

Modélisation et découverte de connaissances dans des données de processus de transformations



Séminaire Réseau IN-OVIVE

Joe RAAD

Encadrantes

Juliette DIBIE, Liliana IBANESCU, Nathalie PERNELLE, Fatiha SAIS

Date : Vendredi 2 décembre 2016

Lieu : AgroParisTech, Paris



I. INTRODUCTION

Modélisation et découverte de connaissances dans des données de processus de transformations

Projet LIONES (CDS Université Paris-Saclay)

Informatique

- Equipe **LInK** de **MIA**
(AgroParisTech & INRA)
- Equipe **LaHDAK** du **LRI**
(Université Paris-Sud)



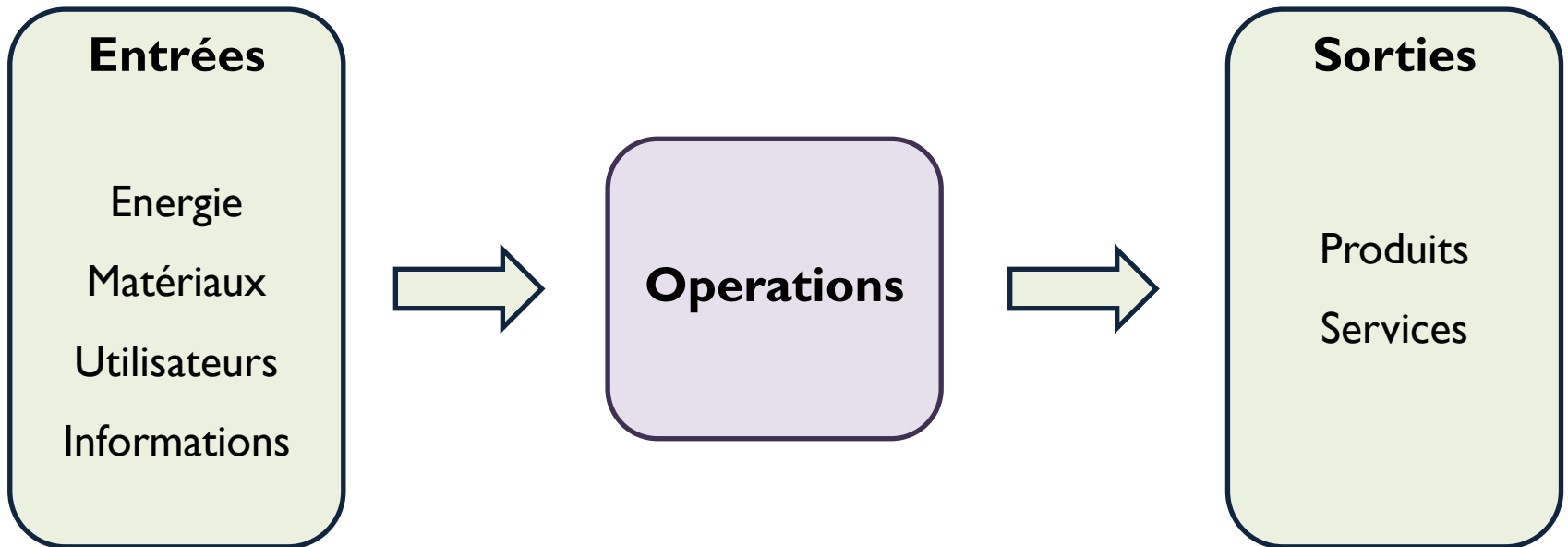
Sciences biologiques

- Equipe **BioMiP** du **GMPA**
(INRA Versailles - Grignon)
- **CSGA**
(INRA Dijon)



I. INTRODUCTION

Qu'est ce qu'un **Processus de Transformation** ?



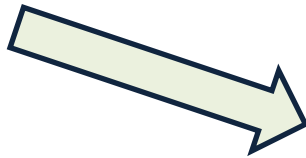
I. INTRODUCTION

Exemple d'un Processus de Transformation

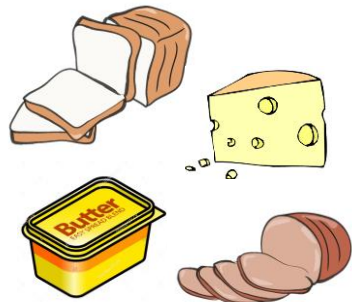
Préparation d'un sandwich



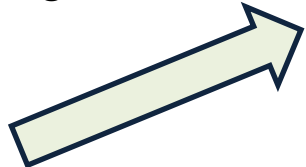
Temps, efforts et
compétences du
chef



Equipements



Ingrédients



Operations

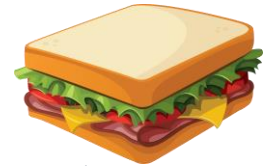
Etaler le beurre
15g

Choisir le pain
2 Pains de mie

Ajouter fromage
1 tranche

Ajouter jambon
2 tranches

Chauffer
7 mins à 60°C



observation



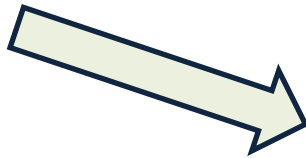
I. INTRODUCTION

Exemple d'un Processus de Transformation

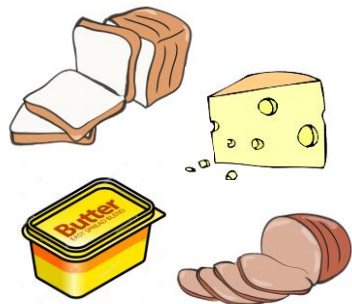
Préparation d'un sandwich (recette grand-mère)



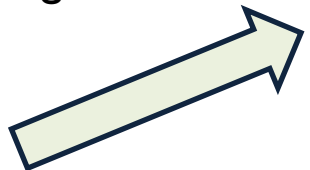
Temps, efforts et
compétences du
chef



Equipements



Ingrédients



Operations

Etaler le beurre
10g

Choisir le pain
1 baguette

Ajouter fromage
2 tranches

Ajouter jambon
2 tranches

Chauffer
4 mins à 100°C



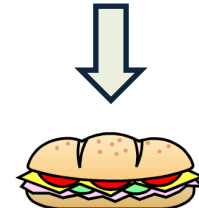
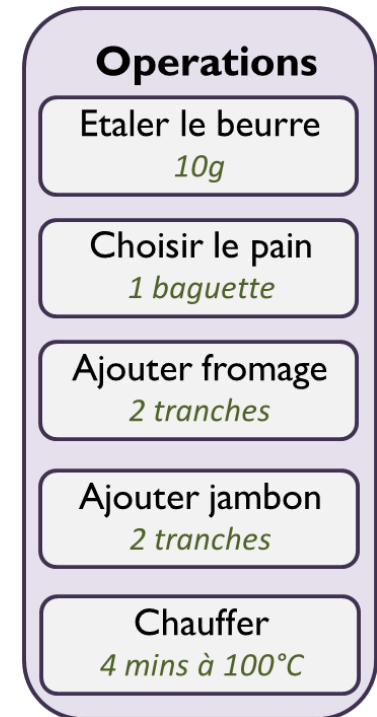
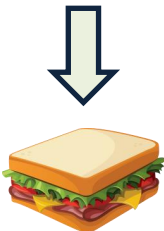
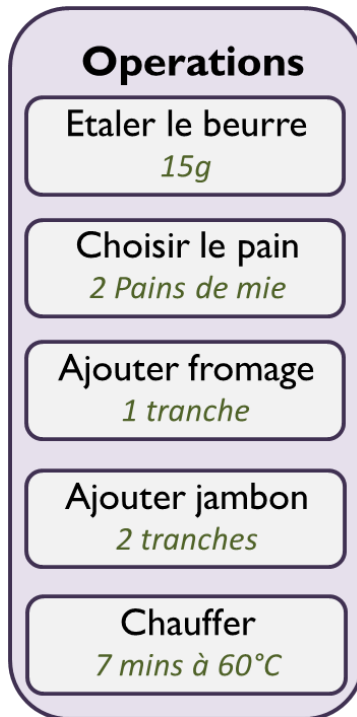
observation



I. INTRODUCTION

Exemple d'un Processus de Transformation

Préparation d'un sandwich



Est-ce que l'ajout de fromages rend le sandwich meilleur ?

Est-ce que chauffer plus longtemps à une plus faible température réduit la qualité du produit ?

Et est-ce que cela a un impact environnemental ?

I. INTRODUCTION

Comment peut-on bénéficier des données expérimentales pour répondre à ces questions ?

Découvrir de nouvelles connaissances sous la forme de règles :

+ fromage → + qualité

baguette ∧ haute_température → + consommation_électrique

Domaines d'applications plus complexes :

- *Processus de fabrication et stabilisation de micro-organismes*
- *Processus de transformation des gels laitiers*

I. INTRODUCTION

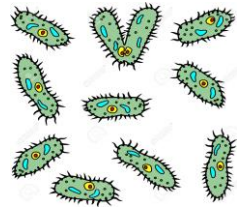
Premier domaine d'application

Fabrication et stabilisation de micro-organismes

Définition

Agents biologiques présents dans de nombreuses applications, notamment la nourriture et le domaine médical

Exemple : bactéries, levures, etc...



Exigences des biologistes

Concevoir des processus de façon à trouver un compromis entre le coût économique, l'impact environnemental, et la qualité du produit



Caractéristique de ce domaine

Données complexes



I. INTRODUCTION

Complexité des données

Incomplètes et Imprécises

Expérimentation 1

- Etape : Fermentation
- Appareil : Fermenteur
- Marque : Servilab
- Volume : [45 – 50]
- Unité : L

Expérimentation 2

- Etape : Fermentation
- Capteur : EMN20-W4
- ID capteur : 144
- Adresse EMN : 78
- Info complémentaire:
Le compresseur est instrumenté
par deux capteurs différents,
LEM D3SP2 et LEM EMN20

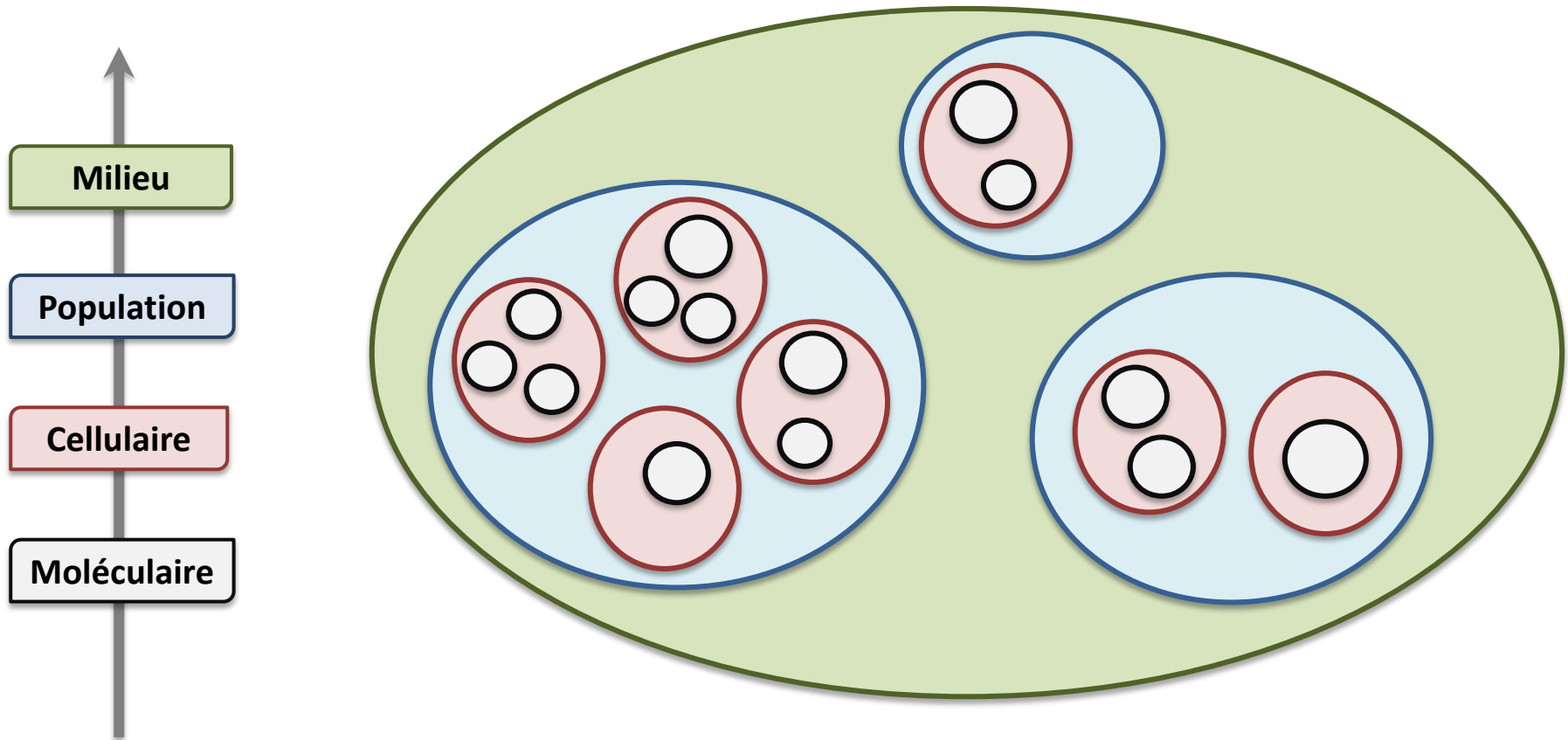
Expérimentation 3

- Etape : Fermentation
- Info complémentaire:
Méthode d'Ellman
(Ellman, 1959)

I. INTRODUCTION

Complexité des données

Multi-échelles



I. INTRODUCTION

Complexité des données

Dynamiques



Solide



Liquide



Gaz



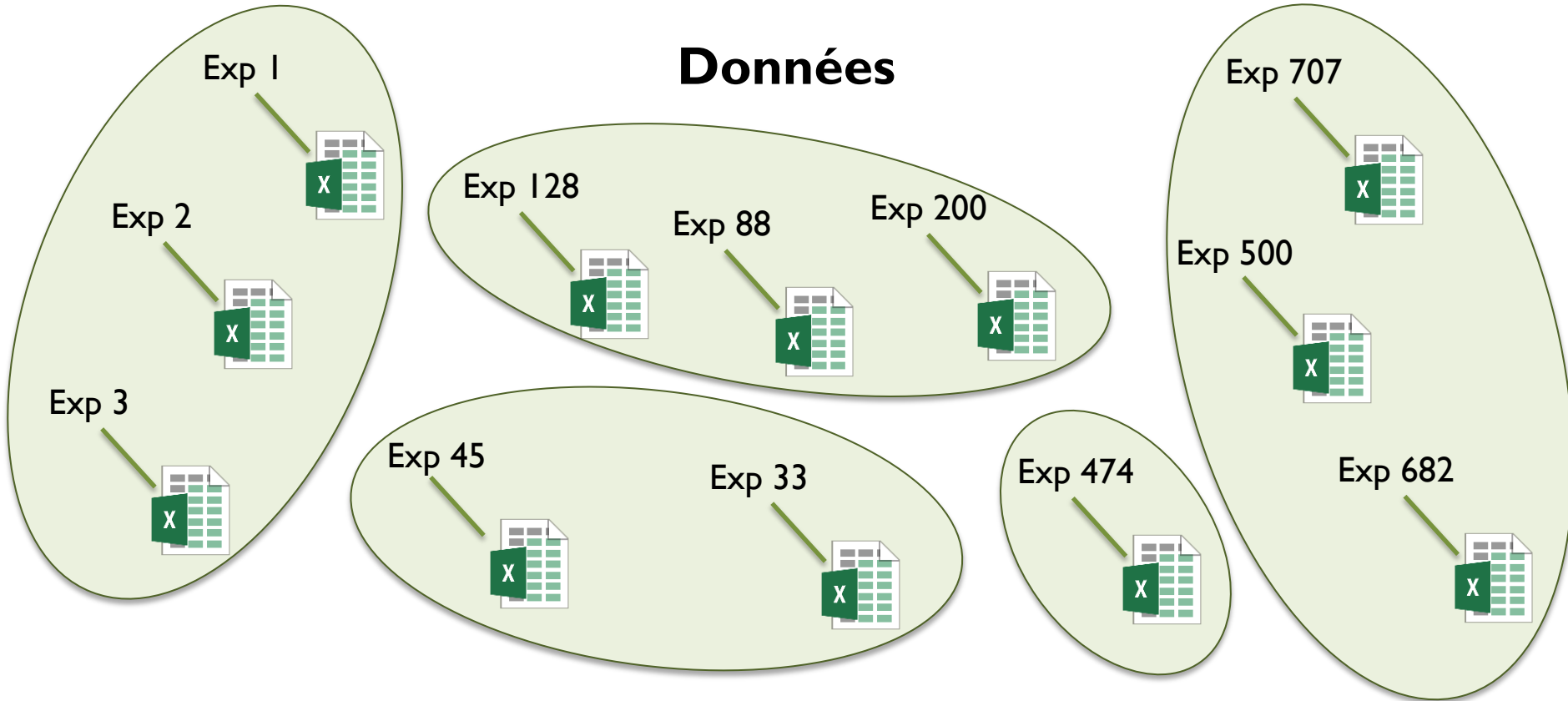
Même Objet ?

I. INTRODUCTION

	A	B	C	D	E	F	G
1	Informations générales						
2	<i>Date de la mesure</i>	2016-01-12					
3	<i>Heure</i>	14:00					
4	<i>Durée de la manipulation</i>	02:30					
5	<i>Étape</i>	Fermentation					
6	<i>Sous étape</i>	Stérilisation du fermenteur et du milieu					
7	<i>Répétition</i>	1					
8	<i>Echelle</i>	Procédé					
9							
10	Descriptif du produit						
11	<i>Code échantillon</i>	2016-01-11-CellExtraDry-001					
12	<i>Nom du fichier descriptif du cycle de vie</i>	2016-01-11-CellExtraDry-001-FicheDescriptifCV					
13							
14	Matériel utilisé						
15	<i>Identifiant</i>	Matériel 8					
16	<i>Fichier</i>	2016-CellExtraDry-Matériel&Méthode					
17							
18	Matériel utilisé						
19	<i>Identifiant</i>	Matériel 27					
20	<i>Fichier</i>	2016-CellExtraDry-Matériel&Méthode					
21							
22	Données calculées						
23	<i>Caractéristiques</i>	<i>Objet</i>	<i>Objet</i>	<i>Objet</i>	<i>Valeur</i>	<i>Unité</i>	<i>Description</i>
24	Energie active	P1			2068.125	Wh	
25	Energie active	P2			0	Wh	Non instrumenté
26	Energie active	P3			-2049.38	Wh	
27	Energie active	P1	P2	P3	18.75	Wh	Somme des énergies actives de P1-P2 et P3
28							
29	Données brutes						
30	<i>Temps relatifs</i>	<i>Frequency</i>	<i>intensity (P1)</i>	<i>intensity (P2)</i>	<i>intensity (P3)</i>	<i>voltage (P1)</i>	<i>voltage (P2)</i>
31	<i>heure</i>	<i>Hz</i>	<i>A</i>	<i>A</i>	<i>A</i>	<i>V</i>	<i>V</i>
32			<i>Minimum</i>	<i>Minimum</i>	<i>Minimum</i>	<i>Minimum</i>	<i>Non instrumenté; Minimum</i>
33	0	50	2.3	0	2.25	403.5833	0
34	0.083681765	50	2.266667	0	2.283333	403.1667	0
35	0.167363531	50	2.3	0	2.283333	403.9167	0
36	0.251045296	50	2.3	0	2.283333	403.0833	0
37	0.334727062	50	2.283333	0	2.3	404.1667	0
38	0.418408827	50	2.283333	0	2.283333	397.9167	0
39	0.502090592	50	2.3	0	2.266667	403.75	0
40	0.585772358	50	2.3	0	2.266667	403.3333	0
41	0.669454123	50.0625	2.283333	0	2.266667	402.5	0
42	0.753135889	50	2.316667	0	2.283333	404.1667	0
43	0.836817654	50	2.3	0	2.283333	403.4167	0

I. INTRODUCTION

Données



707 fichiers

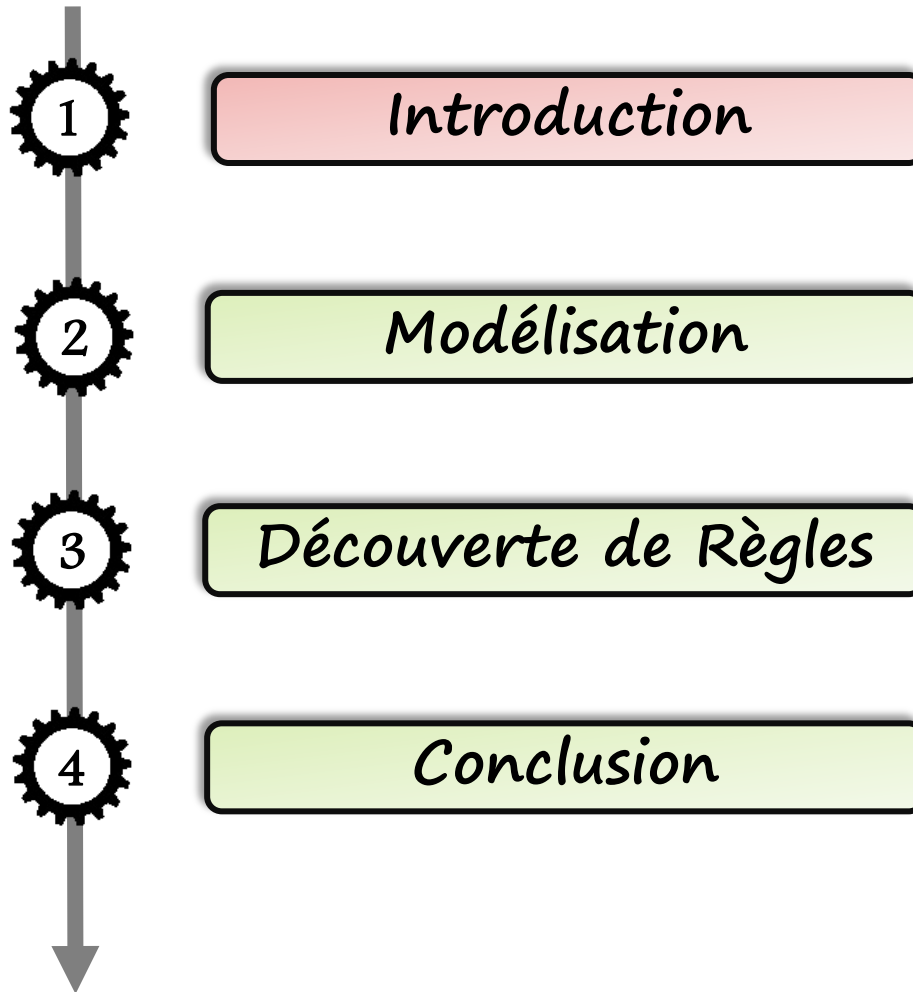
(processus de fabrication et stabilisation de micro-organismes)

10 Processus

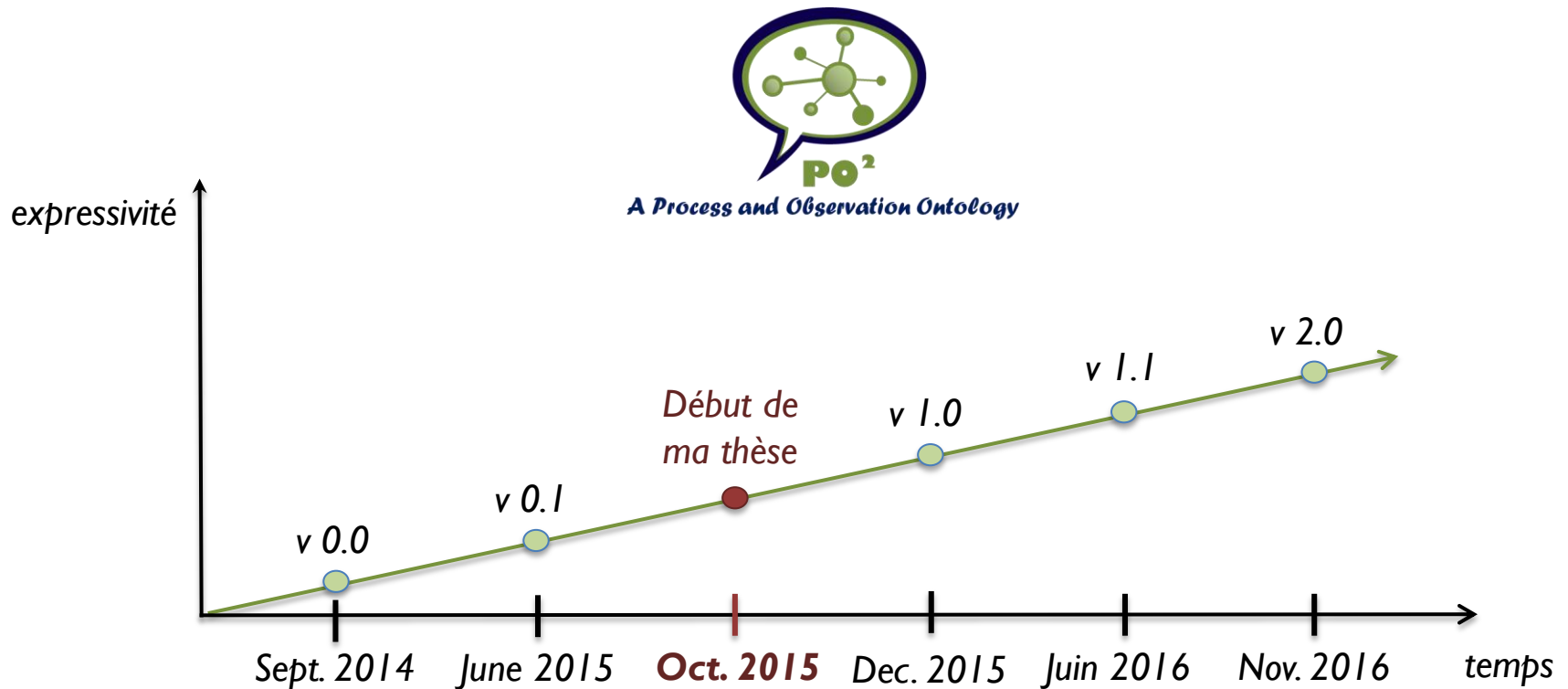
I. INTRODUCTION

Comment peut-on bénéficier des données expérimentales pour découvrir de nouvelles connaissances ?

PLAN



2. MODELISATION

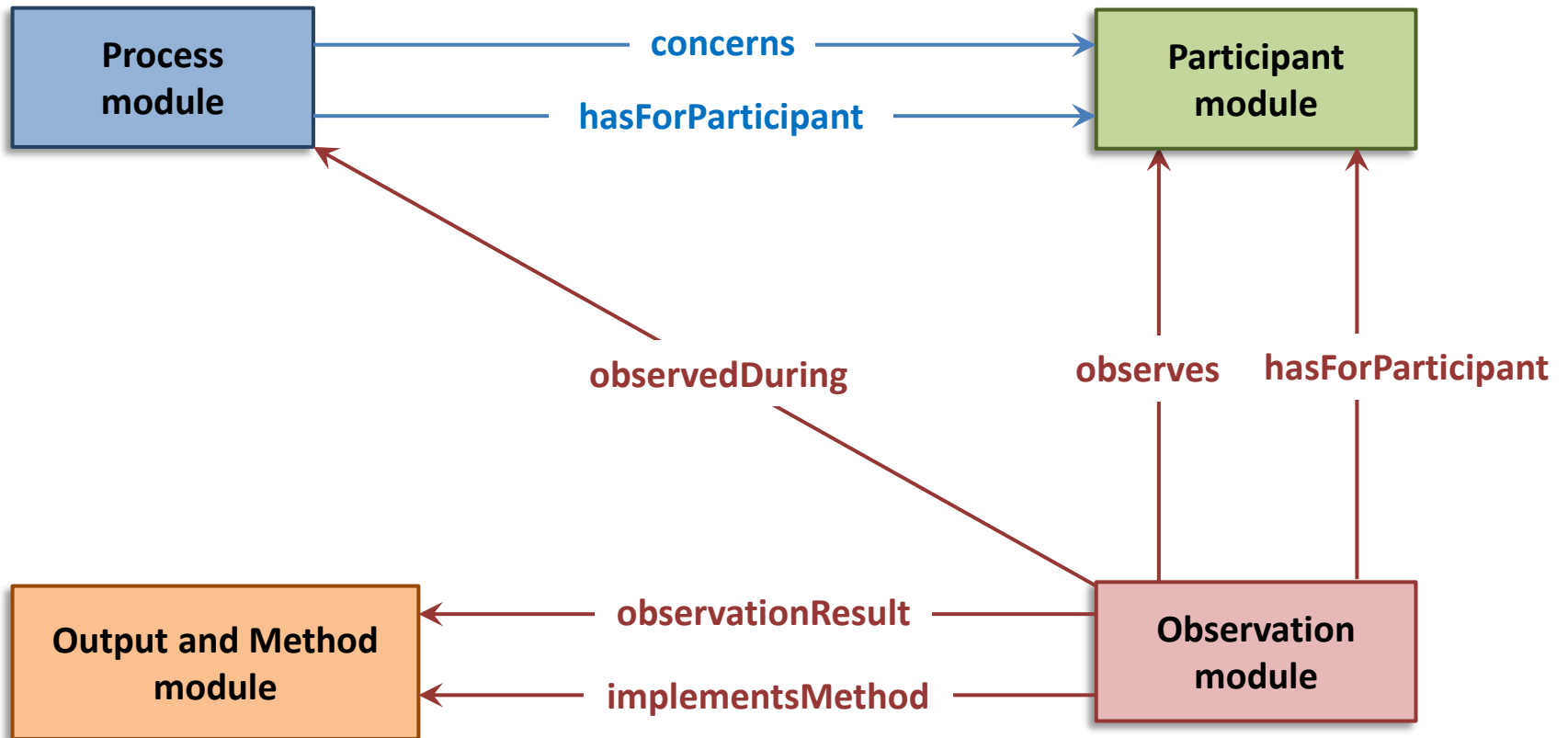


V 1.0 : Dibie, Juliette, et al. "[MS]²O - A Multi-scale and Multi-step Ontology for Transformation Processes: Application to Micro-Organisms." International Conference on Conceptual Structures. Springer International Publishing, 2016.

V 2.0 : Ibanescu, Liliana, et al. "PO² - A Process and Observation Ontology in Food Science. Application to Dairy Gels." Metadata and Semantics Research: 10th International Conference, MTSR 2016, Göttingen, Germany, November 22-25, 2016, Proceedings. Springer International Publishing, 2016.

2. MODELISATION

Modules de PO²



2. MODELISATION

Observation

Multi-échelles

☆ Observation au temps t ,
à l'étape i et à l'échelle e

Analyse
Multi-critères

échelle

Milieu

Population

Cellulaire

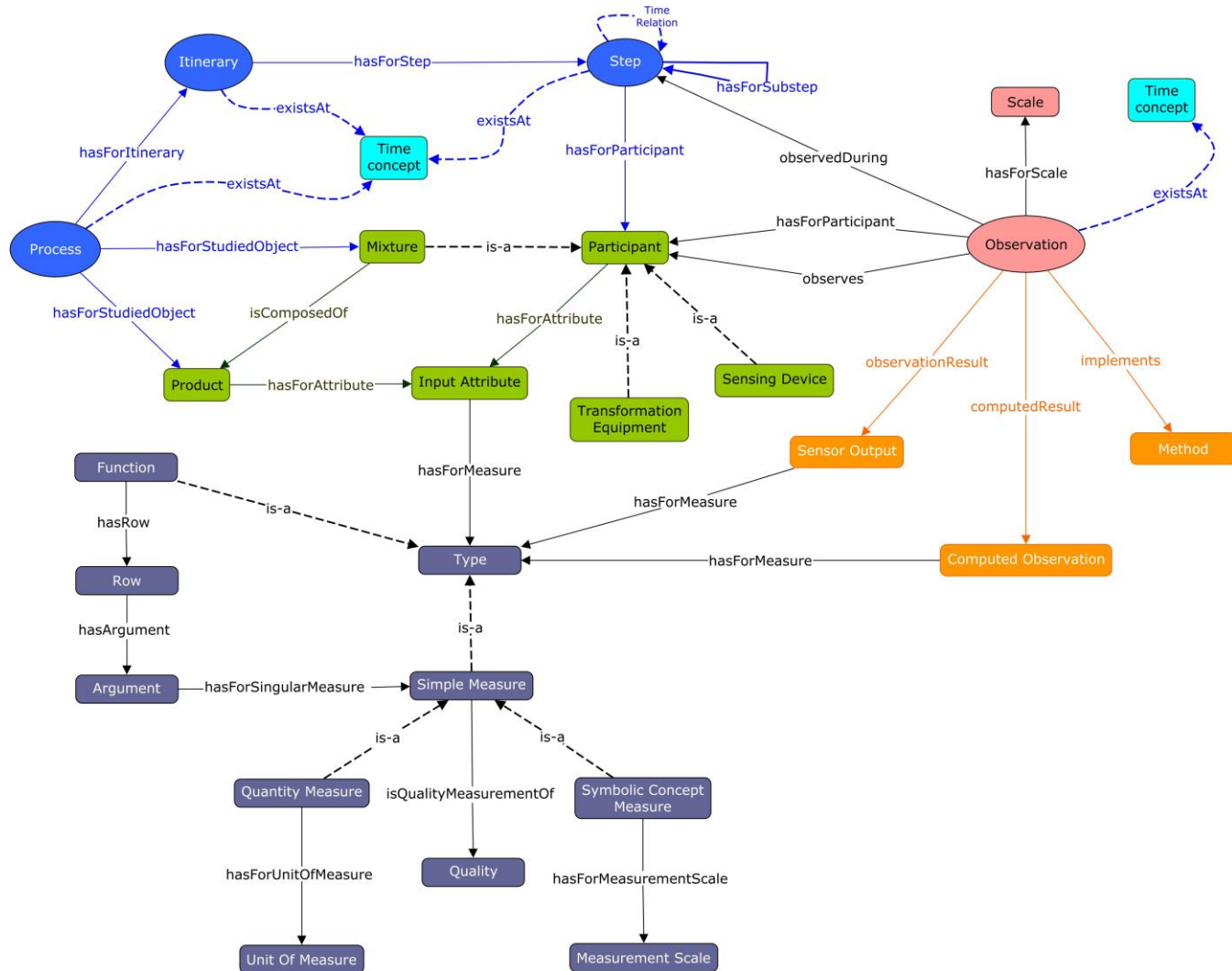
Moléculaire



Multi-étapes

2. MODELISATION

Concepts de PO²



2. MODELISATION



A Process and Observation Ontology

- **Expressive** : capacité à modéliser des processus de transformations et des expérimentations scientifiques
- **Générique** : application à un autre domaine (*transformation des gels laitiers*)
- **Riche** : enrichissement de PO² avec des thésaurus/taxonomies de domaine (*Agrovoc¹*)
- **Alignée** : alignement avec des ontologies existantes (*OWL-Time², OM³, IAO⁴*)
- **Interopérable** : intégration avec une ontologie fondationnelle (*BFO⁵*)

1. <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>

2. <https://www.w3.org/TR/2016/WD-owl-time-20160712/>

3. <http://www.wurvoc.org/vocabularies/om-1.6/>

4. <http://bioportal.bioontology.org/ontologies/IAO>

5. <http://ifomis.uni-saarland.de/bfo/>

2. MODELISATION

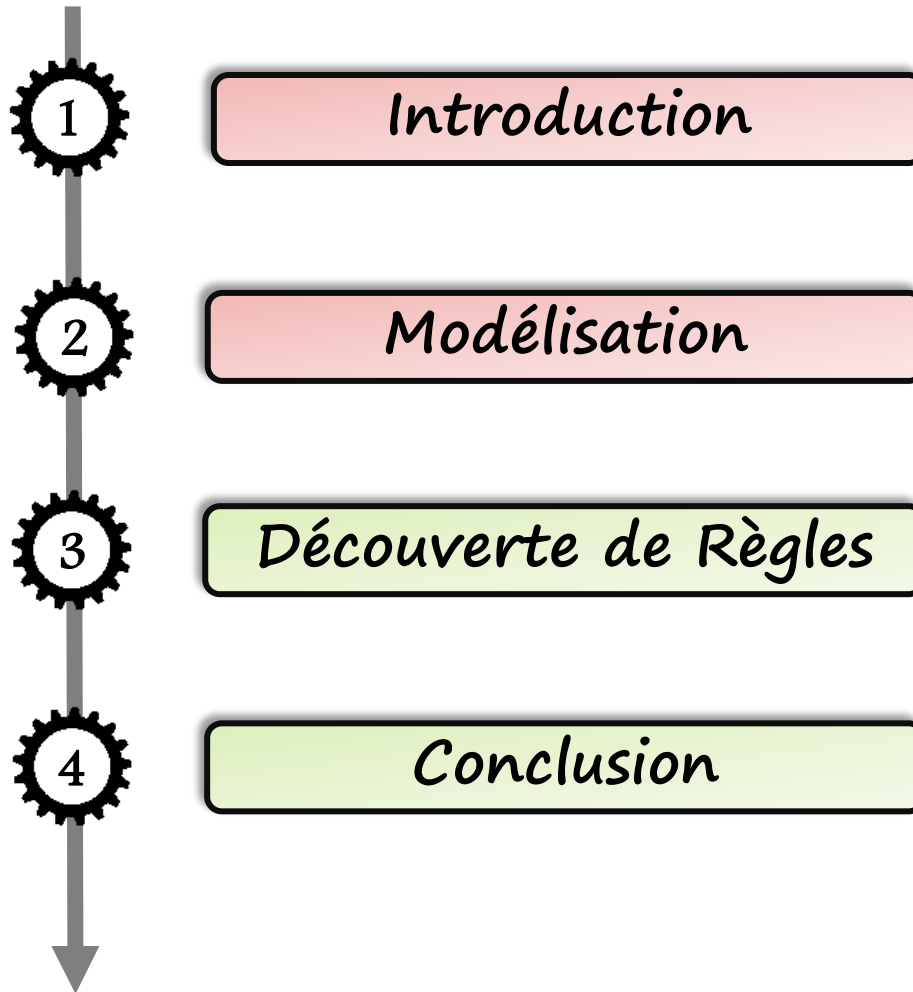


A Process and Observation Ontology

- **Instances** : 326 389
- **Classes** : 4 760
- **Propriétés** : 58
 - (Objet : 28)
 - (Data : 10)
 - (Annotation : 20)

- **Processus** : 10
- **Itinéraires** : 30
- **Observations** : 1 679

PLAN



3. DÉCOUVERTE DE RÈGLES

Comment peut-on bénéficier des données expérimentales pour découvrir de nouvelles connaissances ?

Recette 1

Operations

Etaler le beurre
15g

Choisir le pain
2 Pains de mie

Ajouter fromage
1 tranche

Ajouter jambon
2 tranches

Chauffer
7 mins à 60°C

Recette 2

Operations

Etaler le beurre
10g

Choisir le pain
1 baguette


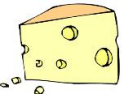


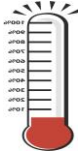






Ajouter fromage
2 tranches

Ajouter jambon
2 tranches

Chauffer
4 mins à 100°C

3. DÉCOUVERTE DE RÈGLES

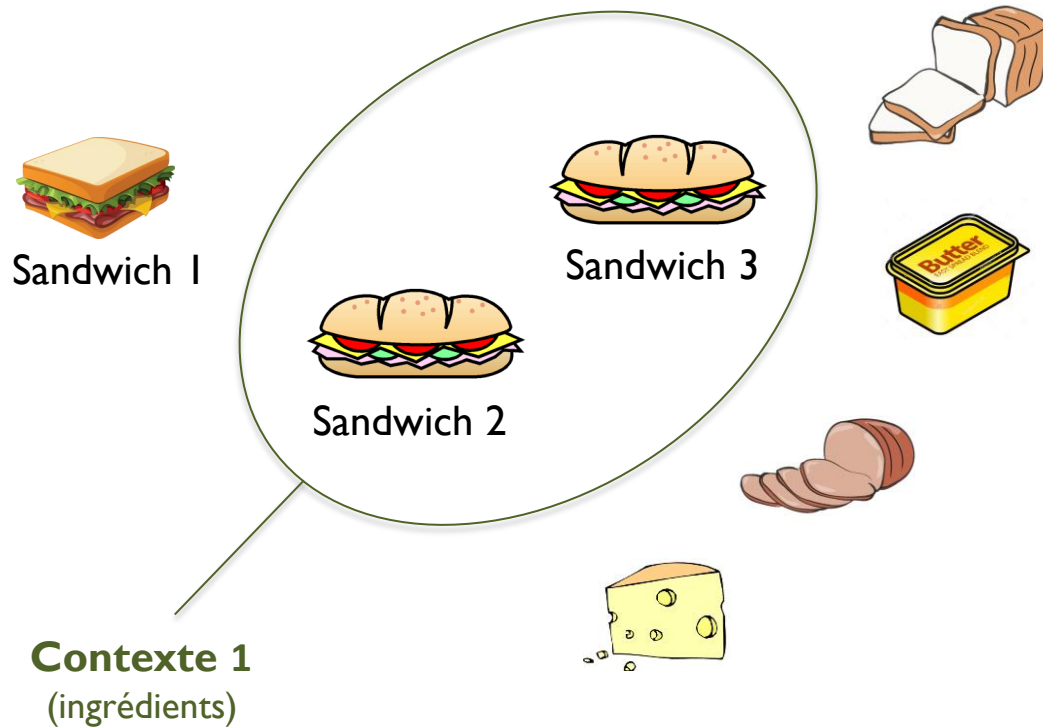
Comment peut-on bénéficier des données expérimentales pour découvrir de nouvelles connaissances ?

								
 Sandwich 1	15 g	1 tr	2 tr	7 mins	60 °C	5 €	8 Kw	5/10
 Sandwich 2	10 g	2 tr	2 tr	4 mins	100 °C	7 €	5 Kw	9/10
 Sandwich 3	10 g	2 tr	2 tr	8 mins	70°C	8 €	8 Kw	9/10

- Est-ce que chauffer plus longtemps à une plus faible température réduit la qualité du produit ? **NON**
- Et est-ce que cela a un impact environnemental ? **OUI**

3. DÉCOUVERTE DE RÈGLES

Interprétation



Règle découverte
 $\text{identiqueContexte1}(s2 \wedge s3) \rightarrow \text{Même qualité}$

3. DÉCOUVERTE DE RÈGLES

Identique Contexte 1 ?

Quels types de liens d'identité existants nous permettent d'établir cette relation entre sandwich2 et sandwich3 ?

- **owl:sameAs**

Définition : relation qui indique que deux individus avec des URI différentes font référence à la même chose

```
<rdf:Description rdf:about="onto1#William_Blythe">  
  <owl:sameAs rdf:resource="onto2#Bill_Clinton"/>  
</rdf:Description>
```



- Très stricte
- Toute propriété présente dans l'une doit figurer dans l'autre
- owl:sameAs(sandwich2, sandwich3) conduit à une inconsistance dans l'ontologie

3. DÉCOUVERTE DE RÈGLES

Identique Contexte 1 ?

Quels types de liens d'identité existants nous permettent d'établir cette relation entre sandwich2 et sandwich3 ?

- **skos:related** et **skos:exactMatch**

skos:related non transitif et sans définition formelle

skos:exactMatch indique un haut degré de confiance sur le fait que les éléments peuvent être utilisés de manière interchangeable dans un grand nombre de contextes mais sans pour autant préciser ces contextes

- Incapacité à définir le contexte d'identité
- Raisonnement très limité

3. DÉCOUVERTE DE RÈGLES

Identique Contexte 1 ?

Quelle est la solution pour représenter des liens d'identité contextuelle ?

- **Liens d'identité spécifiques au domaine**

Définir des liens d'identités spécifiques à chaque contexte dans chaque domaine

MêmeIngrédient(s2, s3) → Même qualité

- Nécessite l'intervention des experts
- Difficile de s'adapter à de nouveaux domaines d'applications

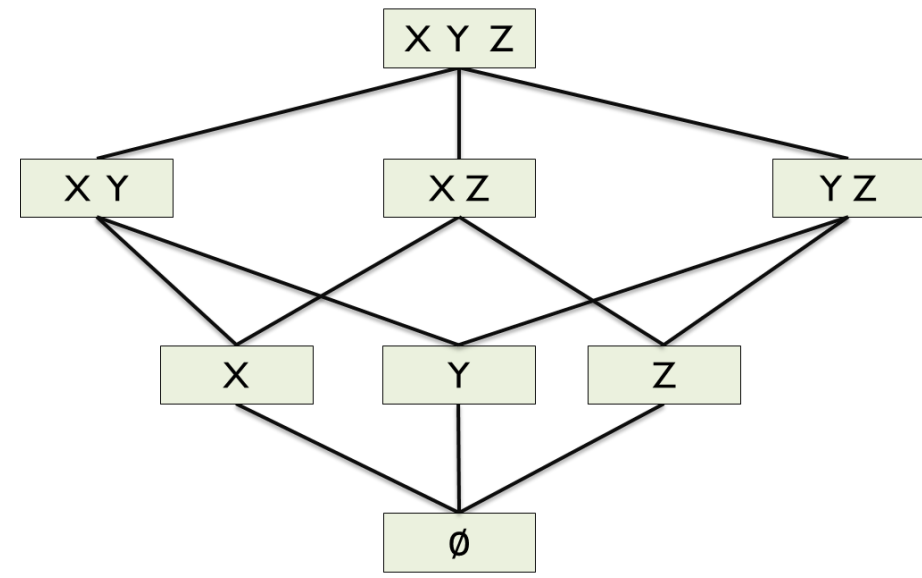
3. DÉCOUVERTE DE RÈGLES

Identique Contexte 1 ?

Quelle est la solution pour représenter des liens d'identité contextuelle ?

- **Relations d'indiscernabilité** [Beek et al., ISWC2016]

Permet de représenter des relations d'identité contextuels en s'appuyant sur un treillis de propriétés pour déterminer le contexte

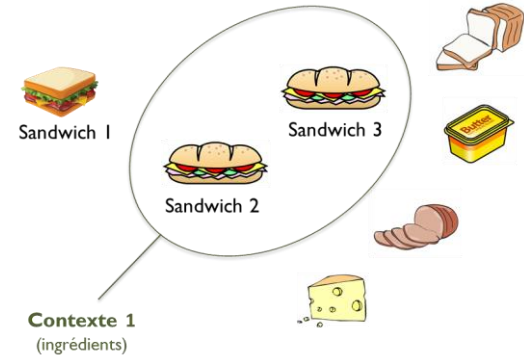
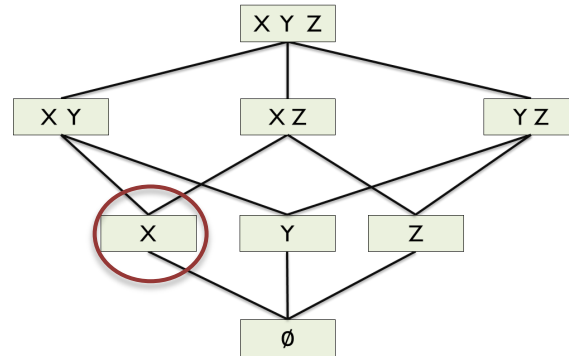


- X = aPourIngredients
- Y = duree
- Z = temperature

3. DÉCOUVERTE DE RÈGLES

Relations d'indiscernabilité [Beek et al., ISWC2016]

- X = aPourIngredients
- Y = duree
- Z = temperature



relationIndiscernabilite{x}(sandwich2, sandwich3) → Même qualité

- Hypothèse du monde clos (toute information non renseignée est considérée fausse)
- Le même contexte est utilisé pour comparer toutes les instances, mêmes si celles-ci peuvent appartenir à différentes classes (sandwich, fromage, etc.)

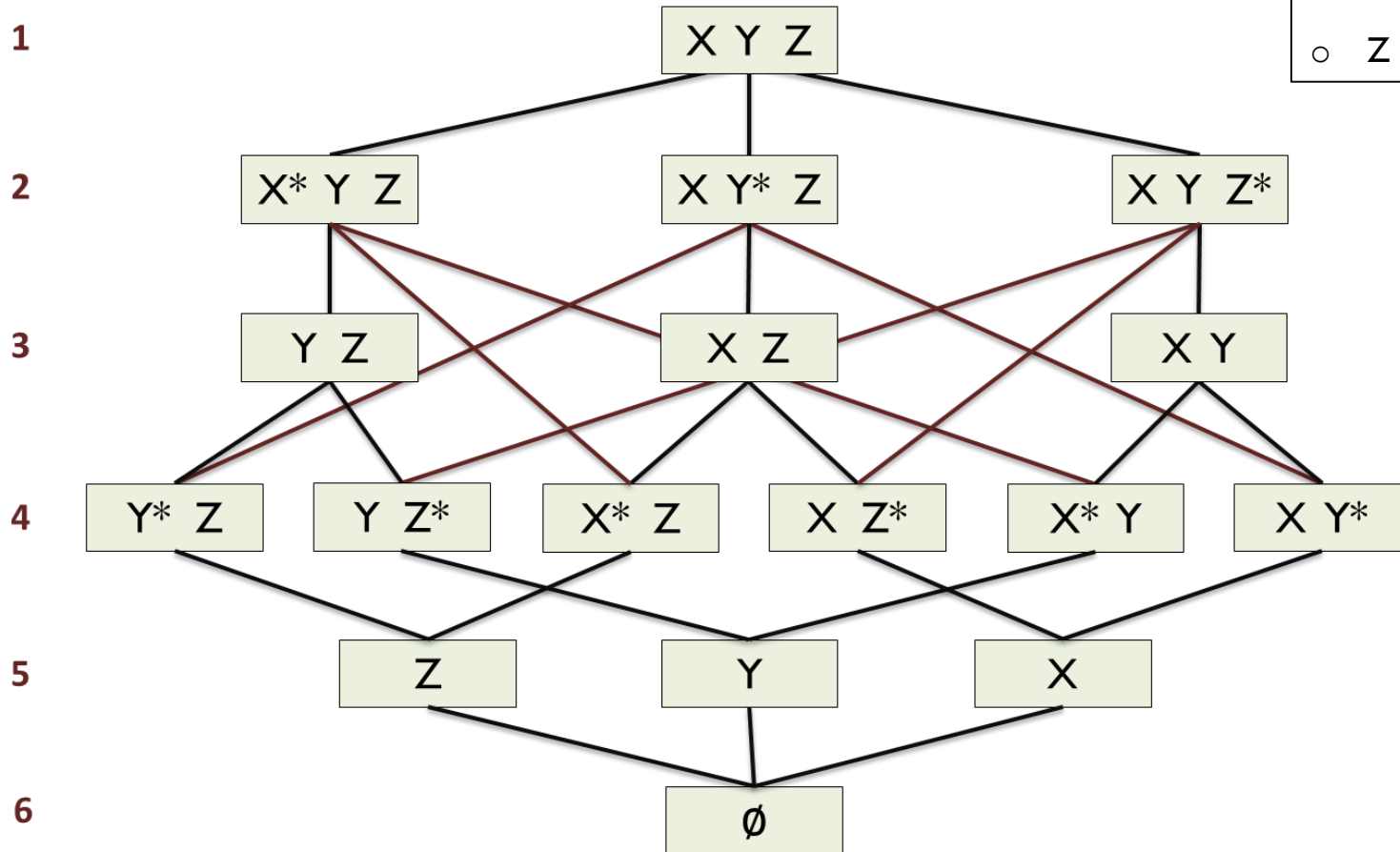
Si on utilise une tranche de fromage congelée, fraîche ou chauffée on aura la même qualité de sandwiches ?!

3. DÉCOUVERTE DE RÈGLES

Proposition d'un nouveau type de lien d'identité contextuelle

< po2 : identicalTo >

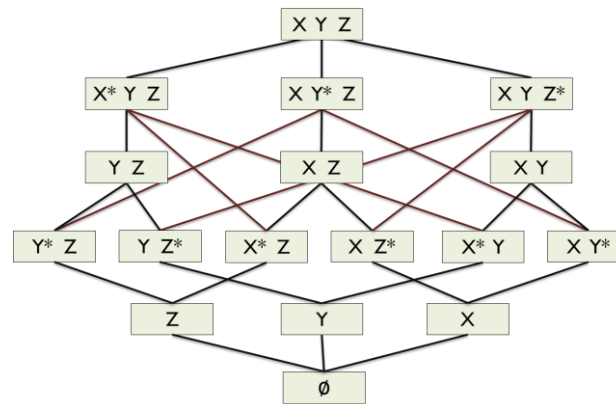
- X = ingrédients
- Y = duree
- Z = temperature







3. DÉCOUVERTE DE RÈGLES

Proposition d'un nouveau type de lien d'identité contextuelle

< po2 : identicalTo >

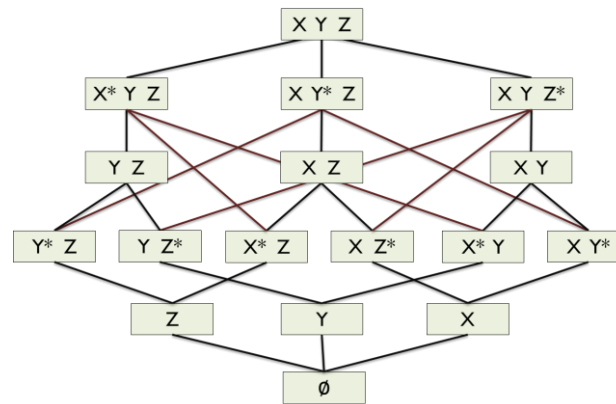


	X	Y	Z
 1	5 g	10 mins	60 °C
 2	5 g	10 mins	60 °C
 3	5 g	10 mins	
 4	5 g	10 mins	100 °C

3. DÉCOUVERTE DE RÈGLES

Proposition d'un nouveau type de lien d'identité contextuelle

< po2 : identicalTo >




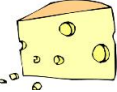


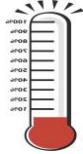







- On peut changer le contexte pour chaque classe

sandwich2 et sandwich3 sont identiques même si on considère pas leurs températures (mais en considérant celle des ingrédients)

- Hypothèse du monde ouvert (toute information non renseignée n'est pas considérée fausse)
- On peut bénéficier des connaissances expertes pour supprimer des branches sans importances et diminuer le temps du calcul (notion des clés)

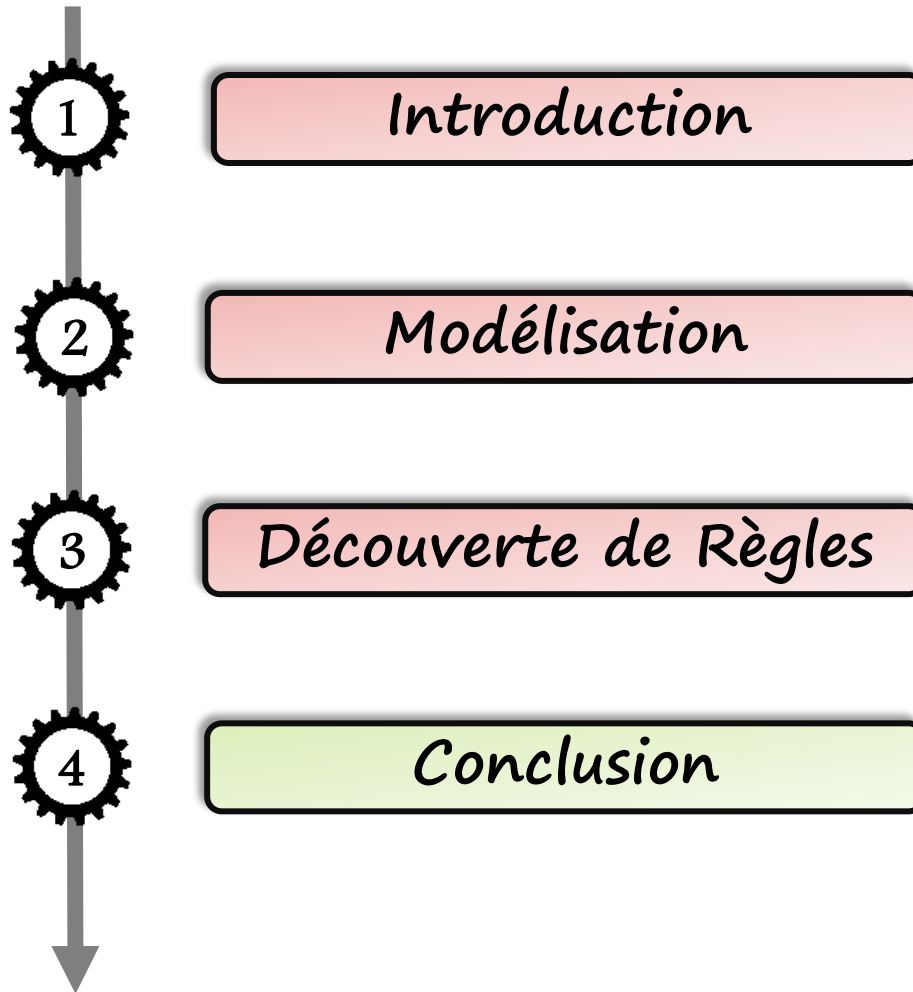
3. DÉCOUVERTE DE RÈGLES

Pourquoi faire des liens d'identité contextuelle ?

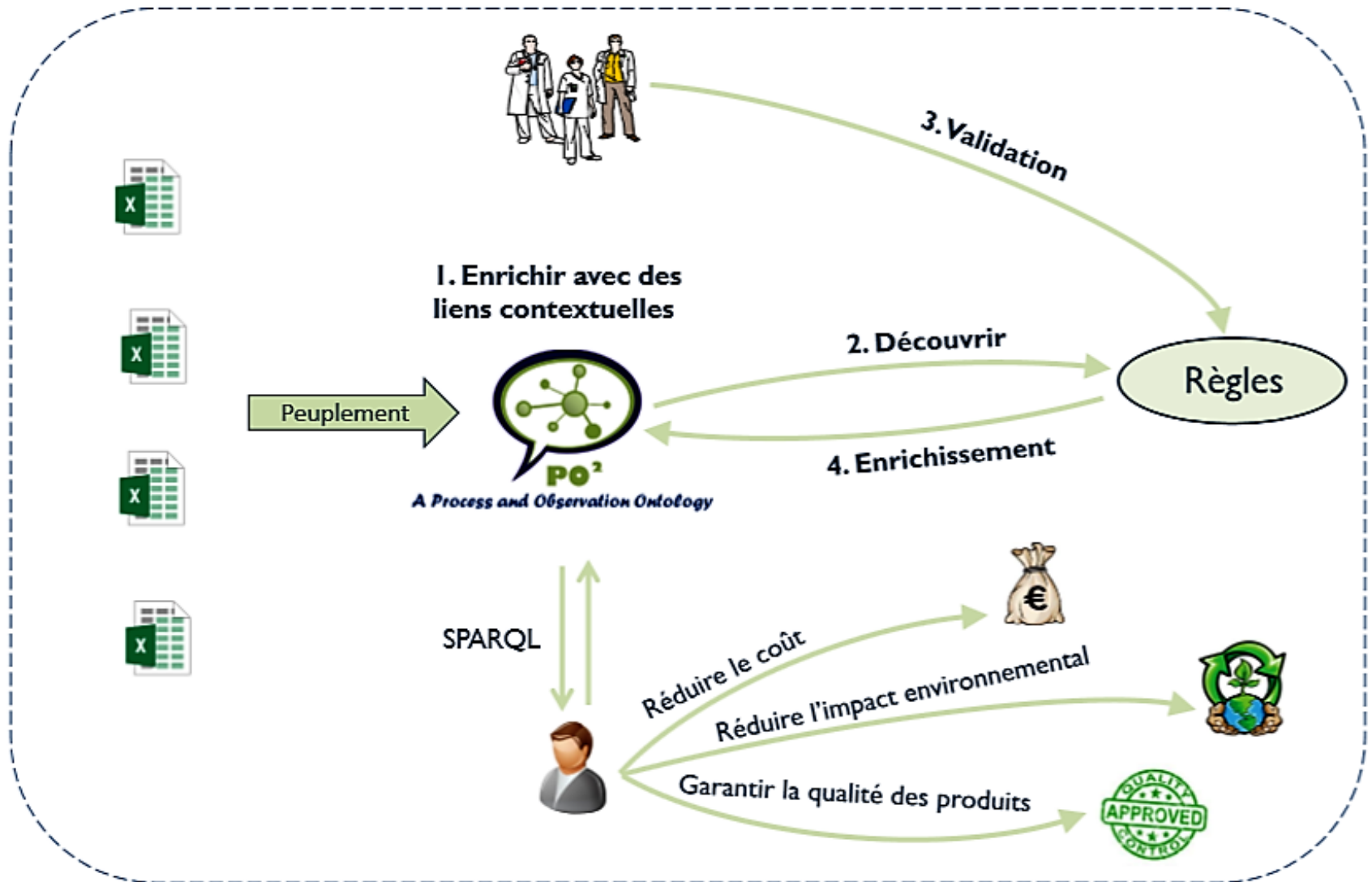
								
 Sandwich 1	15 g	1 tr	2 tr	7 mins	60 °C	5 €	8 Kw	5/10
 Sandwich 2	10 g	2 tr	2 tr	4 mins	100 °C	7 €	5 Kw	9/10
 Sandwich 3	10 g	2 tr	2 tr	8 mins	70°C	8 €	8 Kw	9/10
 Sandwich 4	10 g	2 tr	2 tr	5 mins	80°C			9/10

po2:identicalTo{x}(sandwich2, sandwich3) → Même qualité

PLAN



4. CONCLUSION



Merci pour votre attention...

Questions ???

