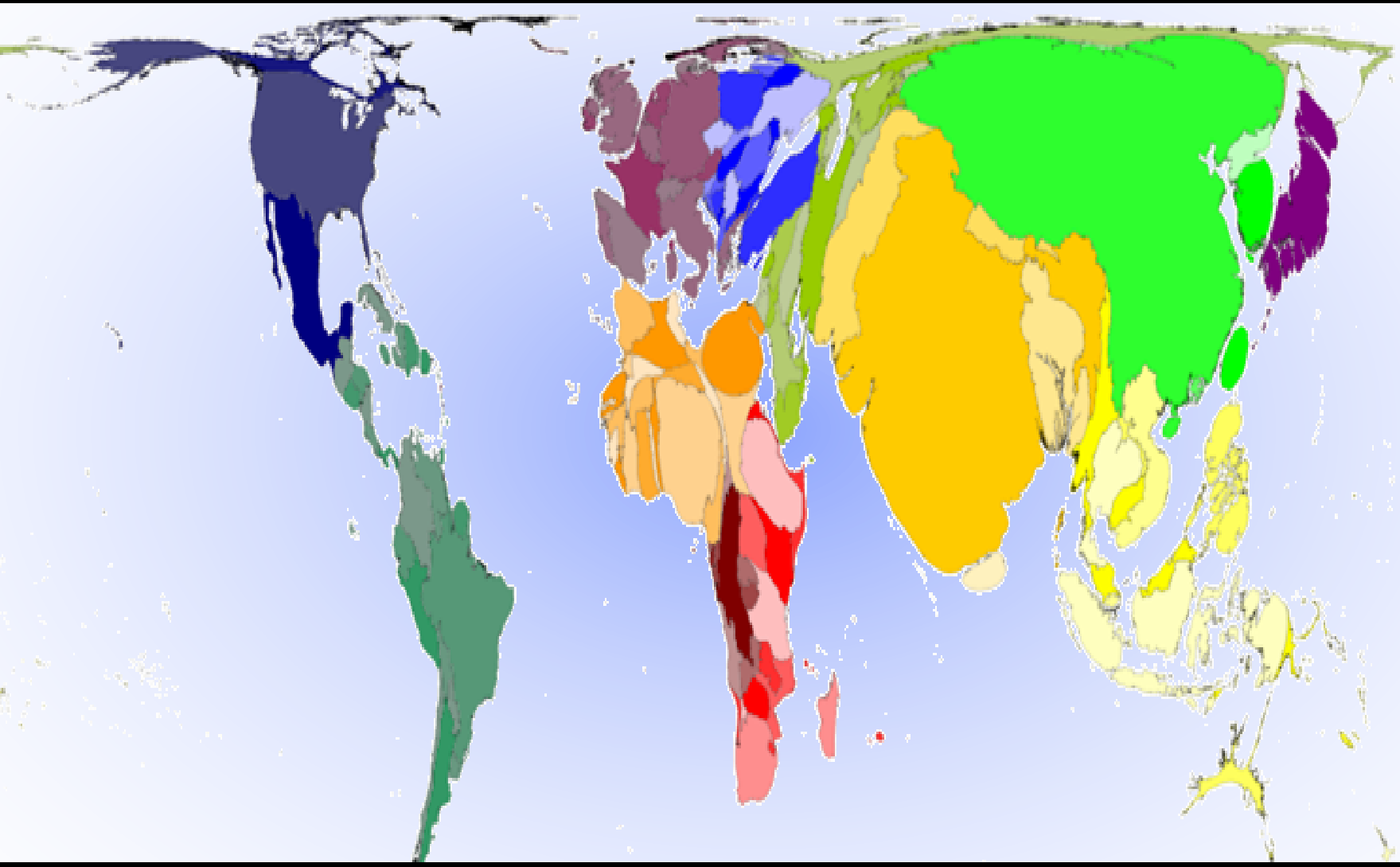
"eau" is French for "water"

"all brains develop inside animals"

"most plant cells have cell walls"

TERMINOLOGY **ONTOLOGY** **RICH KNOWLEDGE**

Redesigning the map



Redesigning the map

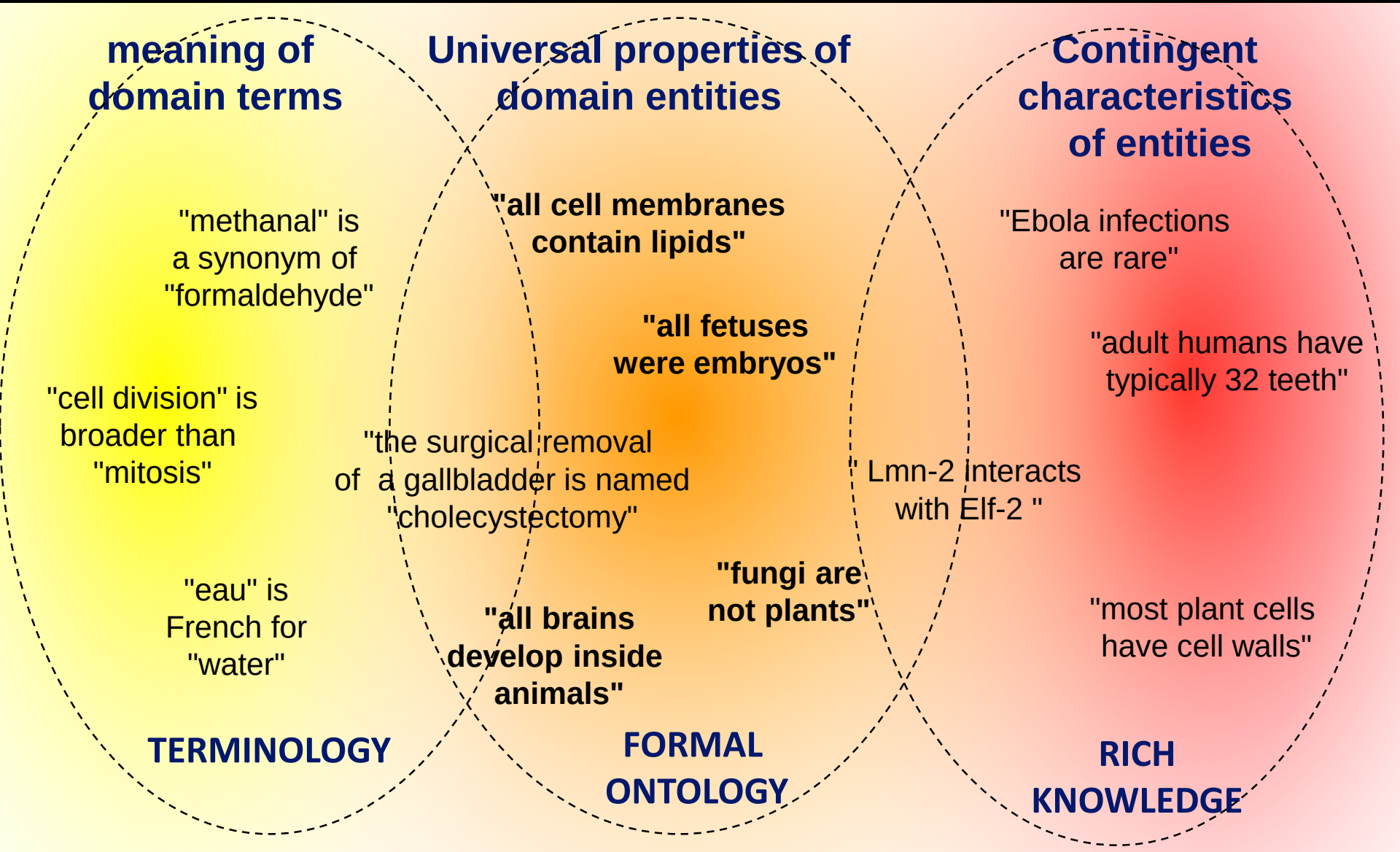
RICH KNOWLEDGE

ONTOLOGY

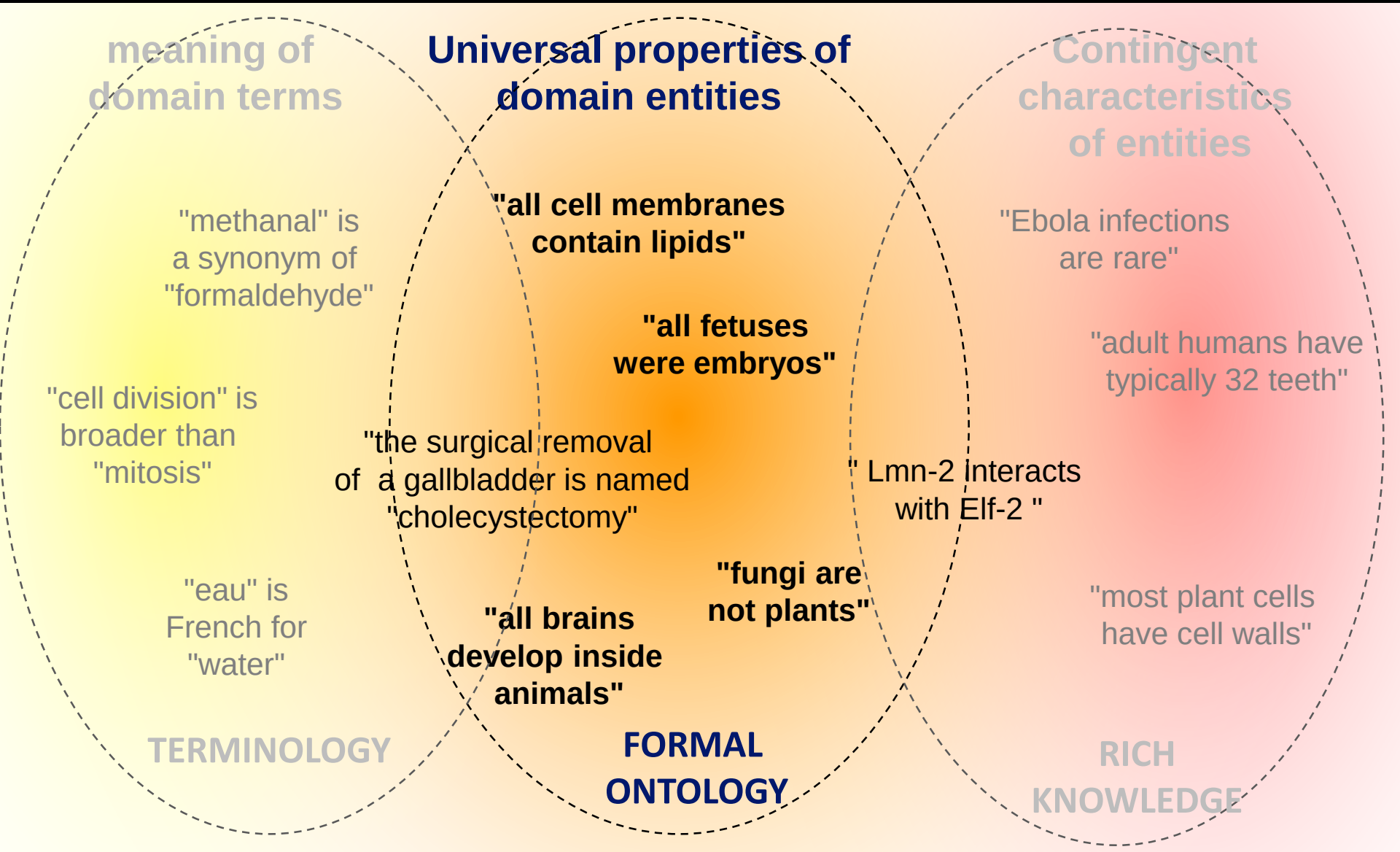
TERMINOLOGY

Alan Rector (2008): " very few interesting items of knowledge that are truly ontological…"
Bill Woods (1975): "conceptual coat rack"

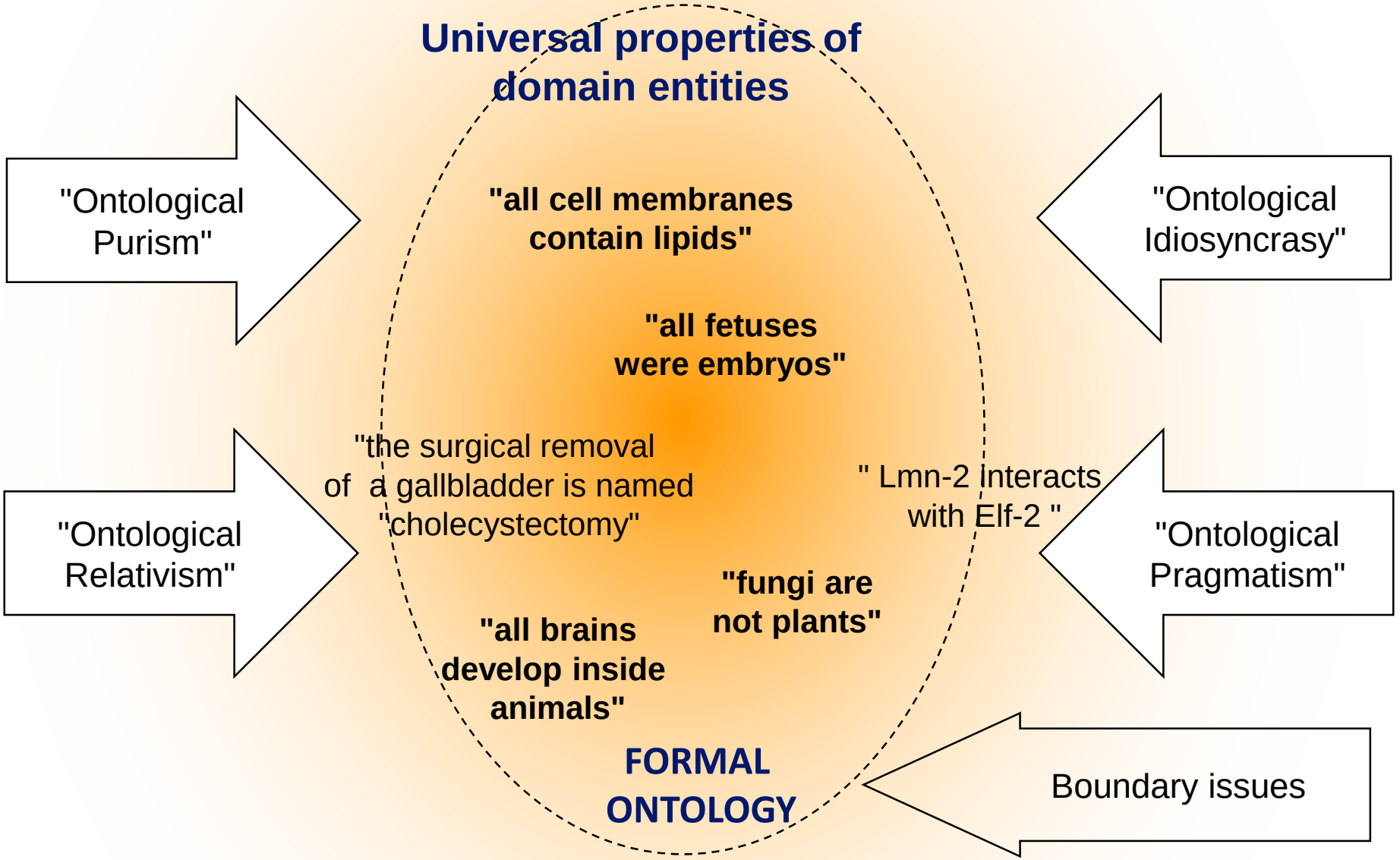
Focusing on Formal Ontology



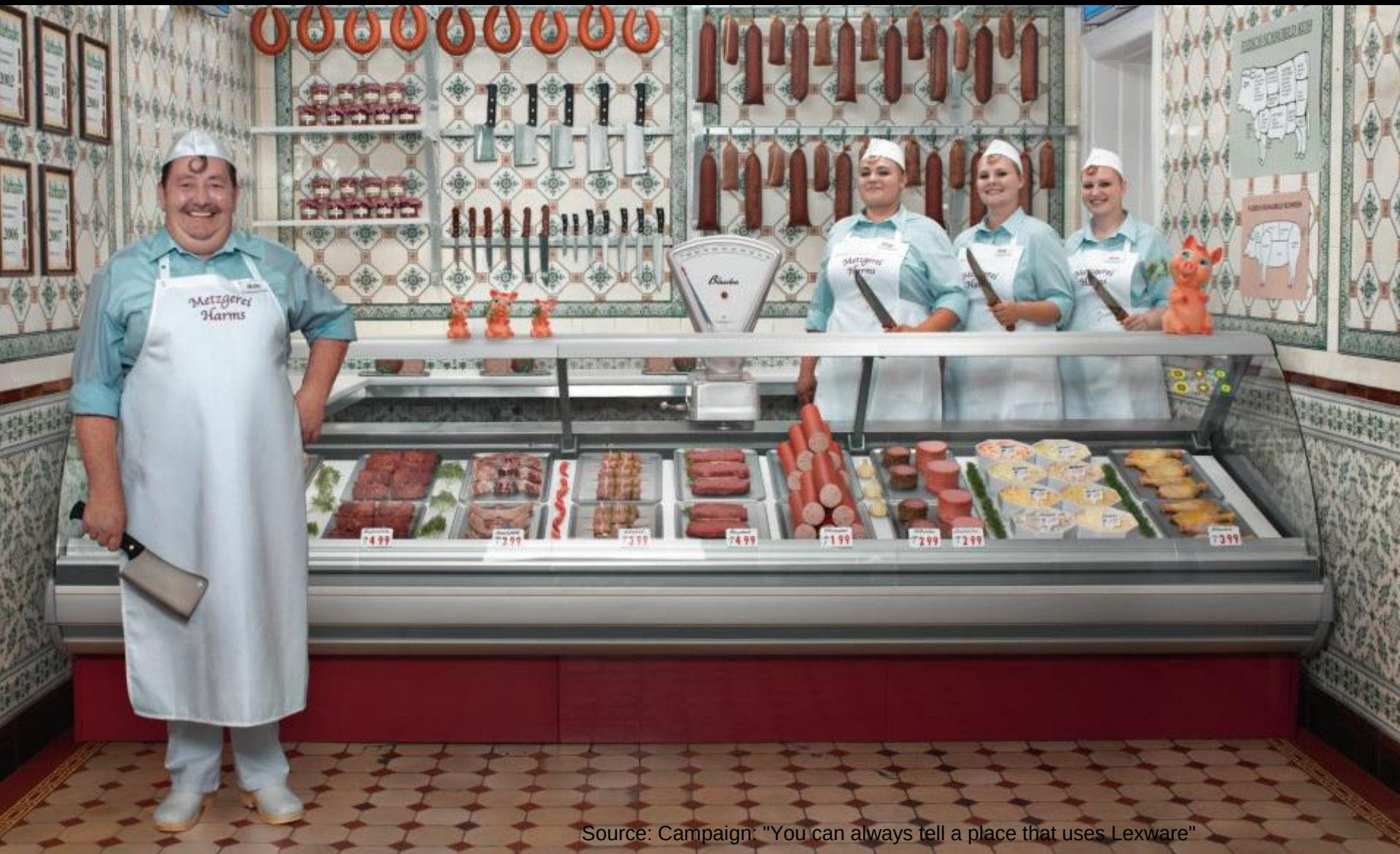
Focusing on Formal Ontology



Focusing on Formal Ontology



Ontological "Purism" (Smith / Ceusters)



Source: Campaign: "You can always tell a place that uses Lexware"

Ontological "Purism" (Smith / Ceusters)

- Ontologies represent universals (types) in reality
- The world is split into universals and individuals and there are objective criteria for this
- Everything is either a continuant or an occurrent
- Ontologies are independent of concrete applications
- Axioms in ontologies state what is universally true for all instances of a type
- Small set of relations
- Relations between continuant individuals are time-indexed
- First-order logics appropriate representation language

Smith B. Beyond Concepts: Ontology as Reality Representation. Proceedings of the International Conference on Formal Ontology in Information Systems, 11, 2004: 39-50.

Ontological Purism: Problems (I)

- OGMS, based on BFO, distinguished:
  - "Disorder" subclassOf Material object
  - "Disease" subclassOf Disposition
  - "Disease course" subclassOf Process
- Medical terms are ambiguous: How to represent "gastric ulcer"?
  - is a piece of anatomically altered stomach wall → material Object
  - is a process (ulceration)

Ontology for General Medical Science (OGMS) https://code.google.com/p/ogms/

Schulz S et al. Scalable representations of diseases in biomedical ontologies. J Biomed Semantics. 2011 May 17;2 Suppl 2:S6.

Ontological Purism: Problems (II)

- BFO 2 uses FOL to introduce ternary relations between continuants
  - located-in (a, b, t_1) AND located-in (b, c, t_1) → located-in (a, c, t_1)
  - located-in (a, b, t_1) AND located-in (b, c, t_2) → ?
- FOL is undecidable
- In Description logics only two-valued relations (object properties)
  - located-in (a, b) AND located-in (b, c) → located-in (a, c)
    If transitive, leads to wrong entailments.
  - Otherwise, incomplete

Smith B et al. Basic Formal Ontology (BFO).

Ontological Relativism (Noy / McGuinness)

Ontological Relativism (Noy / McGuinness)

- Ontologies represent "shared conceptualizations"
- Ontologies + instances = knowledge bases
- Terminologies / vocabularies are kinds of (informal) ontologies
- Whether something is modeled as a class or an instance depends on granularity and context
- Ontologies are built to represent the knowledge needed for specific applications
- Ontology reuse is highlighted but no clear provisions for interoperability taken
- Upper-level ontology not explicitly recommended

Noy NF, McGuinness DL. Ontology Development 101: A Guide to Creating Your First Ontology.
http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html

Ontological Relativism: problems

- Ontologies as shared conceptualizations:
  - Things are represented how they are perceived / known, not as they are (philosophically: ontological realism)
  - Potentially contradictory representations of the same thing
- Example
  - *Glucose* instanceOf *Hexose*
  - What about *L-Glucose*?
- Terminologies, thesauri (e.g. UMLS, MeSH) are also understood as ontologies?
  - How to formally describe them?
  - If not, how to differentiate them?

Noy NF, McGuinness DL. Ontology Development 101: A Guide to Creating Your First Ontology.
http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html

Ontological Idiosyncrasy / Syncretism

Ontological Idiosyncrasy / Syncretism

- Unprincipled, naïve (undisciplined?) approach to ontologies
- Assumptions:
  - informal vocabularies or database schemes wrapped into a formal language (e.g. OWL) become ontologies
  - Everything which represents knowledge in the Semantic Web is an ontology
- The way an ontology is shaped depends on its specific purpose
- "A little semantics goes a long way"
- "Anything goes" with regard to upper-level classes and relations (their need is often questioned)

W Kunierczyk et al. Nontological Engineering. Proceedings of the 4th International Conference FOIS, 2006

Ontological Idiosyncrasy / Syncretism: problems

- Embedding modal, negative, or probabilistic notions. Example: NCI Thesaurus:
  Ureter_Small_Cell_Carcinoma subclassOf
  Disease_May_Have_Finding some *Pain*
- Improper co-ordinations

  Calcium-Activated_Chloride_Channel-2 subClassOf
  Gene_Product_Expressed_In_Tissue some *Lung* and
  Gene_Product_Expressed_In_Tissue some *Mammary_Gland* and
  Gene_Product_Expressed_In_Tissue some *Trachea*
- Weak or non-existing upper level and undefined primitives:
  Relies on implicit human language understanding.
  Barrier to shared conceptualizations.
  Examples:
  - Unclear whether "animal" includes "human"
  - Unclear whether events and processes are the same
  - Unclear whether "part-of" ranges over all times
  - etc…

Schulz S et al. The Pitfalls of Thesaurus Ontologization - the Case of the NCI Thesaurus.
AMIA Annu Symp Proc. 2010 Nov 13;2010:727-31.

Ontological pragmatism: the GoodOD approach

Ontological pragmatism: the GoodOD approach

- Ontologies as formal systems (using OWL DL)
- Ontological engineering supported by
  - clearly defined upper-level categories
  - closed set of basic relations
  - constraining axioms
  - understandable labels
- Criteria of dividing between classes and individuals
- Aristotelian definitions (genus – differentia)
- Naming conventions, design patterns and guidelines
- Upper ontology BioTopLite2
  http://purl.org/biotop/btl2.owl

Schulz S, Boeker M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. In: Furbach, U; Staab, S; editors(s). Informatik 2013. IOS Press; 2013

GoodOd – Good Ontology Design

Guideline on Developing Good Ontologies in the Biomedical Domain with Description Logics

URL: http://www.purl.org/goodod/guideline

Version 1.0
December 2012

Send feedback to:
martin.boeker@uniklinik-freiburg.de
ludger.jansen@uni-rostock.de

Schulz S[1,3], Seddig-Raufie D[1], Grewe N[2], Röhl J[2],
Schober D[1], Boeker M[1], Jansen L[2]

[1]: Institute of Medical Biometry and Medical Informatics,
University Medical Center Freiburg
[2]: Institute of Philosophy, University of Rostock
[3]: Department of Medical Informatics, University of Graz

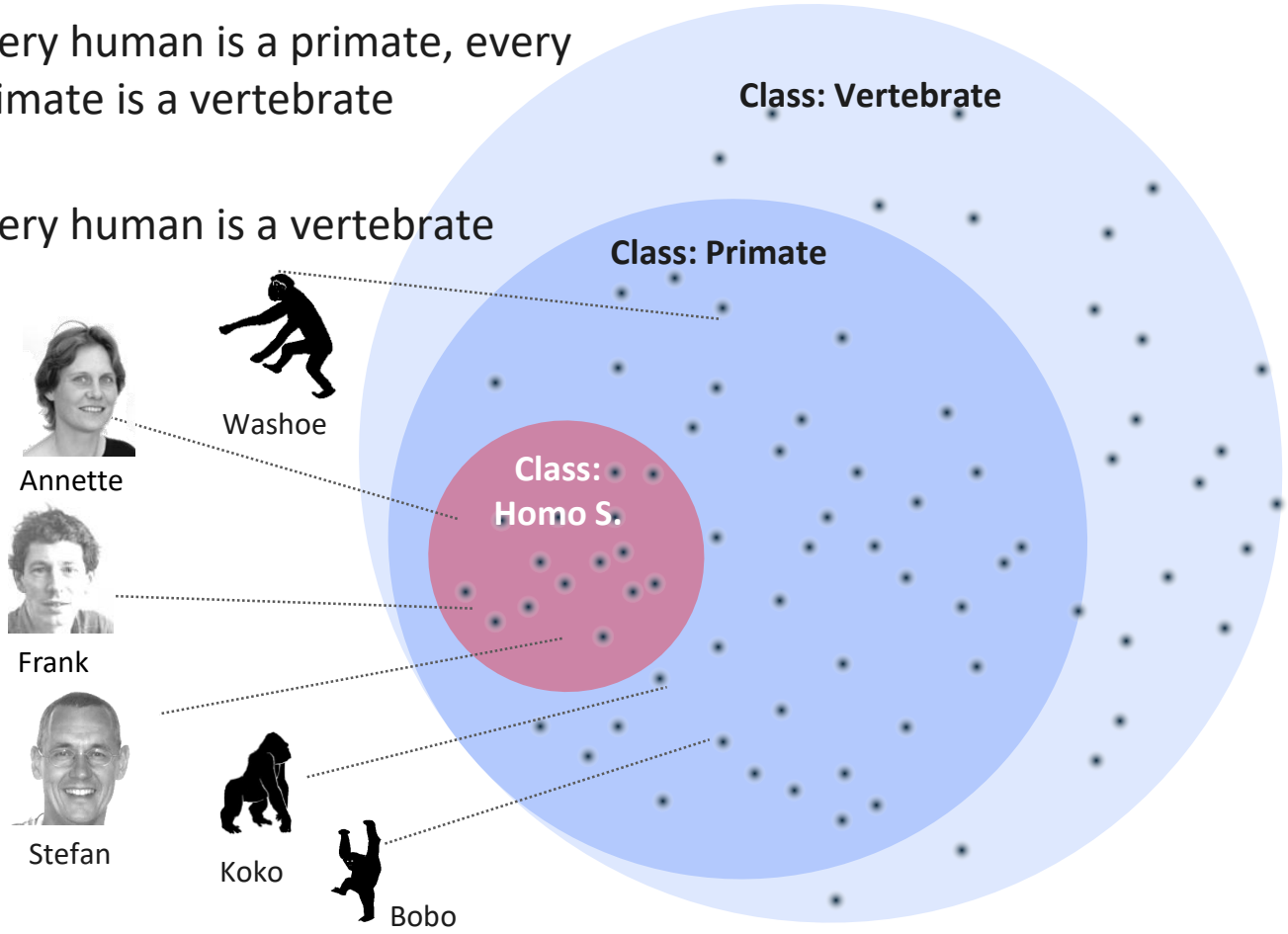
11th December 2012

Class-individual distinction not discretionary

Every human is a primate, every primate is a vertebrate

→

Every human is a vertebrate



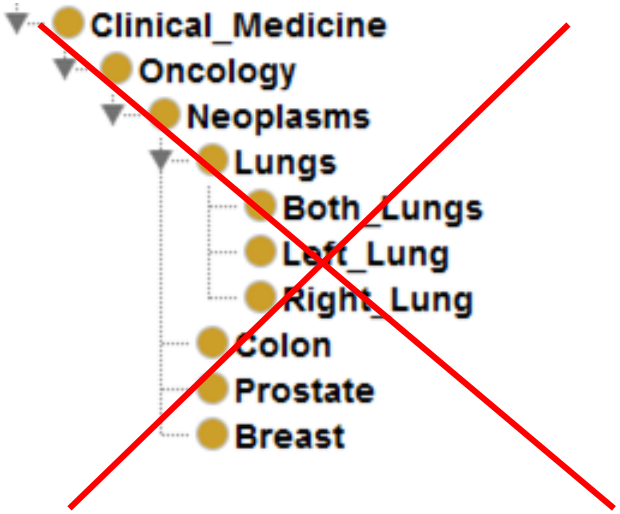
Whether a class is the extension of a universal (type) or not is secondary

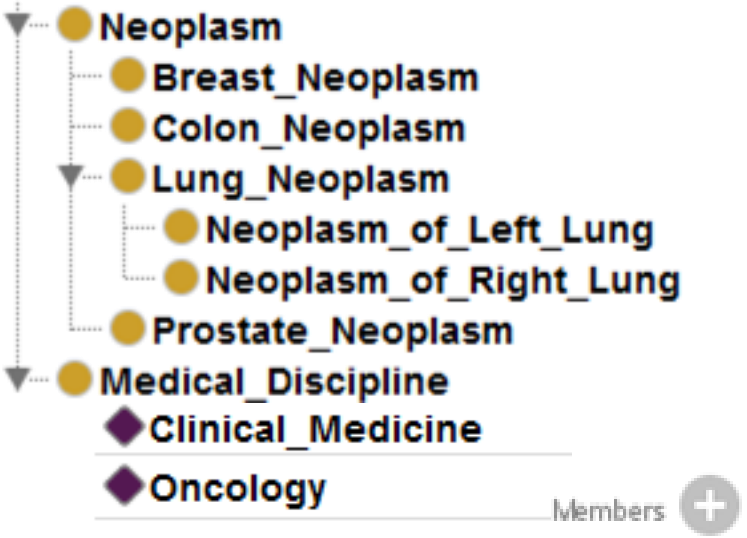
Human subClassOf *Primate*

'A subClassOf B' iff

$\forall$a, t: 'instance of' (A, a, t) $\rightarrow$ 'instance of' (B, a, t)

Intuitive hierarchies ≠ good taxonomies





FOL: $\forall x, t$: **'instance of'** (*X*, x, t) $\leftrightarrow$ **'instance of'** (*Y*, x, t) $\Leftrightarrow$
$\forall t \neg\exists x$: **'instance of'** (*X*, x, t) $\wedge \neg$**'instance of'** (Y, x, t)

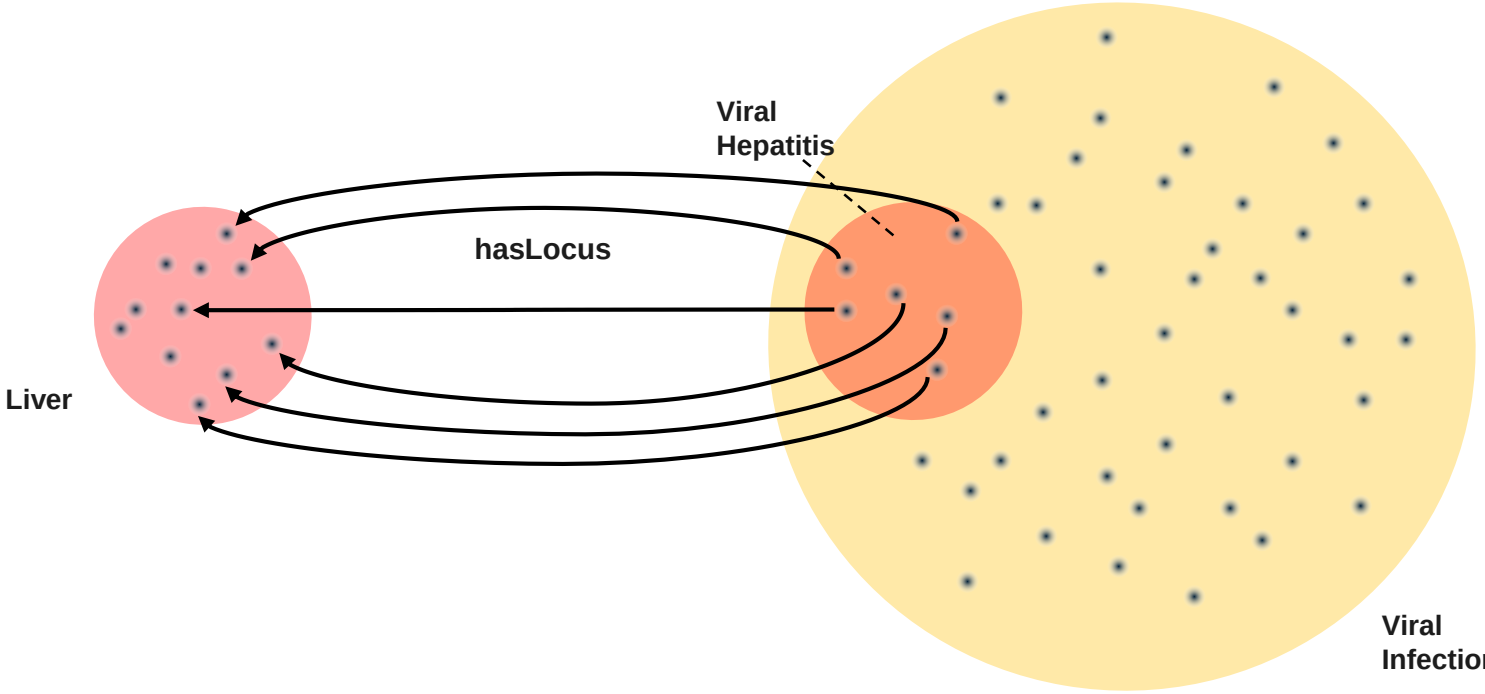
OWL-DL: *X subClassOf Y*
X and not (Y): unsatisfiable

- Test : *there is no neoplasms that is not an oncology*
- *there is no prostate that is not a neoplasm*
- *there is no oncology that is not a clinical medicine*

Labelling !

Guarino N, Welty C. Handbook on ontologies, 2009 – Springer
Schober D et al. .Survey-based naming conventions for use in OBO Foundry ontology development. BMC Bioinformatics. 2009 Apr 27;10:125.

Aristotelian Definitions do not permit exceptions



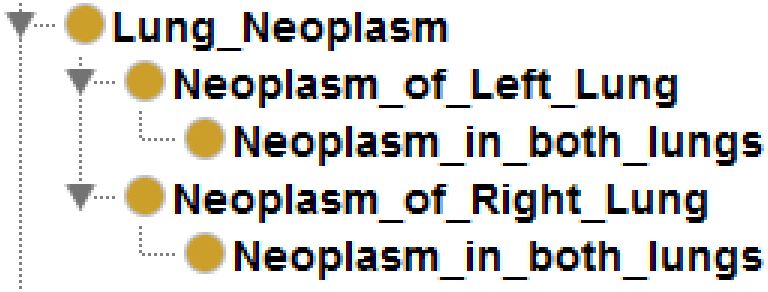
FOL: ∀x, t: **'instance of'** (*'Viral hepatitis*, x, t) ↔ **'instance of'** ('Viral infection', x, t)
∧ ∃z: **'instance of'** (Liver, z, t) ∧ **'is included in'** (x, z, t)

OWL-DL: *'Viral Hepatitis'* equivalentTo *ViralInfection* and *'***is included in***' some Liver*

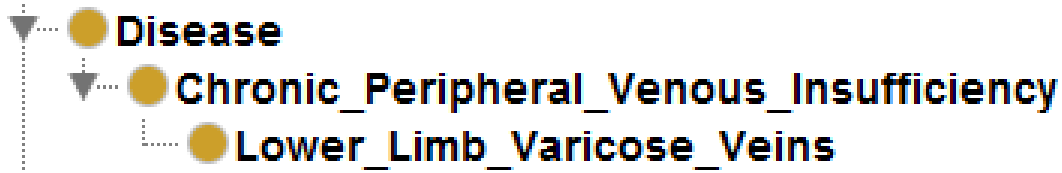
Test :

- *There is no viral hepatitis that is not located in a liver*
- *There in no viral hepatitis that is not a viral infection*

Always investigate the ontological commitment



- Which are exactly the instances?
- Does the label tell us what is meant?
- Is there an implicit context?



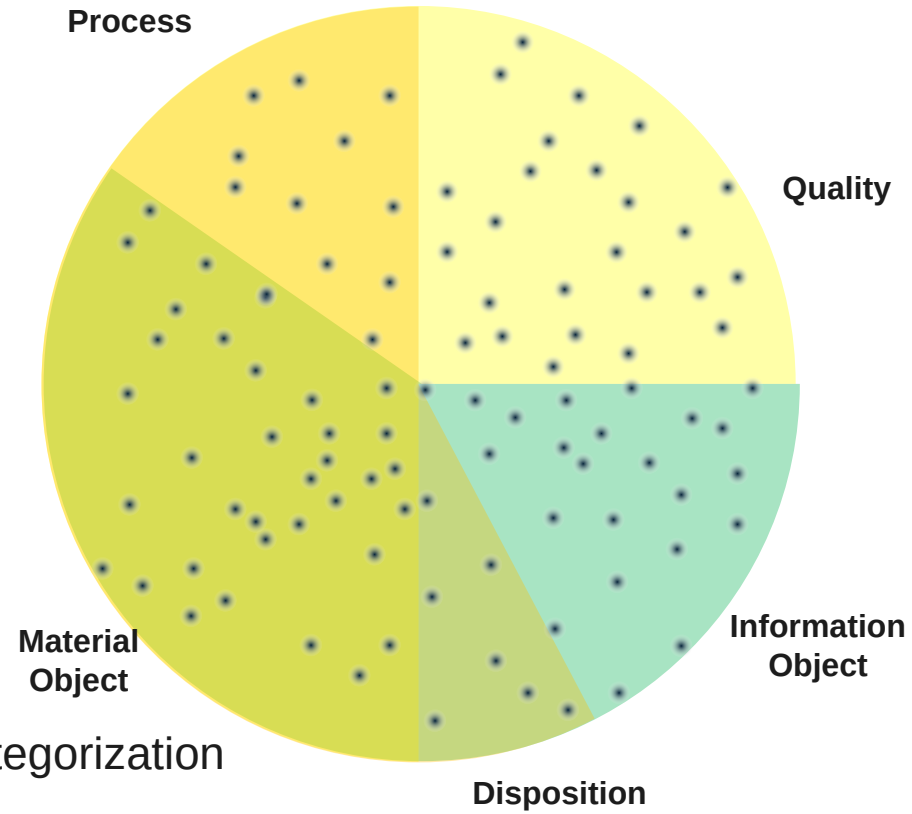
Test :

- *There is no neoplasm in both lungs that is not a neoplasm in the left lung* OR
  There is no patient with neoplasm in both lungs that is not a patient with the neoplasm in the left lung
- *There is no varicose vein in the lower limb that is not a chronic peripheral venous insufficiency* OR
  There is no patient with varicose lower limb veins that is not a patient with a chronic peripheral venous insufficiency

Guarino N, Carrara M, Giaretta P. Formalizing ontological commitment. AAAI, 1994 - aaai.org

Upper level ontologies partition the domain into disjoint and exhaustive categories

- Particular
  - Quality
    - Benign
    - Malignant
  - Information_Object
    - Medical_Discipline
  - Disposition
    - Chronic_Peripheral_Venous_Insufficiency
  - Material_Object
    - Lower_Limb_Varicose_Veins
    - Neoplasm
      - Breast_Neoplasm
      - Colon_Neoplasm
      - Lung_Neoplasm
      - Prostate_Neoplasm
  - Process
    - Growth



- Upper level ontologies enforce a strict categorization
- Constraints on upper-level categories
- Upper level ontology for the biomedical domain **BioTopLite**

Schulz S, Boeker M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. In: Furbach, U; Staab, S; editors(s). Informatik 2013. IOS Press; 2013

BioTopLite provides a small set of toplevel classes, relations, and axioms

Toplevel Categories

- particular
  - condition
    - disposition
    - function
    - 'material object'
    - process
  - 'immaterial object'
  - 'information object'
  - quality
  - role
  - 'temporal region'
  - 'value region'

Basic relations

- 'has participant'
- 'has timestamp'
- includes
- 'is included in'
- 'is preceded by'
- 'is projection of'
- 'is represented by'
- 'participates in'
- precedes
- 'projects onto'
- represents

Constraining axioms

- 'has part' only process
- 'has participant' some Thing
- 'is bearer of' only 'process quality'
- 'is part of' only process
- 'projects onto' only 'temporal region'
- 'projects onto' some 'temporal region'
- 'realization of' only disposition
- includes only (process or 'process quality')

- Precise formulations about generic and defining properties of basic categories of a domain
- Logical Framework (Description logics)
- OWL – DL (Web Ontology Language) complete and decidable language - compromise between expressiveness and performance

- Automated reasoning enables checking consistency, equivalence and subsumption
- Ontologies play an increasing role in new generation of biomedical terminology systems

Schulz S, Boeker M. BioTopLite: An Upper Level Ontology for the Life Sciences. Evolution, Design and Application. In: Furbach, U; Staab, S; editors(s). Informatik 2013. IOS Press; 2013

BioTopLite2: Dealing with ambiguity

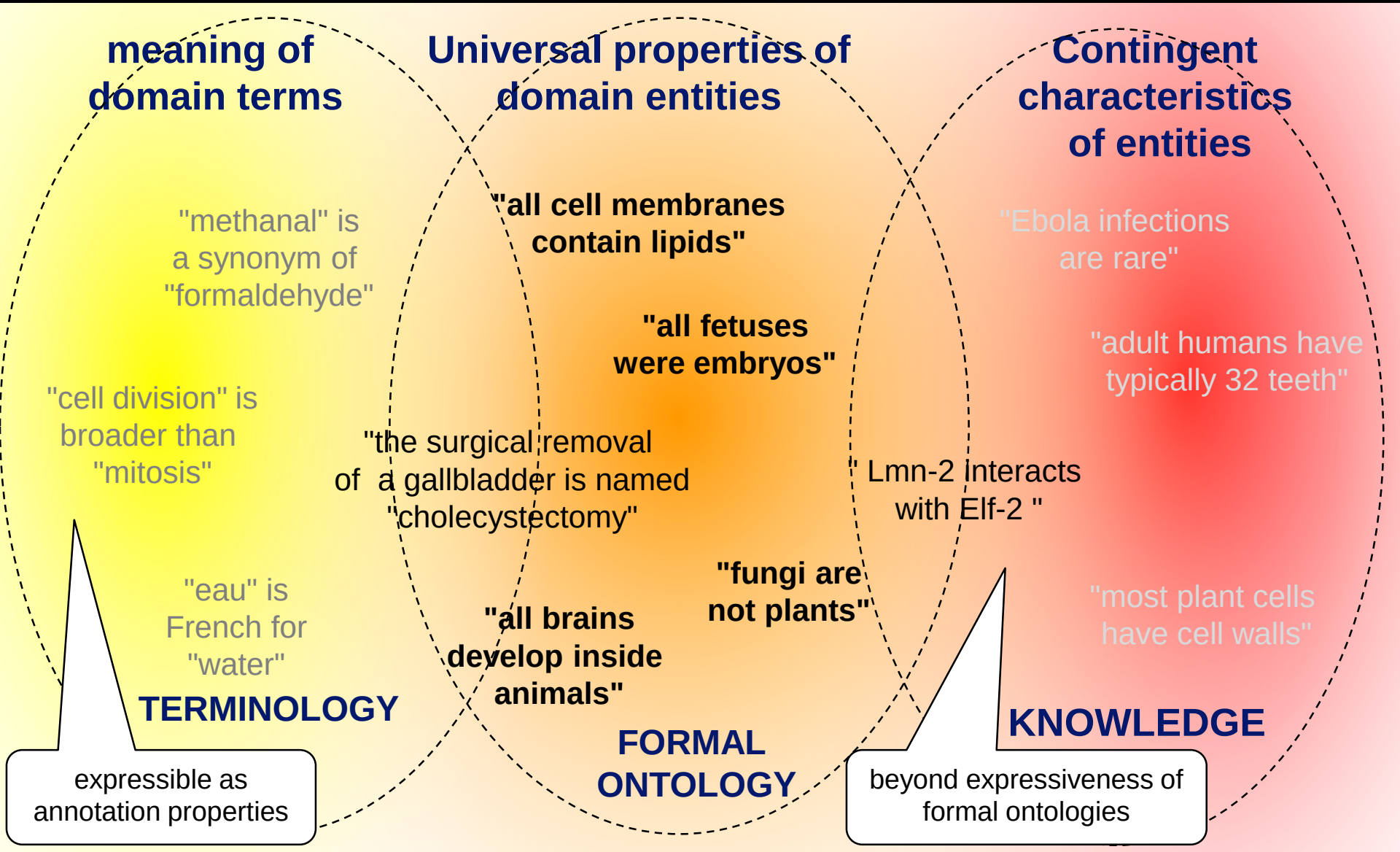
- "Every gastric ulcer is in the stomach wall"
  "Every stomach wall is part of a stomach" →
  "Every gastric ulcer is in the stomach"
- '**is part of**' subPropertyOf '**is included in**'
  (both transitive)
- *Condition* equivalentTo '*Material object*' or
  Disposition or *Process*
- '*Gastric ulcer*' subClassOf *Condition*
  '*Gastric ulcer*' '**is included in**' some '*Stomach wall*'
  '*Stomach wall*' '**is part of**' some *Stomach* →
  '*Gastric ulcer*' '**is included in**' some *Stomach*

Schulz S et al. Scalable representations of diseases in biomedical ontologies. J Biomed Semantics. 2011 May 17;2 Suppl 2:S6.

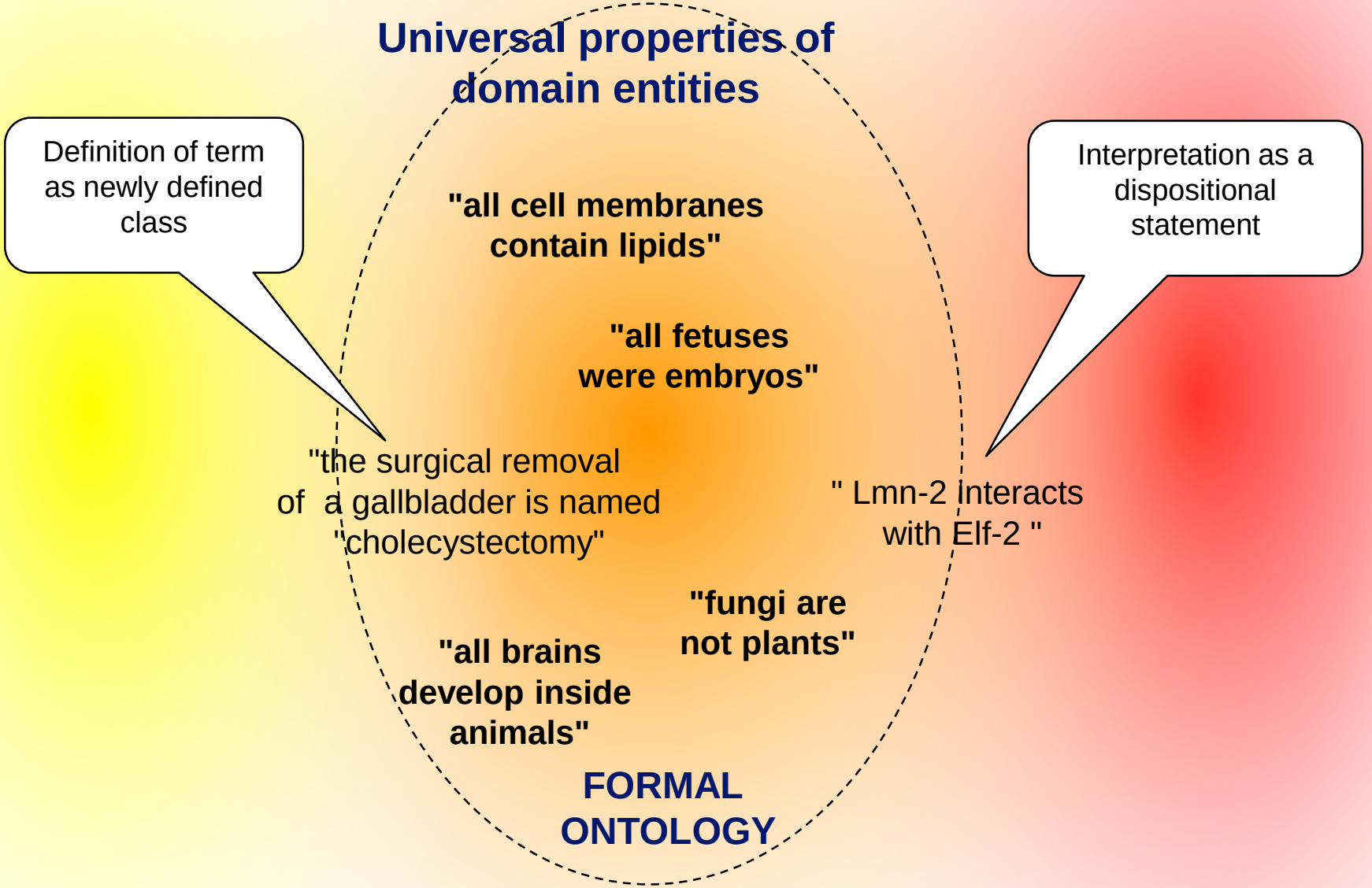
BioTopLite2	BFO2
Ontological Realism "light": no commitment to universalism	Ontological Realism: ontologies describe universals
Description Logic (OWL-DL)	First-Order-Logic (FOL), only OWL-DL class-only and experimental versions
Binary object properties	binary and ternary (time-indexed) relations
Classes: 53	Classes: 36
Relations: 37	Relations (in experimental TR version): 78
Axioms: 527	Axioms (in experimental TR version): 1572
Coverage: domain-independent upper-level classes and relations + few biomedical classes	Coverage: domain-independent classes and relations

BioTopLite2: harmonization intended with BFO2 once its DL version stable

The Boundaries of Formal Ontology



The Boundaries of Formal Ontology



The Boundaries of Formal Ontology

- Towards terminologies:
  - Definition of meaning of a term as new class; expression of ambiguities by disjunction:
    '*Gastric ulcer*' equivalentTo '*Gastric ulcer structure*' or '*Gastric ulcer process*'
  - *Cholecystectomy* equivalentTo '*Surgical removal*' and
    '**has participant**' some *Gallbladder*
- Towards "rich" knowledge bases
  - Qualitative dispositional predicates:
    '*Lmn-2*' subClassOf '**is bearer of**'
    some *Disposition* and '**has realization**'
    only (*Interaction* and '**has participant**' some *Elf-2*)

Schulz S, Jansen L: Molecular interactions: On the ambiguity of ordinary statements in biomedical literature. Applied Ontology, 2009; 4 (1): 21-34
Schulz S, Jansen L. Formal ontologies in biomedical knowledge representation. Yearb Med Inform. 2013;8(1):132-46.

Conclusions

- Domain ontologies are the most sustainable part of the representation of domain knowledge and they should be limited to
- Formal ontologies express what is universally true for all members of a class (all instances of a type)
- Large parts of interesting domain knowledge are not ontological
- Reusable ontologies should be
  - philosophically grounded and expressible in a computable language
  - user-friendly in terms of labelling
- This should be supported
  - by educational material
  - by expressive upper-level ontologies
  - appropriate editor and visualization tools
- Compromises are needed
  - understandability and intuitiveness of toplevel classes and relations
  - representation of ambiguous terms as disjoint classes
  - decidable and tractable logic (e.g. DL only allowing for binary relations)

Stefan Schulz

Medical University of Graz (Austria)

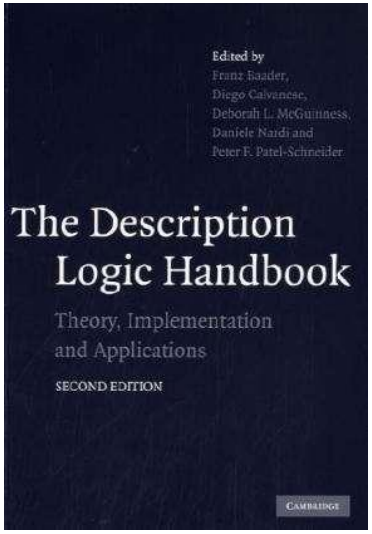
purl.org/steschu

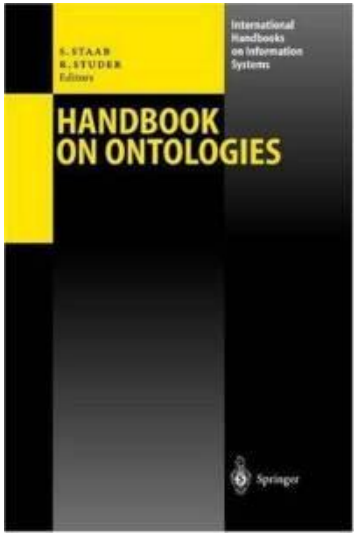
Purism vs. Pragmatism Ontologies as sustainable KR artefacts

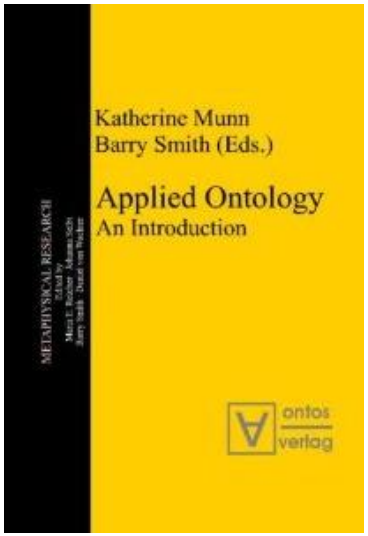
Slides downloadable from

http://user.medunigraz.at/stefan.schulz/presentations.htm

Further readings







Ontology on the Web

- Description Logics: http://dl.kr.org/
- Protégé: http://protege.stanford.edu/
- Bioontology: http://www.bioontology.ch/
- Buffalo Ontology Site: http://ontology.buffalo.edu/smith/
- OBO Foundry: http://obofoundry.org/
- Bioportal: http://bioportal.bioontology.org/
- SNOMED CT: http://www.ihtsdo.org/snomed-ct/ http://terminology.vetmed.vt.edu/sct/menu.cfm
- CO-ODE (Pizza ontology): http://www.co-ode.org/
- GoodOD Guideline: http://www.iph.uni-rostock.de/GoodOD-Guideline.1299.0.html
- BioTop: http://purl.org/biotop