

Ontologie appliquée et données hétérogènes

Adrien Barton

Institut de Recherche en Informatique de Toulouse (IRIT), CNRS
Groupe de Recherche Interdisciplinaire en Informatique de la Santé
(GRIIS), Université de Sherbrooke
adrien.barton@irit.fr

Séminaire « Systèmes Complexes en Sciences Sociales »

Labex SMS

18 décembre 2020

Plan

- Partie I : Des ontologies de domaines pour unifier des données issues de sources hétérogènes
- Partie II : Des ontologies de haut niveau pour unifier des ontologies de domaines hétérogènes
- Partie III : Des théories ontologiques pour articuler des niveaux de réalité hétérogènes



Partie I

Des ontologies de domaines pour unifier des données issues de sources hétérogènes



L'informatique médicale aujourd'hui

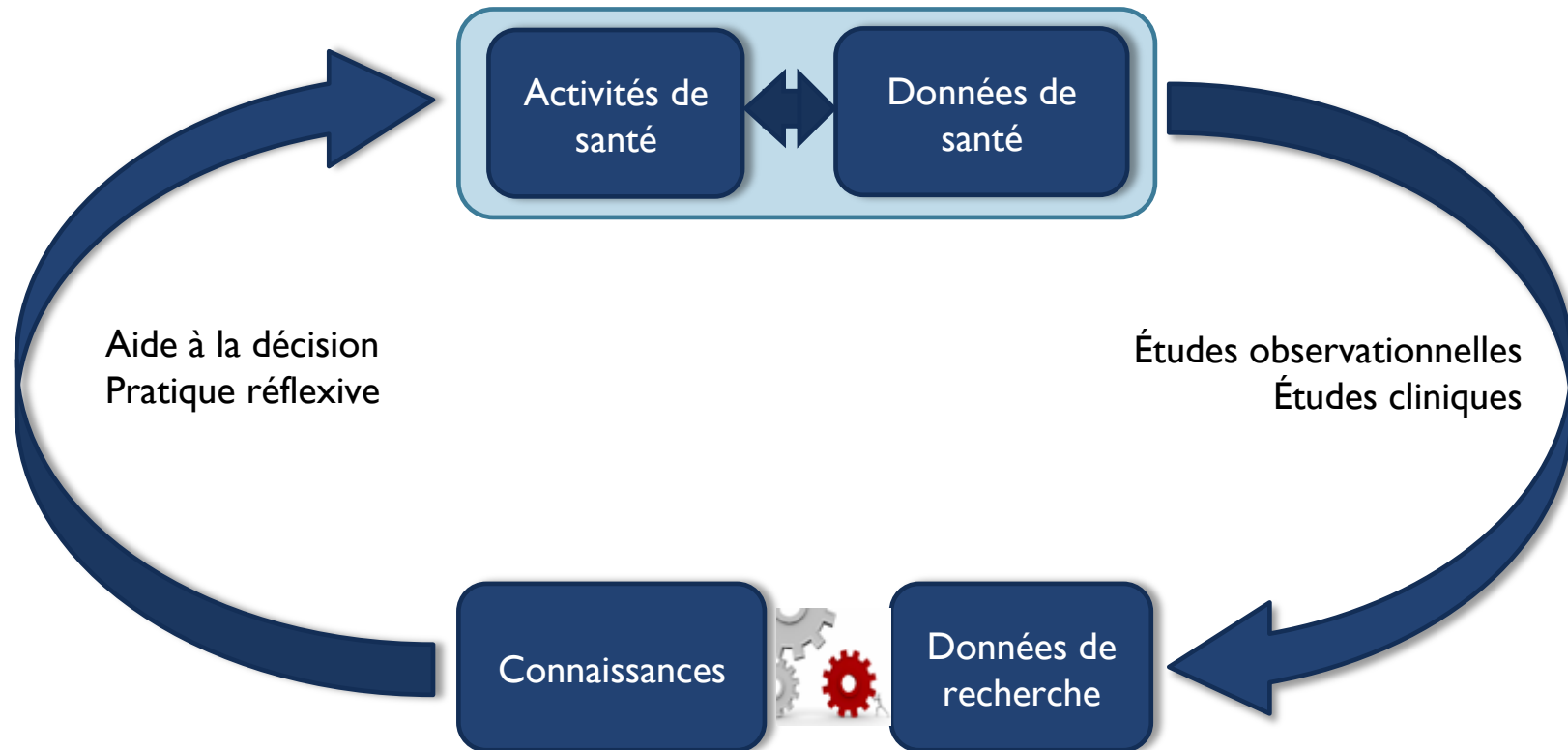
L'exemple du Québec

- Cabinets de médecin
- Hôpitaux
- RAMQ
- Centre Local de Service Communautaire (CLSC)
- Ministère de la Santé et des Services Sociaux (MSSS)
- Cohortes
- Essais cliniques
- Biobanques
- ...



Tour de Babel des systèmes
d'information

Un défi pour la mise au point de systèmes de santé apprenants



L'informatique médicale aujourd'hui

L'exemple du Québec

- Défi : rendre l'information scientifique interopérable, partageable, réutilisable
- Obstacle principal : diversité des données, plutôt que leur quantité
- “Problème de Babel” :
 - Idiosyncrasie technologique
 - Idiosyncrasie humaine
- Solution : terminologies, taxonomies, ontologies

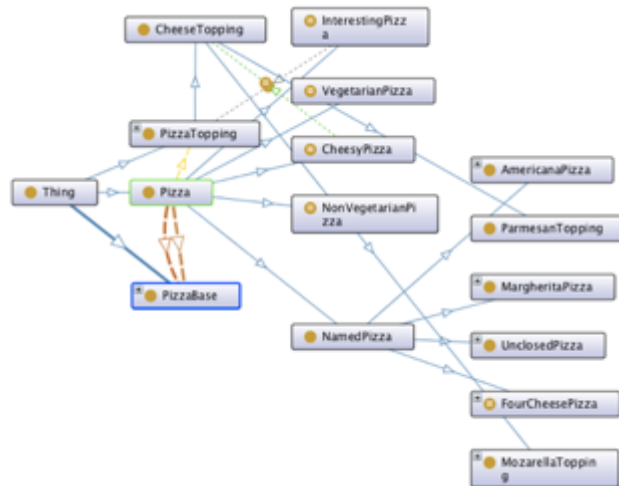


Tour de Babel des systèmes d'information

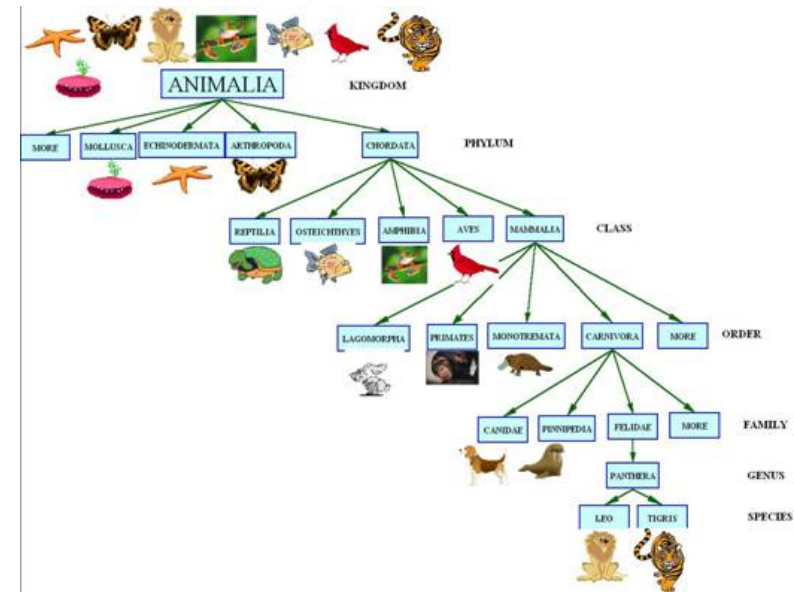
Terminologies, taxonomies et ontologies

Term	Definition
abortifacients	used to stimulate uterine contraction and promote evacuation of the uterus to cause abortion
ACE inhibitor	used to block the enzyme responsible for converting Angiotensin 1 to angiotensin 2 in the lungs; this blocking prevents vasoconstriction
antiarrhythmics	this affects the action potential of cardiac cells and are used to treat arrhythmias and return normal rate and rhythm of the heart muscles
anticoagulant	drugs that inhibit any step of the coagulation process, preventing or slowing clot formation

Terminologie: Contient une liste d'entrées lexicales et de leurs descriptions.



Ontologie: Taxonomie enrichie avec des relations additionnelles.

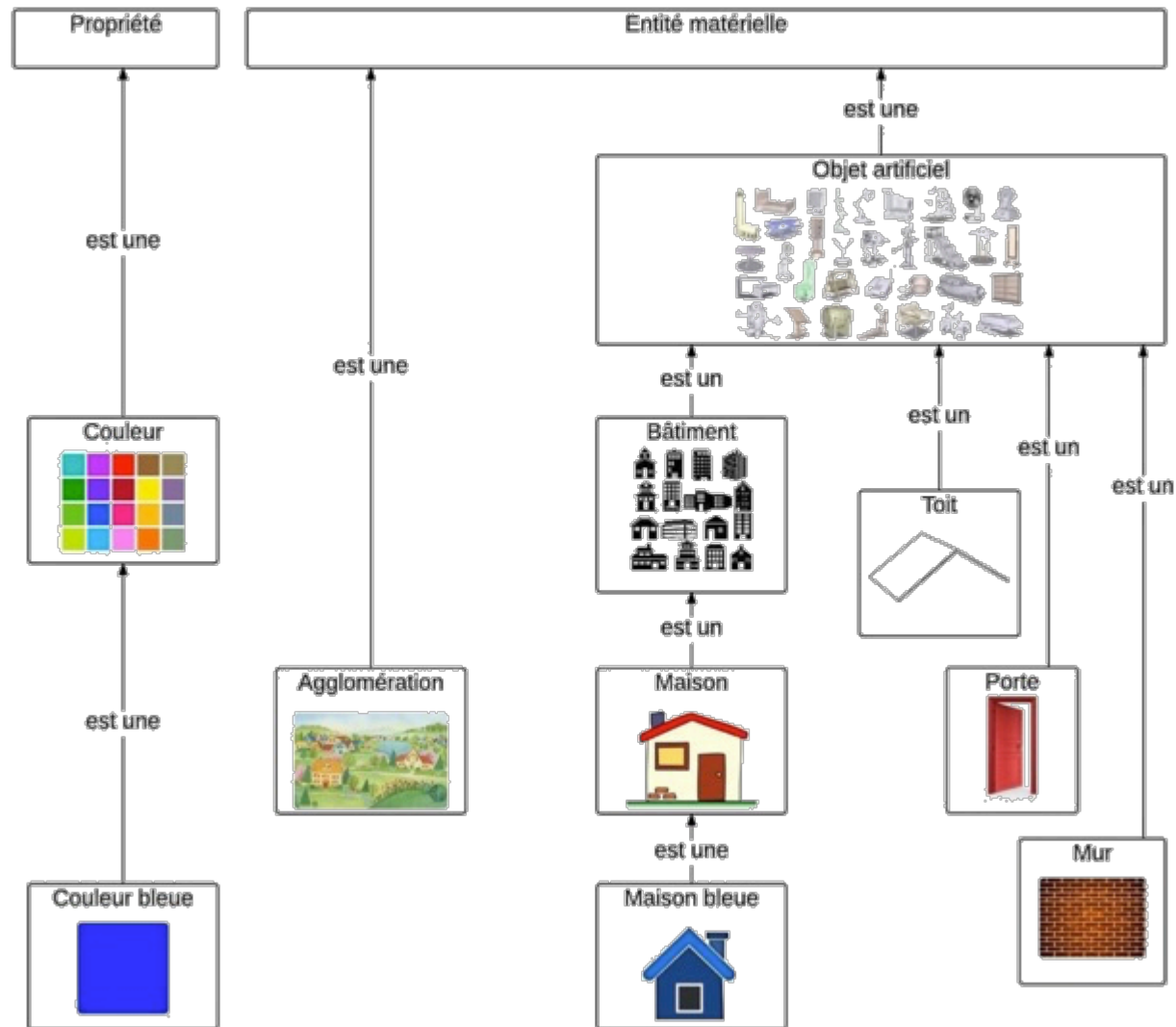


Taxonomie: Hiérarchie (= graphe orienté acyclique) constituée de termes dénotant des classes reliées par des relations de sous-type (*is_a*).

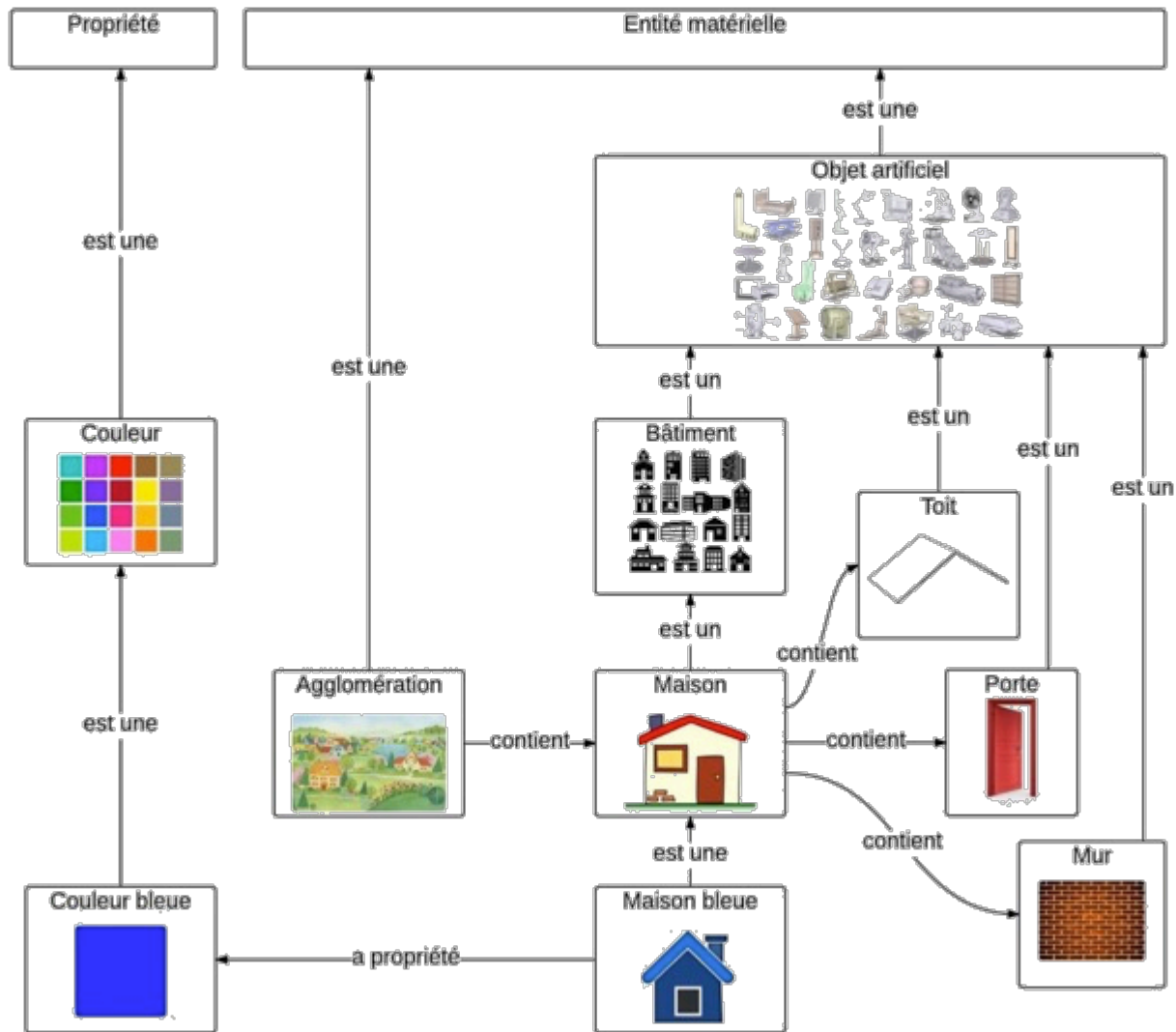
Terminologie

Terme	Définition
Maison	Bâtiment destiné à servir d'habitation à des humains.
Toit	Surface supérieure d'un bâtiment.
Porte	Ouverture limitant un espace clos, permettant la communication entre cet espace et ce qui est extérieur à cet espace.
Mur	Ouvrage de maçonnerie vertical (parfois oblique), élevé sur une certaine longueur pour constituer le côté d'un bâtiment.

Taxonomie



Ontologie

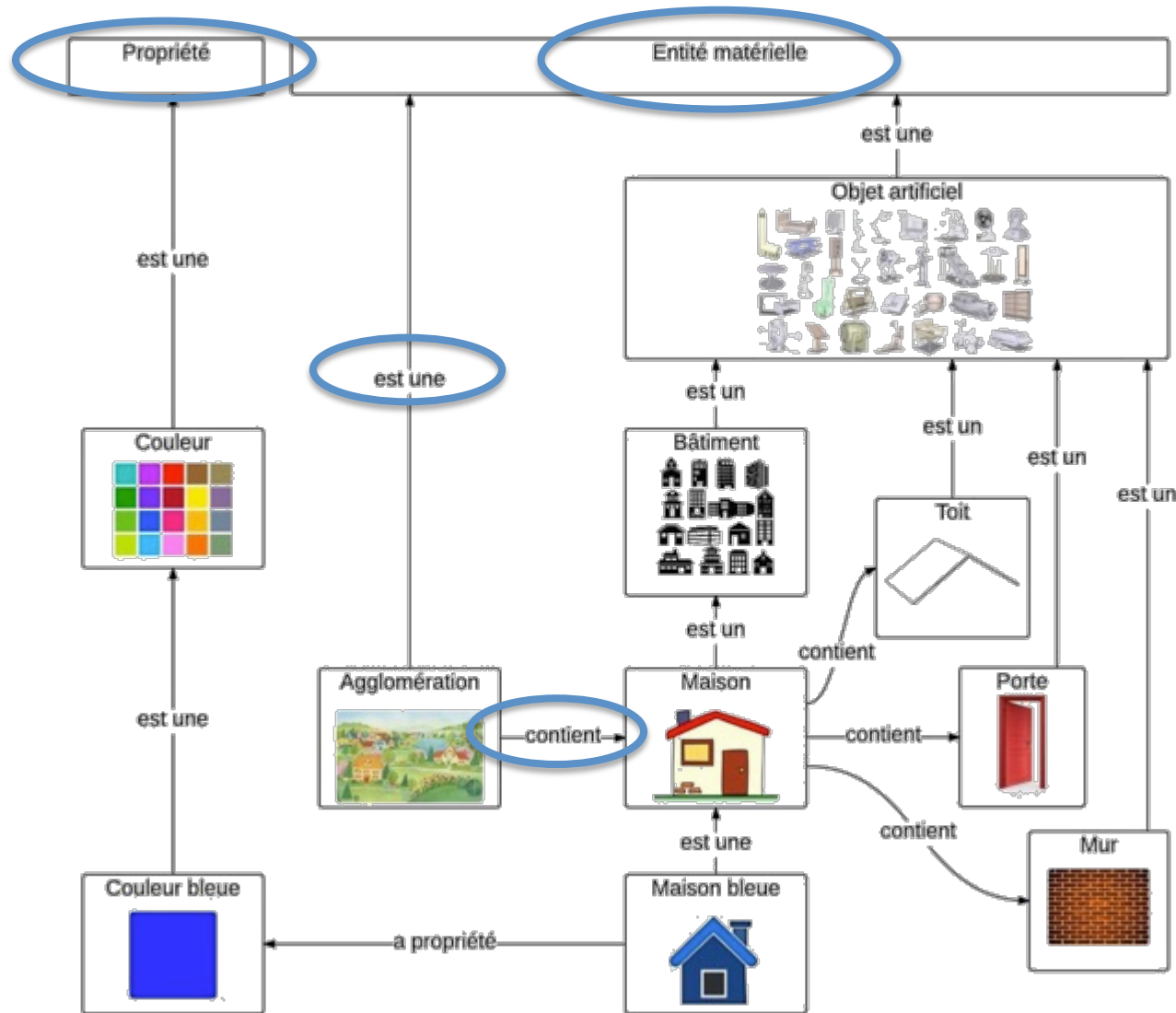


Description: House

SubClass Of +

- **Building**
- **has_part some Door**
- **has_part some Roof**
- **has_part some Wall**

L'apport de la philosophie



Apports de la métaphysique :

- similarité : problème des universaux
- identité et changement
- causalité
- méréologie (relation de tout à partie)
- ...

Apports de la philosophie des sciences :

- maladie
- risque
- probabilités
- ...

Ontologie de domaine

Egalement appelée “ontologie de référence” : représentation canonique des entités et relations d’un domaine donné, visant à l’exhaustivité.

Nous verrons trois autres types d’ontologies.

Exemple d’ontologie de domaine :
Cardiovascular Disease Ontology (CVDO)

CVDO dans Protégé

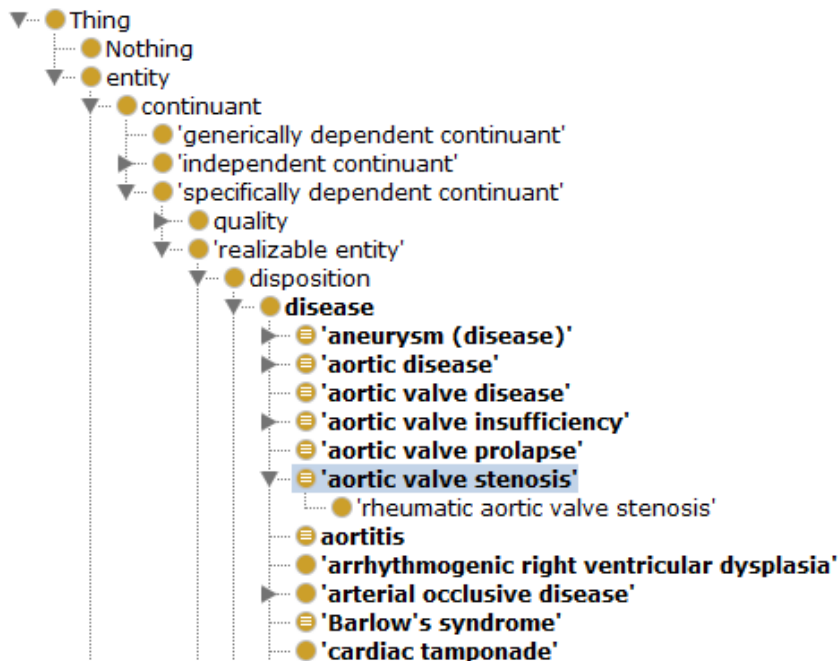
Définition formelle de la sténose aortique valvulaire

Description: 'aortic valve stenosis'

Equivalent To +

- **disease**
 - and ('realized in' some ('disease course' and ('has occurrent part' some 'aortic valve improper opening during left ventricular systole')))
 - and ('has material basis at all times' some 'aortic valve disorder')

Hiérarchie de classes



Quelques relations

- ▶ 'has location at some time'
- ▶ 'has material basis at all times'
- ▶ 'has occurrent part'
- ▶ 'has participant at some time'
- ▶ 'has profile'
- ▶ 'has spatial occupant at some time'
- ▶ 'has spatiotemporal occupant'
- ▶ 'has specific dependent at some time'
- ▶ 'located in at some time'
- ▶ 'material basis of at some time'
- ▶ 'occupies spatial region at some time'
- ▶ 'occupies spatiotemporal region'
- ▶ 'occurs in'
- ▶ 'part of continuant at some time'
- ▶ 'part of occurrent'
- ▶ 'participates in at some time'
- ▶ 'process profile of'
- ▶ 'projects onto spatial region at some time'
- ▶ 'realized in'
- ▶ realizes
- ▶ 'spatial projection of spatiotemporal at some time'
- ▶ 'specifically depends on at some time'

Les ontologies comme légendes pour les bases de données

GlyProt

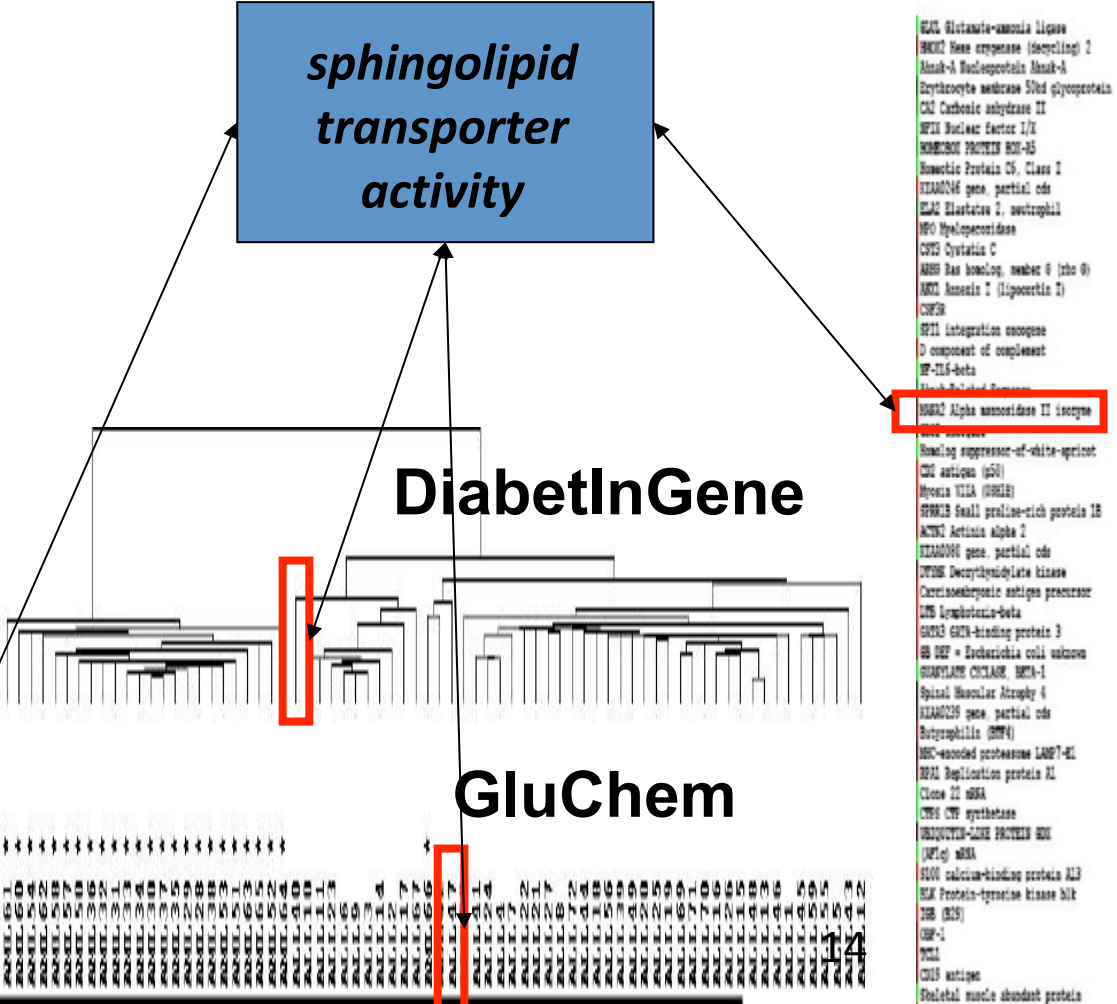
MouseEcotope

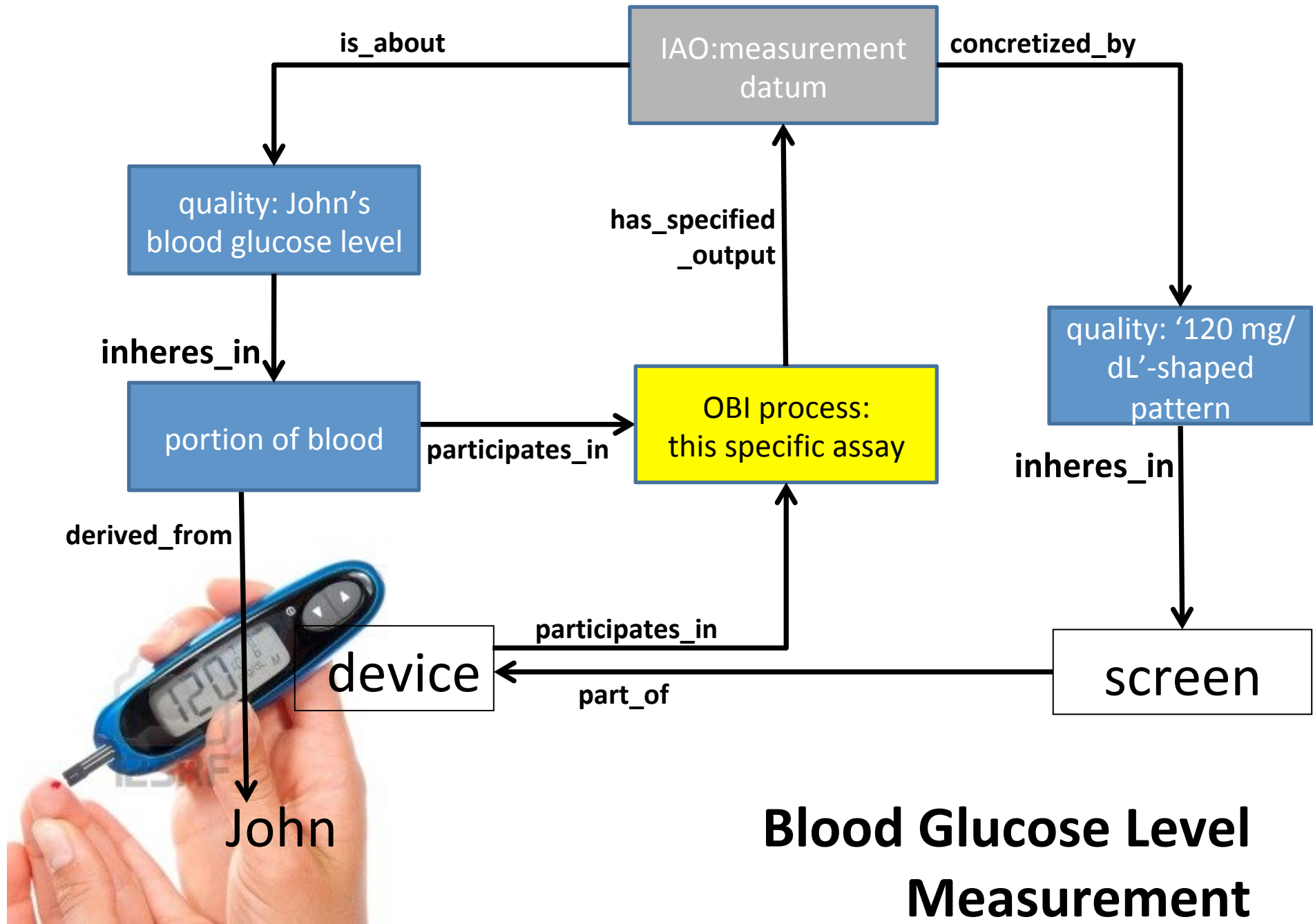
Tool	Statistical model	Correction for multiple experiments	GO Visualization	Microarrays supported	Time to process 200 genes (s)
Onto-Express	χ^2 , Naomani, hypergeometric, Fisher's exact test	Sidak, Holm, Benjamini, FDR	Flat, Tree	172 commercial arrays (Affymetrix, SuperArray, Sigma-Genosys, Clontech, PeptideSera, Openix, Takara, NDA); can also upload a non-defined list	7, 8, 16, 28
GOblast	Fisher's exact test	Relative enrichment	Tree, DAG	uploads from user	77, 123, 223, 340
DAVID	None	None	Not available	Not applicable	15, 17, 27, 54
EASE/online	Fisher's exact test	Benjamini	Not available	27 arrays (Affymetrix only); can also upload a non-defined list	15, 19, 34, 74
GeneGep	Hypergeometric	Benjamini	Flat, no hierarchical structure	Uploads from user	6, 6, 6, 6
FuncAssociate	Fisher's exact test	None	Not available	Uploads from user	22, 27, 29, 50
GOTM	Hypergeometric	None	Tree	37 arrays (Affymetrix only); uploads from user	39, 60, 137,
FatGO	Percentage	Step-down α -FDR (Benjamini and Hochberg, 1995); FDR (Benjamini and Yekutieli, 2001)	Flat, Tree	Uploads from user	15, 49, 69, 105
CLENCW	Hypergeometric, χ^2	None	DAG	Uploads from user	NA
Gblast	χ^2 , Fisher's exact test	FDR, Holm	Not available	Uploads from user	12, 26, 46, 86
GOST/online	Hypergeometric, Naomani, Fisher's exact test	Benjamini, Holm, Hochberg, Benjamini, FDR	Not available	Uploads from user	22, 81, 143, 5
GoMiner	χ^2	None	DAG	27 arrays (Affymetrix only); uploads from user	7, 7, 7, 7
Ontology Tracker	Hypergeometric	FDR	Not available	3 arrays (Affymetrix); uploads from user	NA
eGOs	Binomial	None	Tree	Uploads from user	20, 41, 80, 93

sphingolipid transporter activity

DiabetInGene

GluChem





Moteurs de recherche pour ontologies

Welcome to **BioPortal**, the world's most comprehensive repository of biomedical ontologies

Search for a class

[Advanced Search](#)

Find an ontology

[Browse Ontologies](#)

Ontology Visits (February 2018)

Ontology	Visits (approx.)
CPT	70,000
MEDDRA	15,000
RXNORM	12,000
SNOMEDCT	10,000
NDDF	8,000

[More](#)

BioPortal Statistics

Ontologies	692
Classes	8,847,370
Resources Indexed	48
Indexed Records	39,537,360
Direct Annotations	95,468,433,792
Direct Plus Expanded Annotations	144,789,582,932

Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., ... & Musen, M. A. (2009). BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, 37(suppl_2), W170-W173.

Partie II

Des ontologies de haut niveau pour
unifier les ontologies de domaines
hétérogènes



Difficultés

- Formation de silos de données d'un même domaine : la multiplication de terminologies et ontologies recrée le problème de l'interopérabilité à un plus haut niveau.
- La multiplicité des domaines représentés par des ontologies : médecine, biologie, industrie, géographie, éducation...



Solutions

- Créer des ensembles d'ontologies de domaine :
 - avec une seule ontologie par domaine.
 - unifiées par une ontologie de haut niveau.

- Exemples :

- OBO Foundry

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., ... & Leontis, N. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11), 1251-1255.

- Industrial Ontologies Foundry

Wallace, E., Kiritsis, D., Smith, B., & Will, C. (2018). The Industrial Ontologies Foundry proof-of-concept project. In *IFIP International Conference on Advances in Production Management Systems* (pp. 402-409). Springer, Cham.

The screenshot displays the OBO Foundry website. At the top, there is a navigation menu with links for 'About', 'Principles', 'Ontologies', 'Citation', 'Participate', 'FAQ', and 'Legacy', along with a search bar. Below the navigation is the title 'The OBO Foundry' and a paragraph of introductory text. A table lists various ontologies with columns for 'info', 'name', 'description', and 'format'. The table includes entries for 'Basic Formal Ontology', 'Chemical Entities of Biological Interest', 'Human Disease Ontology', 'Gene Ontology', 'Ontology for Biomedical Investigations', and 'Phenotypic quality'. Below the table, the browser address bar shows 'industrialontologies.org/top-down-wg/'. At the bottom of the page, there is a blue footer with navigation links for 'Home', 'About Us', 'IOF Materials', 'Working Groups', and 'Events'. The Industrial Ontologies Foundry (IOF) logo, featuring a gear and interconnected nodes, is prominently displayed in the center of the page.

Quatre types d'ontologies

- **Ontologie de haut-niveau** (ou formelle) : indépendante du domaine

Exemple : Basic Formal Ontology (BFO)

- **Ontologie de niveau intermédiaire** : utilisable par plusieurs (mais pas toutes) les ontologies de domaine.

Exemple : Ontology for General Medical Science (OGMS)

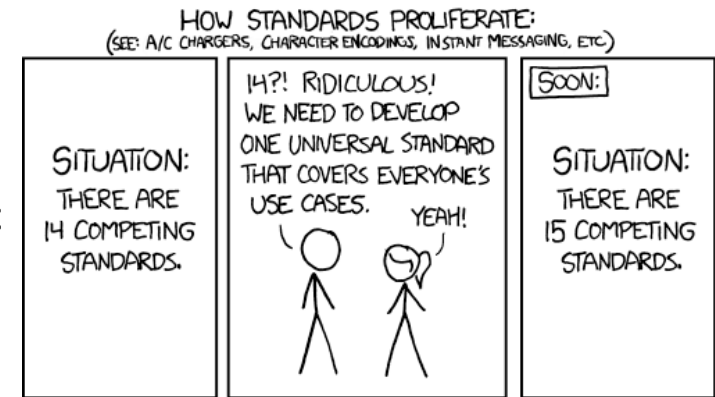
- **Ontologie de domaine** (ou de référence) : représentation canonique des entités et relations d'un domaine donné, visant à l'exhaustivité.

Exemple : Cardiovascular Disease Ontology (CVDO)

- **Ontologie d'application** : créée pour accomplir certains buts spécifiques.

Ontologies de haut niveau

- Plusieurs ontologies de haut niveau proposées :
 - BFO (Basic Formal Ontology)
 - DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering)
 - UFO (Unified Foundational Ontology)
 - GFO (General Formal Ontology)
 - ...



<https://xkcd.com/927/>

- Différence entre ces ontologies : acceptation ou non des universaux
- Point commun de ces ontologies : orientation générale **anti-réductionniste** :
 - Orientation **perspectivaliste** : représenter des perspectives complémentaires sur chaque domaine (partie II)
 - Orientation **adéquatiste** : rendre compte de la réalité à différents niveaux de granularité (partie III)

Universaux et particuliers :

Le carré ontologique d'Aristote

Angelelli I. *Studies on Gottlob Frege and Traditional Philosophy*. Dordrecht: Reidel, 1967

	Substance	Accident
Universel	<i>Humain</i> <i>Chat</i> <i>Table</i>	<i>Rougeur</i> <i>Mal de tête</i> <i>Peur</i>
Particulier	<i>cet humain</i> <i>ce chat</i> <i>cette table</i>	<i>cette rougeur</i> <i>ce mal de tête</i> <i>cette peur</i>

Logique de prédicat standard – $F(a)$, $R(a,b)$...

Armstrong, D. M. (1978). *Nominalism and Realism: Volume 1: Universals and Scientific Realism*. CUP Archive.

	Substance	Accident
Universel		Attributs F, G, R
Particulier	Individus a, b, c	

Nominalisme bicatégoriel

Rodriguez-Pereyra, G. (2019) Nominalism in Metaphysics. *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2019/entries/nominalism-metaphysics/>>.

	Substance	Accident
Universel		
Particulier	<i>cet humain</i> <i>ce chat</i> <i>cette table</i>	<i>cette rougeur</i> <i>ce mal de tête</i> <i>cette peur</i>

Perpectivalisme

Johnson, W. 1921. *Logic: Part I*. Cambridge : Cambridge University Press.

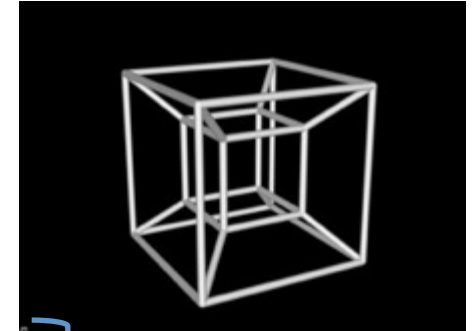
Perspective 3D (continuants / endurants)



Mitochondries
Cellules
Organes
Organismes
Espèces
Familles
Sociétés
Populations
Embryons

existent
pleinement à
tout instant de
leur existence

Perspective 4D (occurrents / perdurants)



Circulation sanguine
Division de cellule
Enfance
Vieillesse
Croissance
Apprentissage
Reproduction
Mort

n'existent pas
pleinement à
un instant : ils
ont des parties
temporelles

Perspectivalisme

- Il peut y avoir des perspectives alternatives fournissant des représentations également vraies qui capturent différentes caractéristiques importantes de la même réalité.
- Perspectivalisme \neq Relativisme

Approche perspectivaliste : Le carré ontologique étendu

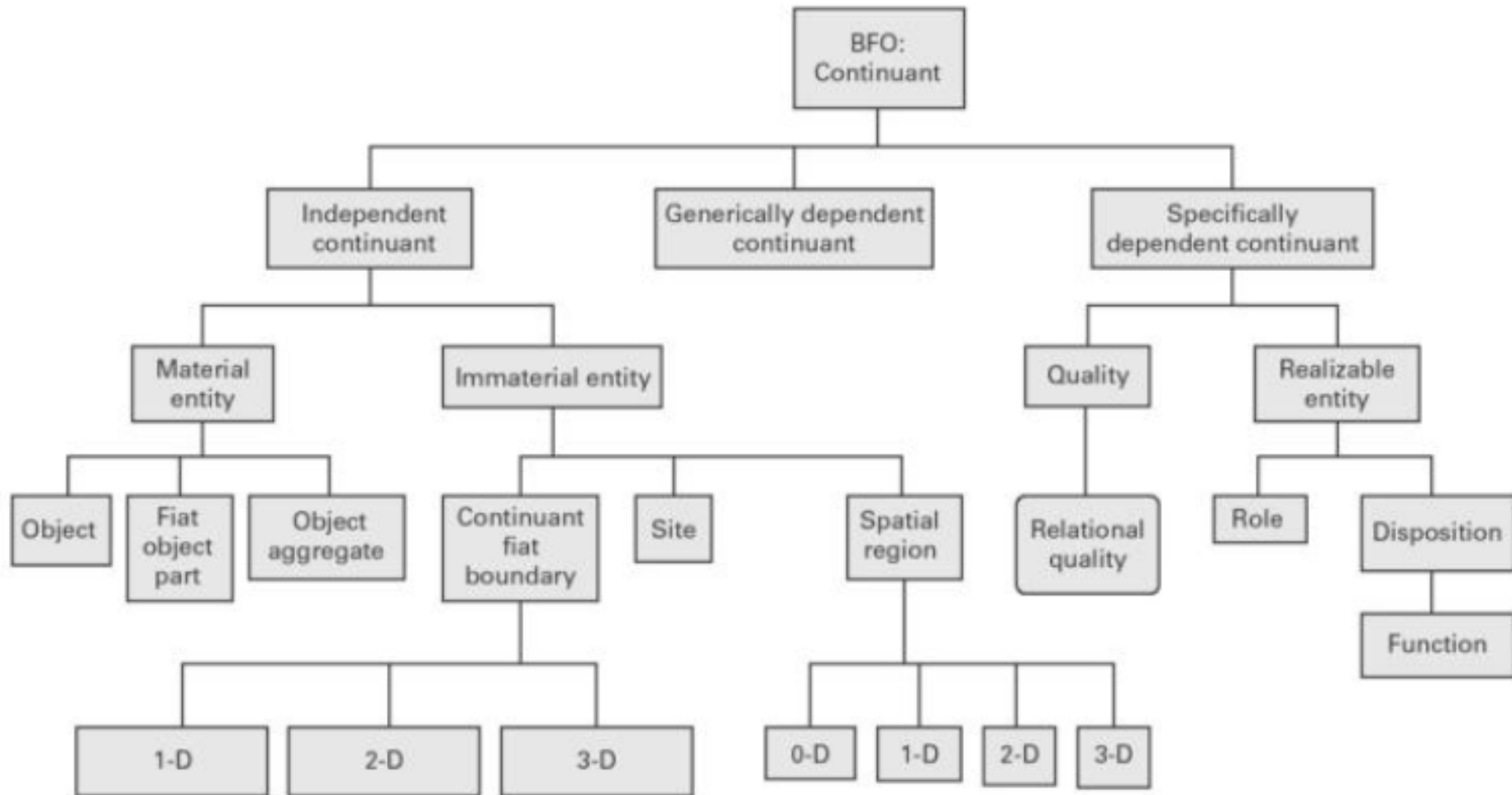
Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with basic formal ontology*. Mit Press.

	Substances	Accidents	Processus
Universaux	<i>Humain</i> <i>Chat</i> <i>Table</i>	<i>Rougeur</i> <i>Mal de tête</i> <i>Peur</i>	<i>Cours</i> <i>Dîner</i> <i>Circulation sanguine</i>
Particuliers	<i>cet humain</i> <i>ce chat</i> <i>cette table</i>	<i>cette rougeur</i> <i>ce mal de tête</i> <i>cette peur</i>	<i>ce cours</i> <i>ce dîner</i> <i>cette circulation sanguine</i>

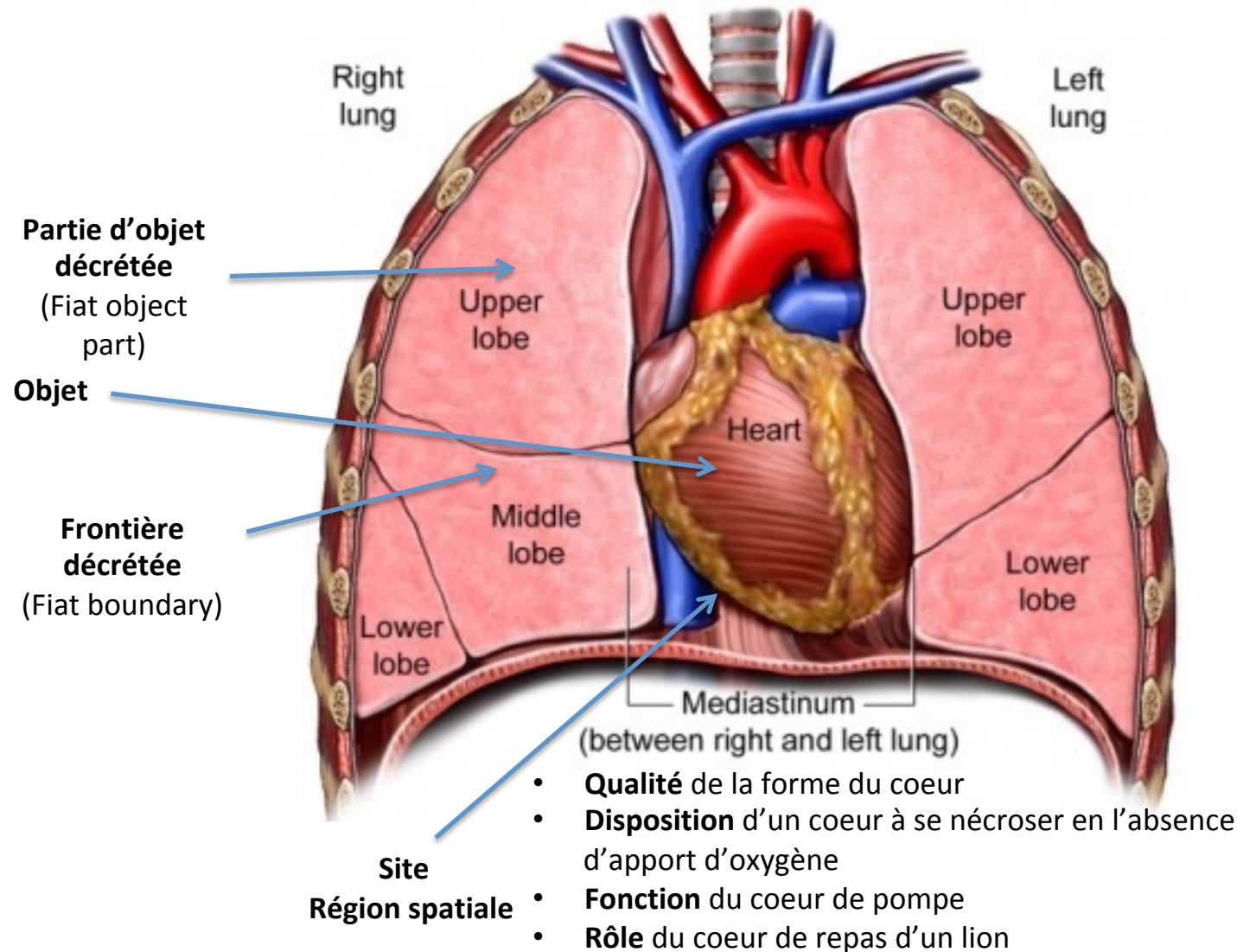
Approche anti-perspectivaliste : Métaphysique des processus

	Substances	Accidents	Processus
Universaux			
Particuliers			<i>cette vie d'individu</i>

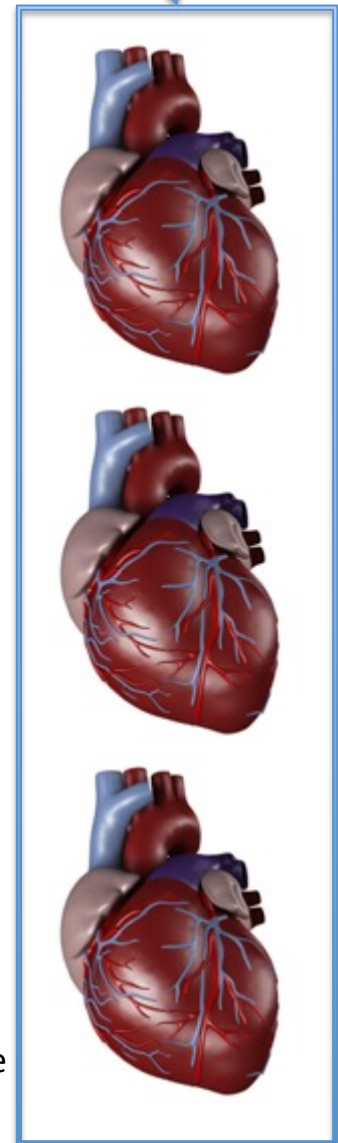
BFO, continuant



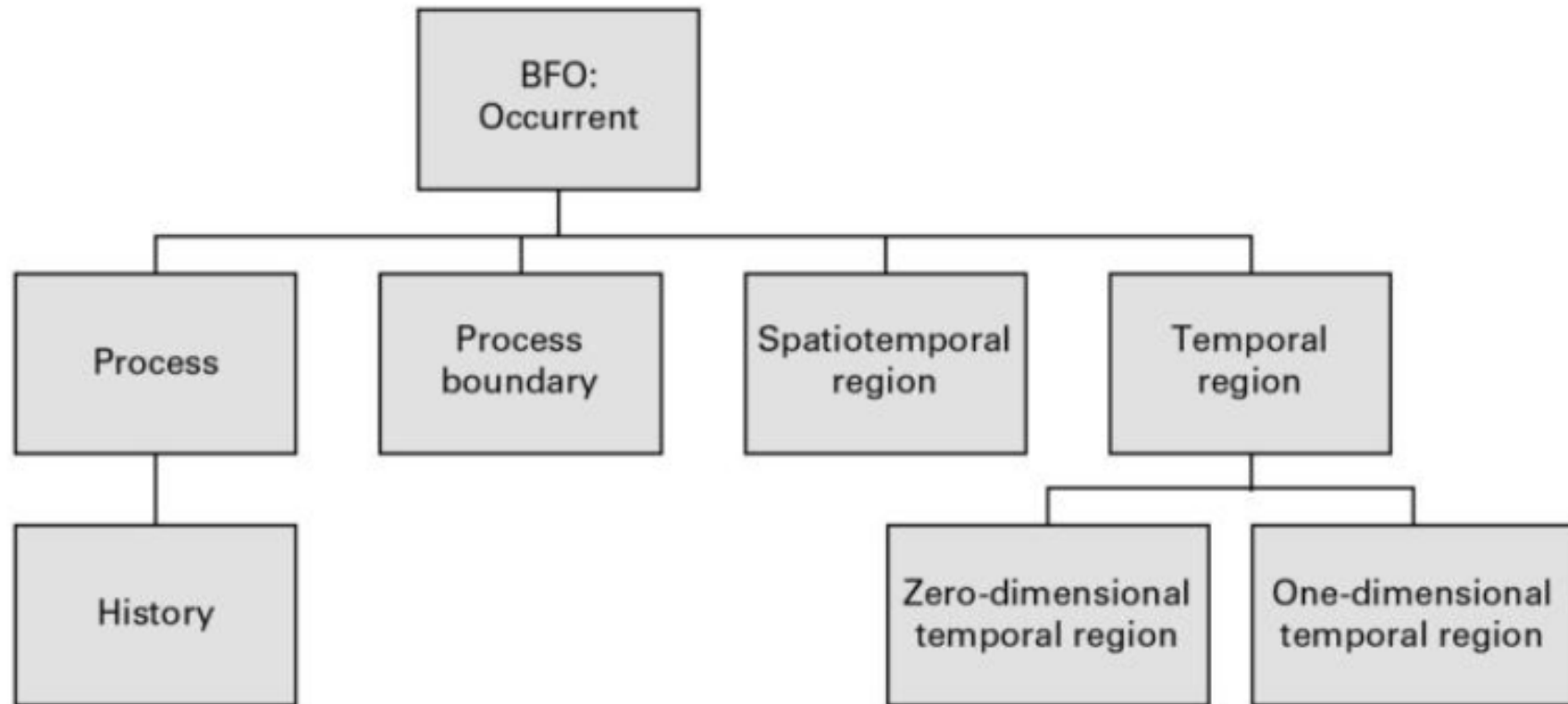
Continuants : exemples



Aggrégat
d'objets



BFO, occurrents

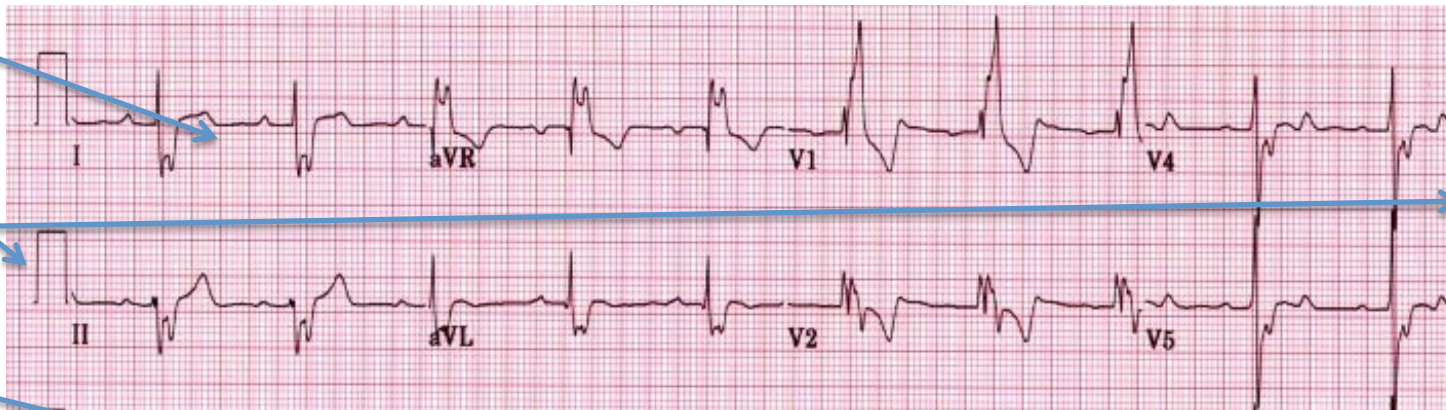


Occurrents : exemples



Processus de réalisation d'ECG

Processus de mesure d'activité électrique



Frontière de processus

Région spatio-temporelle occupée par le coeur

Instant temporel



t1

Histoire du coeur

Intervalle temporel

t2

Partie III

Des théories ontologiques pour
articuler des niveaux de réalité
hétérogènes



Exemples de description à différents niveaux de réalité

Statue
vs. Quantité d'argile



Morceau de papier imprimé
vs. Billet



Presse-papier
vs. Caillou
vs. Matière du caillou



Constitution

Johnston, M. (1992). Constitution is not Identity. *Mind*, 101(401), 89-105.

Quelle est la relation entre :

- Une statue et l'argile dont elle est constituée ?
- Un bateau et le bois, clous, tissu, etc. dont il est fait ?
- Une bactérie et la matière organique dont elle est faite ?

Certaines ontologies (e.g. DOLCE) sont multiplicativistes:

- la statue est un objet physique
- l'argile est une quantité de matière
- la statue est différente de la quantité de matière dont elle est constituée

→ théories formelles de la constitution

Claudio Masolo, Laure Vieu, Roberta Ferrario, Stefano Borgo and Daniele Porello. Collectives, Composites and Pluralities. *Formal Ontology in Information Systems: Proceedings of the 11th International Conference (FOIS 2020)* (Vol. 330). IOS Press.



Approches non-multiplicativistes de la constitution

Noonan, H. W. (1993). Constitution is identity. *Mind*, 102(405), 133-146.

Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with basic formal ontology*. Mit Press.

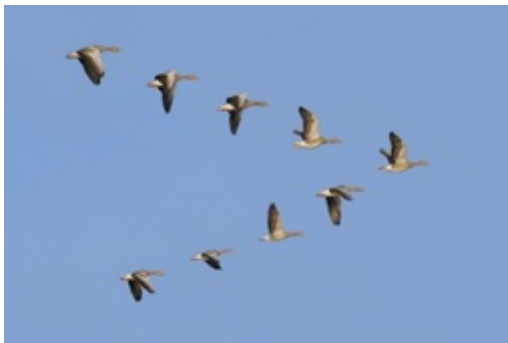
- La portion d'argile est identique à la statue pendant le temps où cette statue existe.
- Le bateau est la somme méréologique des portions de bois, clous, tissus etc. pendant le temps où ils sont combinés en une forme-de-bateau.



Aggrégat d'objets

a est un agrégat d'objet =

Il y a une partition de a en objets mutuellement exhaustifs et disjoints deux à deux.



Une quantité d'argile peut être identifiée à un agrégat de molécules.

Les deux tables d'Eddington



Table 1

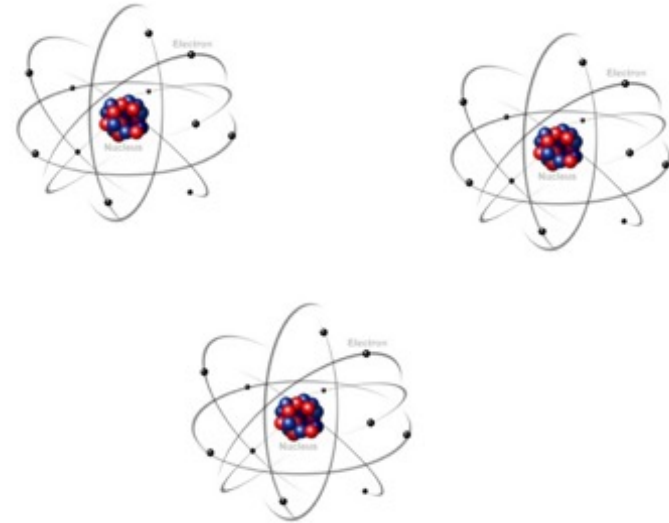


Table 2

La table 1 et la table 2 existent à deux niveaux de granularité différents.

- Pour Eddington, seule la table 2 existe.
- Pour les ontologies appliquées multiplicativistes, la table 1 et la table 2 existent et ne sont pas identiques.
- Pour les ontologies appliquées non-multiplicativistes, la table 1 et la table 2 existent et sont identiques.

Rôles



Différentes possibilités de représentation :

- Un rôle de billet inhère dans le morceau de papier.

Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with basic formal ontology*. Mit Press.

- Il existe deux objets : le morceau de papier (objet matériel) et le morceau-de-papier-qua-billet (« qua entity »).

Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gangemi, A., & Guarino, N. (2004). *Social Roles and their Descriptions*. Principles of Knowledge Representation and Reasoning (KR 04), 267–277.

Fondation métaphysique (« Grounding »)

Schaffer, J. (2009) On What Grounds What. In *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford University Press. pp. 347-383.

- Intuitivement :

b est fondé sur a

ssi

b existe en vertu de l'existence de a.

- Exemple :

- Le presse-papier est fondé sur le caillou.
- Le caillou est fondé sur la quantité de matière du caillou.

- Formalisation des notions de fondation métaphysique et de niveau de réalité :

Masolo, C. (2010). *Understanding Ontological Levels*. Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010), pp. 258-268.



Intrication physique/mental/social

Ferrario, R., & Porello, D. (2015). Towards a conceptualization of sociomaterial entanglement. *International and Interdisciplinary Conference on Modeling and Using Context* (pp. 32-46)

- Modules ontologiques pour différents domaines (physique, mental, social)
- Un objet manifeste une **intrication ontologique** s'il est (métaphysiquement) fondé sur (au moins) deux entités qui appartiennent à des modules différents.
- Exemple: Un caillou-qua-presse-papier est fondé sur deux entités :
 - Une entité **physique** : le caillou
 - Une entité **mentale** : l'intention de la personne qui s'en sert comme presse-papier
- Les auteurs proposent une caractérisation formelle de cette notion d'intrication.



Conclusion

- Partie I : Des ontologies de domaines pour unifier des données issues de sources hétérogènes
- Partie II : Des ontologies de haut niveau pour unifier des ontologies de domaines hétérogènes
- Partie III : Des théories ontologiques pour articuler des niveaux de réalité hétérogènes



Ontologie appliquée et données hétérogènes



Merci pour votre attention !

Adrien Barton

IRIT, CNRS & GRIIS, Université de Sherbrooke

adrien.barton@irit.fr

Remerciements : Barry Smith, Quentin Rano