ExerSys: leveraging expert's knowledge to improve recommender systems for nutrition

Cristina Manfredotti

Maître de conférences, HDR

UFR Informatique, Département MMIP, AgroParisTech

Equipe Ekinocs, MIA Paris-Saclay, INRAe

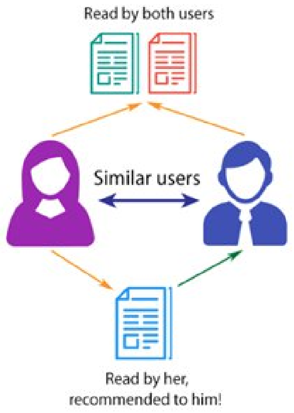
The exersys* project

Develop a RecSys for meals recommendation combining

- User preferences *thanks to « classic » recsys approaches*
- Nutritional constraints *thanks to knowledge graphs*

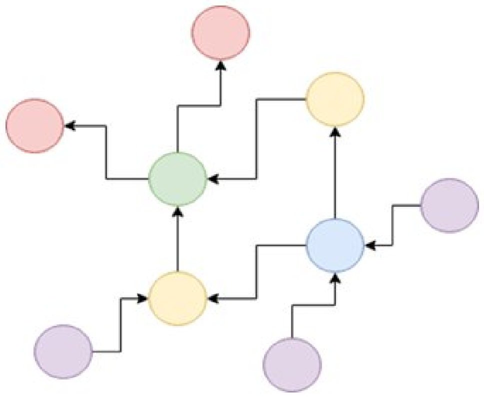
1 **Axe 1:**

Recommandation basée sur les **préférences utilisateurs** par **apprentissage automatique**



2 **Axe 2:**

Intégration de **contraintes nutritionnelles** grâce aux **graphes de connaissance**

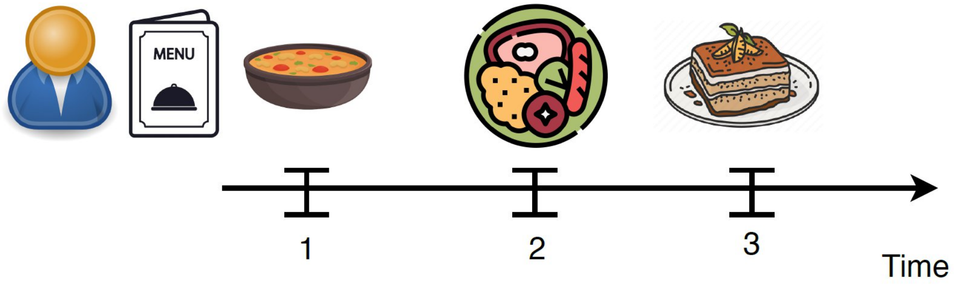


* Exersys : project financed by

Plan

- ExerSys genesis

- The (ongoing) ExerSys project



- The ExerSys project followUP

ExerSys (and not) collaborations

- Stéphane Dervaux
- Fatiha Saïs
- Vincent Guigue
- Paolo Viappiani
- Nicolas Darcel
- Liliana Ibanescu
- Juliette Dibie
- Pierre-Henri Wuillemin
- Patrice Buche

- Alexandre Combeau
- Noémie Jacquet
- Ayoub Hammal
- Nageeta Kumari
- Lisa Bouhanna
- Mélanie Munch

DATAIA, Dept MIA, AgroParisTech, ED ABIES

MIA
PARIS-SACLAY
EKINOCS

ExerSys Genesis

- Before APT : probabilistic models that take into account (not formalised) expert's knowledge

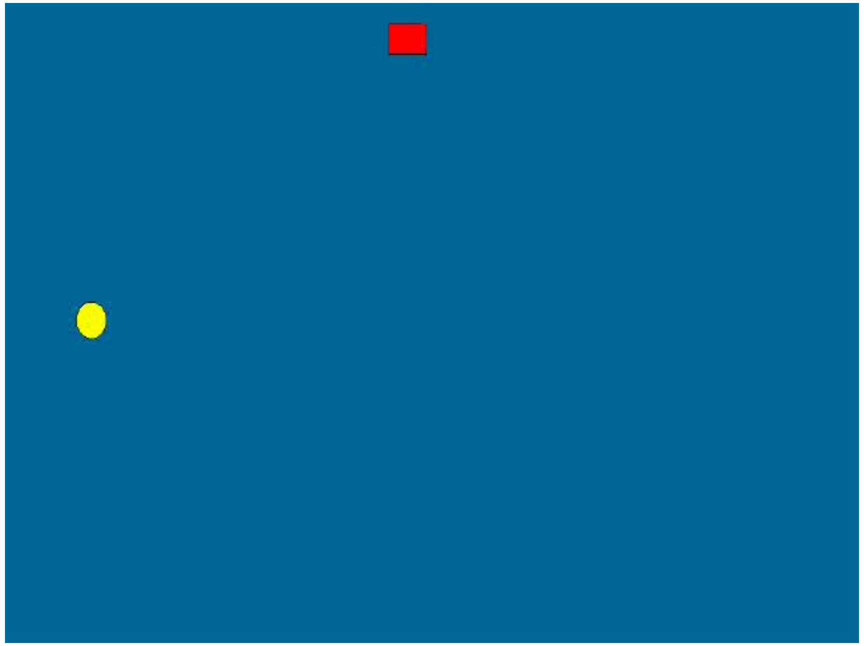
ExerSys Genesis

- Before APT : probabilistic models that take into account (not formalised) expert’s knowledge

ExerSys Genesis

- Before APT : probabilistic models (**PRM**) that take into account (not formalised) expert's knowledge

Thanks to Davide Piazza



PRMs

A BN is the representation of a joint probability over a set of random variables that uses a DAG to encode probabilistic relations between variables

Combine advantages of relational data bases & **Bayesian networks**:

- natural domain modeling: objects, properties, relations;
- generalization over a variety of situations;
- compact, natural probability models.

➢ **Relational Schema** and **Relational slots** (and slot chain)

PRM system => (big) BN

ExerSys Genesis

- Before APT : probabilistic models that take into account (not formalised) expert's knowledge

- Seminaire IN-OVIVE (Montpellier, dec 2014) : similarity between BN and Ontologies

Ontologies

A KB = ontology + knowledge graph
Ontology = classes, properties, axioms
Knowledge graph = data organised following the ontology

- Represent the knowledge on a domain with
  - classes,
  - relations (or properties) between these classes and
  - instances of these classes.
- Used as a common and standardized vocabulary for representing a domain.
- Data can be collected in a **knowledge base** that organises them according to the structure defined by an ontology.

ExerSys Genesis

- Before APT : probabilistic models that take into account (not formalised) expert’s knowledge

- Seminaire IN-OVIVE (Montpellier, dec 2014) : similarity between BN and Ontologies

ExerSys Genesis

- Before APT : probabilistic models that take into account (not formalised) expert’s knowledge

- Seminaire IN-OVIVE (Montpellier, dec 2014) : similarity between BN and Ontologies

Expert’s knowledge can be structured (formalised) in an ontology and taken into account by the probabilistic model

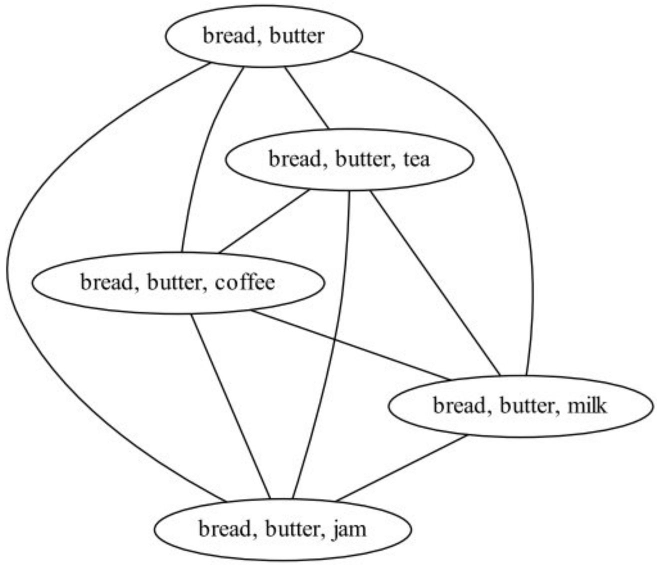
ExerSys Genesis

- Couplage ontologies and Probabilistic Relational Models :
  - PRM learning easier if we have a KG to start with
  - Model explanation
  - Causal verification

ExerSys Genesis

- Couplage ontologies and Probabilistic Relational Models :
  - PRM learning easier if we have a KG to start with
  - Model explanation
  - Causal verification

- Nutrition domain :
  - Substitutability
  - Recommender systems

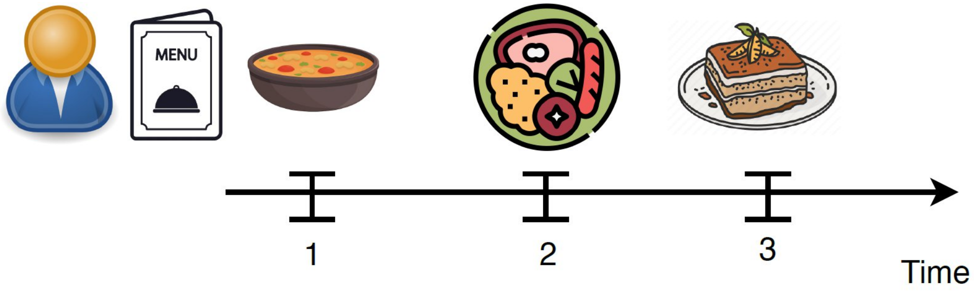


ExerSys Genesis

- Couplage ontologies and Probabilistic Relational Models :
  - PRM learning easier if we have a KG to start with
  - Model explanation
  - Causal verification

- Nutrition domain :
  - Substitutability
  - Recommender systems

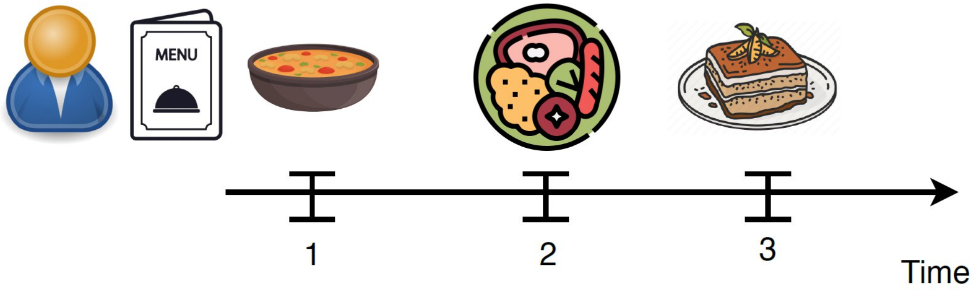
- The ExerSys projets puts those together



RecSys in nutrition : importance & originality

- Unhealthy diet => diseases
- Ineffectiveness of healthy diet campaigns
- ➢ Need of long-term, health-aware recs

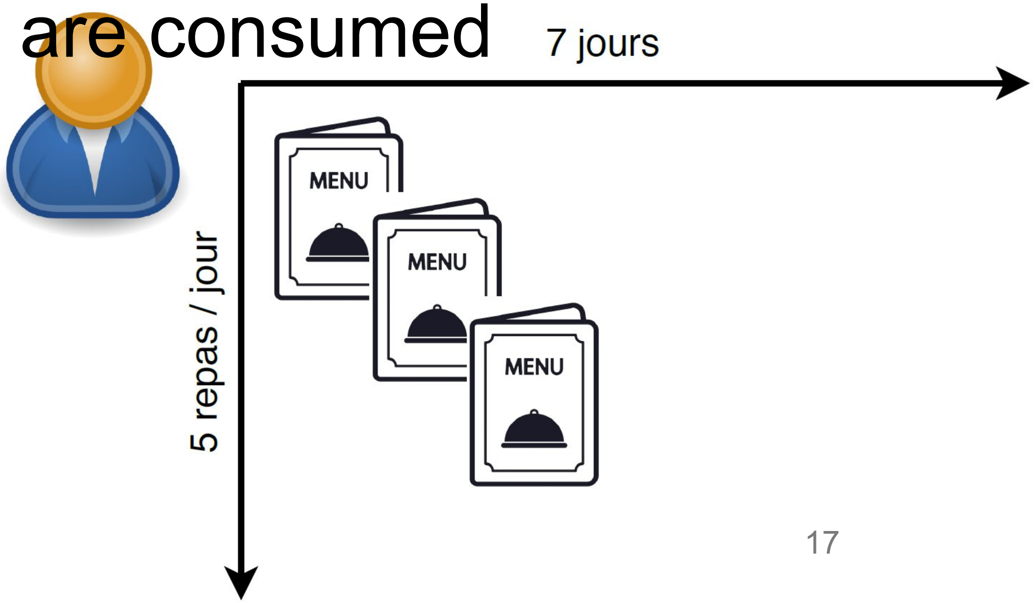
- Meal rec ≠ dish rec ≠ item rec
  - Dish and Meal structure
  - Sequentiality : coherent sequence of dishes
  - Food items frequence
  - Rules : control over nutritional quantity/quality
  - Context : sex, age, home/office, friends/family, ...



RecSys in the nutrition domain : INCA2 dataset

- Quantity : 2624 adults individuals, **7 days**, **5 meals**
- 280 food items organized in 44 **groups** and 110 sub-groups of food items
- **Context factors** : place, company, eater carachteristics
- Structured, good quality, sequences of real consumptions over 7 days
- Consumptions are entered in order as they are consumed

(+ transversal food items at the end)



- **Meal mean lenght** : **7/8 items**

RecSys in the nutrition domain : INCA3 dataset

- Quantity : 3157 adults individuals, **3 days non-consecutive**, **5 meals**
- food items organized in FoodEx **groups** and sub-groups
- **Context factors** : place, company, eater carachteristics, **food preparation**, **storage habits**
- Structured, good quality, sequences of real consumptions over **24 hours**
- Consumptions are entered in order as they are consumed

(+ transversal food items at the end)

INCA2 dataset : our choices

We did not consider consumed quantities of food items

- Consumption => implicit preference,
- Binary interaction matrix : food items/users

We separately modeled breakfast and lunch&dinner

- breakfast : repetitive, 11% new food items on the 7th day
- lunch&dinner : longer, 51% new food items on the 7th day (**almost cold start**)

- Sliding week
  - ➢ no biais for the week-end

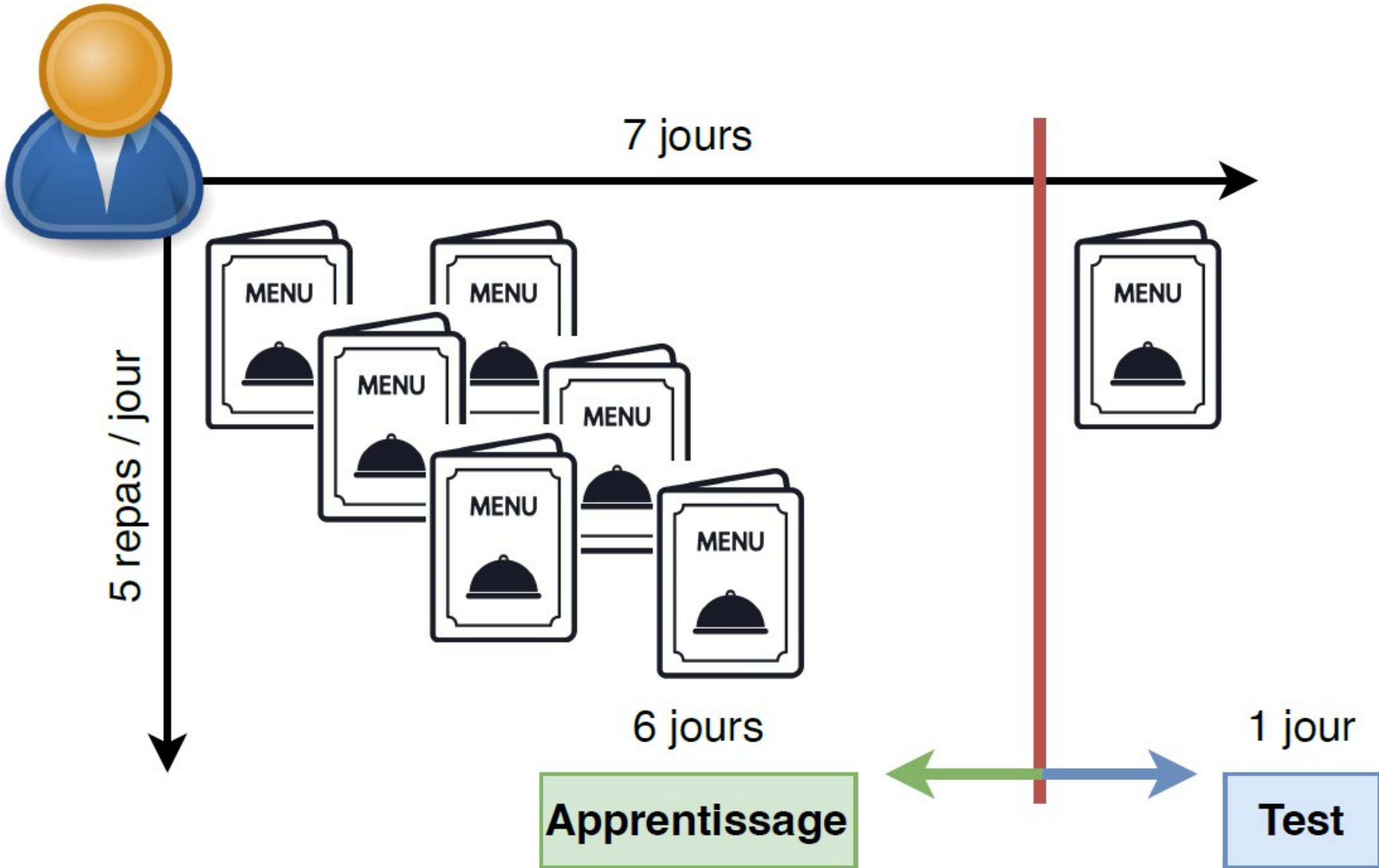
INCA2 : our choices

We did not consider consumed quantities of food items

- Consumption -> implicit preference,
- Binary interaction matrix: «food items»/«users»

We separately modeled bre

- breakfast : repetitive, 11%
- lunch&dinner : longer, 51%
  day (quasi cold start)



- Sliding week
  - ➢ **no biais for the week-end**

The exersys project

Develop a RecSys for meals recommendation combining

1 **Axe 1:**

Recommendation based on **user's preferences** learnt by **machine learning methods**

PDJ

- ❑ Collaborative filtering approach : ***'FiltreCollab'***

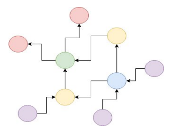
DEJ/ DIN

- ❑ Sequence generation by RNNs : ***'GenSeqRNN'***

Génération de recommandati

2 **Axe 2:**

Nutritional constraints integration by **knowledge graphs**



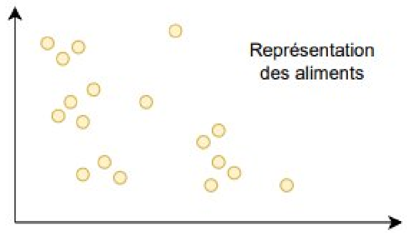
The exersys project : filtreCollab

Recommendation generation

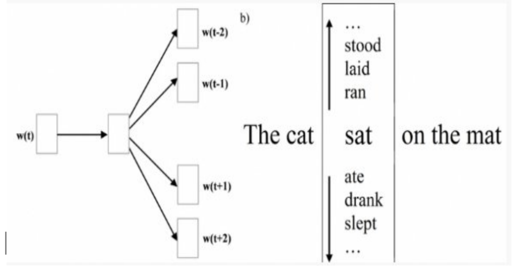
1. Learn a representation space for food items
2. Categories of closed food items
3. Modeling and food items sampling by category
4. Meal recommendation

PDJ

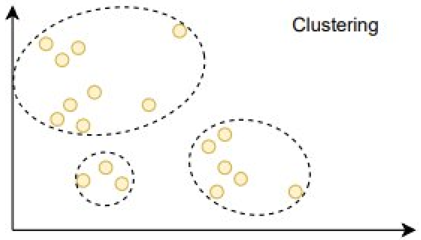
Word2Vec learnt on breakfast



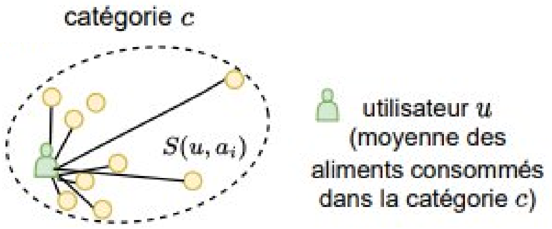
1 food item = 1 10 dimensions vector



K-means, **k categories**



Food item recommendation :
Sampling following a multinomia distribution



$S(u, a_i)$ similarité cosine entre u et a_i

$$P(a_1 \mid c) = \frac{e^{\alpha \cdot S(u,a_1)}}{\sum_i e^{\alpha \cdot S(u,a_i)}}$$

c catégorie consommée

Meal recommendation :

- **Sampling**
(1 food item per category already consumed and in β% of cases otherwise)
- **Association/exclusion rules**

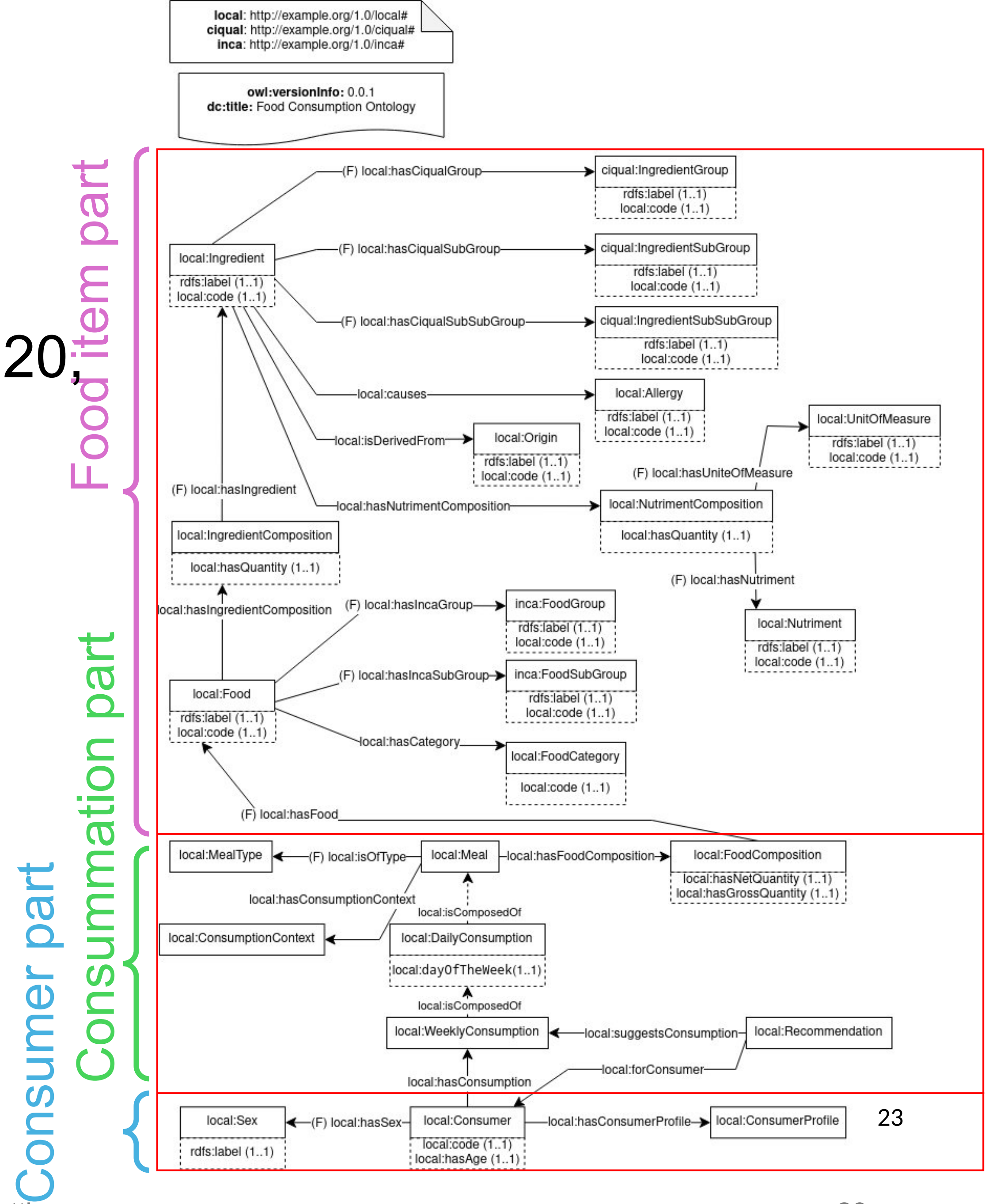
MIA PARIS-SACLAY
EKINOCS

The exersys project : the knowledge graph

Multiple sources of data : INCA 2, CIQUAL 2020, Open Food Facts

Modeling

- consumer’s profiles
- nutrition constraints
- Association rules
- Exclusion rules
- Cardinality rules



N. Jacquet, C. Manfredotti, V. Guigue, F. Saïs, P. Viappiani : An EXplainable RecommandER SYStem for the Nutrition Domain, combining Knowledge Graphs and Machine Learning. CoCoA-BeANS workshop: Cognitive and COmputational Approaches of Behaviour and Nutrition Studies. Paris, May 11-12, 2023.

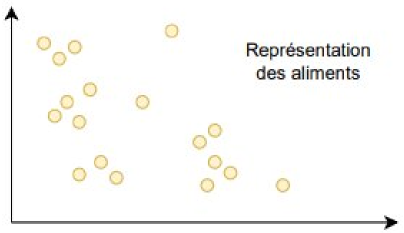
The exersys project : filtreCollab

Recommendation generation

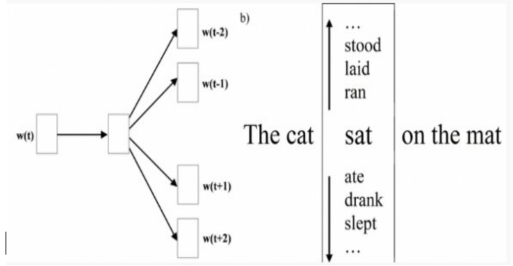
PDJ

1. Learn a representation space for food items
2. Categories of closed food items
3. Modeling and food items sampling by category
4. Meal recommendation

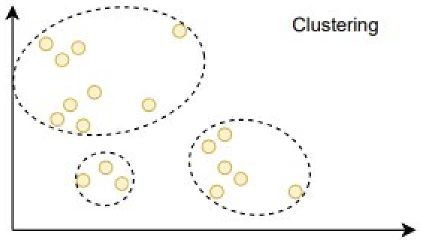
Word2Vec learnt on breakfast



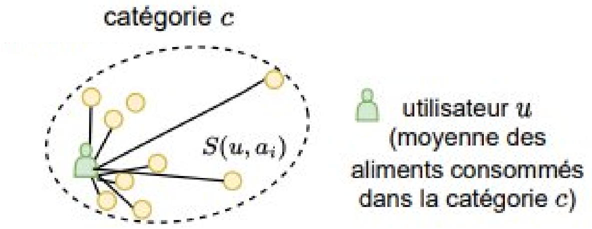
1 food item = 1 10 dimensions vector



K-means, **k categories**



Food item recommendation :
Sampling following a multinomia distribution



$S(u, a_i)$ similarité cosine entre u et a_i

$$P(a_1 \mid c) = \frac{e^{\alpha \cdot S(u,a_1)}}{\sum_i e^{\alpha \cdot S(u,a_i)}}$$

c catégorie consommée

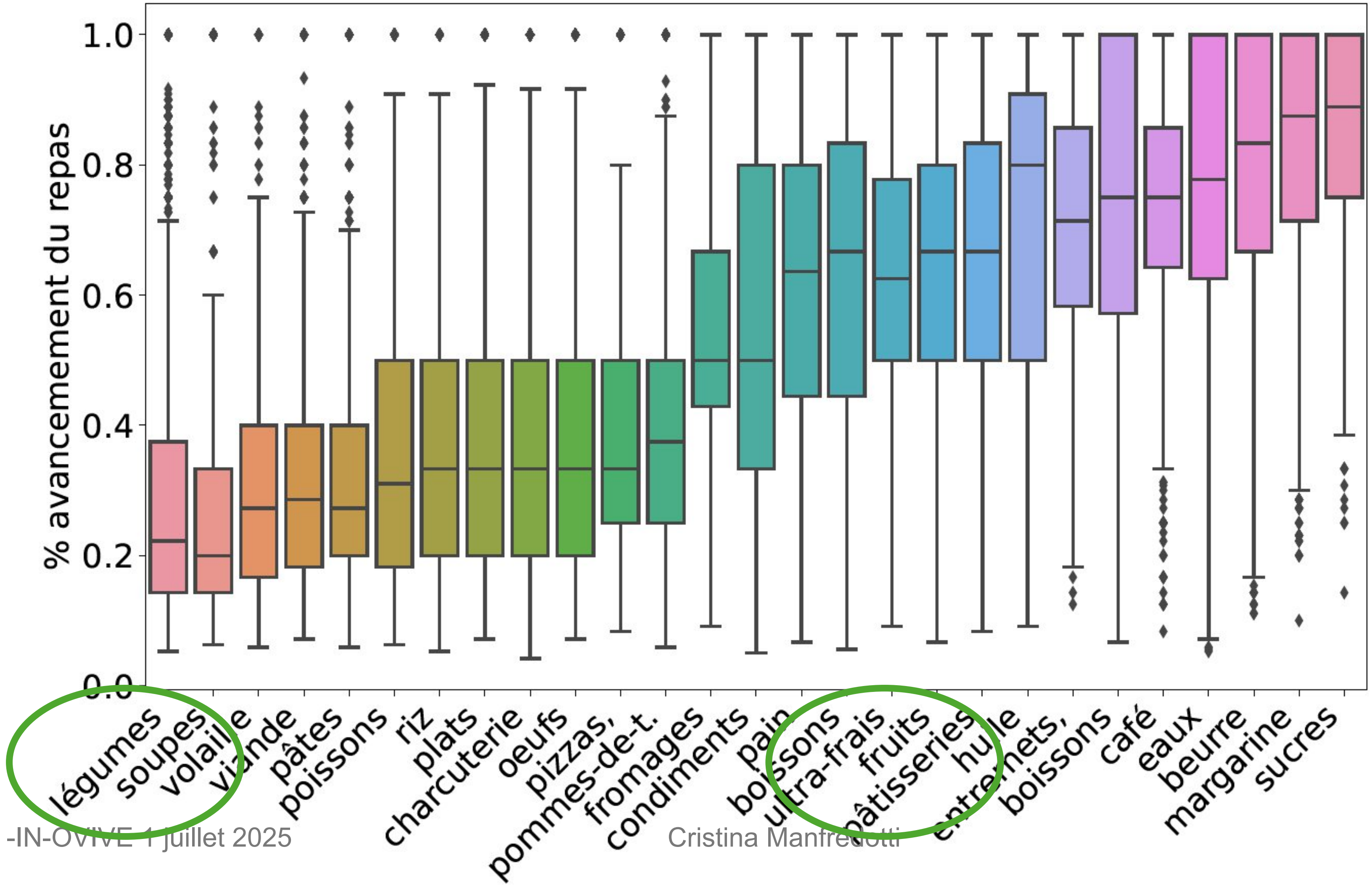
Meal recommendation :

- **Sampling**
(1 food item per category already consumed and in β% of cases otherwise)
- **Association/exclusion rules**

MIA PARIS-SACLAY
EKINOCS

INCA2 : sequentiality

Food items temporal distribution during lunch and dinner.



The exersys project : GenSeqRNN

1 **Axe 1:**

Recommendation based on **user's preferences** learnt by **machine learning methods**

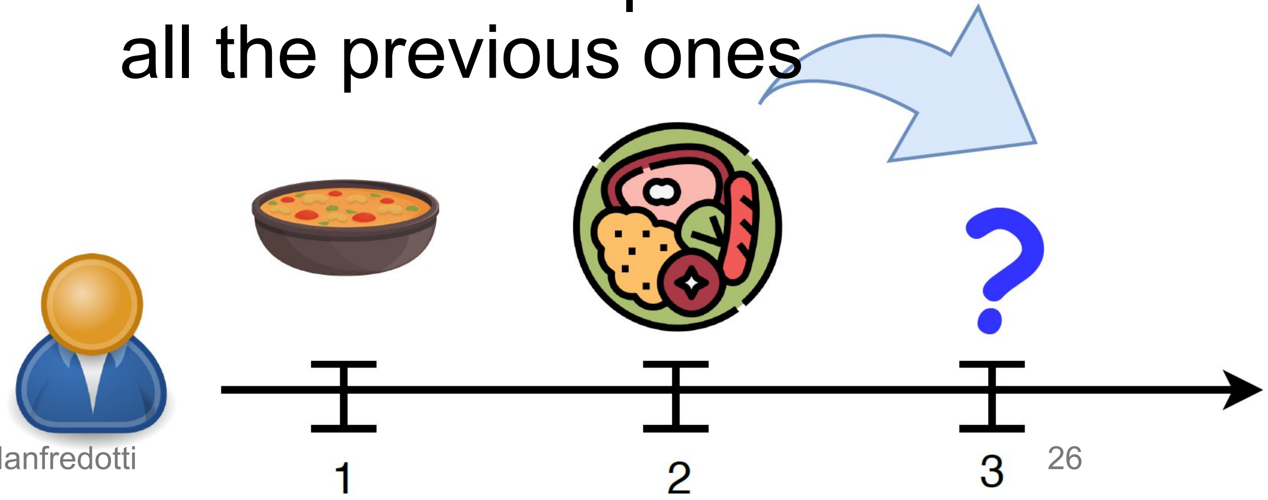
- ❑ Collaborative filtering approach : '*FiltreCollab*'

- ❑ Sequence generation by RNNs : '***GenSeqRNN***'

Recurrent Neural Networks:

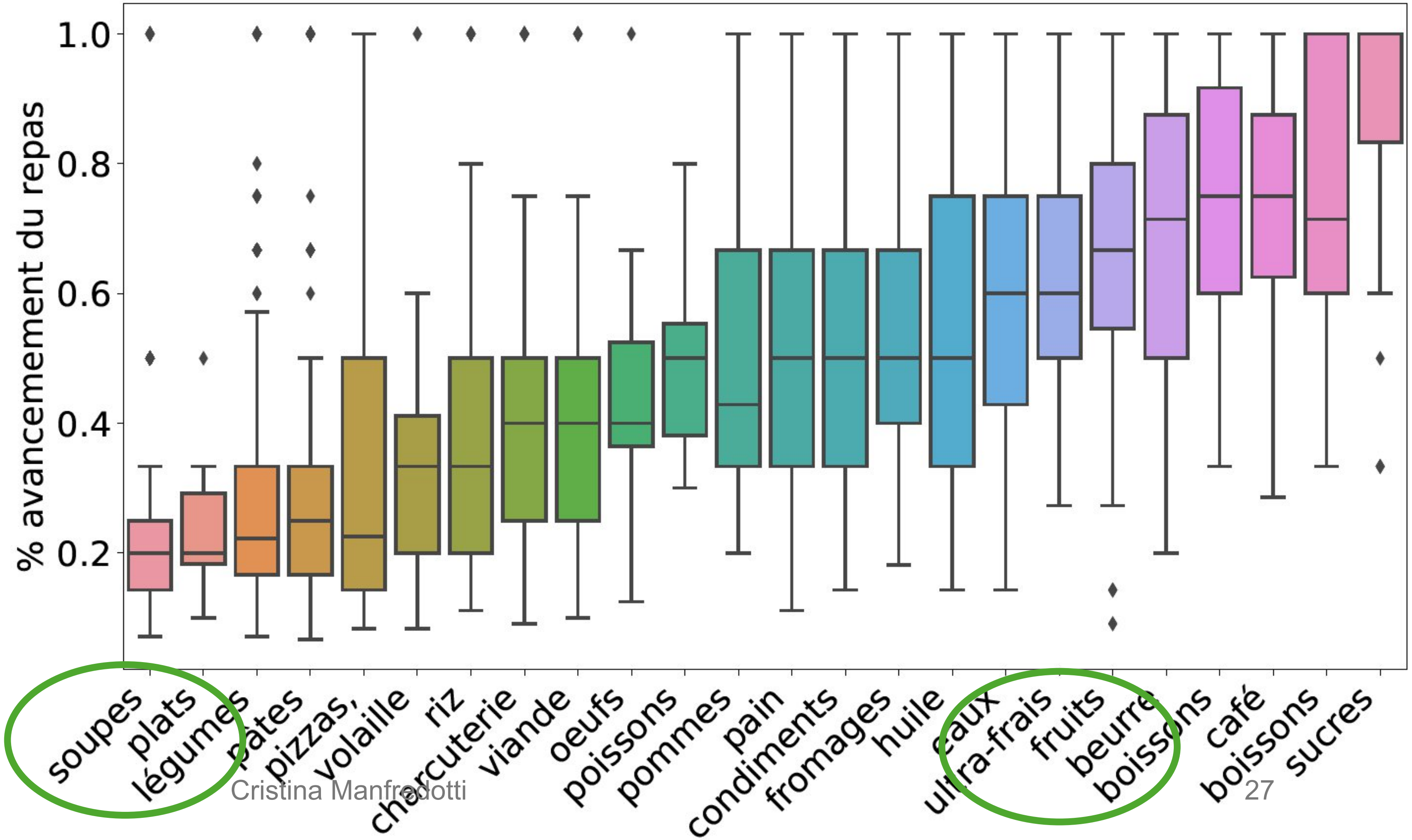
$$h_t = g(W_1 a_t + W_2 h_{t-1})$$
$$\hat{p} = softmax(W_3 h_t)$$

- Next food item prediction wrt all the previous ones



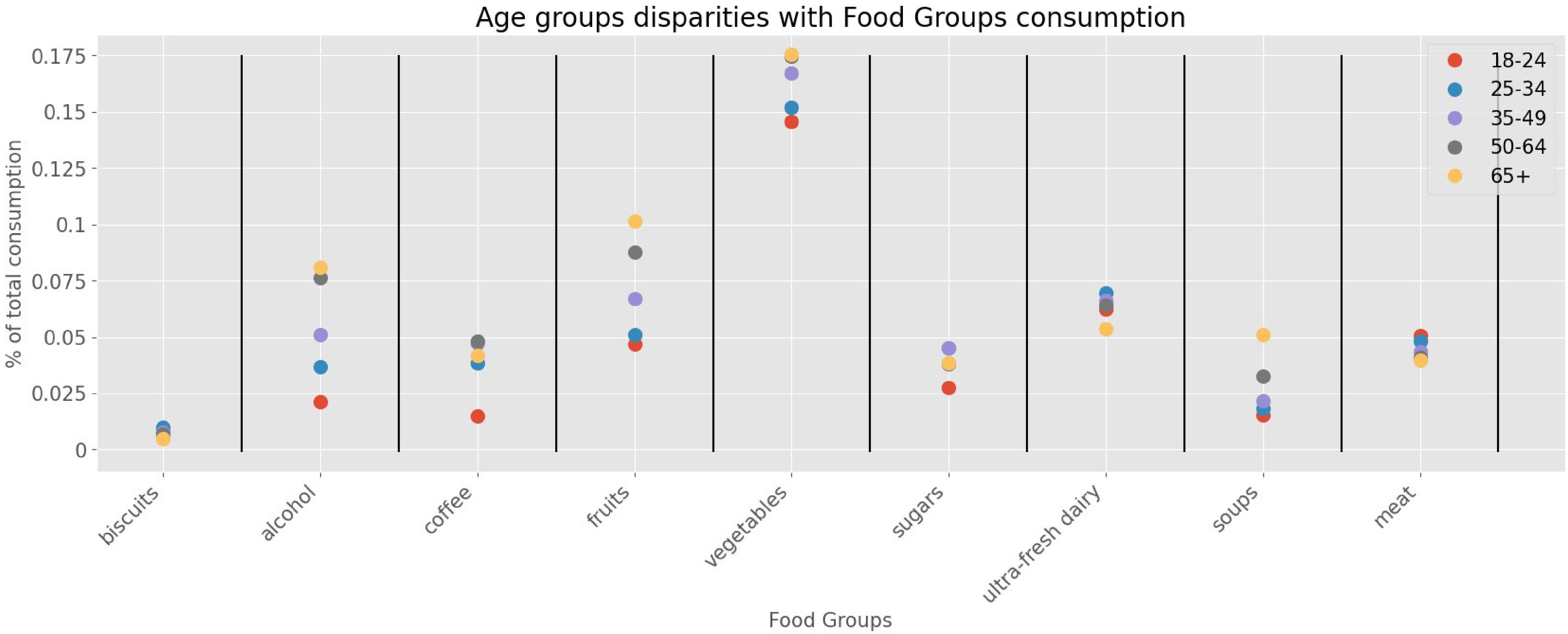
Test : sequentiality

Food items temporal distribution in the **RNN prediction in lunch and dinner**.

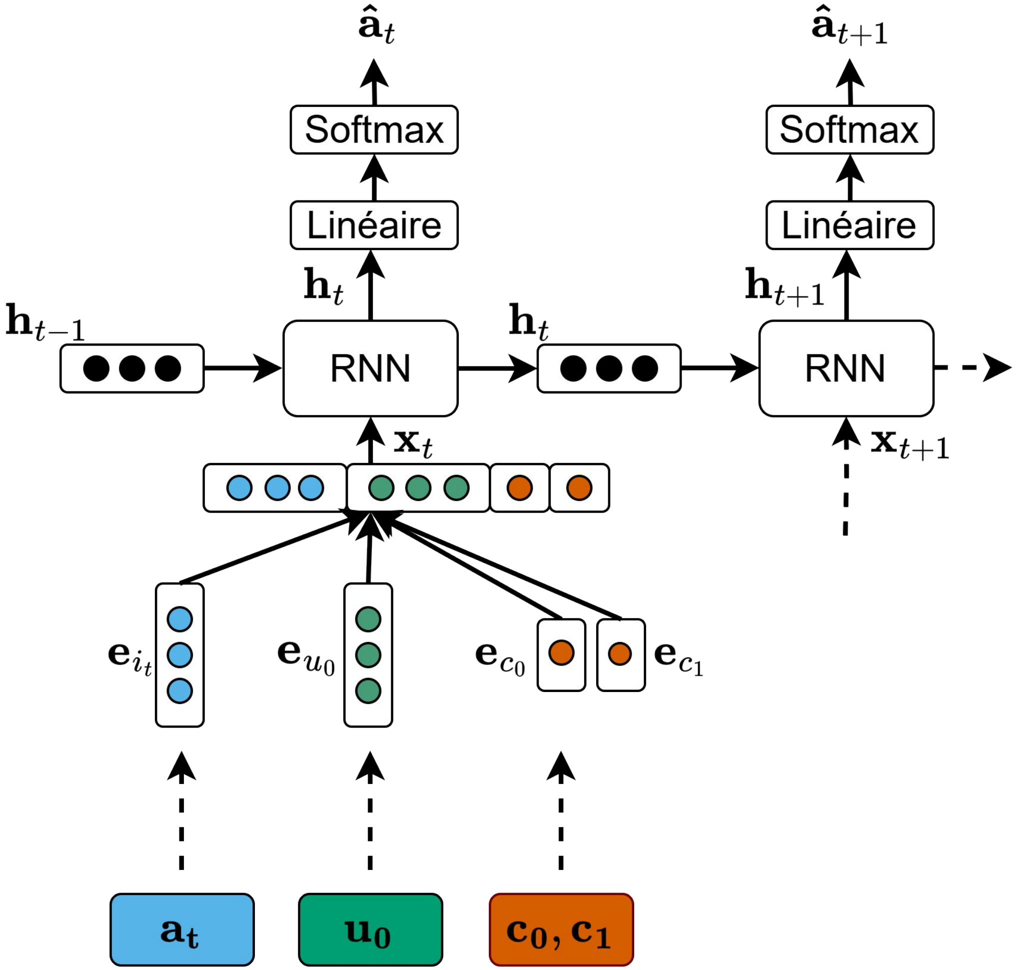


N. Jacquet, V. Guigue, C. Manfredotti, F. Saïs, S. Dervaux, P. Viappiani: Modélisation du caractère séquentiel des repas pour améliorer la performance d'un système de recommandation alimentaire. 24eme Conférence francophone sur l'Extraction et la Gestion des Connaissances (EGC 2024), Jan 2024, Dijon, France.

INCA 2 : context as a separator of consumptions



Sequentiality : context and user integration



$$X = \{m_{u,d,r}\}$$

$$m_{u,d,r} = [a_1, \ldots, a_t, \ldots a_T]$$

$$a \in \mathcal{A} \mapsto \mathbf{a} \in \mathbb{R}^z$$

$$u \in \mathcal{U} \mapsto \mathbf{u} \in \mathbb{R}^z$$

$$c \in \mathcal{C} \mapsto \mathbf{c} \in \mathbb{R}^z$$

$$\mathbf{h}_t = g(W_a\, \mathbf{a}_t + W_u\, \mathbf{u}_i + \sum_j W_{c_j}\, \mathbf{c}_j + U\, \mathbf{h}_{t-1})$$

The exersys project : user & context integration

Performance : prediction of the next food item

Modèle	MRR	Acc
(Item)	0.329	0.243
(Item,User)	0.405	0.328
(Item,User,Age)	0.413	0.336
(Item,User,Typm)	**0.429**	**0.355**
(Item,User,Typm,Age)	**0.429**	0.354

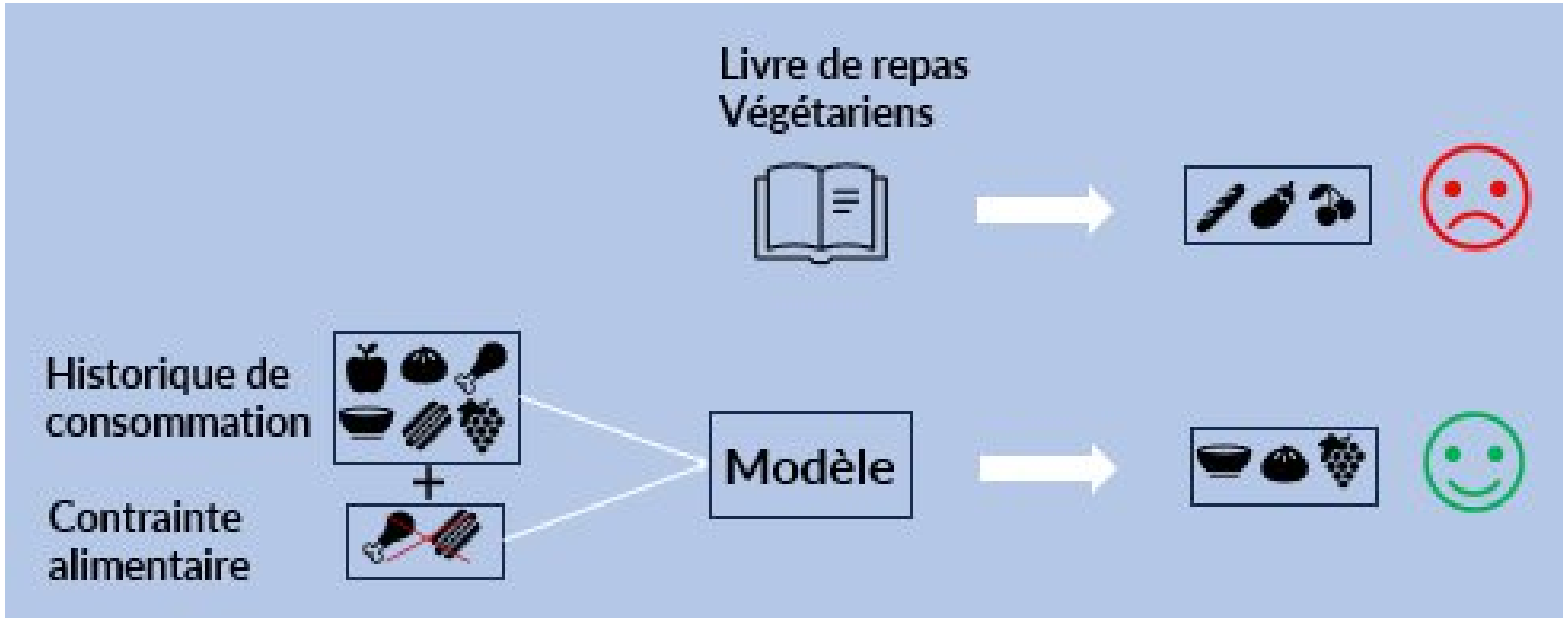
The exersys project : user & context integration

Exactitude dysparity between food items wrt the type of meal:
Exactitude augment for the first 2 or 3 items

Modèle	Petit-déjeuner	Déjeuner	Dîner
	Exactitude du premier élément		
(Item,User)	64.8	2.5	3.6
(Item,User,Typm)	75.9	8.0	13.3
	Exactitude du deuxième élément		
(Item,User)	63.9	7.5	8.0
(Item,User,Typm)	65.6	9.2	11.3
	Exactitude du troisième élément		
(Item,User)	63.5	12.5	16.0
(Item,User,Typm)	63.3	13.6	16.8
	Exactitude globale des éléments		
(Item,User)	65.6	20.9	20.7
(Item,User,Typm)	68.4	22.5	24.4

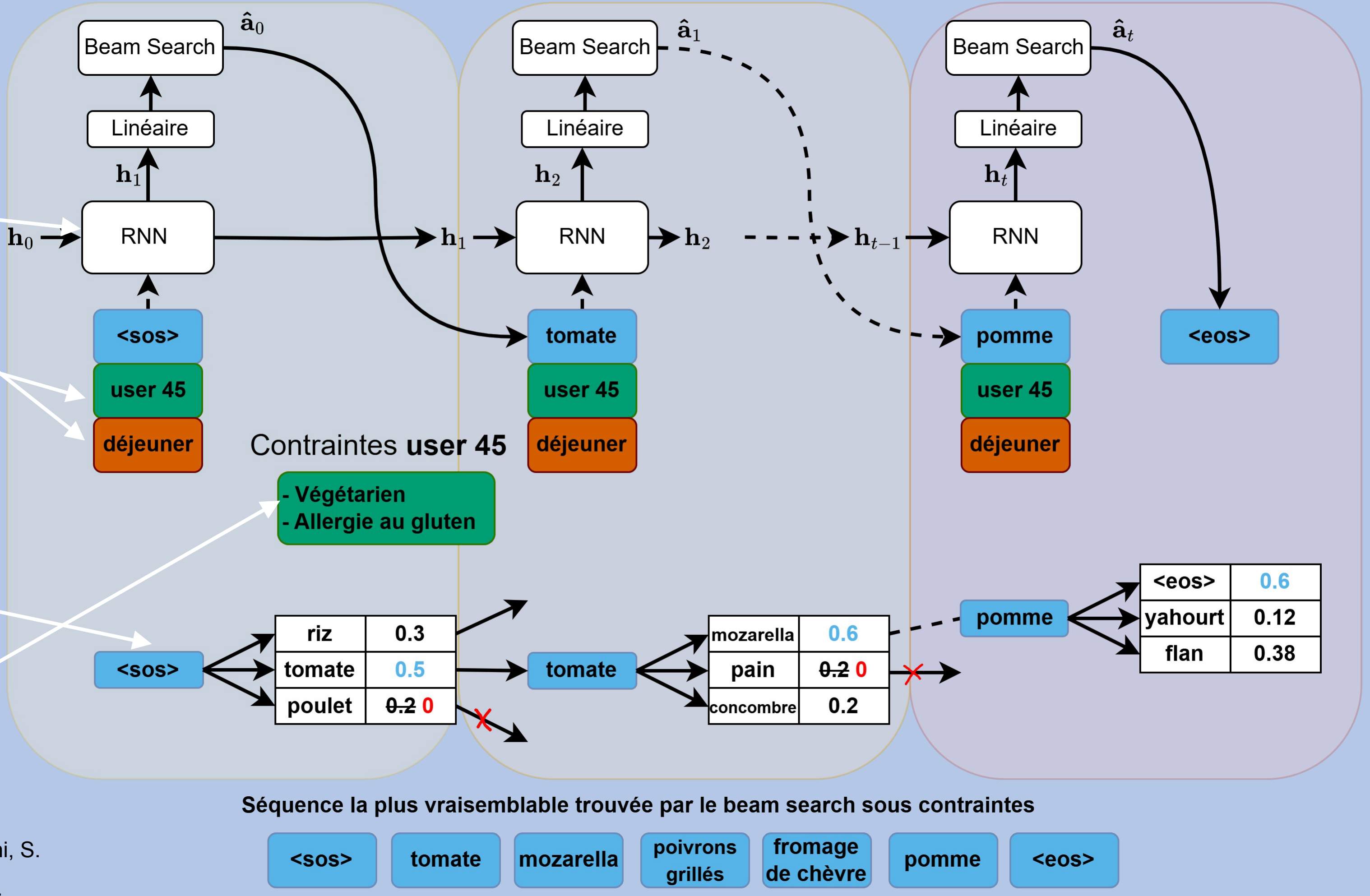
The exersys project : individual recommendation

A veggie meal generated following user's preferences is better than a generic veggie meal



The exersys project : individual recommendation

The **model** generates a meal based on
- The consumption **context**
- **User's preferences**



Beam Search creats a **tree of all possible meals** and save the most log-likelihood

It can be **constrained**

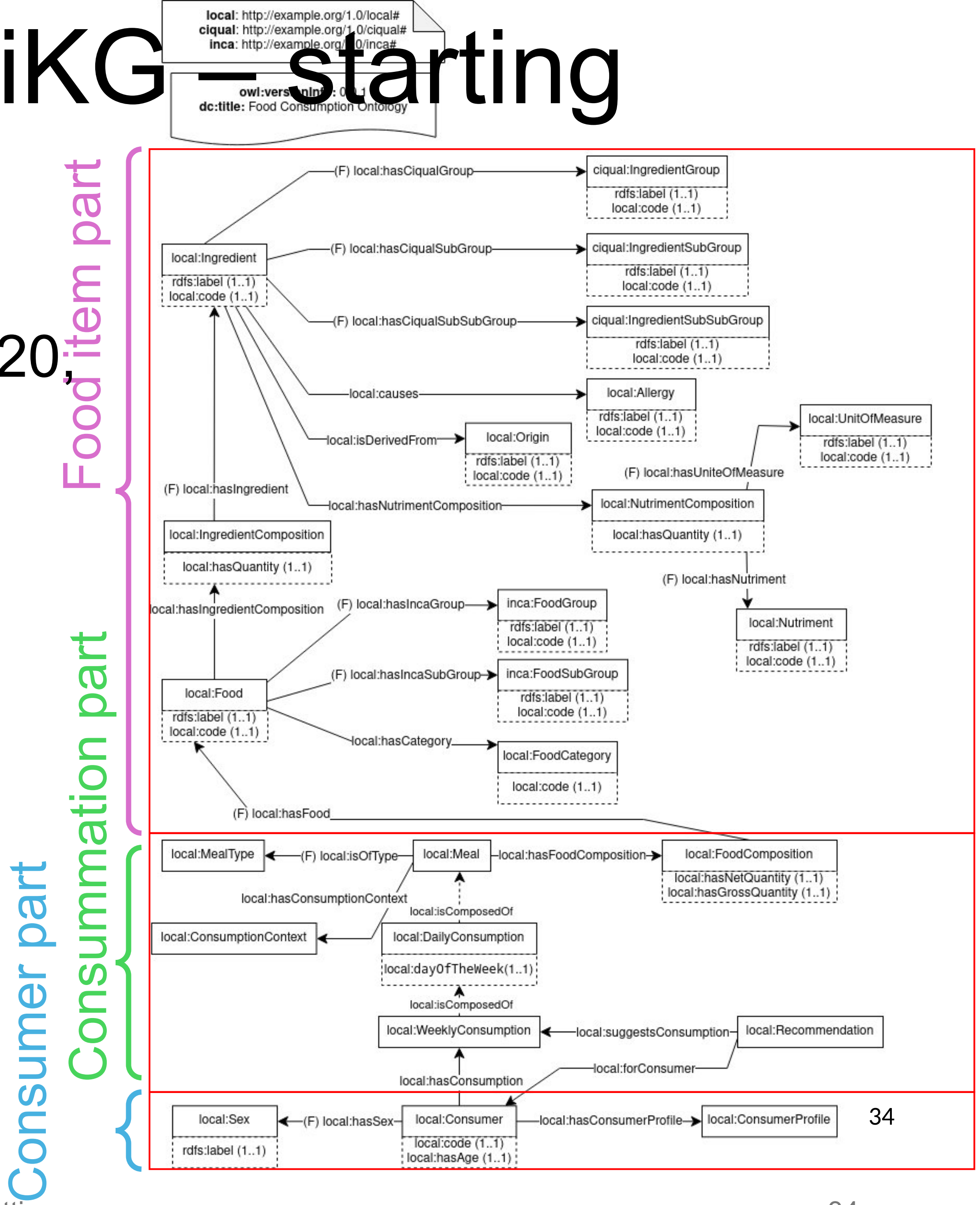
A. Combeau, V. Guigue, M. Cristina, F. Sais, P. Viappiani, S. Dervaux :
Génération séquentielle prenant en compte des informations contextuelles en nutrition. CAP 2025, Juin 2025, Dijon, France.

The Exersys Project : NutriKG – starting point

Multiple sources of data : INCA 2, CIQUAL 2020, Open Food Facts

Modeling

- consumer's profiles
- nutrition constraints
- Association rules
- Exclusion rules
- Cardinality rules



The Exersys Project : NutriKG

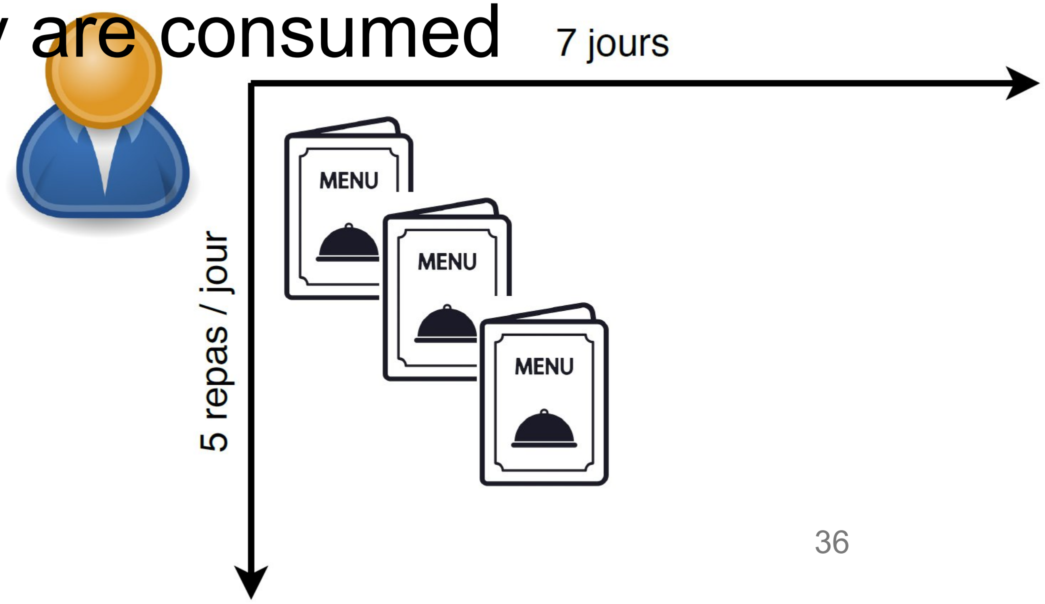
- puts together INCA2 and INCA3 datasets

DataSet	#individus adults	#jours	#conso	Taille conso
INCA2	2624	7	80052	5.8
INCA3	3157	3	34964	5.7

RecSys in the nutrition domain : INCA2 dataset

- Quantity : 2624 adults individuals, **7 days**, **5 meals**
- 280 food items organized in 44 **groups** and 110 sub-groups of food items
- **Context factors** : place, company, eater carachteristics
- Structured, good quality, sequences of real consumptions over 7 days
- Consumptions are entered in order as they are consumed

(+ transversal food items at the end)



- **Meal mean lenght** : **7/8 items**

RecSys in the nutrition domain : INCA3 dataset

- Quantity : 3157 adults individuals, **3 days non-consecutive**, **5 meals**
- food items organized in FoodEx **groups** and sub-groups
- **Context factors** : place, company, eater carachteristics, **food preparation**, **storage habits**
- Structured, good quality, sequences of real consumptions over **24 hours**
- Consumptions are entered in order as they are consumed

MIA
PARIS-SACLAY
EKINOCS

The Exersys Project : NutriKG

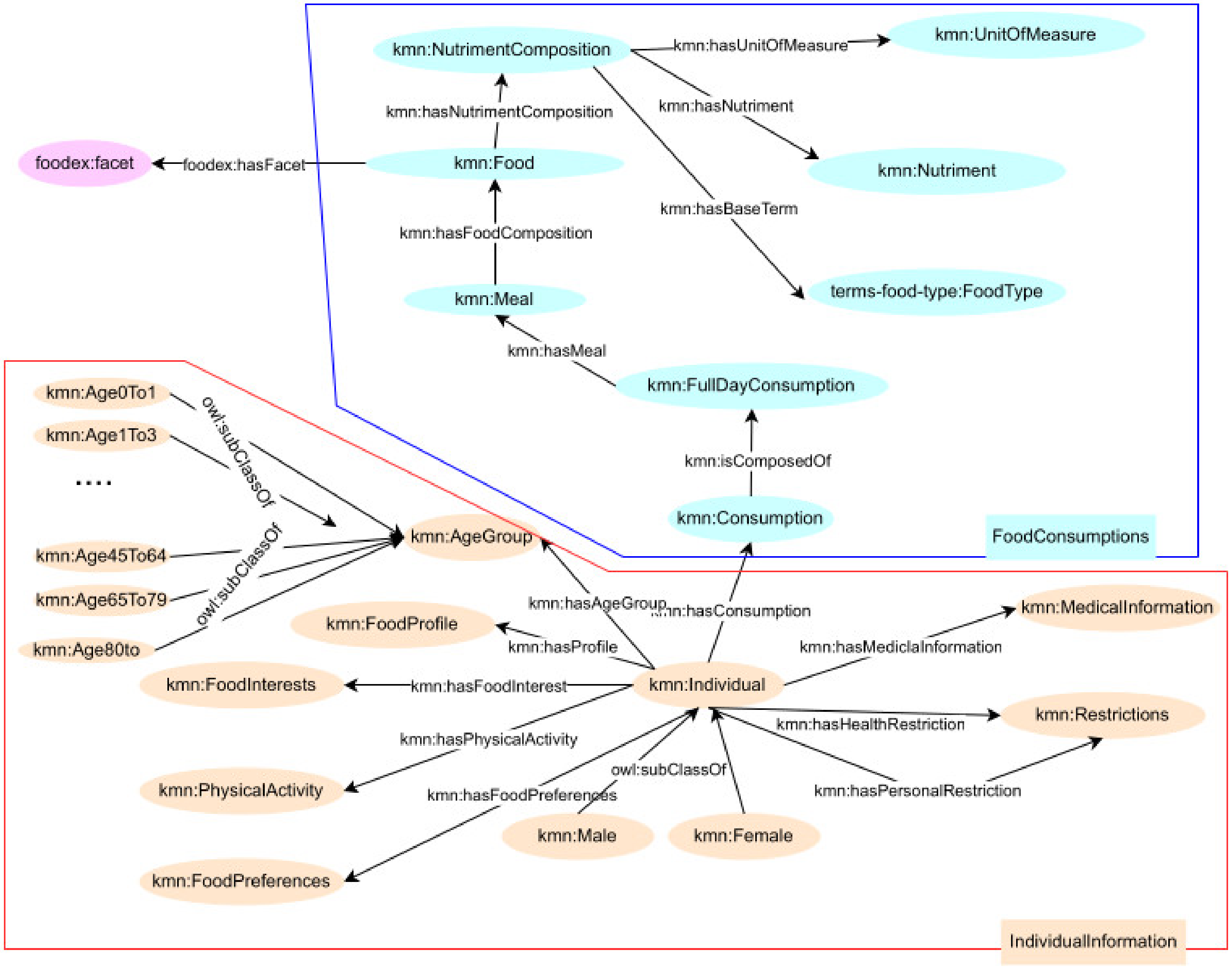
- puts together INCA2 and INCA3 datasets

DataSet	#individus adults	#jours	#conso	Taille conso
INCA2	2624	7	80052	5.8
INCA3	3157	3	34964	5.7

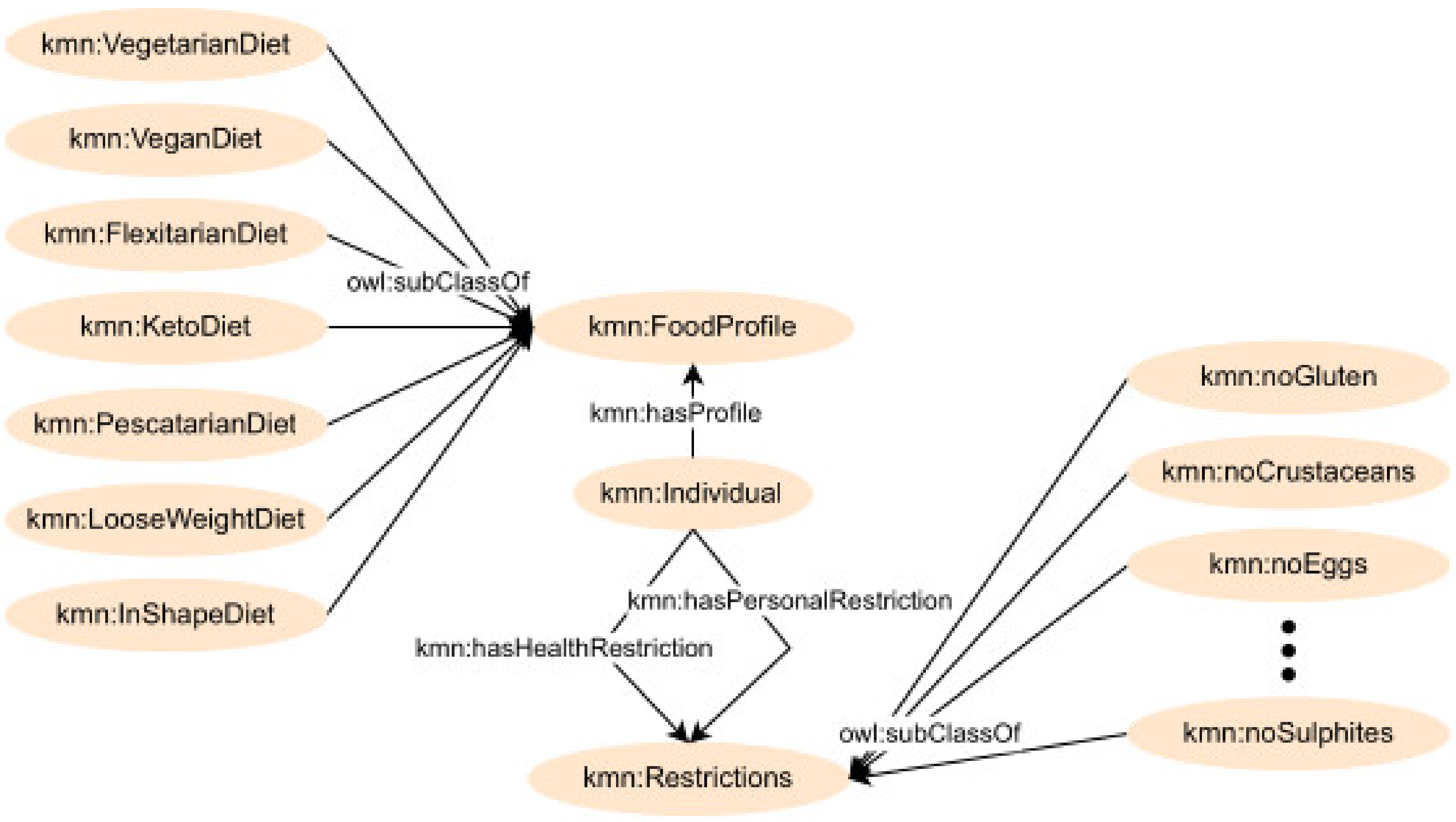
- In the FoodEx2 reference
- KG enriched by SWRL rules and a SHACL validation schema

A. Combeau, F. Sais, S. Dervaux, N. Kumari, C. Manfredotti, V. Guigue, P. Viappiani : NutriKG - un graphe de connaissances pour modéliser les préférences et les besoins nutritionnels IC 2025, Juin 2025, Dijon, France.

NutriKG : ontolog



NutriKG : restrictions and user profiles



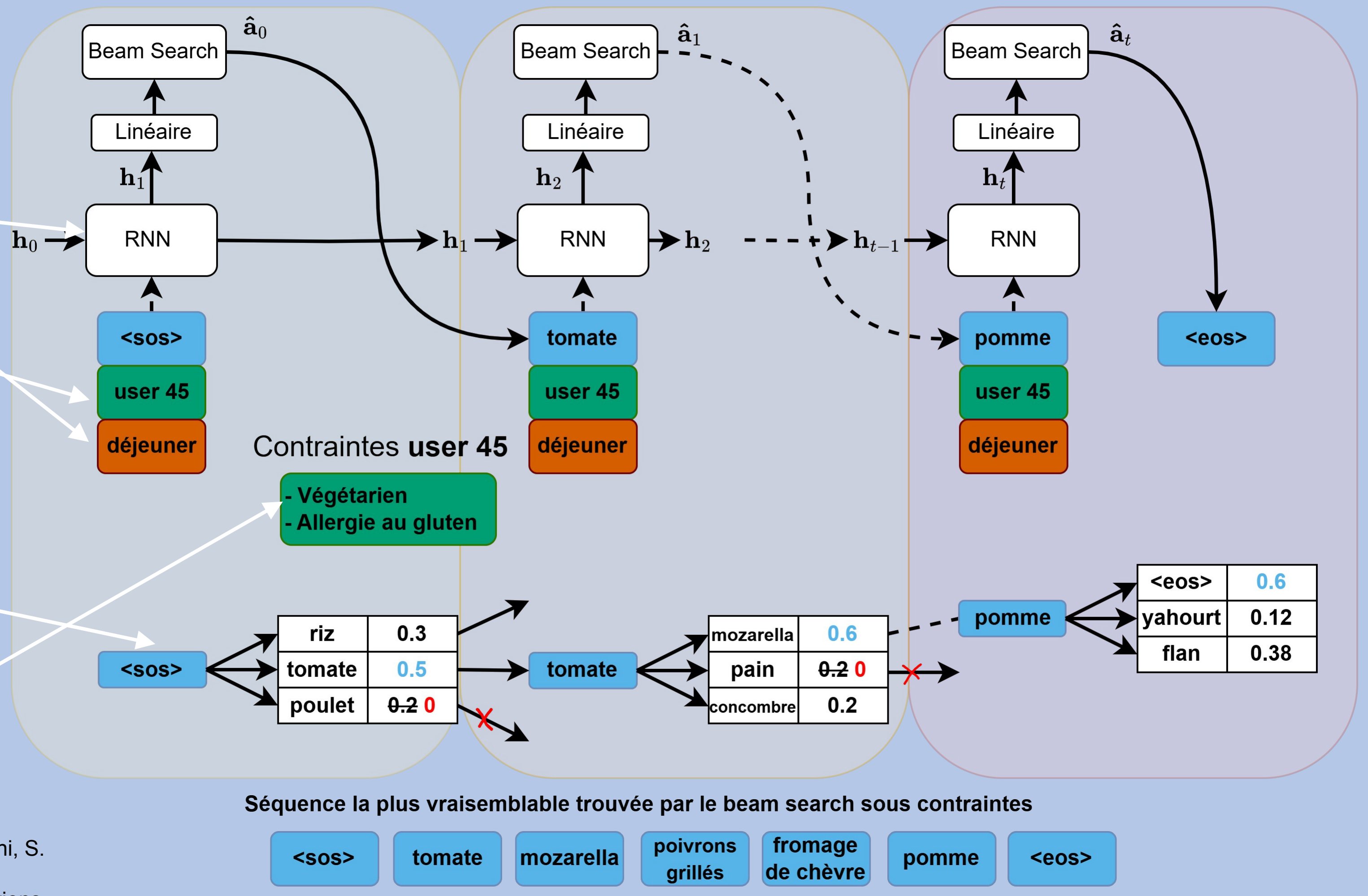
The exersys project : individual recommendation

The **model** generates a meal based on

- The consumption **context**
- **User's preferences**

Beam Search creats a **tree ofall possible meals** and save the most log-likelihood

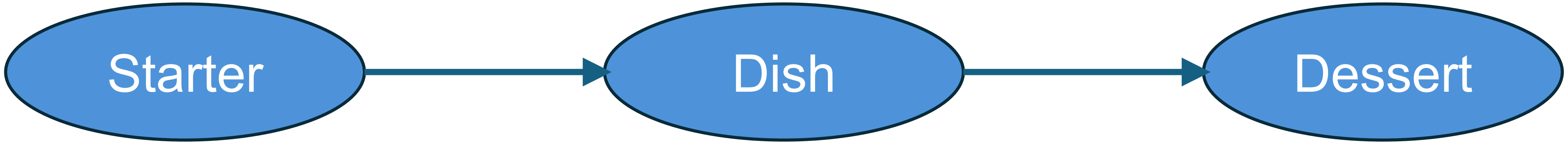
It can be **constrained**

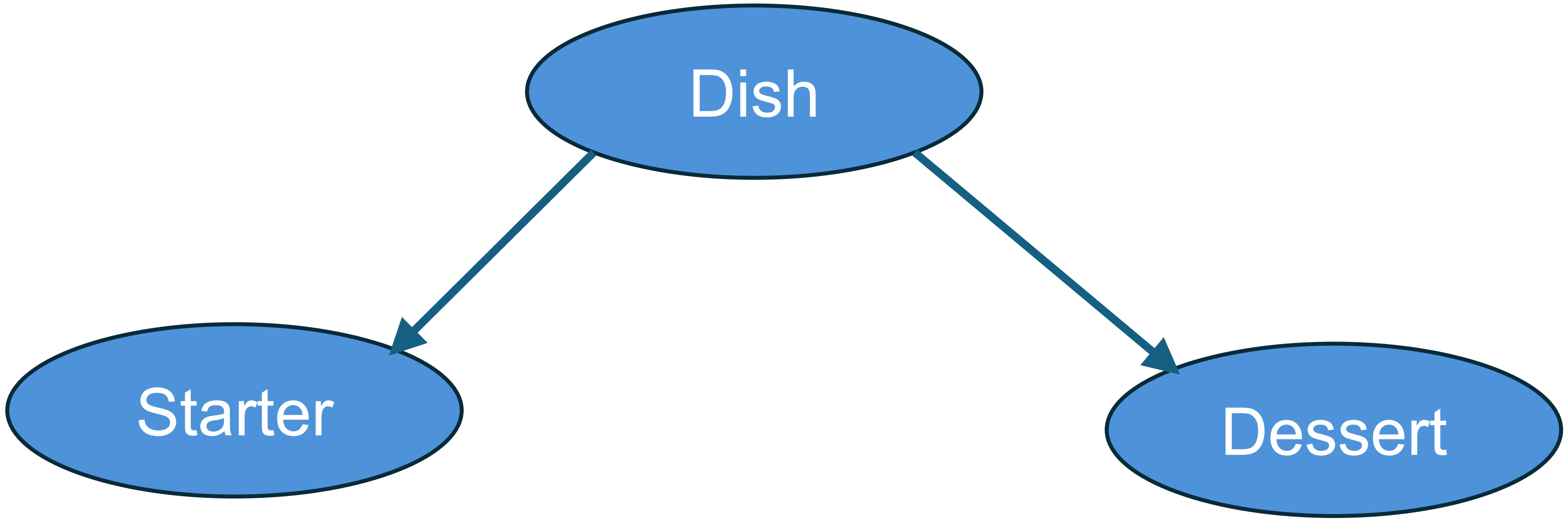


A. Combeau, V. Guigue, M. Cristina, F. Sais, P. Viappiani, S. Dervaux :
Génération séquentielle prenant en compte des informations contextuelles en nutrition. CAP 2025, Juin 2025, Dijon, France.

The exersys satellite project : sequentiality vs causality

1.





2.

Meal-Dinner_consumption-7_consumer-INCA2-110207

Nom des consommations	Ordre de prise de l’aliment
“plain omelete cooked in butter”	1
“plain cooked bacon”	2
“boiled potato”	3
“non light salted butter added during cooking”	4
“escarole, raw”	5
“vinegar”	6
“fresh clementine or mandarin”	7
[illegible]	[illegible]

Meal classification : 14 rules

→ Rules order very important

Table: explicit food items groups classification

Entrée	Plat principal	Dessert	Boisson
38 – Soupes	17 – Viandes	8 - Pâtisseries	31 - eau
40 - Entrées	18 – Volailles et gibiers	10 – Ultra-frais laitier	32 - BRSA
	19 - Abats	28 - Glaces	33 – boissons alcoolisées
	36 – Pizzas, quiches	29 - Chocolat	34 - café
	37 – Sandwiches, casse croutes	42 – Compotes et fruits cuits	35 – boissons chaudes
	39 – Plats composés	41 - Entremets	

Example : Consommateur 110207 (INCA2) - Dîner 7

Rule 1 : explicit food items classification

Boisson	**“contrex still water”**	**8**

Rule 2 : « entrée » non explicit based on nutrition

teneur en fibre, eau, énergie, qte + position dans le repas

Entrée	**“boiled potato”**	**3**
Entrée	**“escarole, raw”**	**5**
Boisson	“contrex still water”	8

Rule 3 : non explicit dessert classification

Les seuls concernés sont les fruits frais + sec
Pourquoi ?

- tous les autres dessert sauf les fruits sont classés dès la règle 1 en explicite
- les fruits peuvent être aussi bien des entrées que des desserts (pamplemousse, melon, etc)

Entrée	"boiled potato"	3
Entrée	"escarole, raw"	5
Dessert	**"fresh clementine or mandarin"**	**7**
Boisson	"contrex still water"	8

2 autres règles suivent la règle 3 mais n'ont pas eu d'influence sur cette classification

Rule 6 : eggs

grp 12, 2 rôle possibles : Entrée ou Plat selon la qte + la présence ou non d’une entrée après l’œuf étudié

Entrée	**“plain omelete cooked in butter”**	**1**
Entrée	“boiled potato”	3
Entrée	“escarole, raw”	5
Dessert	“fresh clementine or mandarin”	7
Boisson	“contrex still water”	8

MIA
PARIS-SACLAY
EKINOCS

Rule 7 : charcuterie

grp 17, 2 rôles possibles : Entrée ou Plat, selon la qte + présence ou non d'une entrée après la charcuterie étudiée

Entrée	“plain omelete cooked in butter”	1
Entrée	**“plain cooked bacon”**	**2**
Entrée	“boiled potato”	3
Entrée	“escarole, raw”	5
Dessert	“fresh clementine or mandarin”	7
Boisson	“contrex still water”	8

Rule 9 : managing dish absence

l'entrée la plus calorique, qui n'est pas un fruit, change de rôle et devient Plat

Entrée	"plain omelete cooked in butter"	1
Plat	**"plain cooked bacon"**	**2**
Entrée	"boiled potato"	3
Entrée	"escarole, raw"	5
Dessert	"fresh clementine or mandarin"	7
Boisson	"contrex still water"	8

Rule 13 : managing positional errors

les entrées qui suivent un plat sont converties en plat

Entrée	“plain omelete cooked in butter”	1
Plat	“plain cooked bacon”	2
Plat	**“boiled potato”**	**3**
Plat	**“escarole, raw”**	**5**
Dessert	“fresh clementine or mandarin”	7
Boisson	“contrex still water”	8

pas de ligne 4 et 6 !

Non classified food items

⇢ ils font partie d'un groupe : les bases alimentaires

⇢ on ne leur affecte pas de rôle

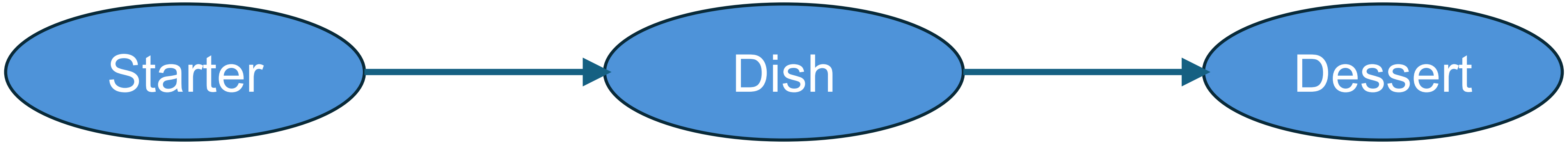
Bases alimentaires INCA2 :	Beurre, huile, margarine, autres graisses, sucres et dérivés (sauf bonbons), pain, condiments et sauces

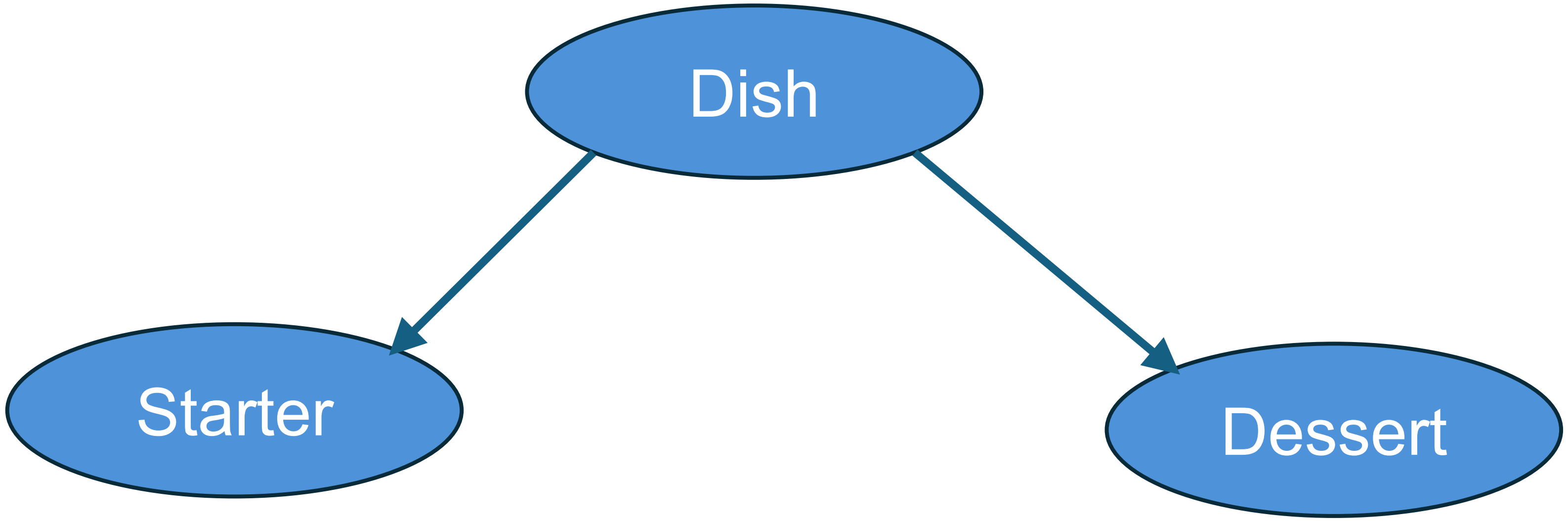
Dans notre exemple :

	"non light salted butter added during cooking"	4
	"vinegar"	6

The exersys satellite project : sequentiality vs causality

1.



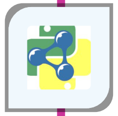


2.

The exersys project

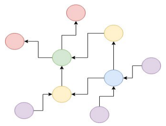
Developing a RecSys for meals recommendation

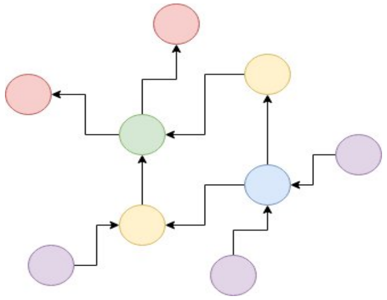
- **<u>Axe 1:</u>**
  Recommendation based on **user's preferences** learnt by **machine learning methods**



Génération de recommandati

<u>Axe 2:</u>
Nutritional constraints integration by **knowledge graphs**



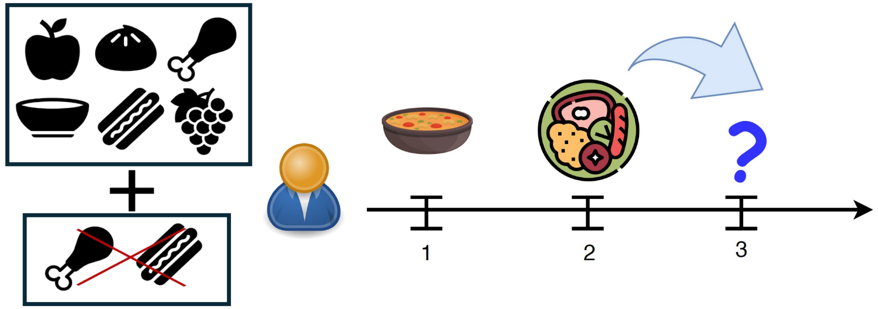


CLASSIFICATIO N

Sequentiality
vs
Causality

Historique de consommation

Contrainte alimentair e



NutriKG : meal classification

RDF

- Format standard pour représenter des données liées .
- Basé sur des triplets :

Sujet $\xrightarrow{\text{Prédicat}}$ Objet

?food kmn:energy ?energy

Graphe de connaissance : donner un sens aux données

+ 800 000 lignes
de consommations

Ontologies (OWL)

- Définir des règles sémantiques (classes, hiérarchie)

```
kmn:Male a owl:Class ;
    rdfs:label "Male" ;
    rdfs:subClassOf kmn:Individual .
```

SPARQL

- Langage de requête utilisé pour extraire des connaissances du graphe

```
SELECT ?food
WHERE {
   ?food kmn:codgr ?codgr .
    FILTER(?codgr IN (12,17)}
```

The exersys project : Sequence generation with user

$$\mathbf{h}_t = g(W_a\,\mathbf{a}_t + W_u\,\mathbf{u}_i + U\,\mathbf{h}_{t-1})$$

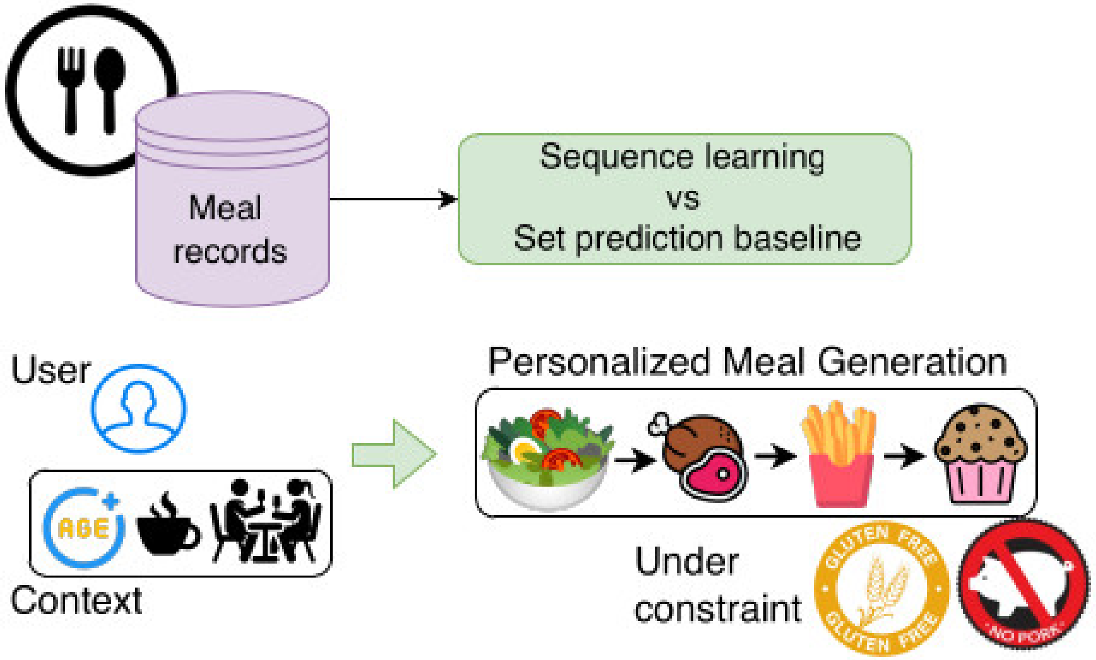
$$\hat{p} = f(\mathbf{h}_t) = \mathtt{softmax}(V\,\mathbf{h}_t)$$

$$V \in \mathbb{R}^{|\mathcal{A}|\times h}$$

$$\hat{a}_{t+1} = \arg\max \hat{p}$$

MIA PARIS-SACLAY EKINOCS

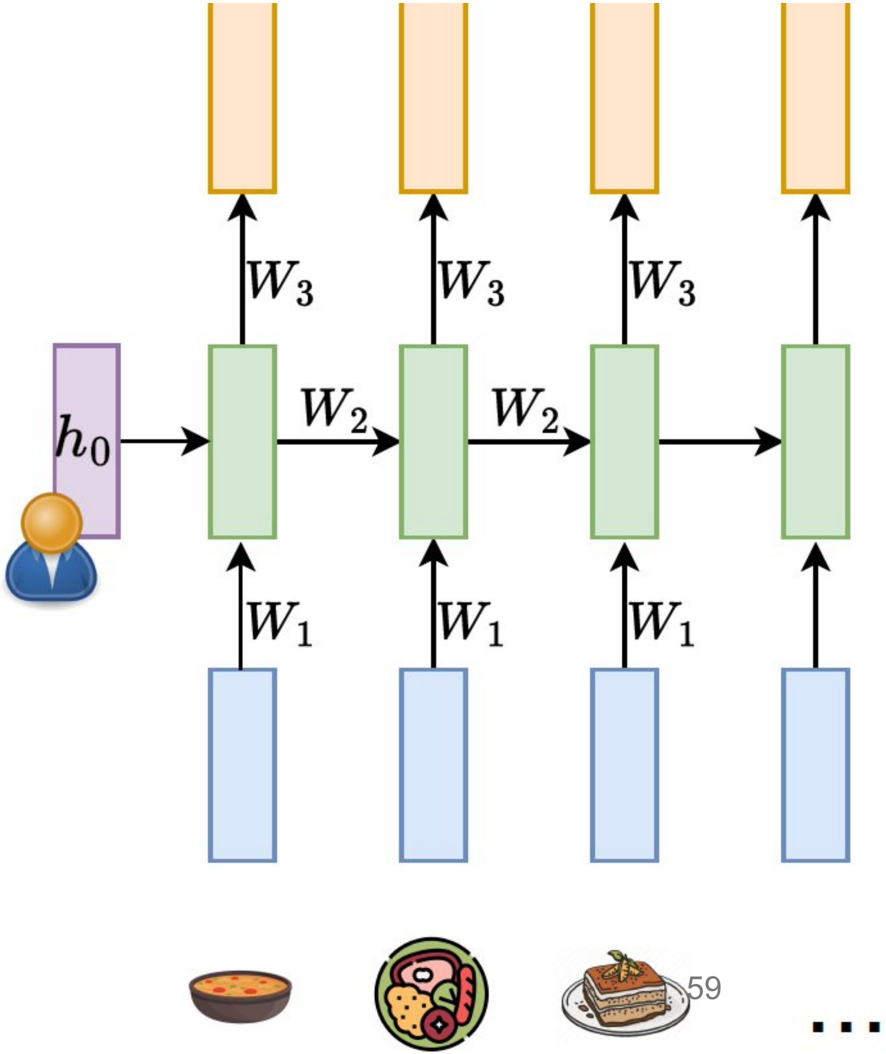
The exersys project : Sequence generation



Sequentiality modeling : RNN

Sequence prediction :

- Reconstructing a full sequence
- How to take into consideration the user ? The context ? First item prediction ?



User integration :

1. User = initial hidden state
   - First item prediction
2. User => concatenated with food items
   - Elegant architecture and less sensible to forgetting

Sequentiality modeling : user integration

$$X = \{m_{u,d,r}\}$$

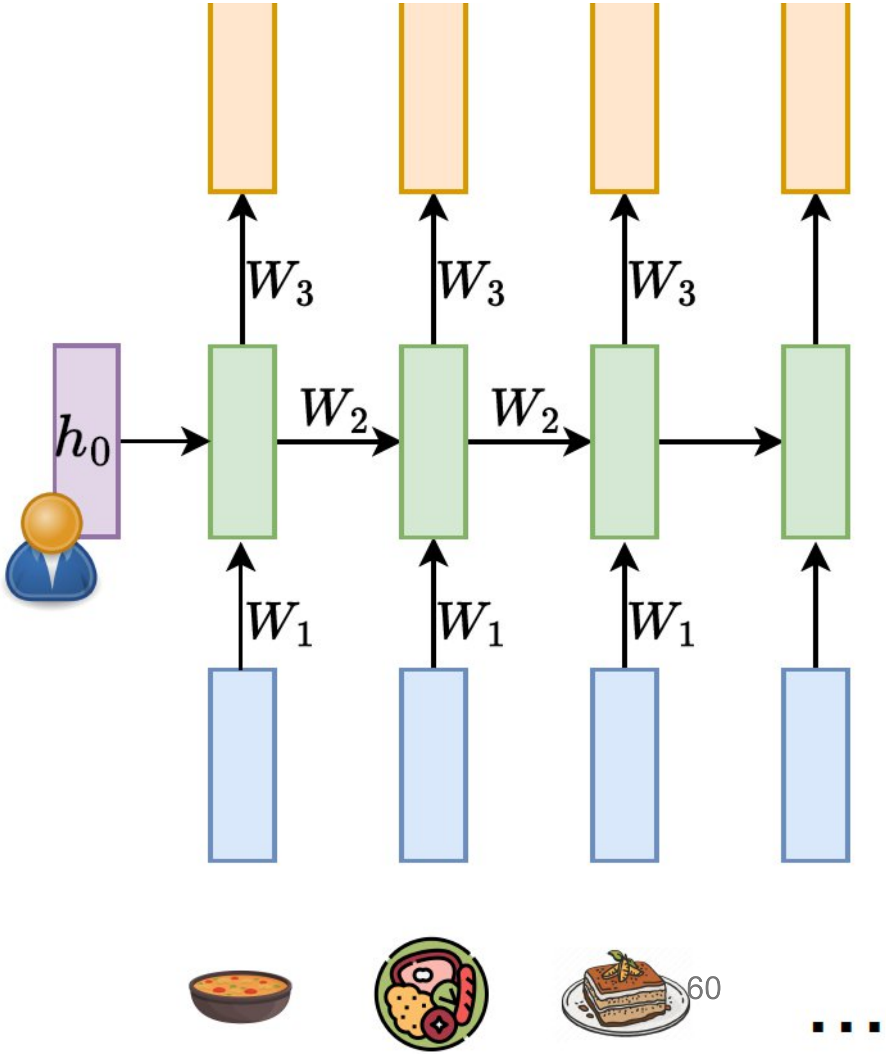
$$m_{u,d,r} = [a_1, \ldots, a_t, \ldots a_T]$$

$$a \in \mathcal{A} \mapsto \mathbf{a} \in \mathbb{R}^z \qquad u \in \mathcal{U} \mapsto \mathbf{u} \in \mathbb{R}^z$$

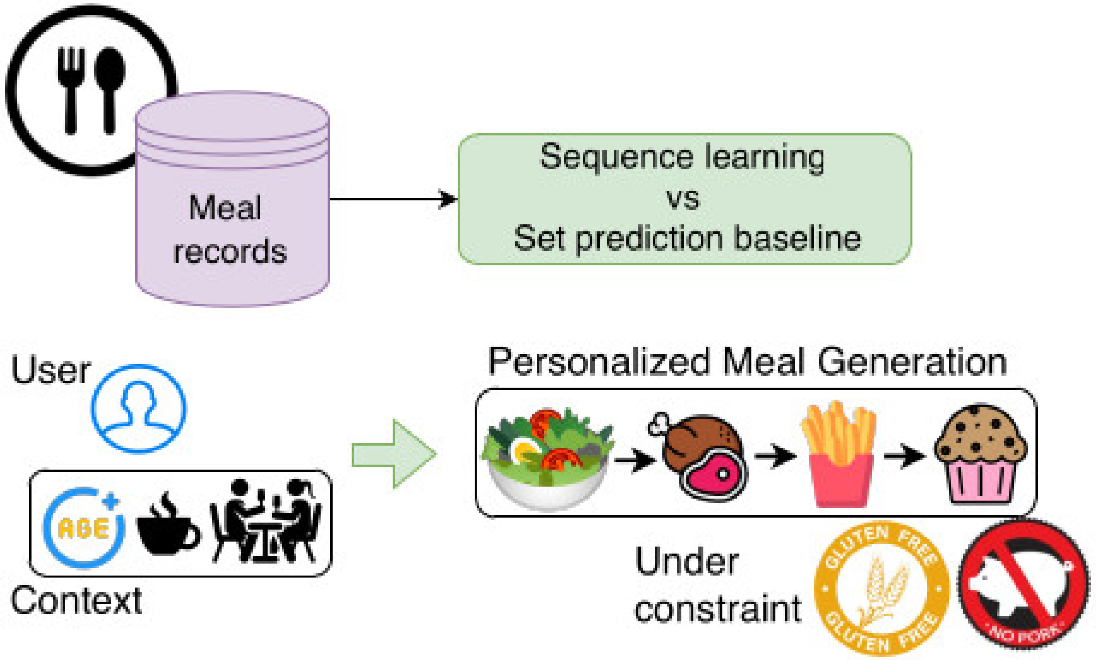
$$\mathbf{h}_t = g(W_a\, \mathbf{a}_t + W_u\, \mathbf{u}_i + U\, \mathbf{h}_{t-1})$$

$$\hat{p} = f(\mathbf{h}_t) = \mathtt{softmax}(V\, \mathbf{h}_t)$$

$$V \in \mathbb{R}^{|\mathcal{A}| \times h} \qquad \hat{a}_{t+1} = \arg\max \hat{p}$$



The exersys project : context integration



$\forall \left(u_i, m_{u_i,d,r}\right)$ -> context $c \in C$

Contexts :

- User or meal caractheristics
- age or meal type

$$c \in C \mapsto \mathbf{c} \in \mathbb{R}^z$$

$$\mathbf{h}_t = g(W_a\,\mathbf{a}_t + W_u\,\mathbf{u}_i + \sum_j W_{c_j}\,\mathbf{c}_j + U\,\mathbf{h}_{t-1})$$

Test : exact prediction

Métriques	Tx	Tx-Top3	Tx-CAT	Tx	Tx-Top3	Tx-CAT
Modèles	DEJ+DIN			PT-DEJ		
Sans utilisateur	0.11	0.25	0.20	0.38	0.60	0.47
Util. $= \mathbf{h}_0$	0.13	0.27	0.24	0.66	0.78	0.73
Concat. Util.+Alim.	0.16	0.30	0.24	0.71	0.83	0.77
Aléatoire	0.003	0.01	0.023	0.018	0.055	0.023

Tx : good prediction rate for the next food item,

Tx-Top3 : good prediction rate in top3.

Tx-Cat : good prediction rate for the INCA2 category.

The exersys project : ongoing work

- Currently :
  - considering more context elements (age, type of meal, etc.)
- Enrich the dataset
  - INCA2 + INCA3
  - with massive online food websites (Marmiton) ??
- Inject **EK** to the recommender system
  - guiding the prediction with rules and axioms deduced by the ontology
- Curriculum Learning for full sequence generation

RecSys in the nutrition domain : future work

- Improve the explicability of the prediction
  - better system-human interaction
  - possible help from visualisation

- Probabilistic models to model dependencies between courses

Plan

- Background : how I got here and what I learnt from this exercice
- Accomplished research
  - Experts' knowledge and PRMs
  - Experts' knowledge for RecSys
- Drawbacks and Future Works

MIA
PARIS-SACLAY
EKINOCS

Experts' knowledge : common points

- How **EK** can **improve** the learning of a probabilistic model for inference and prediction and the recommandation of an item in a RecSys for nutrition

- We dealt with applications where **data are scarse**, **EK** helps in enriching them :
  - They specify them
  - They constraints them

- **EK** help in **explicability** and add **rules** (=> causality)

MIA
PARIS-SACLAY
EKINOCS

Experts' knowledge : future works

- Relying more on the knowledge graphs
  - Exersys => *planned*
  - PRMs => *to investigate*

- Data linking
  - INCA2 + INCA3
  - Transfer learning

- Expert inclusion and visualisation

- Using the mapping for RecSys

Journal papers:

- J. Vandeputte, P. Herold, M. Kuslii, P. Viappiani, L. Muller, C. Martin, O. Davidenko, F. Delaere, C. Manfredotti, A. Cornuéjols, N. Darcel. Principles and Validations of an Artificial Intelligence-Based Recommender System Suggesting Acceptable Food Changes. Journal of Nutrition, 153, 2 (2023).
- M. Münch, P. Buche, C. Manfredotti, P-H. Wuillemin, H. Angellier-Coussy. Formalizing Contextual Expert Knowledge for Causal Discovery in linked Knowledge Graphs about Transformation Processes: Application to processing of bio-composites for food packaging International Journal of Metadata, Semantics and Ontologies IJMSO-349696
- M. Münch, P. Buche, S. Dervaux, J. Dibie, L. Ibanescu, C. Manfredotti, P-H. Wuillemin, H. Angellier-Coussy. Combining ontology and probabilistic models for the design of bio-based product transformation processes. Expert Systems for Applications 203: 117406 (2022)
- L. Cattelani, C. Manfredotti, E. Messina. A Particle Filtering Approach for Tracking an Unknown Number of Objects with Dynamic Relations. Journal of Mathematical Modeling and Algorithms in OR 13(1): 3-21 (2014)
- E. Fersini, E. Messina, F. Archetti, C. Manfredotti. Combining Gene Expression Profiles and Drug Activity Patterns Analysis: A Relational Clustering Approach. Journal of Mathematical Modeling and Algorithms, 9(3): 275-289 (2010)

Conference papers:

- N. Jacquet, V. Guigue, C. Manfredotti, F. Saïs, S. Dervaux, P. Viappiani. Modélisation du caractère séquentiel des repas pour améliorer la performance d'un système de recommandation alimentaire. 24eme Conférence francophone sur l'Extraction et la Gestion des Connaissances (EGC 2024), Jan 2024, Dijon, France.
- M. Munch, P. Buche, C. Manfredotti, P-H. Wuillemin, H. Angellier-Coussy. A process reverse engineering approach using Process and Observation Ontology and Probabilistic Relational Models: application to processing of bio-composites for food packaging. 15th International Conference on Metadata and Semantics Research 2021.
- C. Manfredotti, P. Viappiani. A Bayesian Interpretation of the Monty Hall Problem with Epistemic Uncertainty. MDAI 2021: 93-105
- M. Munch, J. Dibie, P-H. Wuillemin, C. Manfredotti. Towards Interactive Causal Relation Discovery Driven by an Ontology. FLAIRS Conference 2019: 504-508
- M. Munch, P-H. Wuillemin, J. Dibie, C. Manfredotti, T. Allard, S. Buchin, E. Guichard. Identifying Control Parameters in Cheese Fabrication Process Using Precedence Constraints. DS 2018: 421-434
- M. Munch, P-H. Wuillemin, C. Manfredotti, J. Dibie, S. Dervaux. Learning Probabilistic Relational Models Using an Ontology of Transformation Processes. OTM Conferences (2) 2017: 198-215
- M. Bouyrie, C. Manfredotti, N. Peyrièras, A. Cornuéjols. Denoising 3D Microscopy Images of Cell Nuclei using Shape Priors on an Anisotropic Grid. nternational Conference on Pattern Recognition Applications and Methods (ICPRAM)} 2016: 291-298
- C. Gonzales, S. Dubuisson, C. Manfredotti. A New Algorithm for Learning Non-Stationary Dynamic Bayesian Networks With Application to Event Detection. Florida Artificial Intelligence Research Society Conference (FLAIRS) Conference 2015: 564-569
- C. Manfredotti, C. Baudrit, J. Dibie, P-H. Wuillemin. Mapping Ontology with Probabilistic Relational Models. International Conference on Knowledge Engineering and Ontology Development (KEOD) 2015, pp. 171-178.
- C. Manfredotti, K. Steenstrup Pedersen, H. J. Hamilton, S. Zilles. Learning Models of Activities Involving Interacting Objects. Proceedings of Advances in Intelligent Data Analysis (IDA) 2013, pp. 285-297.
- L. Cattelani, C. Manfredotti, E. Messina. Multiple Object Tracking with Relations. International Conference on Pattern Recognition Applications and Methods (ICPRAM), 2012: 459-466.
- C. Manfredotti, D. J. Fleet, H. J. Hamilton, S. Zilles. Simultaneous Tracking and Activity Recognition. Proceedings of the 23rd IEEE International Conference on Tools with Artificial Intelligence ICTAI 2011: 189-196.
- C. Manfredotti, D. J. Fleet, E. Messina. Relations to improve Multi-Target Tracking in an Activity Recognition System. 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP-09), December, 2009.
- C. Manfredotti, E. Messina. Relational Dynamic Bayesian Networks to Improve Multi-target Tracking. Advanced Concepts for Intelligent Vision Systems (ACIVS)} 2009: 528-539
- C. Manfredotti. Modeling and Inference with Relational Dynamic Bayesian Networks. Canadian Conference on AI}, 2009, pp. 287-290.
- F. Archetti, C. Manfredotti, V. Messina, D. G. Sorrenti. Foreground-to-Ghost Discrimination in Single-Difference Pre-processing. Advanced Concepts for Intelligent Vision Systems (ACIVS)} 2006: 263-274
- F. Archetti, C. Manfredotti, M. Matteucci, E. Messina, D. G. Sorrenti. Parallel first-order Markov Chain for on-line Anomaly Detection in traffic video surveillance. IET Conference on Crime and Security: the Technical Fight}, June, 2006.

Workshop papers:

- N. Jacquet, C. Manfredotti, V. Guigue, F. Saïs, P. Viappiani An EXplainable RecommandER SYStem for the Nutrition Domain, combining Knowledge Graphs and Machine Learning CoCoA-BeANS workshop: Cognitive and COmputational Approaches of Behaviour and Nutrition Studies. Paris, May 11-12, 2023.
- M. Munch, C. Manfredotti, L. Ibanescu, P. Buche. An ontology-based pipeline to support the design of technical itineraries: application to composite food packaging. Integrated Food Ontology Workshop 2021, Sept 2021, Bolzano, Italy.
- M. Caillat, N. Darcel, C. Manfredotti, P. Viappiani. Bayesian Vote Elicitation for Group Recommendations. DA2PL 2020, Nov 2020, Trento, Italy.
- M. Munch, J. Dibie-Barthélemy, P-H. Wuillemin, C. Manfredotti. Interactive Causal Discovery in Knowledge Graphs. PROFILES/SEMEX@ISWC 2019: 78-93
- S. Akkoyunlu, C. Manfredotti, A. Cornuéjols, N. Darcel, F. Delaere. Exploring Eating Behaviours Modelling for User Clustering. HealthRecSys@RecSys 2018: 46-51
- S. Akkoyunlu, C. Manfredotti, A. Cornuéjols, N. Darcel, F. Delaere. Investigating Substitutability of Food Items in Consumption Data. HealthRecSys@RecSys 2017: 27-31
- L. Cattelani, C. Manfredotti, E. Messina. Multiple objects tracking with probabilistic relationships. 1st interdisciplinary workshop on Mathematics of Filtering and its Applications (MFA2011), July, 2011.
- C. Manfredotti, D. J. Fleet, H. J. Hamilton, S. Zilles. Relational Particle Filtering. NIPS workshop on Monte Carlo Methods for Bayesian Inference in Modern Day Applications, December 2010.
- C. Manfredotti, S. Zilles, H. J. Hamilton. Learning RDBNs for Activity Recognition. NIPS Workshop on Learning and Planning in Batch Time Series Data, December 2010.
- C. Manfredotti, E.Messina, D. Fleet. Relations as Context to improve Multi Target Tracking and Activity Recognition. First International Workshop on Logic-Based Interpretation of Context: Modeling and Applications, September, 2009.
- E. Fersini, C. Manfredotti, E. Messina, F. Archetti. Relational Clustering for Gene Expression Profiles and Drug Activity Pattern Analysis. SysBioHealth Symposium, October, 2007.

Mapping ontology and PRM: remarks

- Motivations :
  - Need to reason with uncertainty in transformation processes
  - Similarity between ontologies and PRMS
- The experts
  - give the ontology
  - structures the variables with his hypothesis
  - afterwards, critiques the model
- The approaches
  - Verify the expert's hypothesis
  - Suggest new experiments
  - Able to do prediction

Mapping ontology and PRM: ideas for the future

- The expert's validation step could be investigated more
  - Exploring algorithms able to learn a network locally
  - Visualisation and human machine interaction
  - (automatic) Explanation, (also) to suggest new experiments

- The ontology could be used more
  - Answering competency questions to structure de domain giving more insights to the learning

- Transfer Learning, Data linking, Graph Matching

The exersys project : ongoing work

- Currently :
  - considering more context elements (age, type of meal, etc.)

- Enrich the dataset
  - INCA2 + INCA3
  - with massive online food websites (Marmiton) ??

- Inject **EK** to the recommender system
  - guiding the prediction with rules and axioms deduced by the ontology

- Curriculum Learning for full sequence generation

RecSys in the nutrition domain : future work

- Improve the explicability of the prediction
  - better system-human interaction
  - possible help from visualisation

- Probabilistic models to model dependencies between courses

EK and PRMs : why ?

@ AgroParisTech :
- necessity of reasoning about transformation processes
- already an ontology structuring this domain

➢ PRMs to reason on uncertainty in transformation processes

- Similarity between PRMs and Ontologies => mapping

PRMs

A BN is the representation of a joint probability over a set of random variables that uses a DAG to encode probabilistic relations between variables

Combine advantages of relational data bases & **Bayesian networks**:

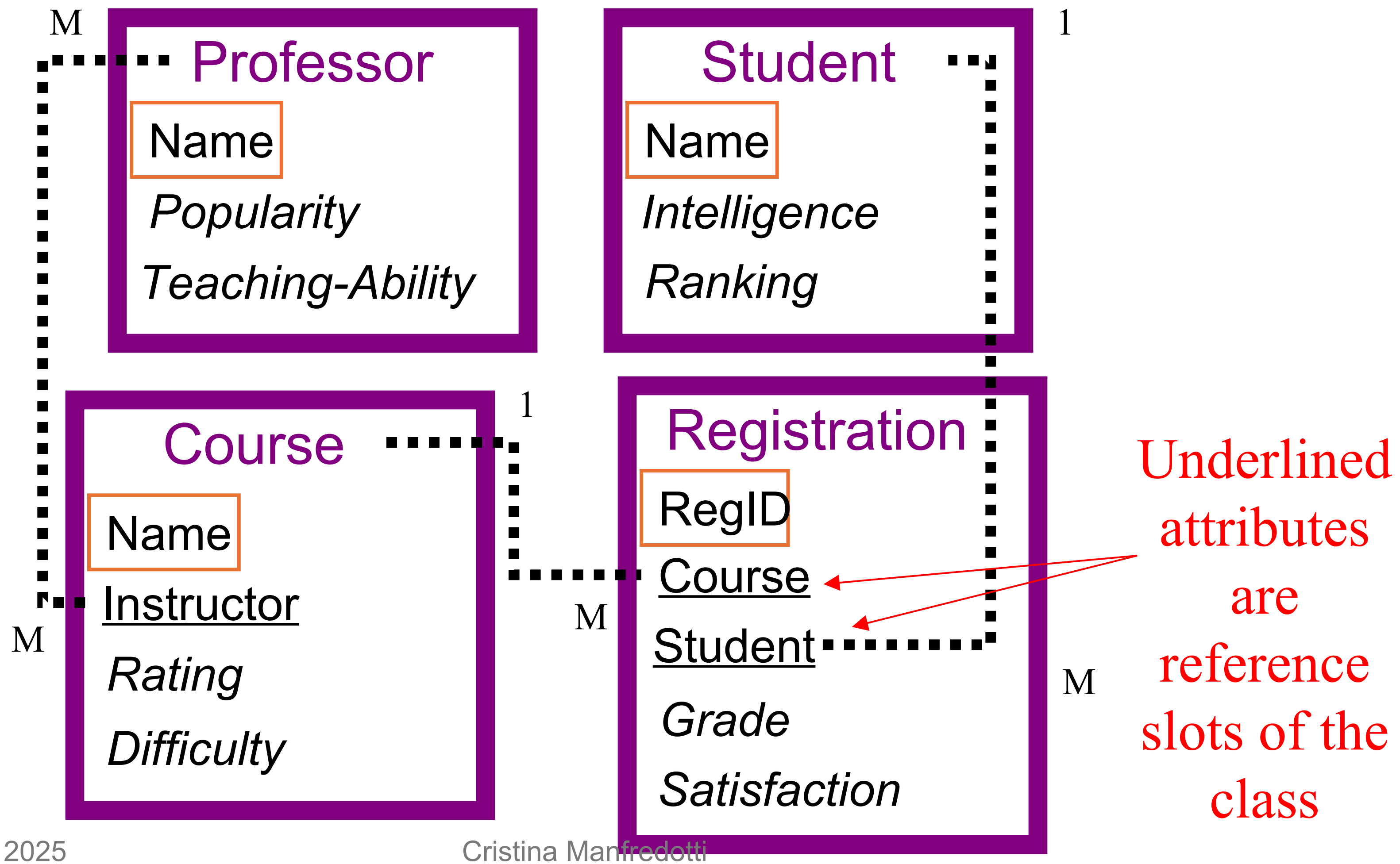
- natural domain modeling: objects, properties, relations;
- generalization over a variety of situations;
- compact, natural probability models.

➢ **Relational Schema** and **Relational slots** (and slot chain)

PRM system => (big) BN

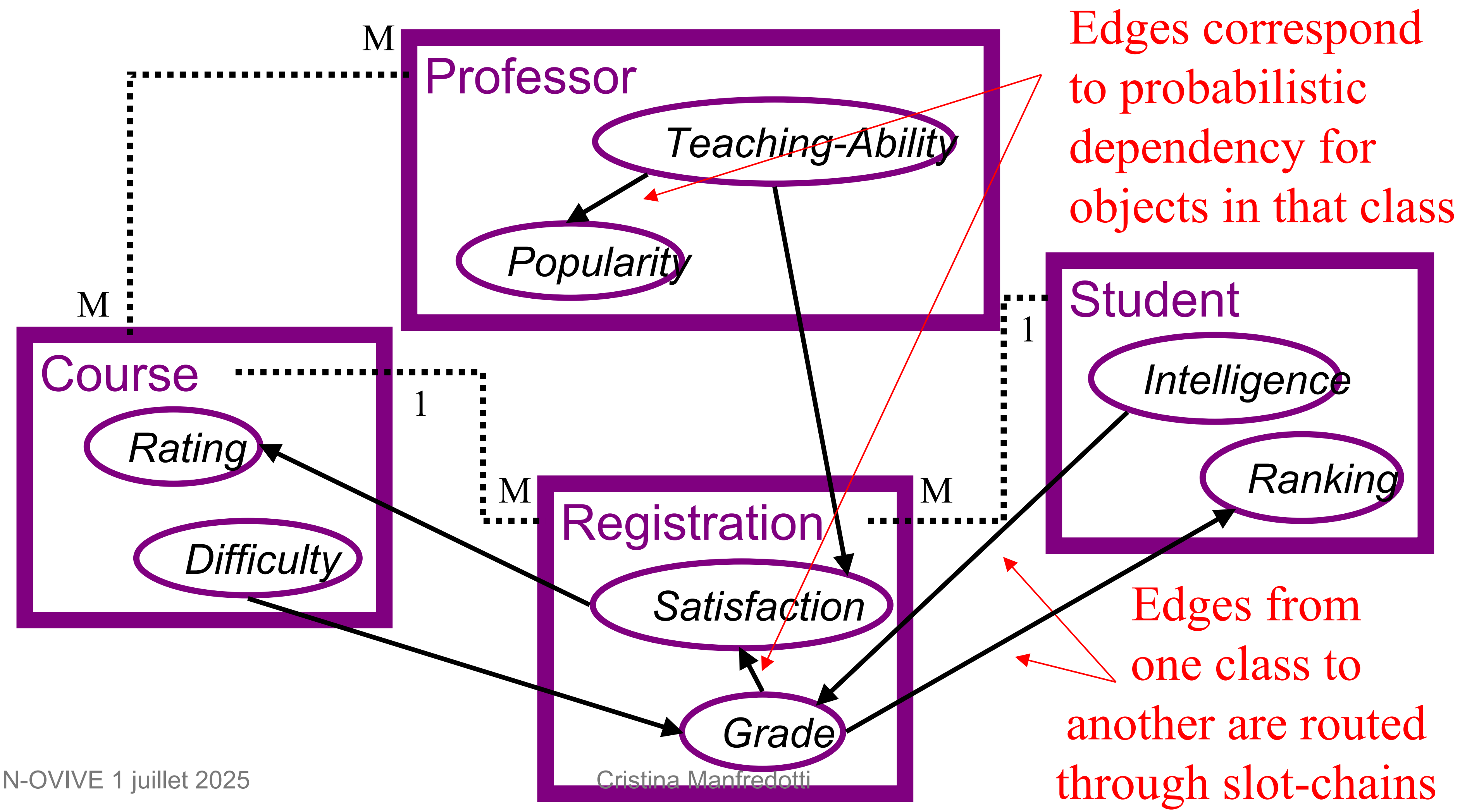
PRM – Relational Schema [Getoor 01]

University Domain Example



PRM definition

the University Domain



PRMs

A BN is the representation of a joint probability over a set of random variables that uses a DAG to encode probabilistic relations between variables

Combine advantages of relational data bases & **Bayesian networks**:

- natural domain modeling: objects, properties, relations;
- generalization over a variety of situations;
- compact, natural probability models.

- Relational Schema
- Relational slots (and slot chain)

PRM system => (big) BN

EK and PRMs : why ?

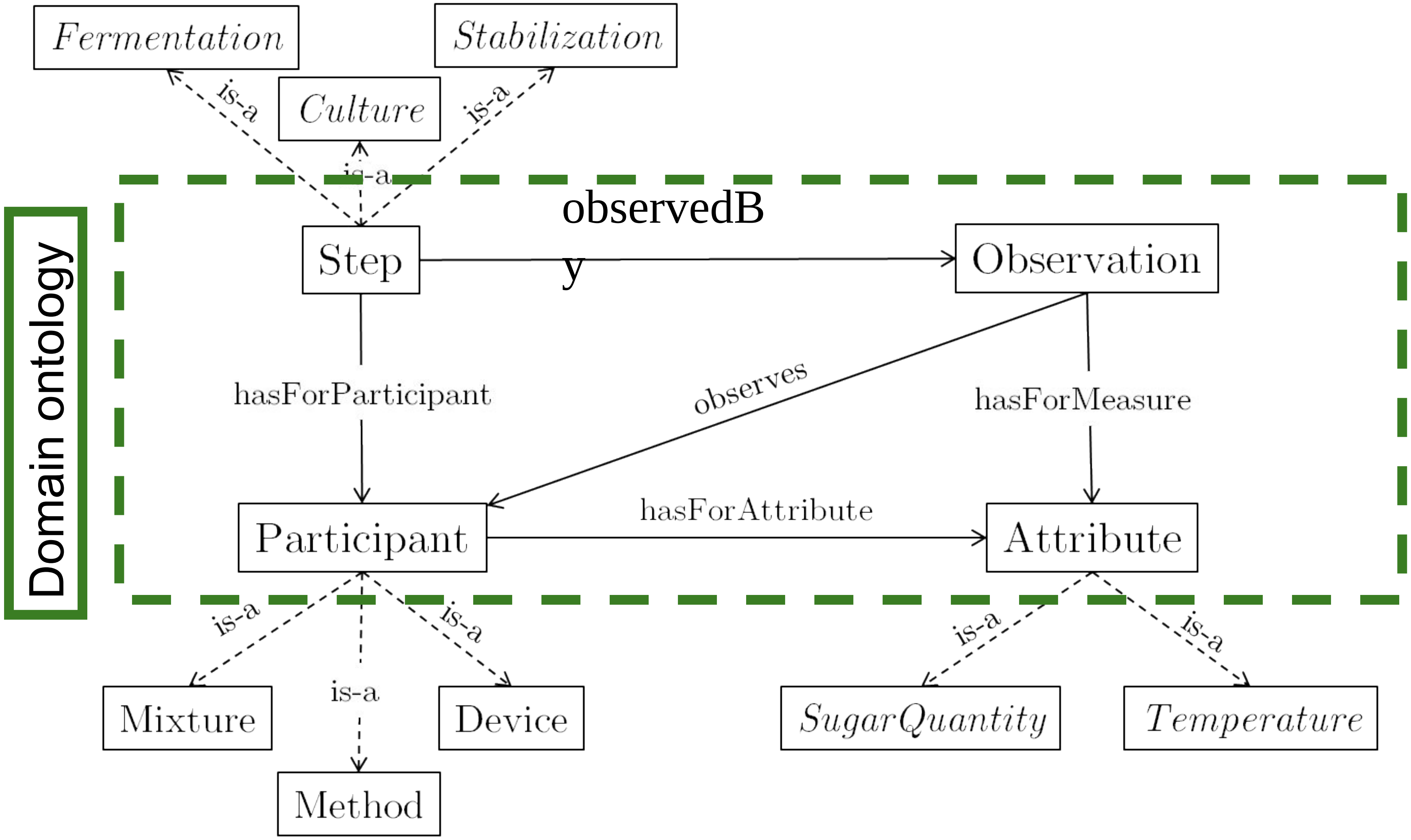
@ AgroParisTech :

- necessity of reasoning about transformation processes
- already an ontology structuring this domain

- PRMs to reason on uncertainty in transformation processes

- **Similarity** between PRMs and Ontologies -> mapping

The Process and Observation Ontology [Ibanescu16]



Modeling Cheese fabrication Process

QUESTION: *Which attributes can define the final product? How?*

*Thermophile lactic bacteria

Type of milk

Temperature

Starters*

. . .

Quality

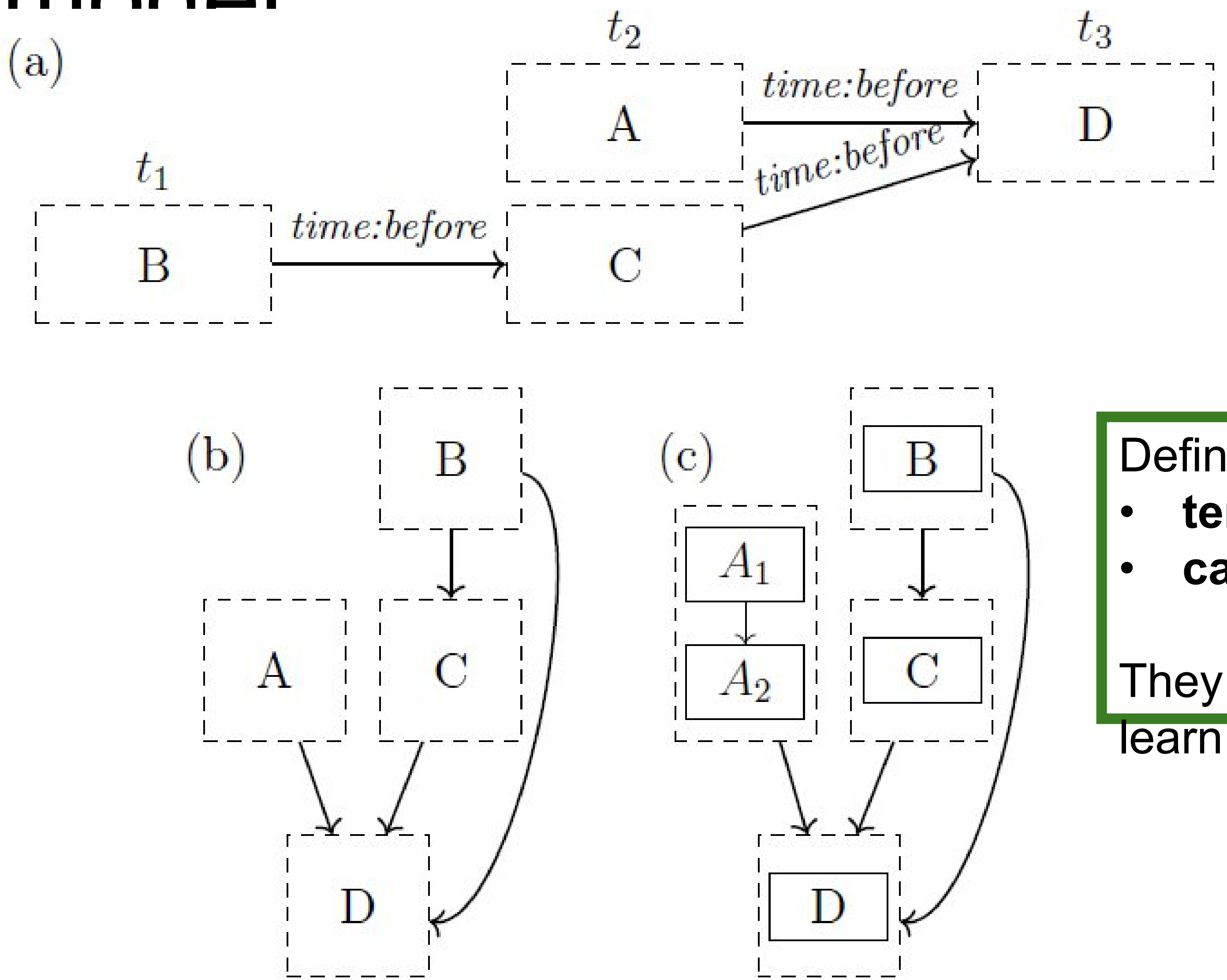
Taste

Solidity

Odour

TrueFood project : to what extent the characteristics of some hard cooked cheese is affected by the use of various composition of different milks and by the use of different technological conditions.

Mapping ontology and PRM : the stack model



Defined by **two constraints**:

- **temporal** (given by the ontology)
- **causal** (given by the expert)

They define the ordering with which we learn

- M. Munch, J. Dibie, P-H. Wuillemin, C. Manfredotti. Towards Interactive Causal Relation Discovery Driven by an Ontology. FLAIRS Conference 2019: 504-508
- M. Munch, J. Dibie-Barthélemy, P-H. Wuillemin, C. Manfredotti. Interactive Causal Discovery in Knowledge Graphs. PROFILES/SEMEX@ISWC 2019: 78-93

Cheese fabrication modeling

Construction of the **relational schema** with the expert

Step in the vat

Type of milk Temperature Starters

Hardening time Cloting

time

Ripening

5 different measures of concentration

Mastication

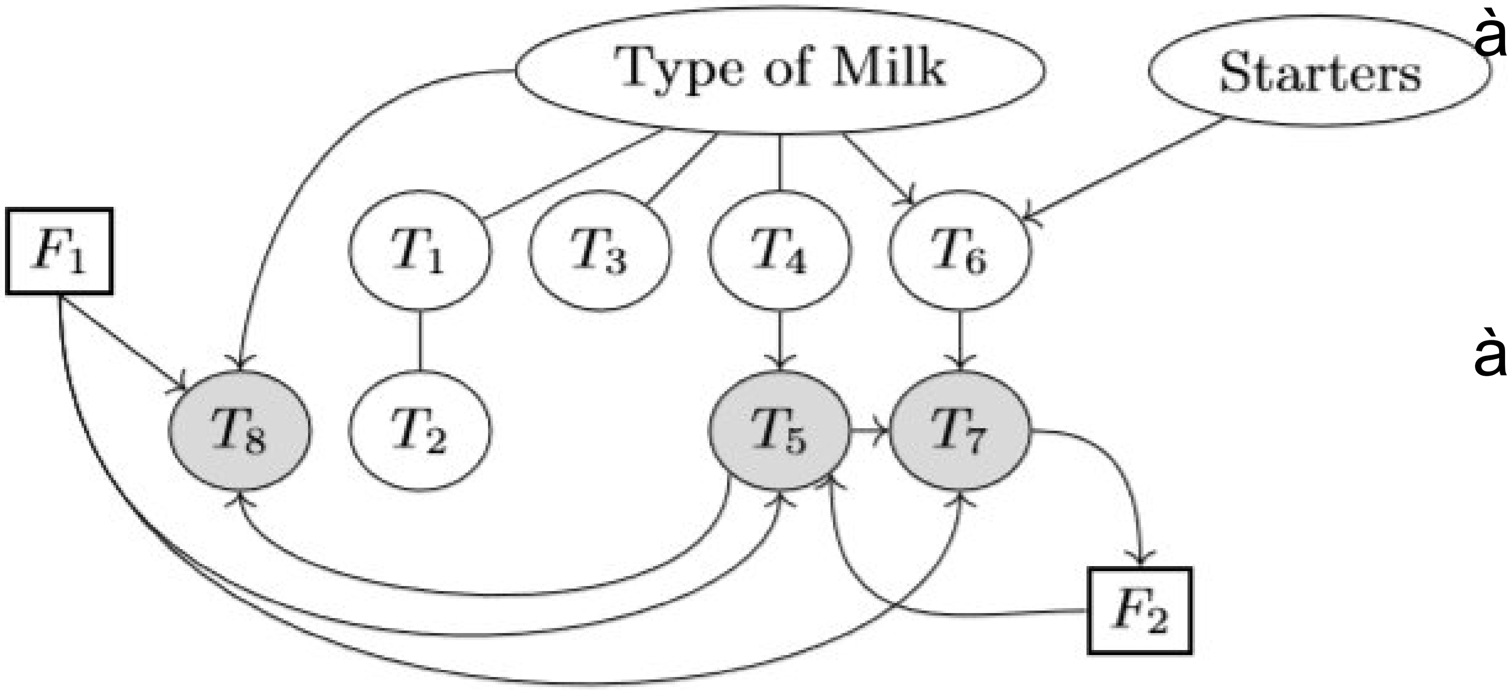
39 different scores

à 9 Texture attributes
à 30 Flavor attributes

Interpretation

Excerpt of the Essential Graph:

- T texture attributes
- F groups of flavor attributes

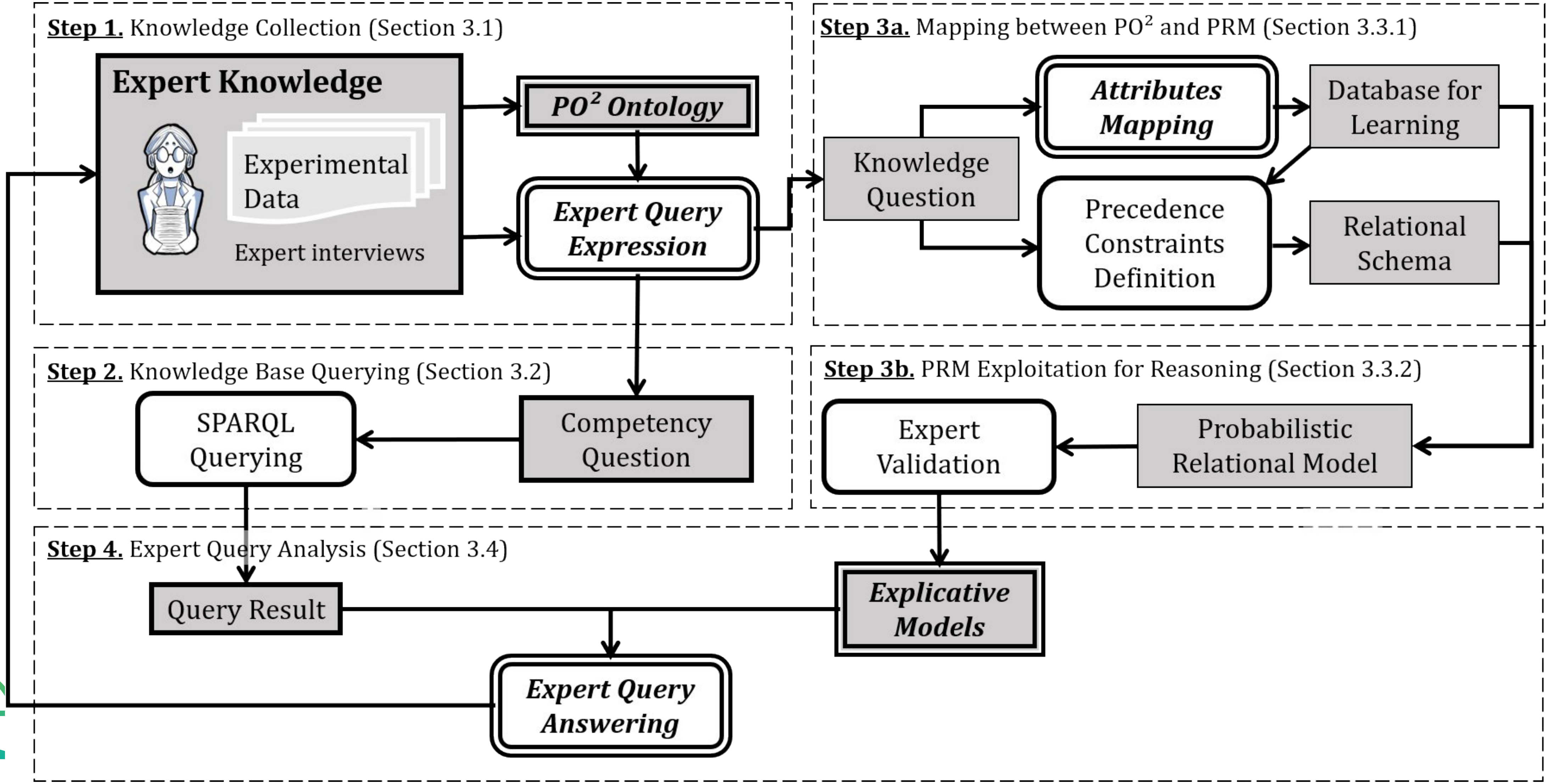


à The type of milk mostly explain all **texture attributes**

à The **flavor attributes** are hardly explained by the control parameters

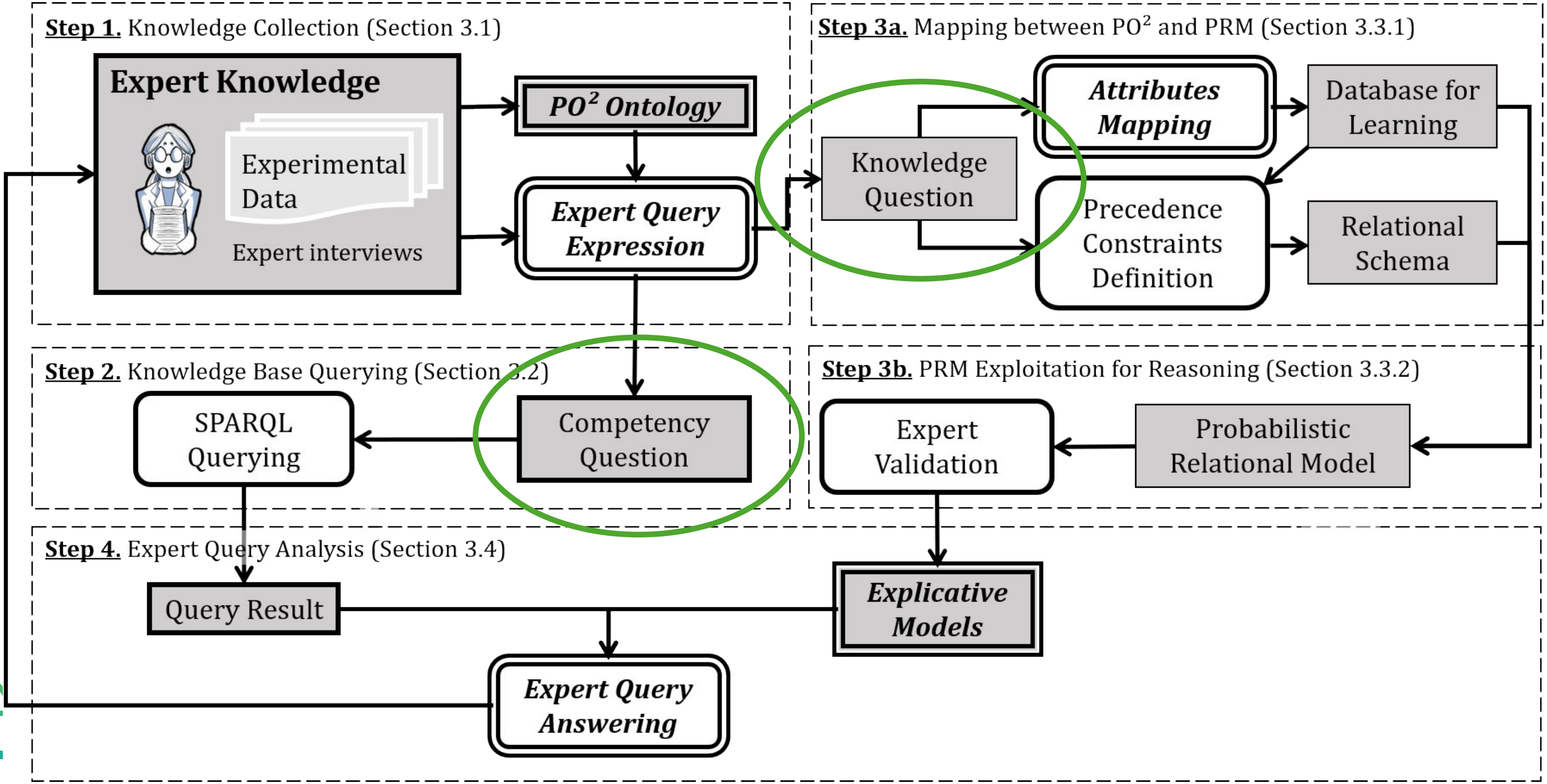
M.Munch, P-H. Wuillemin, J. Dibie, C. Manfredotti, T. Allard, S. Buchin, E. Guichard. Identifying Control Parameters in Cheese Fabrication Process Using Precedence Constraints. DS 2018: 421-434

Mapping ontology and PRM : POND

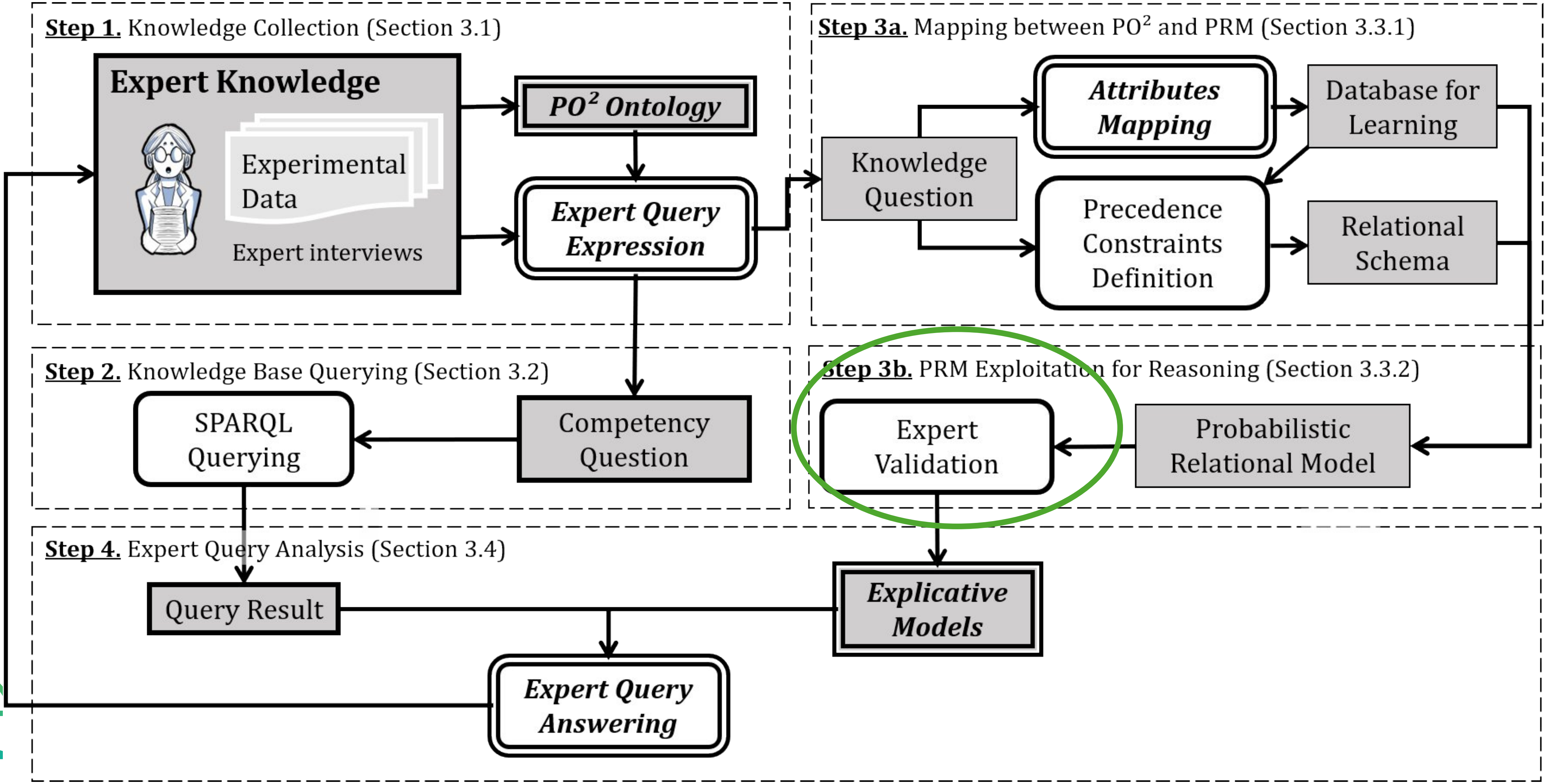


MIA
PARIS-SACLAY
EKINOCS

Mapping ontology and PRM : POND



Mapping ontology and PRM : POND



MIA
PARIS-SACLAY
EKINOCS

Food packaging biocomposite manufacturing

- PHBV : bacterial bio-polymer, biodegradable, expensive

- **Idea** : Mix it with lignocellulosic fillers (LF)

LF = Dry fractions of organic residues

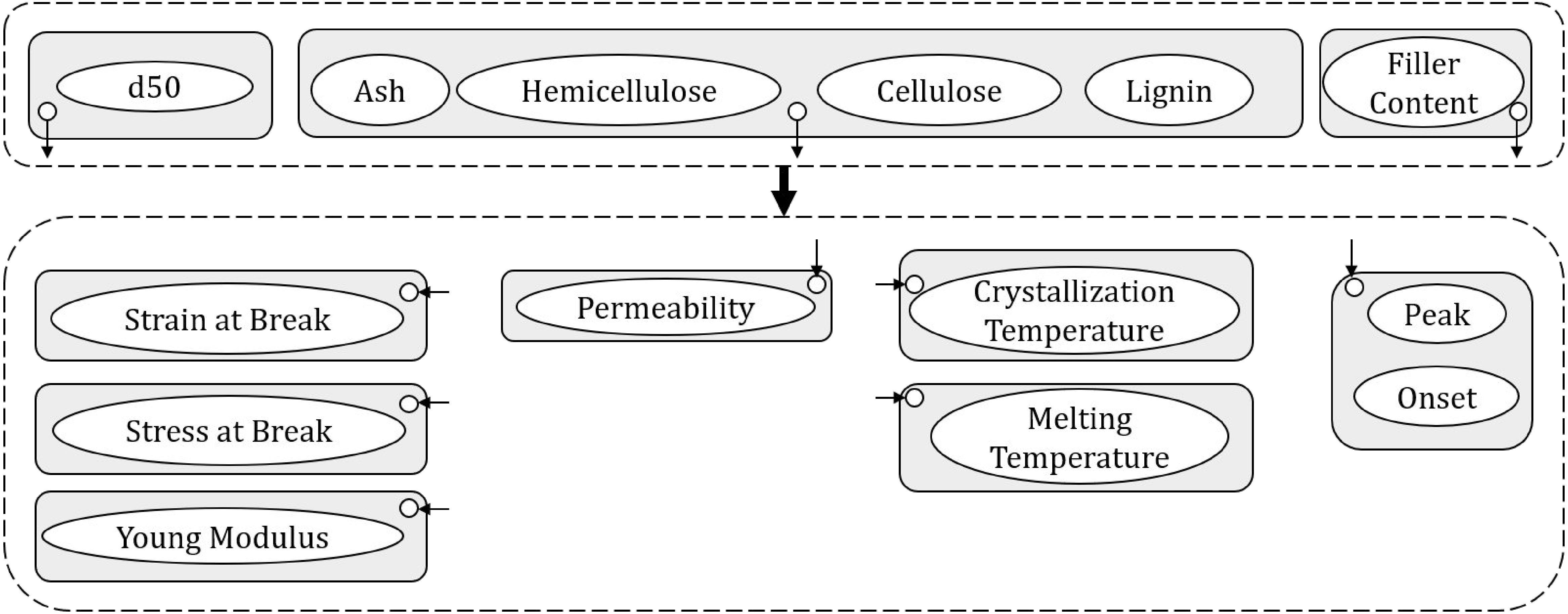
➢ It reduces the overall cost

➢ But modulates the functional properties

Question : find the right compromise between the maximum acceptable filler content (wrt overall product), filler size and resulting properties

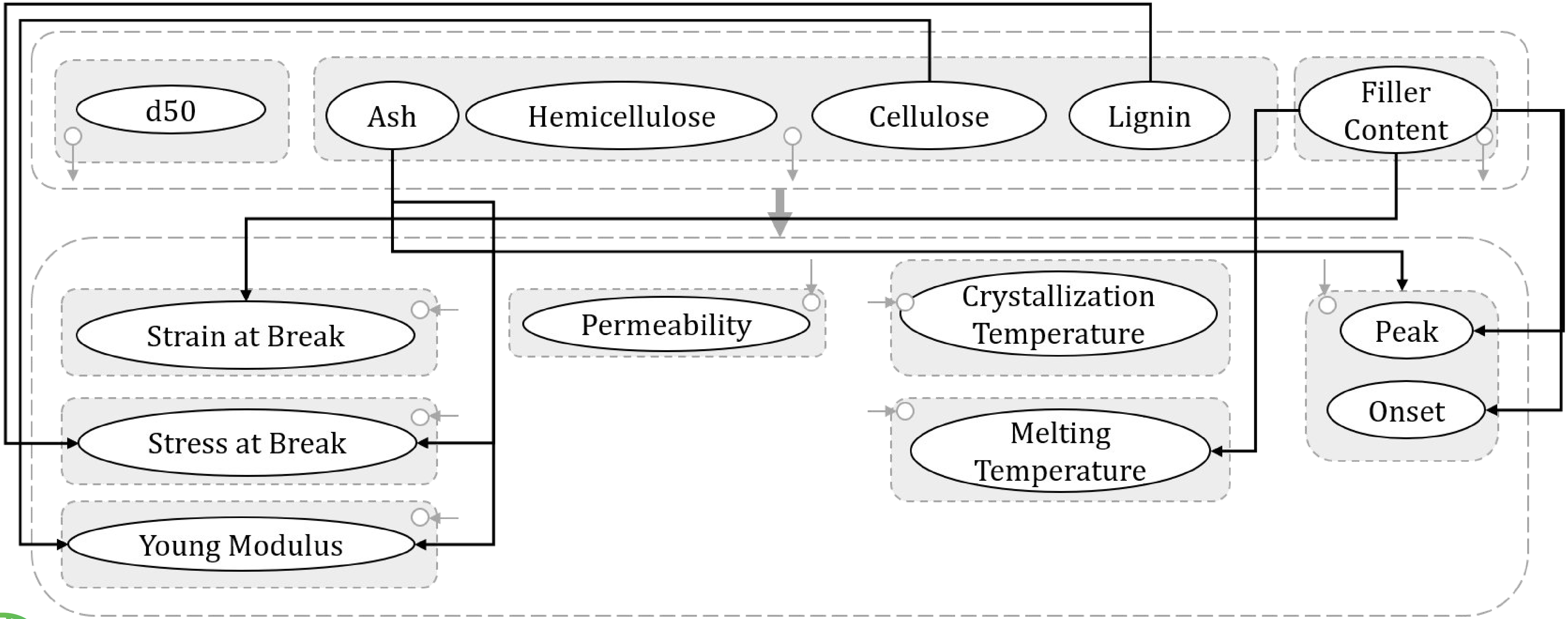
MIA PARIS-SACLAY EKINOCS

Mapping ontology and PRM : POND



QUESTION: *Which parameters explain the thermal degradation temperatures ?*

Mapping ontology and PRM : POND



Mapping ontology and PRM : POND

Filler Content	Peak Temperature]0.72; 0.95]	]0.95; 1]	]1; 1.16]
]2; 4]	0.09	**0.82***	0.09
]4; 11]	0.38	**0.52***	0.1
]11; 21]	**0.91***	0.09	0
]21; 50]	**0.6***	0.2	0.2

Conditional Probability Table showing the influence of the Filler Content over the Peak Temperature distribution. * shows the maximum likelihood.

Mapping ontology and PRM: remarks

- Motivations :
  - Need to reason with uncertainty in transformation processes
  - Similarity between ontologies and PRMS
- The experts
  - give the ontology
  - structures the variables with his hypothesis
  - afterwards, critiques the model
- The approaches
  - Verify the expert’s hypothesis
  - Suggest new experiments
  - Able to do prediction

Mapping ontology and PRM: ideas for the future

- The expert's validation step could be investigated more
  - Exploring algorithms able to learn a network locally
  - Visualisation and human machine interaction
  - (automatic) Explanation, (also) to suggest new experiments

- The ontology could be used more
  - Answering competency questions to structure de domain giving more insights to the learning

- Transfer Learning, Data linking, Graph Matching

RecSys in the nutrition domain : substitutions

Substitutability score : takes into account associativity as well

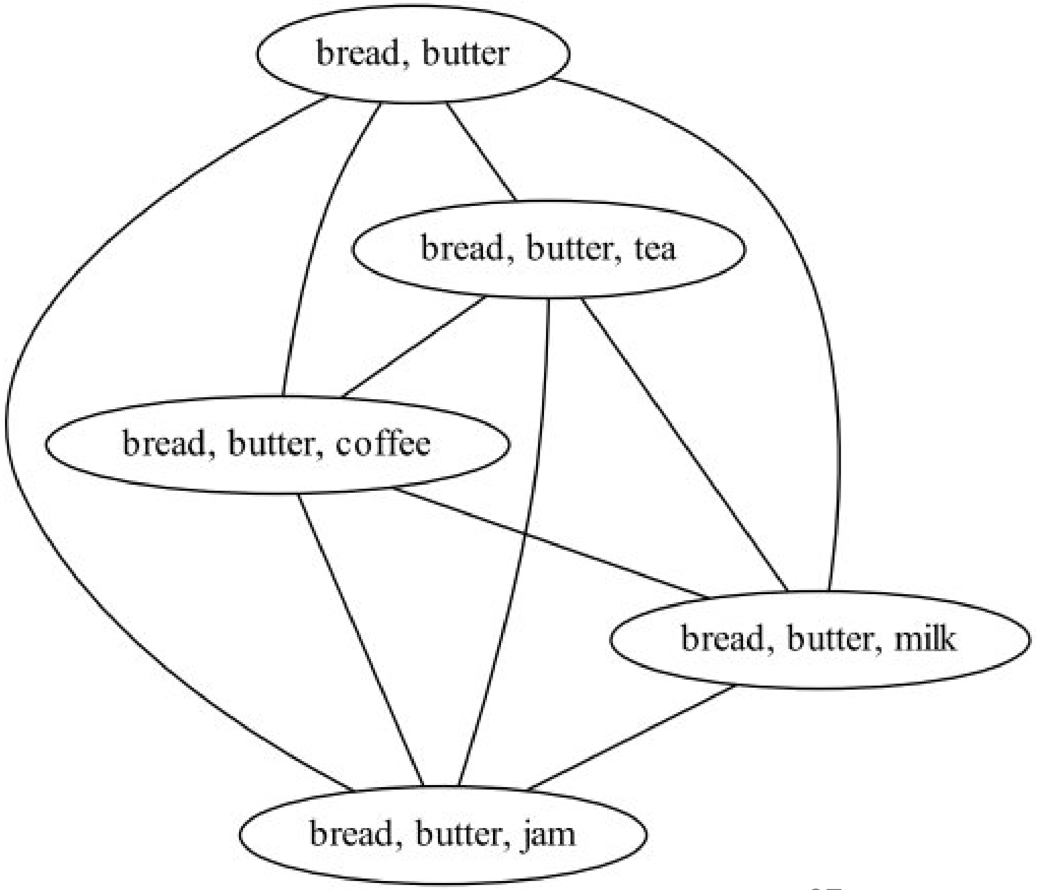
meal graph $G = (V,E)$:

- Meal context $\{bread, butter\}$
- substitutable set $\{coffee, tea, milk, jam, nothing\}$.

Discovering substitutable sets

=

mining maximal cliques in a graph



$$f(x,y) = \frac{|C_x \cap C_y|}{\ldots}$$