Towards decision support system for the agri-food sector using heterogeneous scientific data annotated with an ontology and Bayesian networks: a proof of concept applied to milk microfiltration

Cédric Baudrit, **Patrice Buche**, Nadine Leconte, Christophe Fernandez, Maëllis Belna, Geneviève Gésan-Guiziou
UMR IATE Montpellier, I2M Bordeaux, STLO Rennes
CATI DIISCICO

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INRAE

Motivation and issues

Decision tasks based on the **scientific literature in the agri-food sector** (here milk microfiltration) because:

- Key role in the sector (here dairy sector)
- Need to improve prediction of unit operation performances and optimize plant design (here microfiltration)
- Existing models only available on restricted operating conditions

Issues

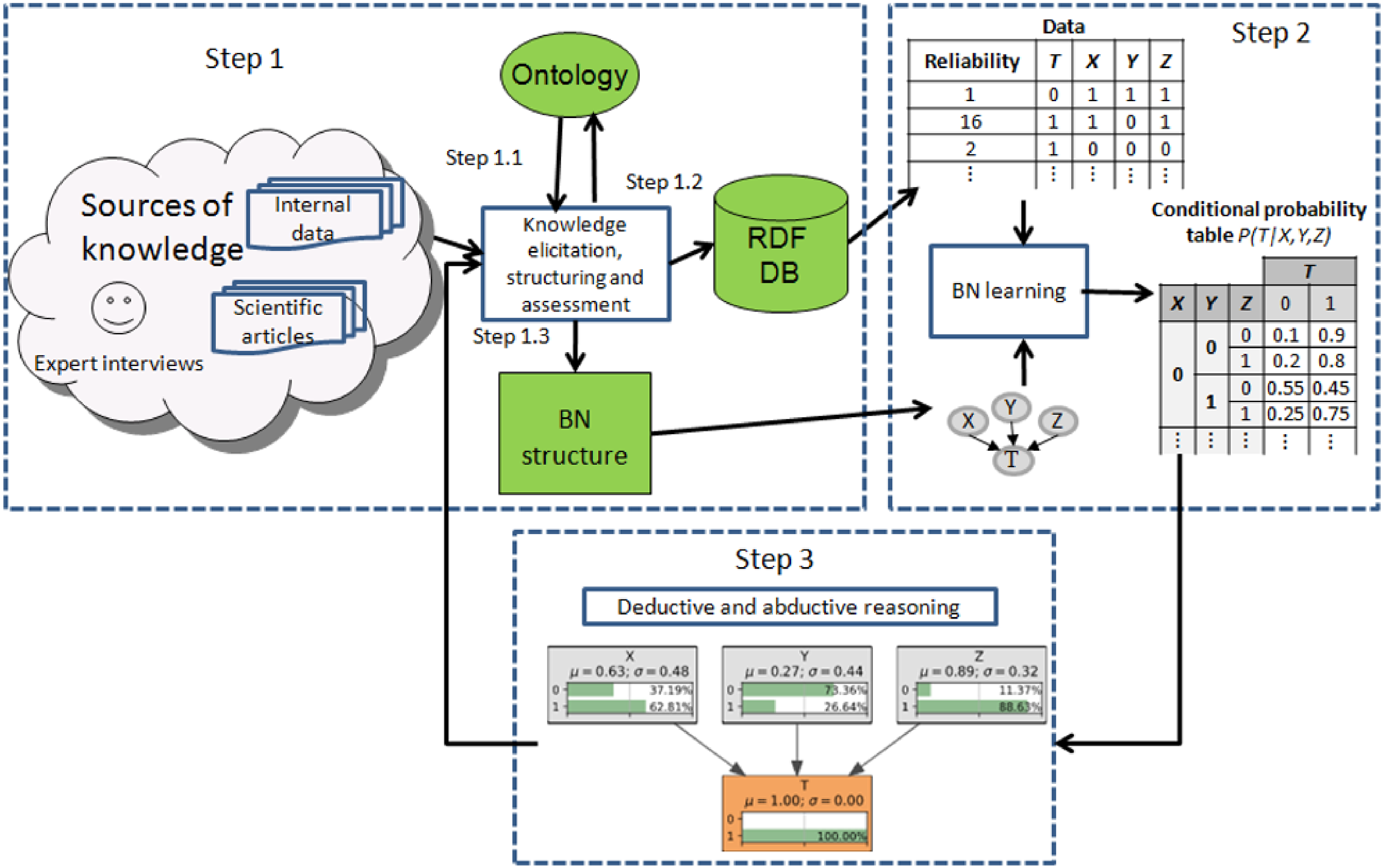
- Scientific data **heterogeneously** structured, available in textual format
- Data **uncertainty** due to processes involving biological material
- Data **reliability** due to the way experiments have been conducted

Proposal

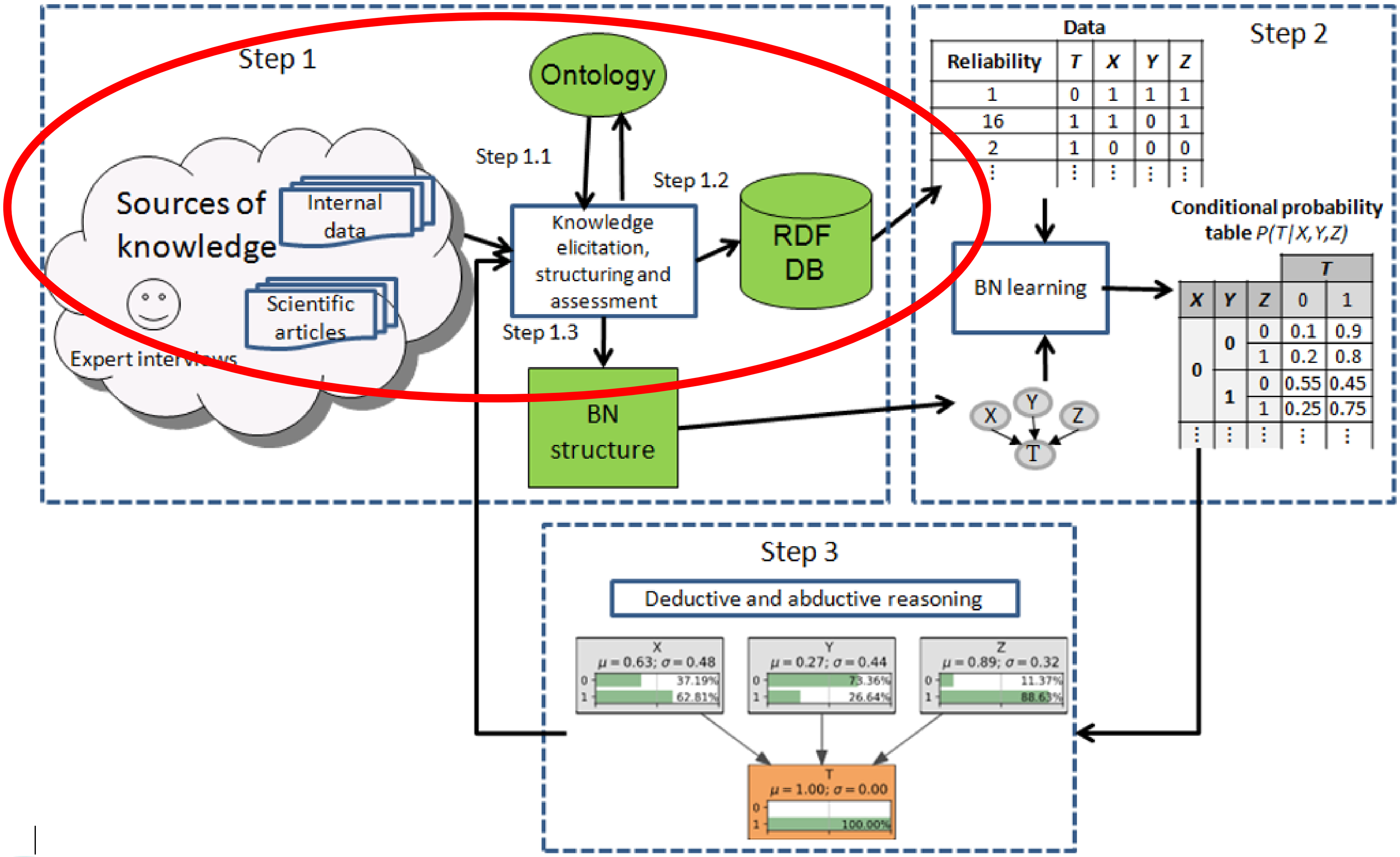
A **Decision Support System (DSS)** based on ontology and Bayesian networks which deals with the preceeding issues in order to:

- **structure** and **integrate** heterogeneous experimental data sources by using **ontologies**
- **provide open access** to annotated data
- **assess** data source reliability
- **learn Bayesian networks using uncertain data** taking into account data reliability

Decision Support System workflow



Decision Support System workflow



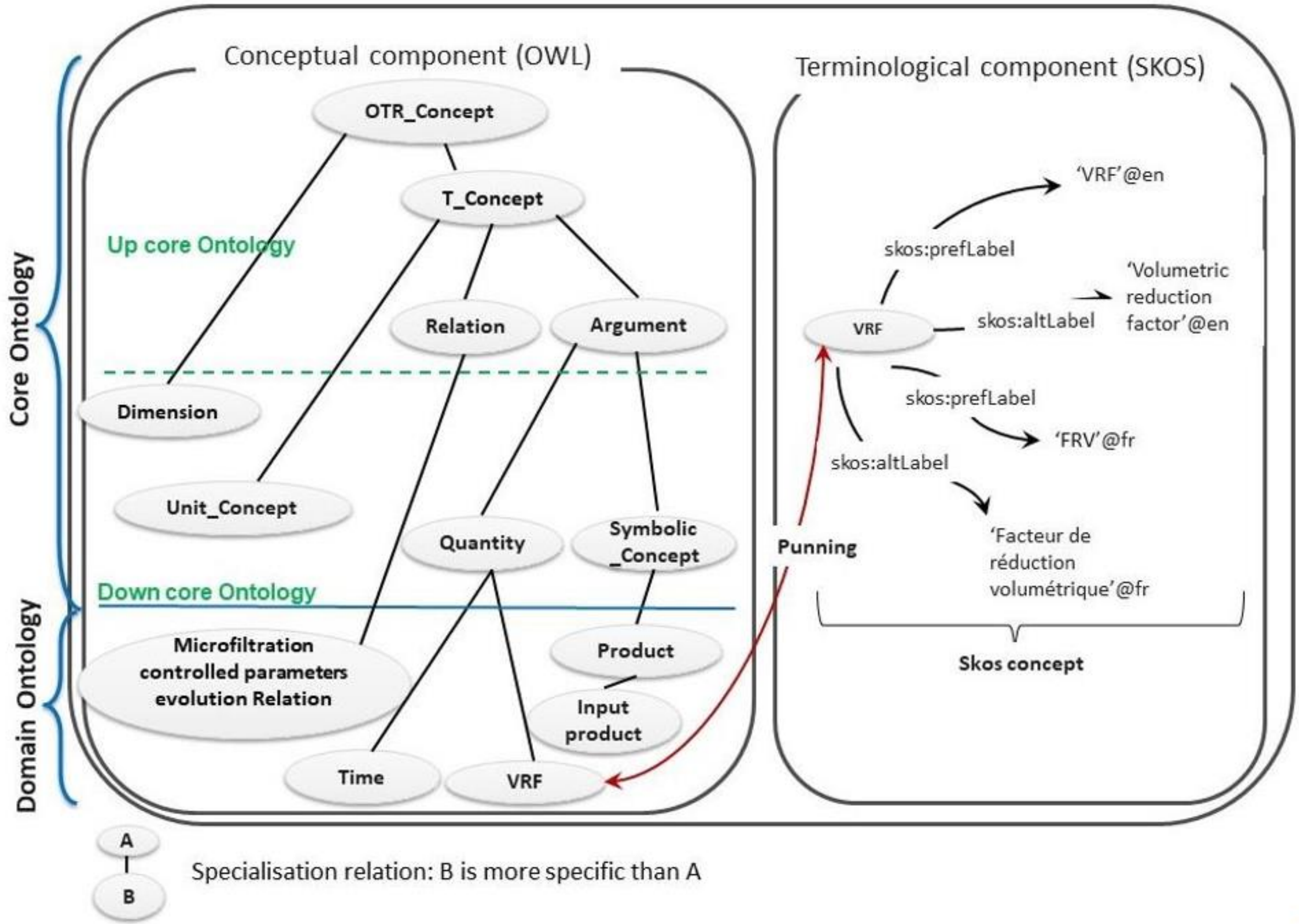
Key ideas to integrate data

Representing knowledge using international standards to represent ontologies and data (semantic web languages as OWL/RDF)

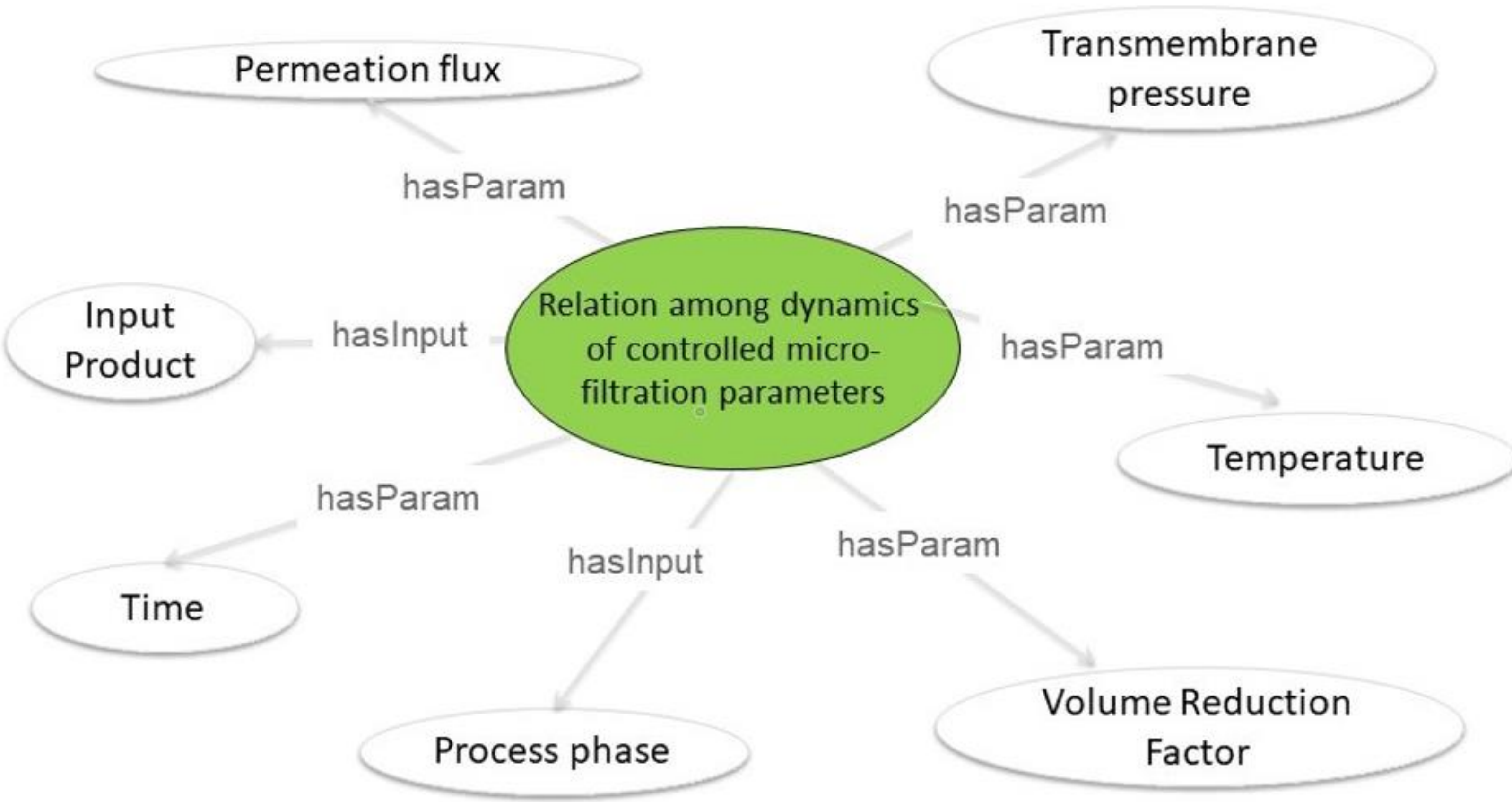
Using an ontological model to provide a reading grid for article annotation

Developing end-user oriented Web applications for ontology and data management based on semantic web languages

naRyQ Termino-Ontological Resource (TOR)



OWL Relation concept to model unit operation involved in a bioprocess



https://doi.org/10.15454/5MQMKG
Milk Microfiltration v1.5

Key ideas to integrate data

Representing knowledge using international standards to represent ontologies and data (semantic web languages as OWL/RDF)

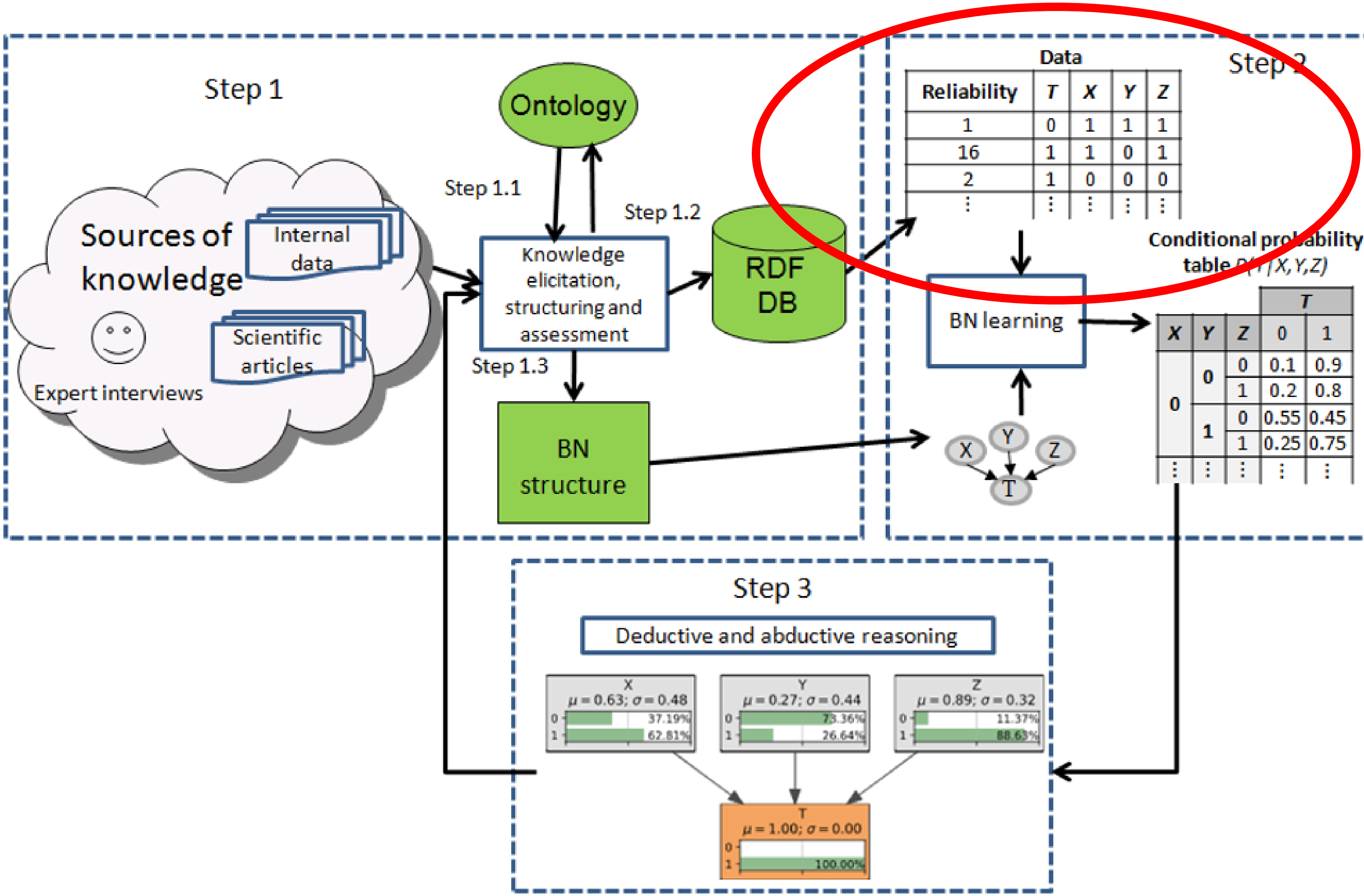
Using an ontological model to provide a reading grid for article annotation

Developing end-user oriented Web applications for ontology and data management based on semantic web languages

Relation concept instanciation in @Web import template CSV file

	P	S	V	X	Y	AC	AD	AE	AM	AO	AP	AT	AW	AY	BE	BI
1	Experience	Process step	Time_	Time_UNIT	Process phase	VRF_AVG	VRF_STD	VRF_UNIT	Temperatu	Tem	TMP_MIN	TMP_UNIT	Jp_AVG	Jp_UNIT	Crossflow ve	Crossf
370	1	5	90	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	71,8369131	kPa	58,9	l.h-1.m-2	6,9	m.s-1
371	1	5	95	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	73,6860843	kPa	58,9	l.h-1.m-2	6,9	m.s-1
372	1	5	100	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	74,7661331	kPa	58,9	l.h-1.m-2	6,9	m.s-1
373	1	5	105	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	75,5516231	kPa	58,9	l.h-1.m-2	6,9	m.s-1
374	1	5	110	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	76,0425544	kPa	58,9	l.h-1.m-2	6,9	m.s-1
375	1	5	115	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	76,9262307	kPa	58,9	l.h-1.m-2	6,9	m.s-1
376	1	5	120	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	78,6935832	kPa	58,9	l.h-1.m-2	6,9	m.s-1
377	1	5	125	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	79,4790732	kPa	58,9	l.h-1.m-2	6,9	m.s-1
378	1	5	130	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	78,6935832	kPa	58,9	l.h-1.m-2	6,9	m.s-1
379	1	5	135	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	80,559122	kPa	58,9	l.h-1.m-2	6,9	m.s-1
380	1	5	140	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	81,4427983	kPa	58,9	l.h-1.m-2	6,9	m.s-1
381	1	5	145	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	83,1119646	kPa	58,9	l.h-1.m-2	6,9	m.s-1
382	1	5	150	min	Constant phase (constant VRF)	2,5	0,1	One	50,1	°C	82,8174058	kPa	58,9	l.h-1.m-2	6,9	m.s-1

Decision Support System workflow



Key ideas to assess data source reliability

Defining meta data to evaluate (milk microfiltration) data source reliability

Using a model to merge metadata quotations

Developing end-user oriented Web applications

Meta data quotations (milk microfiltration)

@Web | Ontology | Documents | Query

Login :
Password :
Sign in / Sign up

- BIOREFINERY - V 95
- DURUM_WHEAT - V 44
- FOOD_COMPO - V 2
- FOOD_TEXTURE - V 56
- MICROFILTRATION - V 70
  - Download
  - Generate template import
  - View
  - Explore
  - View judgements on criteria
- TRANSMAT - V 107
- VALORCARN - V 19
- Unit Ontology

	unknown	not at all reliable	not at all or hardly	hardly reliable	hardly or average	average reliable	average or reliable	reliable	reliable or very	very reliable
MF-Number of repetitions - Compositional analysis										
> 3 or more										++
> 2								+		
> 1				-						
MF-Source Type										
> Book								+		
> Proceeding				-						
> Journal article superior IF quarter (in at least one area)										++
> Journal article second IF quarter (in at least one area)								+		
> Journal article third IF quarter (in at least one area)				-						
> Journal article inferior IF quarter (in at least one area)		--								
> Encyclopedia article						-+				
> Patent				-						
> Report				-						
> Thesis						-+				
> Science popularization article				-						

See Buche et al. 2021 (https://doi.org/10.1016/j.dib.2021.107063) for a complete description of the 8 meta data

Key ideas to assess data source reliability

Defining meta data to evaluate (milk microfiltration) data source reliability

Using a model to merge metadata quotations

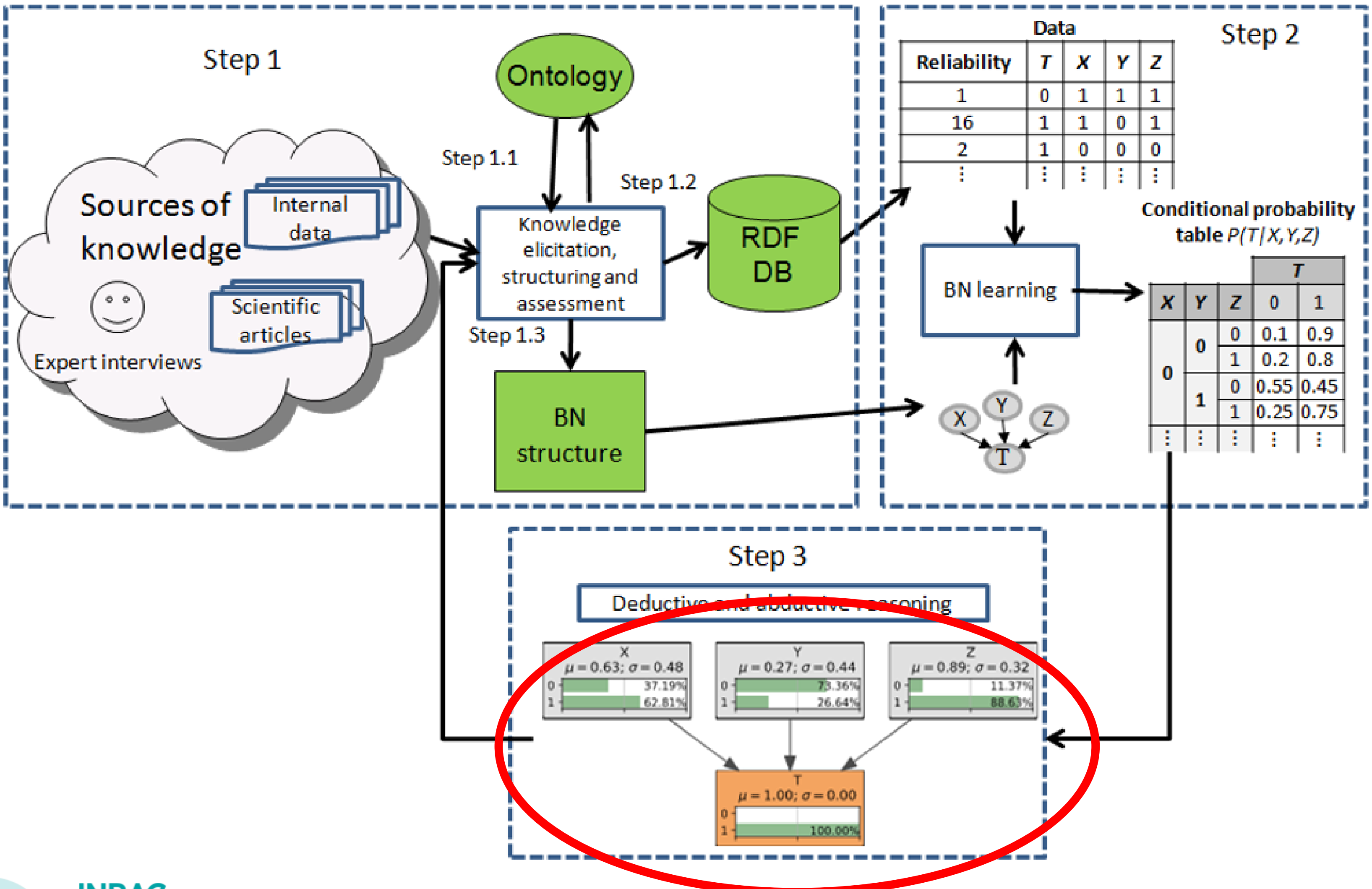
Developing end-user oriented Web applications

Model with belief functions

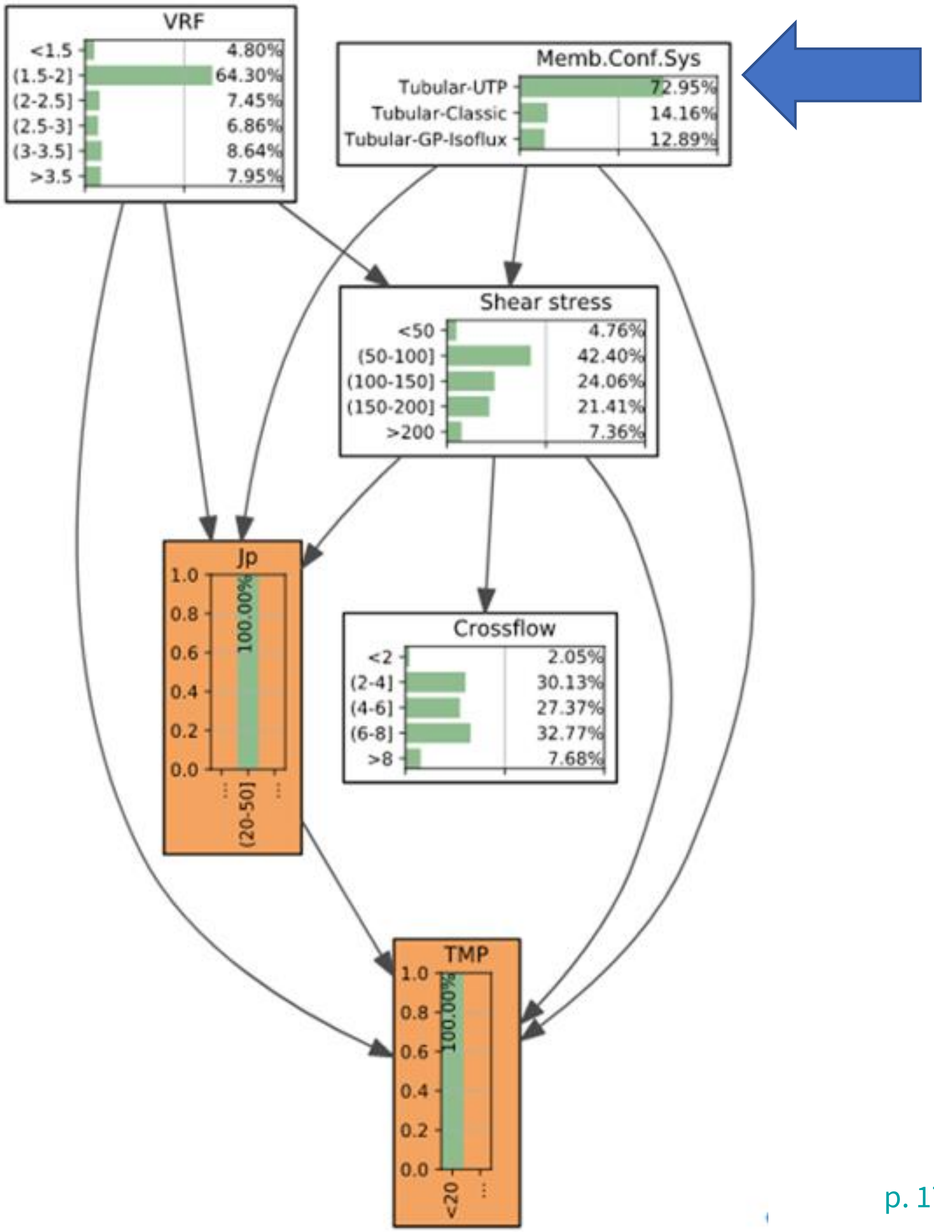
- Inputs: metadata associated with the data source and expert quotations associated with metadata values
- Model with belief functions: merging quotations associated with metadata for a given data source (able to model ignorance, identify and manage conflicts)
- Output: estimated reliability $\left[\underline{E_{d_i}}, \overline{E_{d_i}}\right]$
- Usefulness: ranking of sources based on interval comparisons

d_i	(#cit vs age)	repetitions	source type	$\left[\underline{\mathbb{E}_{d_i}}, \overline{\mathbb{E}_{d_i}}\right]$
d_1	-	No	Journal paper	[1.40,4.66]
d_2	(79,3)	Yes	Journal paper	[4.67,4.97]
d_3	-	No	Technical sheet	[1.05,1.40]

Decision Support System workflow



Decision support using Bayesian network learned with data



Including data reliability in model's predictive accuracy based on leave-one-out cross-validation for
prediction of Jp given shear stress, VRF constraints and the type of membrane used **increases the overall accuracy from 42% to 70%**

Conclusion-perspectives

Conclusion

- Proposition of a generic ontological model (and associated tool) to guide manual annotation of scientific papers
- Reuse of annotated scientific data available in a SPARQL endpoint to learn a BN model for DSS
- Reliability assessment of data sources enhances model's predictive accuracy
- Preliminary version of a DSS dedicated to Milk Microfiltration taking into account large operating conditions

Perspectives

- Speeding up annotation using NLP guided by ontology (cf Martin's speech)
- impacts assessment of process parameters on milk quality parameters

Questions ?

References

Buche, P., Dervaux, S., Leconte, N., Belna, M., Granger-Delacroix, M., Garnier-Lambrouin, F., Gesan-Guiziou, G. (2021). Milk microfiltration process dataset annotated from a collection of scientific papers. Data in Brief, 36 doi:10.1016/j.dib.2021.107063

Baudrit Cédric, Buche, P., Leconte, N., Belna, M., Fernandez C., Gesan-Guiziou, G. (2021). Decision support tool for the agri-food sector using data annotated by ontology and Bayesian network: a proof of concept applied to milk microfiltration (submitted to IJAEIS)

patrice.buche@inrae.fr

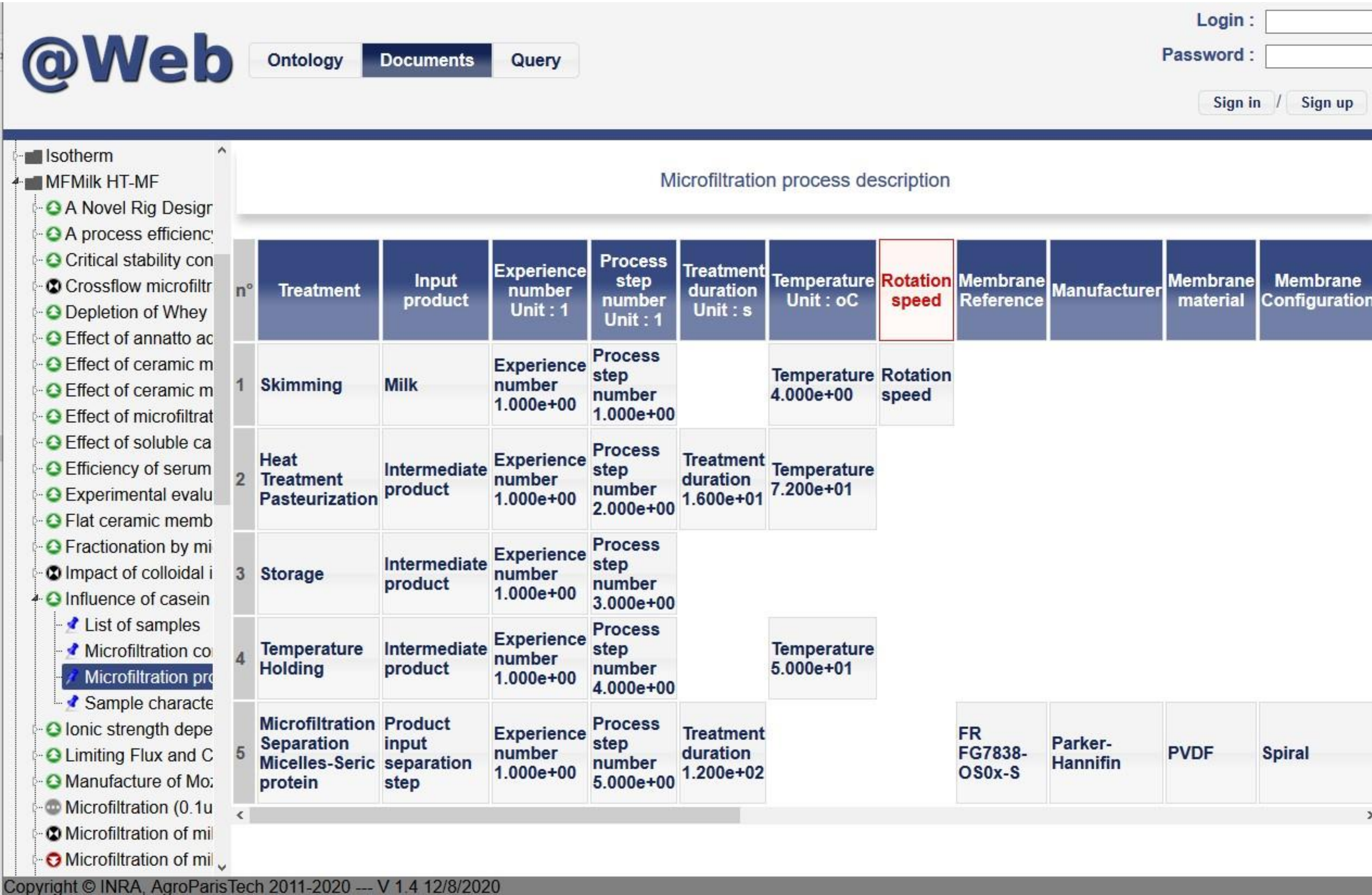
Key ideas to integrate data

Representing knowledge using international standards to represent ontologies and data (semantic web languages as OWL/RDF)

Using an ontological model to provide a reading grid for article annotation

Developing end-user oriented Web applications for ontology and data management based on semantic web languages

Relation concept instanciation in @Web web app



@Web

Ontology | Documents | Query

Login :

Password :

Sign in / Sign up

- Isotherm
- MFMilk HT-MF
  - A Novel Rig Desigr
  - A process efficienc
  - Critical stability con
  - Crossflow microfiltr
  - Depletion of Whey
  - Effect of annatto ac
  - Effect of ceramic m
  - Effect of ceramic m
  - Effect of microfiltrat
  - Effect of soluble ca
  - Efficiency of serum
  - Experimental evalu
  - Flat ceramic memb
  - Fractionation by mi
  - Impact of colloidal i
  - Influence of casein
    - List of samples
    - Microfiltration co
    - Microfiltration pro
    - Sample characte
  - Ionic strength depe
  - Limiting Flux and C
  - Manufacture of Mo:
  - Microfiltration (0.1u
  - Microfiltration of mi
  - Microfiltration of mi

Microfiltration process description

n°	Treatment	Input product	Experience number Unit : 1	Process step number Unit : 1	Treatment duration Unit : s	Temperature Unit : oC	Rotation speed	Membrane Reference	Manufacturer	Membrane material	Membrane Configuration
1	Skimming	Milk	Experience number 1.000e+00	Process step number 1.000e+00		Temperature 4.000e+00	Rotation speed				
2	Heat Treatment Pasteurization	Intermediate product	Experience number 1.000e+00	Process step number 2.000e+00	Treatment duration 1.600e+01	Temperature 7.200e+01					
3	Storage	Intermediate product	Experience number 1.000e+00	Process step number 3.000e+00							
4	Temperature Holding	Intermediate product	Experience number 1.000e+00	Process step number 4.000e+00		Temperature 5.000e+01					
5	Microfiltration Separation Micelles-Seric protein	Product input separation step	Experience number 1.000e+00	Process step number 5.000e+00	Treatment duration 1.200e+02			FR FG7838-OS0x-S	Parker-Hannifin	PVDF	Spiral

Key ideas to assess data source reliability

Defining meta data to evaluate (milk microfiltration) data source reliability

Using a model to merge metadata quotations

Developing end-user oriented Web applications

Data source reliability assessment

Document's general information

Document's name :
Optimization of protein fractionation by skim milk microfiltration: Choice of ceramic membrane pore size and filtration temperature

Associate topic : MFMilk HT-MF

Associate ontology : MICROFILTRATION

Accepted Tables : 5

Rejected Tables : 0

Untreated Tables : 0

Authors :
Jorgensen, C. E., Abrahamsen, R. K., Rukke, E.-O., Johansen, A.-G., Schüller, R. B., Skeie, S. B.

Journal : Journal of Dairy Science

Year : 2016

Volume : 99

Issue : 2016P 6164-6179.

Identifier : http://doi.org/10.3168/jds.2016-11090

Document criteria values

Criterion MF-Number of replicates - Compositional analysis

MF-Number of replicates - Compositional analysis : 3 or more

Criterion MF-Source Type

MF-Source Type : Journal article first IF quartile (in at least one area)

Criterion MF-Assay type

MF-Assay type : As a function of time (<2h)

Criterion MF-Number of replicates - Microfiltration essays

MF-Number of replicates - Microfiltration essays : 2 or more

Criterion MF-Operating mode

MF-Operating mode : Batch (UTP/GP)

Criterion MF-Number of automatic parameter controls during MF

MF-Number of automatic parameter controls during MF : 4 or more

Criterion MF-Initial product state

MF-Initial product state : Liquid milk

Criterion MF-Protein analysis method

MF-Protein analysis method : Kjeldahl

Reliability model →

Reliability assessment document information

Reliability results

Low expectation : 4.7 ; High expectation : 4.99

Known criteria values rate : 100.0 %

Last assessment date (yyyy-mm-dd) : 2020-03-09