

# Méthodologie de mise en oeuvre de workflows scientifiques

Yuan Lin

yuan.lin@supagro.inra.fr  
UMR MISTEA  
INRA, SupAgro, Montpellier



# Contexte & Objectif

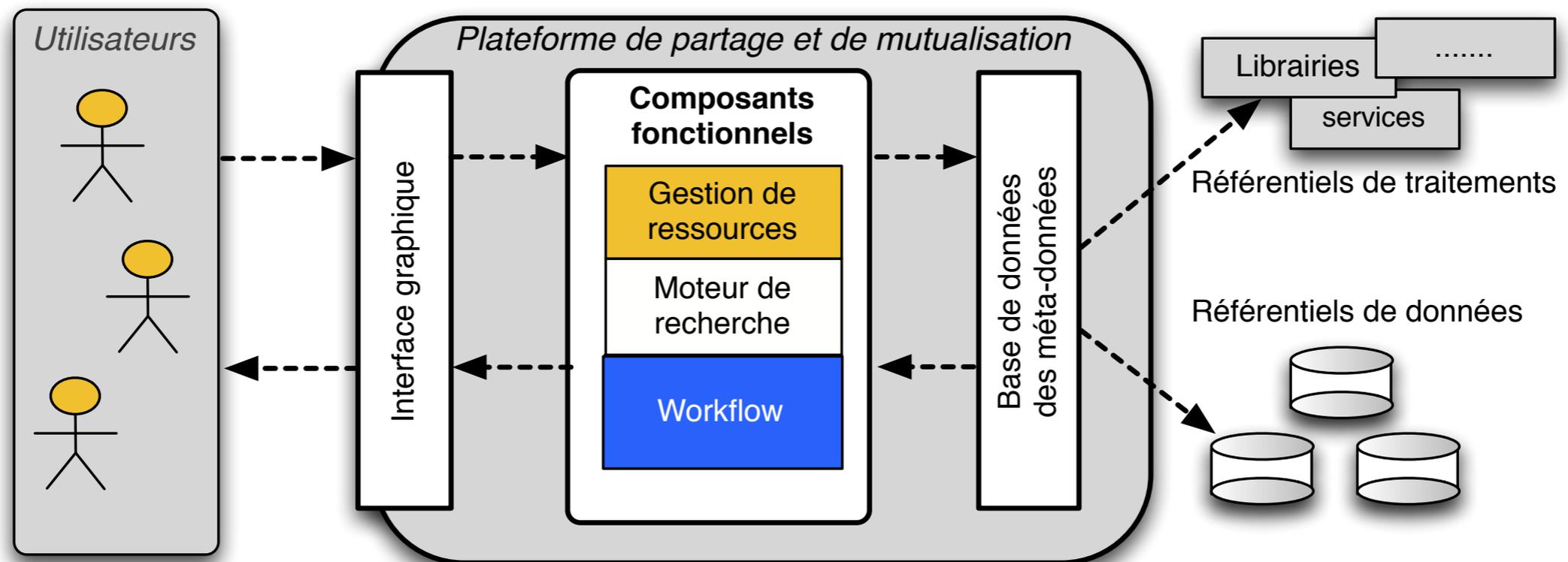
- Divers domaines scientifiques
- Ressources existantes disponibles
  - Partage et mutualisation
  - Connaissance / expertise
  - Hétérogénéité, complexité
- Processus de validation d'expérience
  - Automatisation, réutilisation

# Plan

- Vision générale
  - Différentes phases
- Perspectives + Conclusion

# Vision générale

Un environnement d'expérimentation de workflow intégré dans une plateforme de partage et de mutualisation



# MIDWeb



Geographic catalog

Language ▾ Login

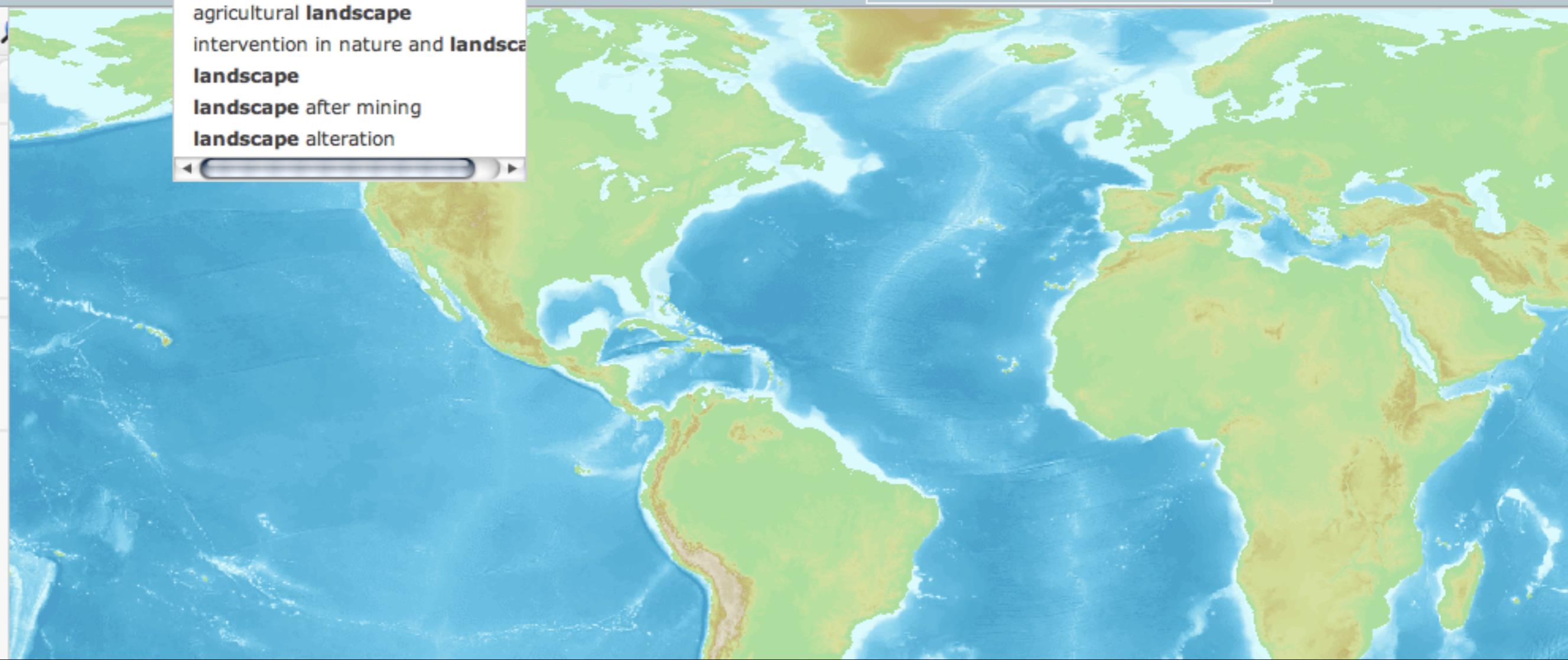
Advanced »

landscape

Search

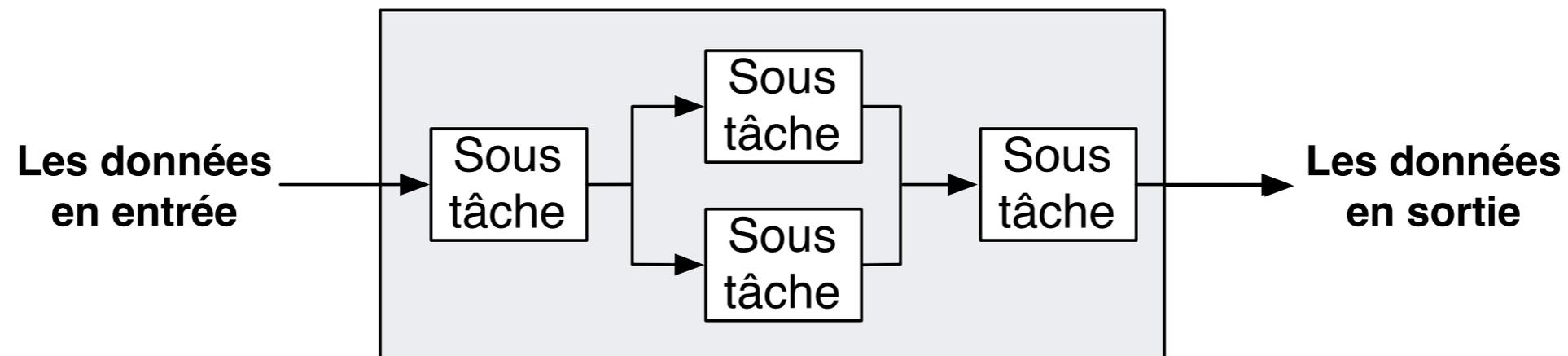


- agricultural **landscape**
- intervention in nature and **landscape**
- landscape**
- landscape** after mining
- landscape** alteration



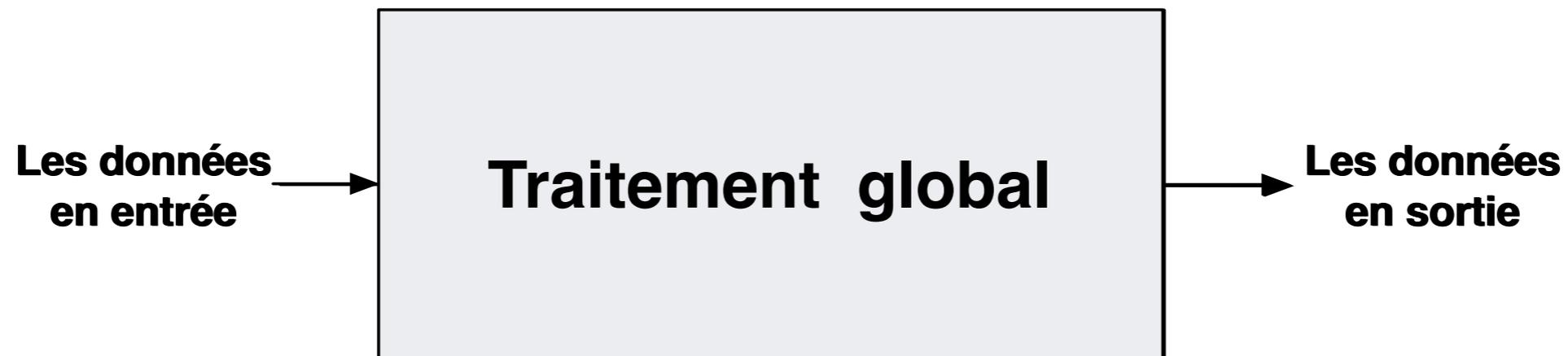
# Notions de workflow

- The automation of a business process, in whole or part, during which documents, informations or tasks are passed from one participant to another for action, according to a set of procedural rules. (WFMC)
- A scientific workflow attempts to capture a series of analytical steps, which describe the design process of computational experiments. [Adam 2008]
- Workflows are activities involving the coordinated execution of multiple tasks performed by different processing entities. [Rusinkiewicz 1995]



# Notions de workflow

- The automation of a business process, in whole or part, during which documents, informations or tasks are passed from one participant to another for action, according to a set of procedural rules. (WFMC)
- A scientific workflow attempts to capture a series of analytical steps, which describe the design process of computational experiments. [Adam 2008]
- Workflows are activities involving the coordinated execution of multiple tasks performed by different processing entities. [Rusinkiewicz 1995]



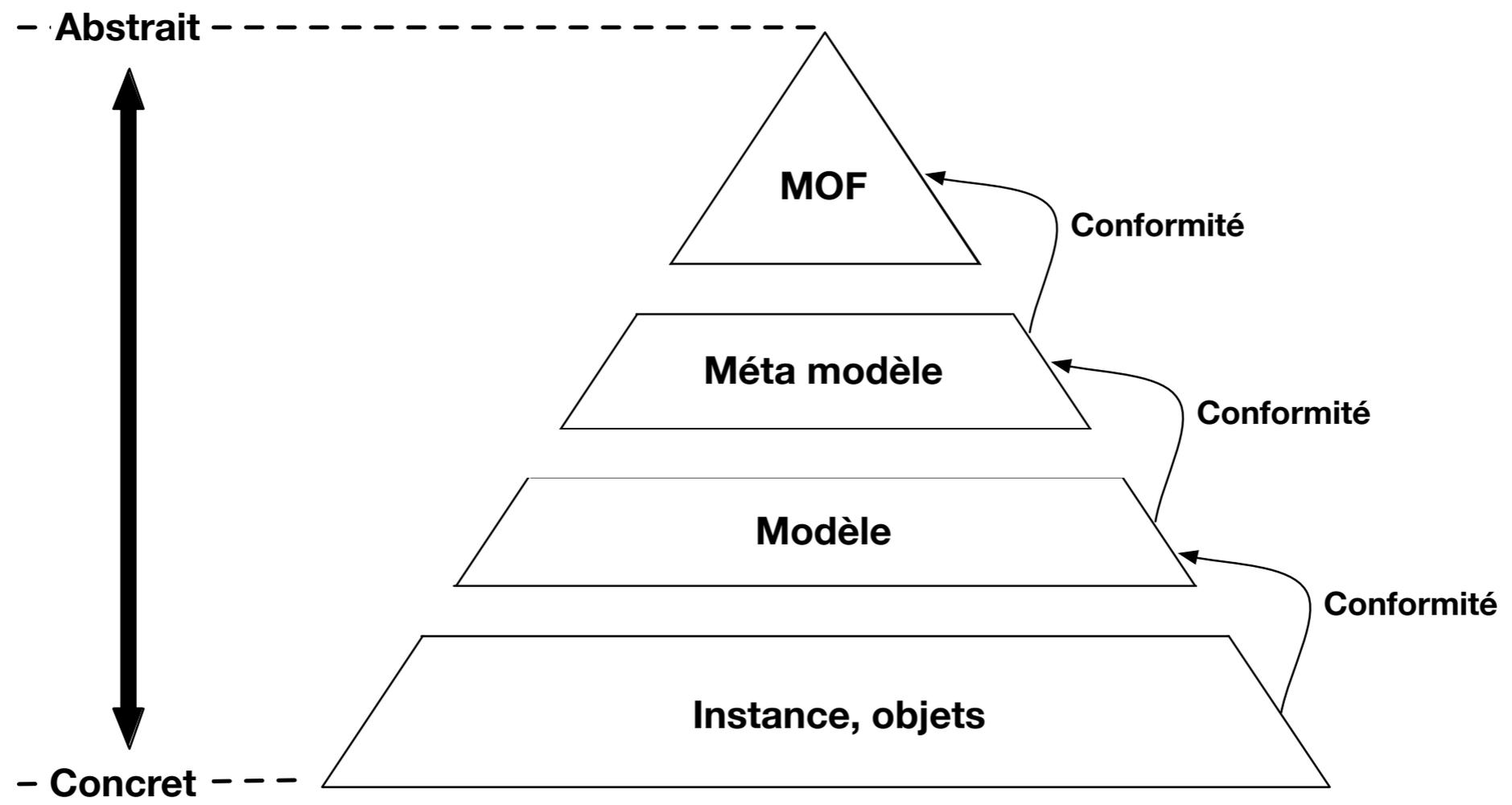
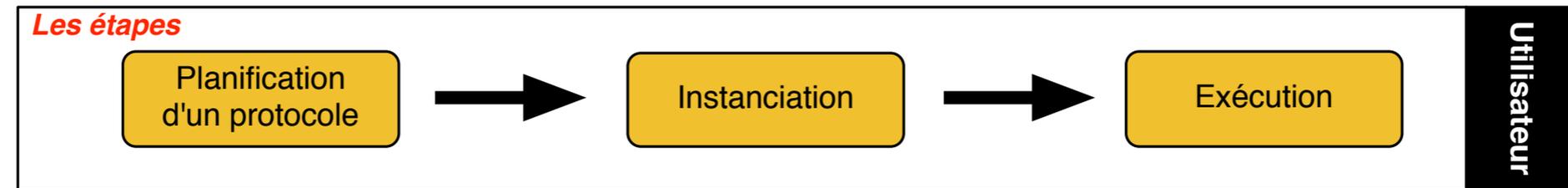
**Focus : Workflow Scientifique**

# Notions de workflow

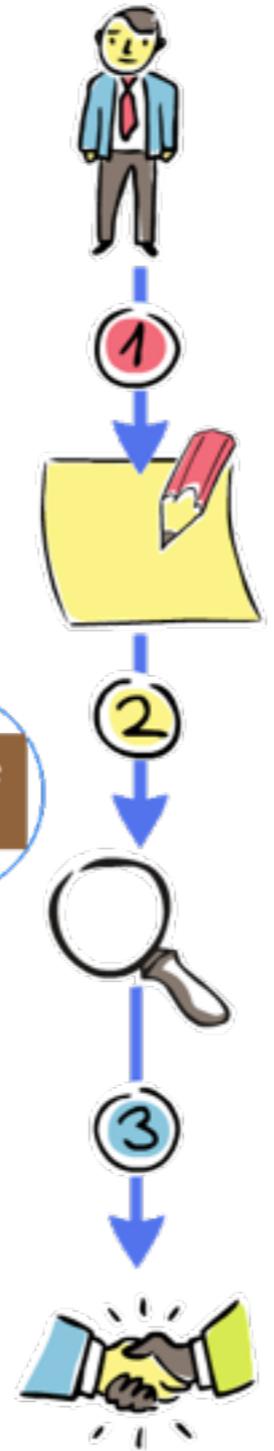
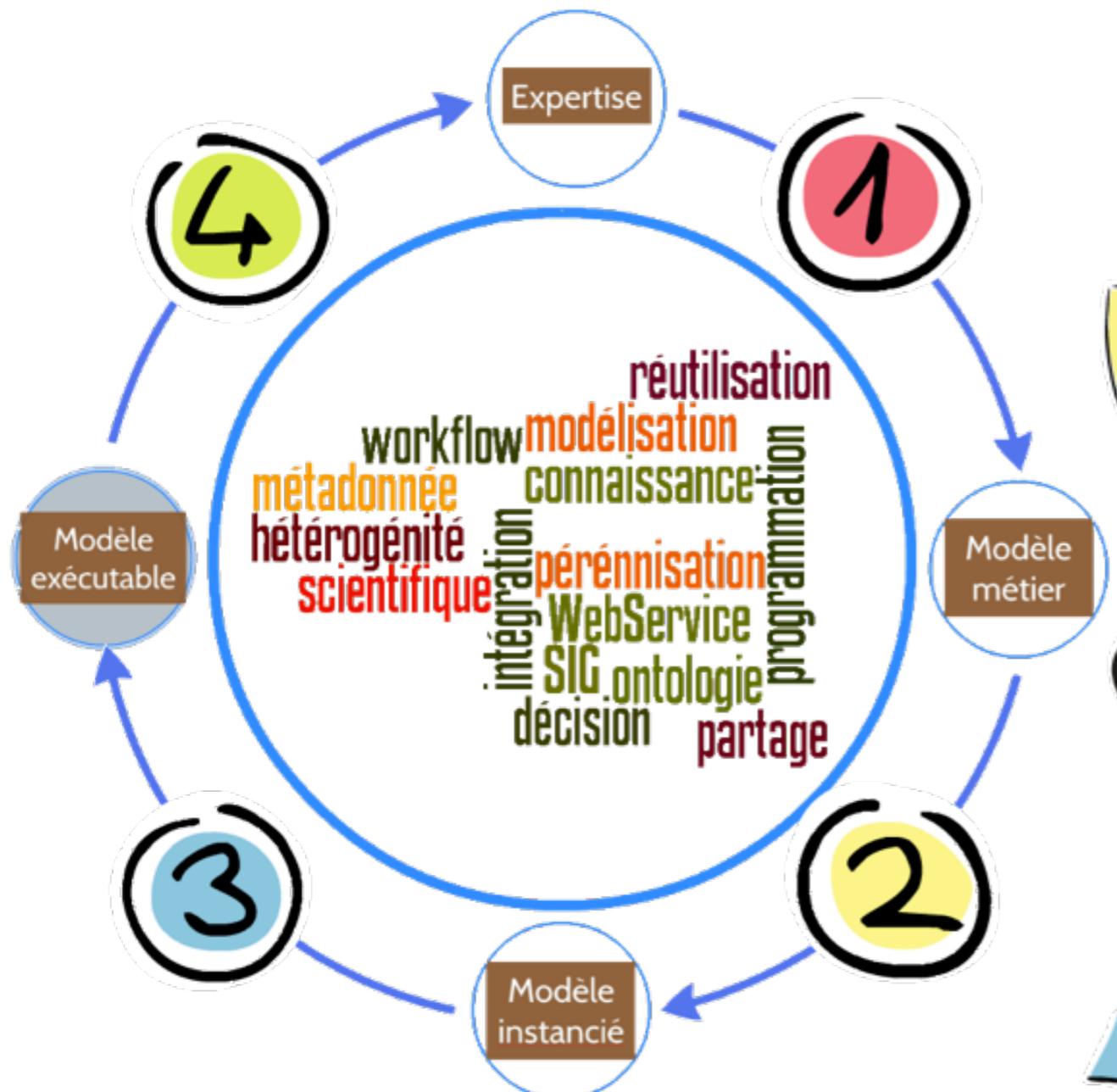
Workflow scientifique != Business process

Critères	Scientifique	Business
Objectif	Spécifique, propre à un individu ou une équipe limitée	Entreprise, organisation importante
Données	Volumineuses, complexes, souvent onéreuses	De faible volume, structures simples
Traitements	Nombreux, disponibles sous diverses versions, paramétrables, autonomes du point de vue exécution	Stéréotypés, propriétaires
Acteurs	Peu d'intervention	Interventions importantes
Exécution	Type <i>orchestration</i> , exécution centralisée, parallèles, pas à pas	Type <i>chorégraphie</i> , contrôle humain
Modèles	Relativement simples, data-driven	Complexes, contrôle drivers, control-driven
Environnement	Plateforme complète avec gestion de ressources, de workflows et d'expertise	Environnement spécifique pour les développeurs, standard pour les utilisateurs

# Vision usage

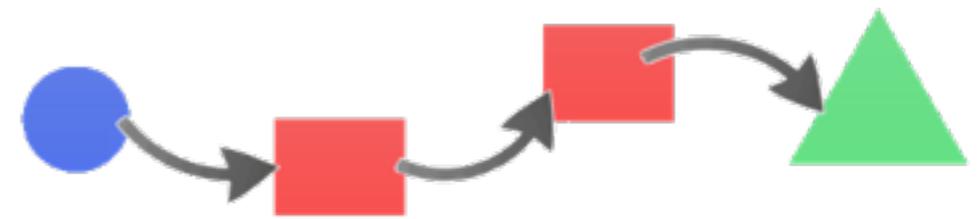


# Vision usage



Calculer la température moyenne du mois d'octobre à Montpellier

- Calculer la somme
- Division

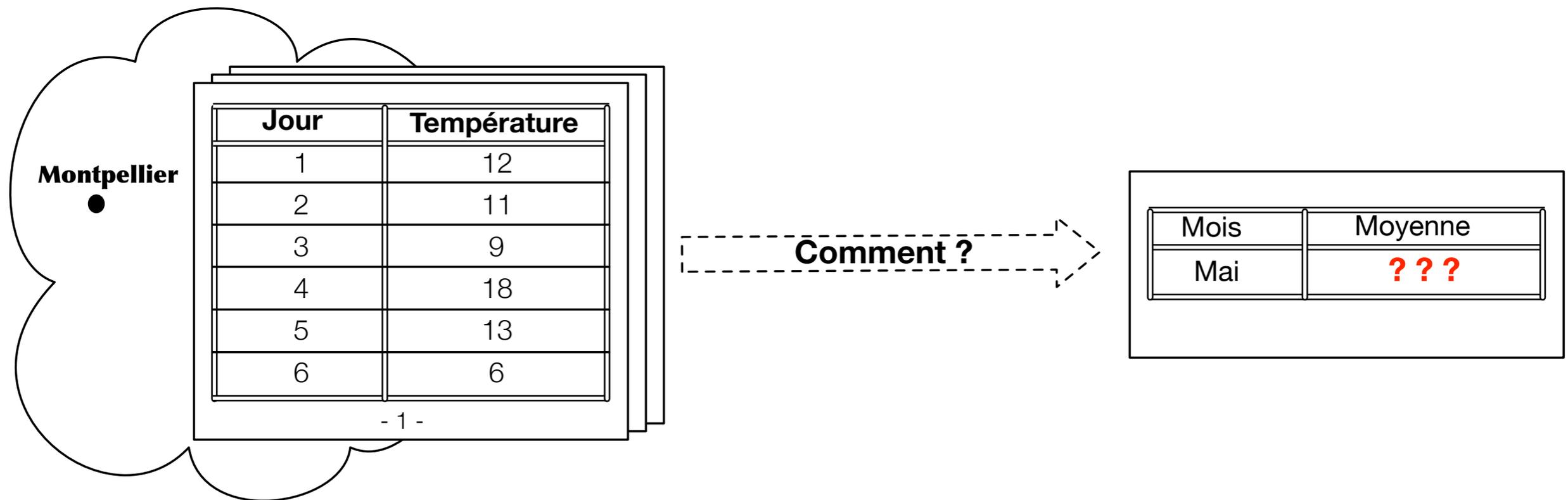


Traitements = {RoueA, RoueB, ...}  
Links = {RoueA -> RoueB, ...}

```
Traitement T1 = new Traitement("RoueA");
Traitement T2 = new Traitement("RoueB");
Link L1 = new Link(T1, T2);
```

# Phase 1

Expertise  $\longrightarrow$  Abstrait métier



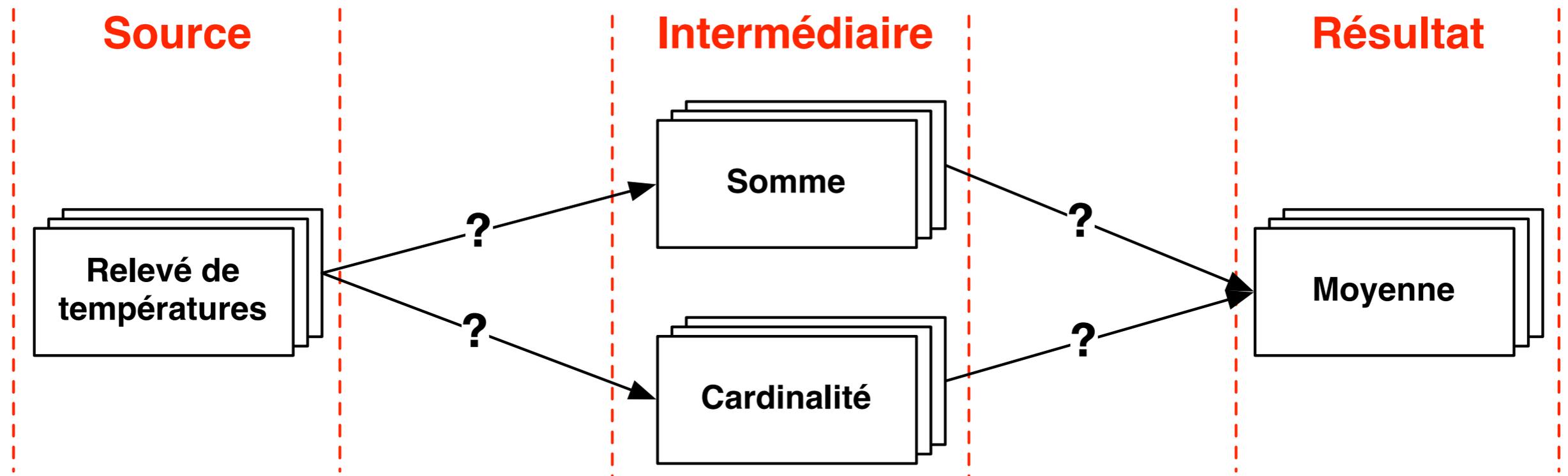
# Phase 1

Expertise  $\longrightarrow$  Abstrait métier



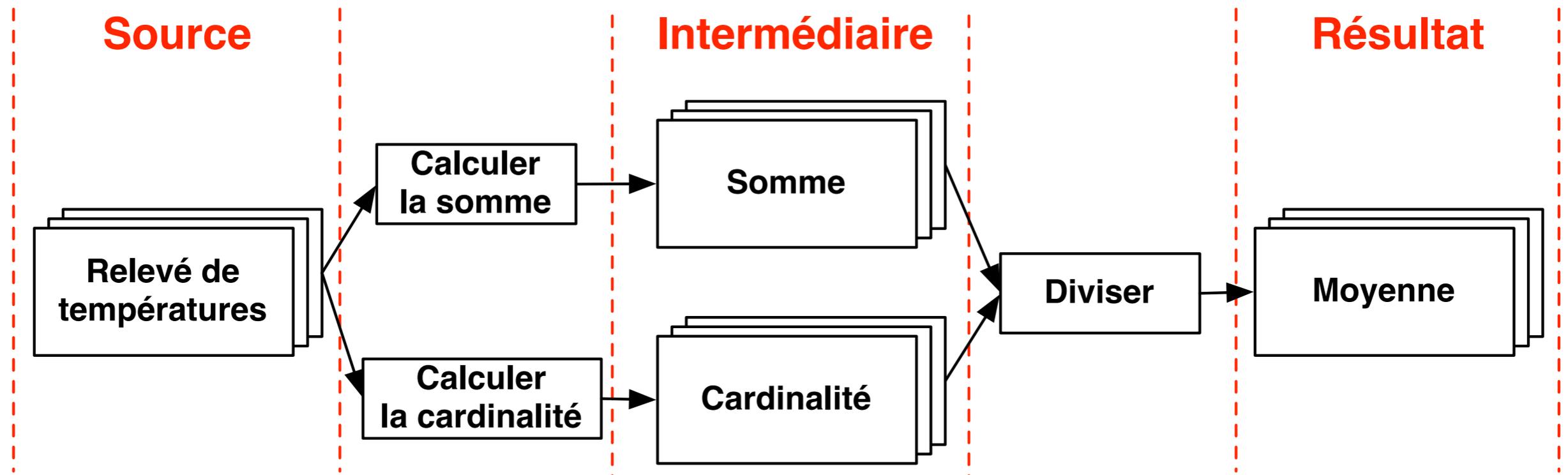
# Phase 1

Expertise  $\longrightarrow$  Abstrait métier



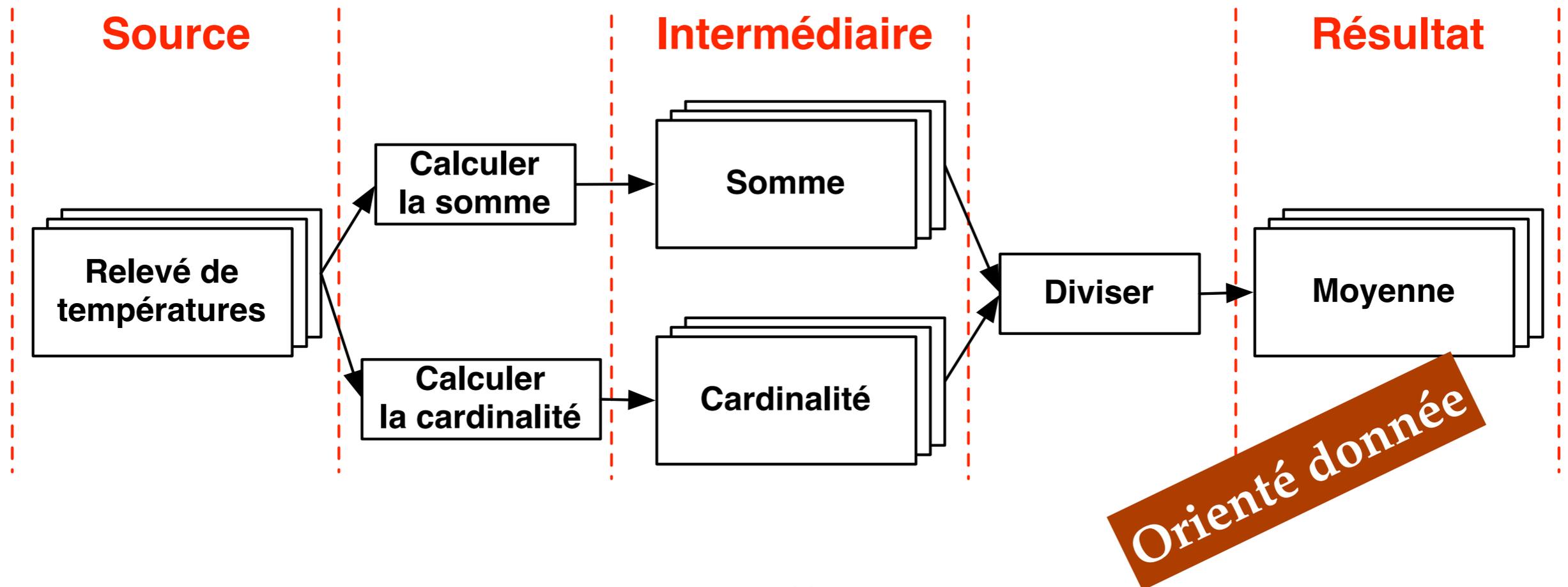
# Phase 1

Expertise  $\longrightarrow$  Abstrait métier



# Phase 1

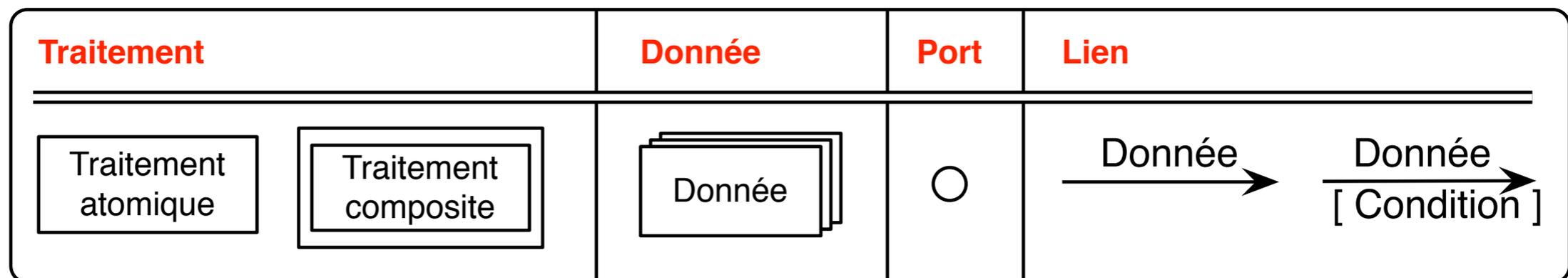
Expertise  $\longrightarrow$  Abstrait métier



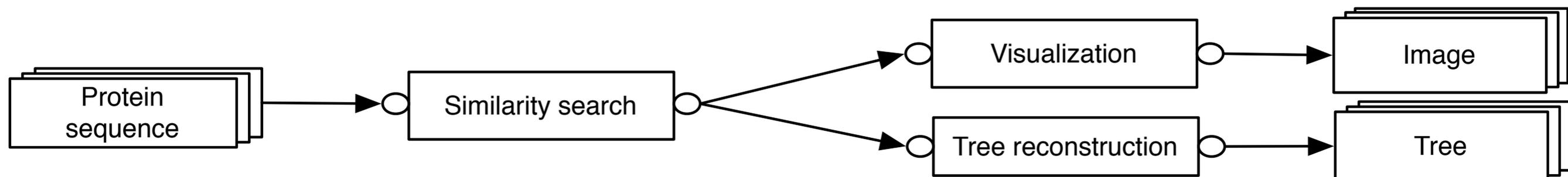
# Phase 1

## Expertise $\longrightarrow$ Abstrait métier

- Etat de l'art (UML, BPEL, Taverna, Kepler, etc.)
- Notre proposition (Langage graphique associé)

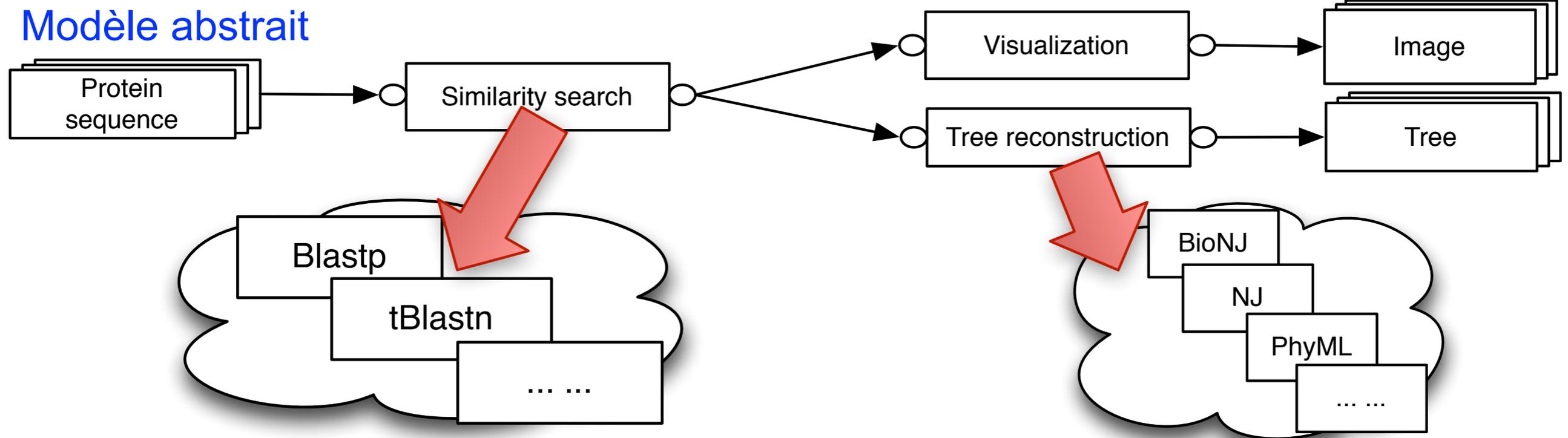


### Exemple



# Phase 2

Abstrait métier  $\longrightarrow$  Instancié

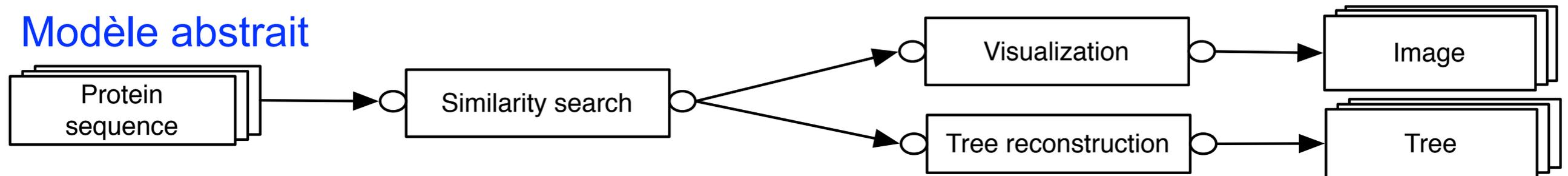


- Organisation des ressources pour la recherche et la localisation

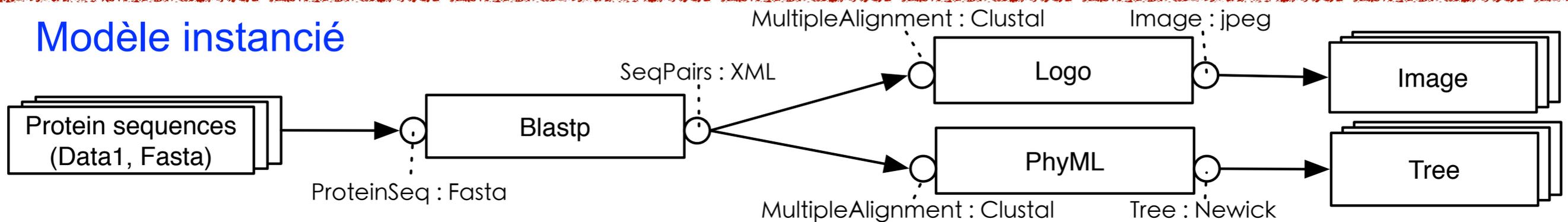
# Phase 2

## Abstrait métier $\longrightarrow$ Instancié

### Modèle abstrait



### Modèle instancié



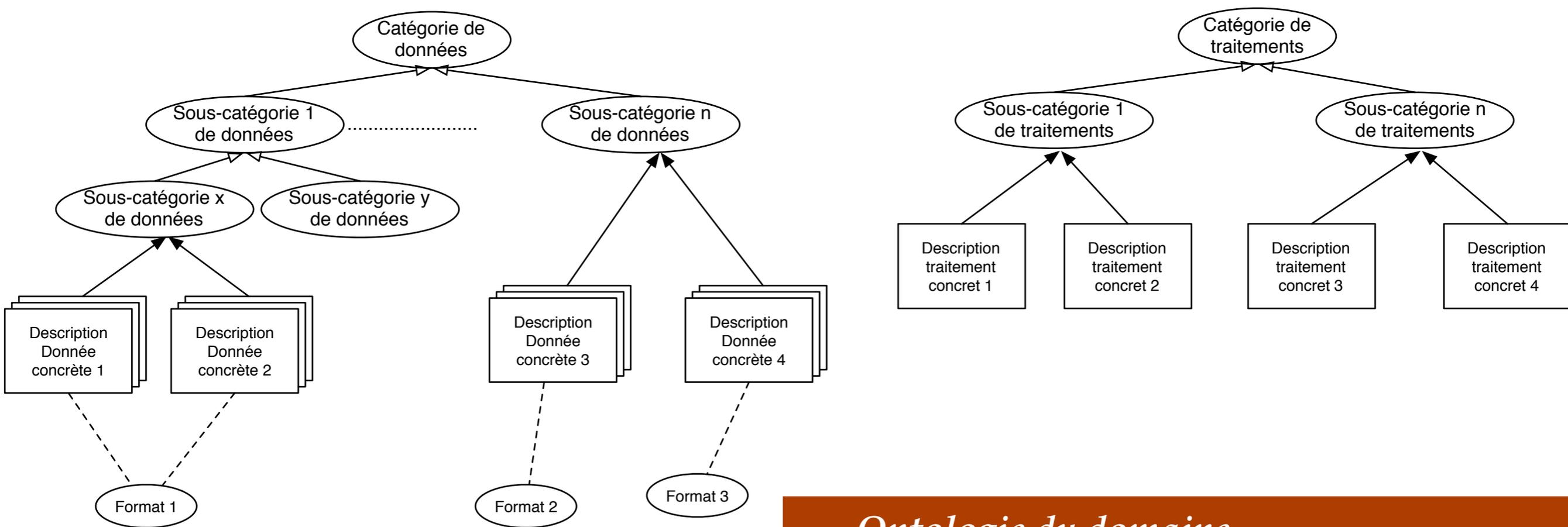
**Valide ?**

- Organisation des ressources pour la recherche et la localisation
- Validation du modèle concret obtenu

# Phase 2

## Abstrait métier $\rightarrow$ Instancié

### Le contexte de travail

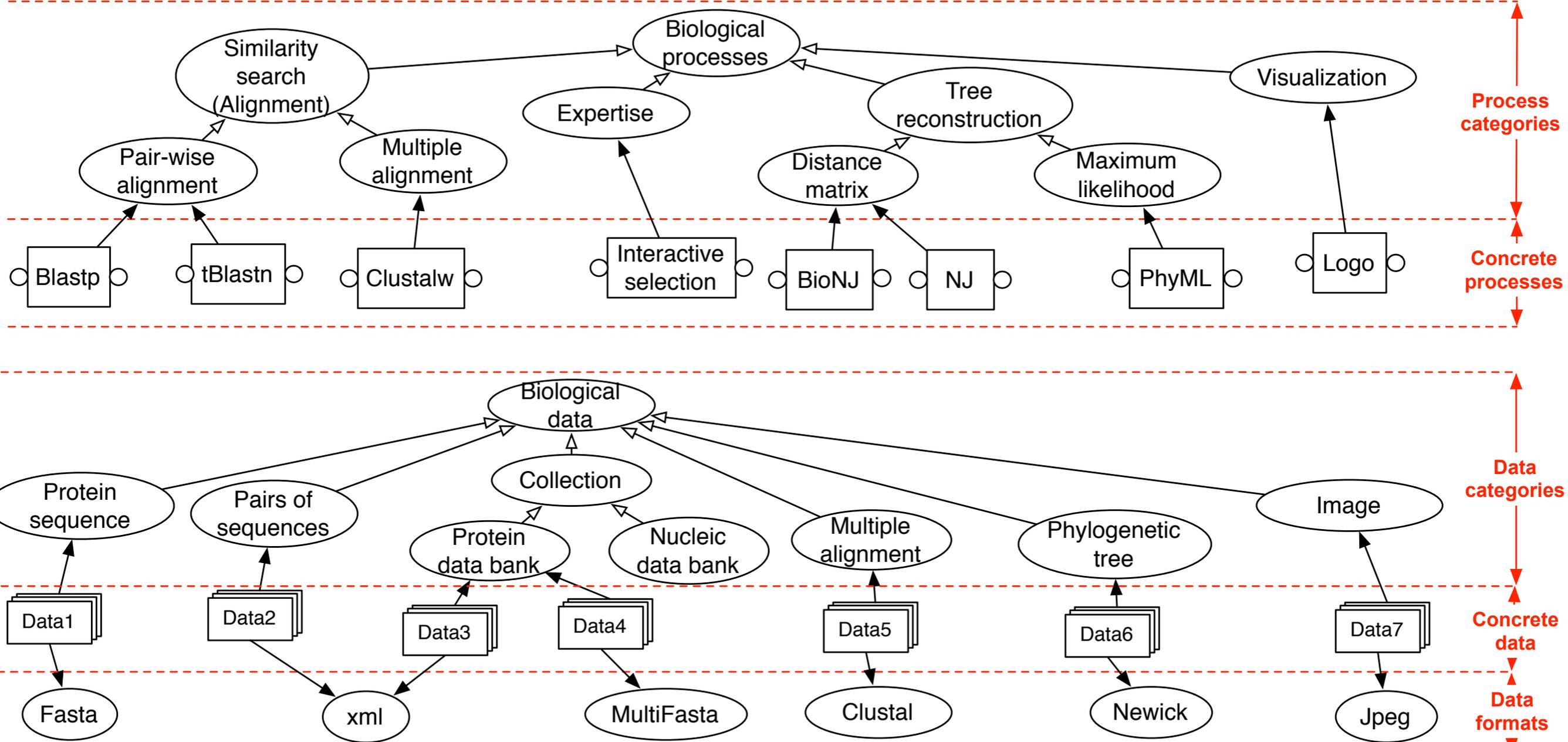


- *Ontologie du domaine*

- *Enregistrer dans la base de connaissances*

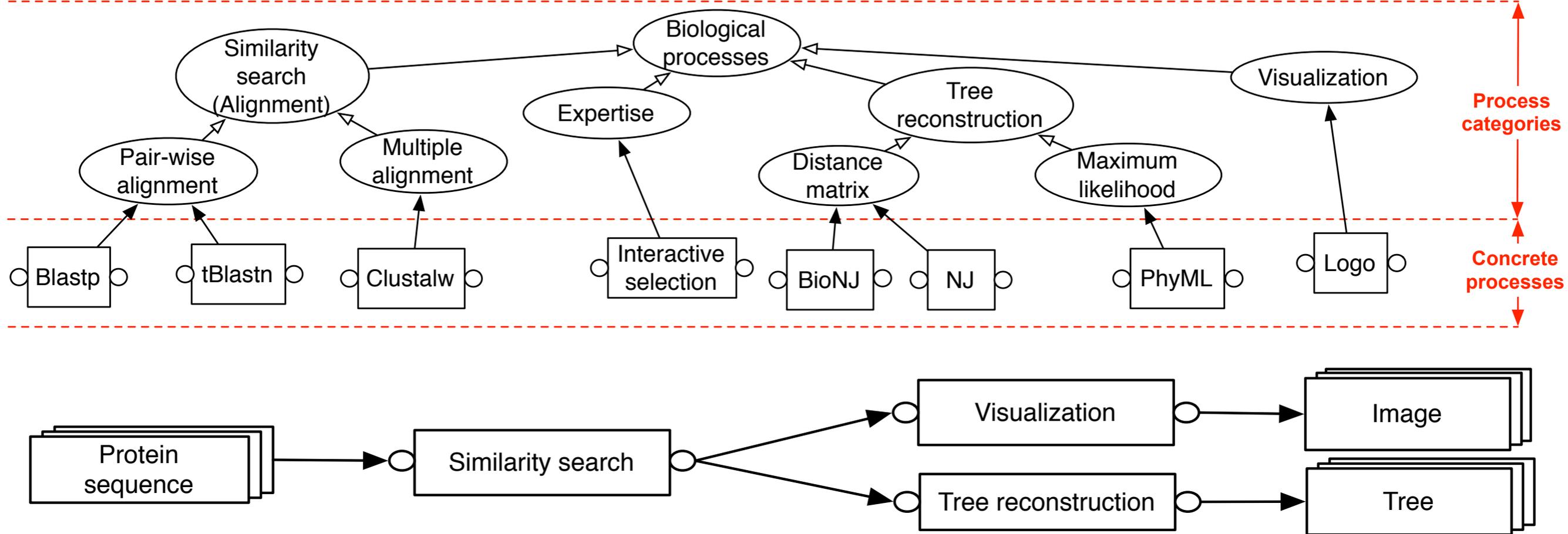
# Phase 2

## Abstrait métier $\rightarrow$ Instancié



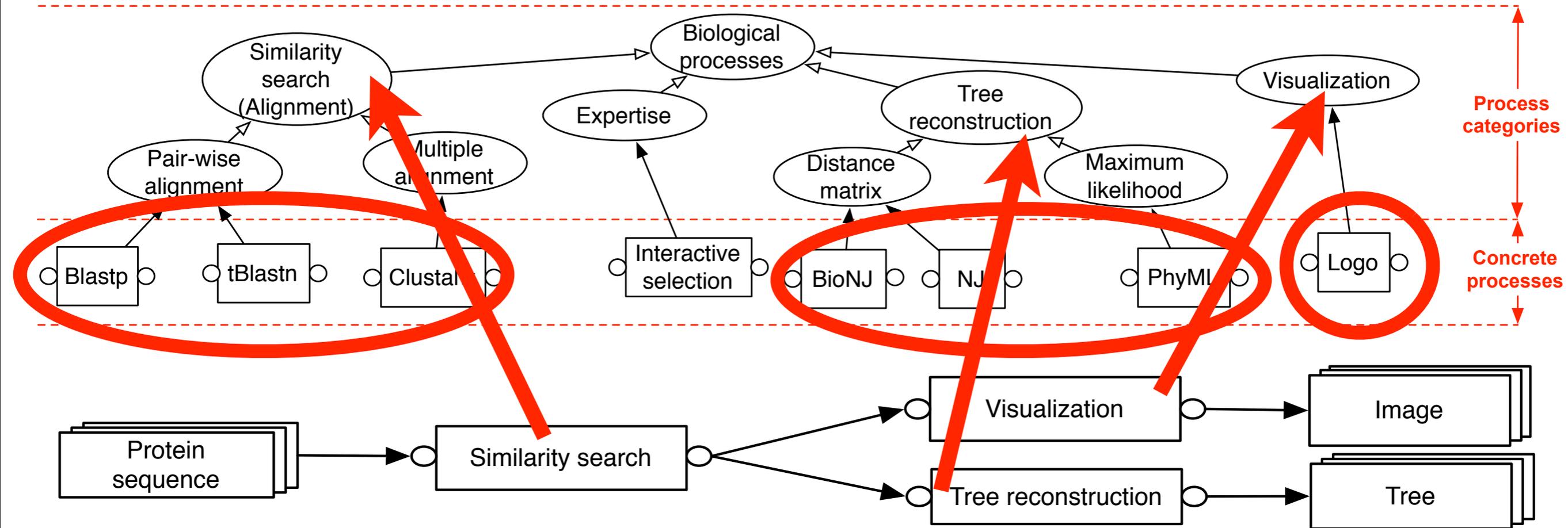
# Phase 2

## Abstrait métier —> Instancié



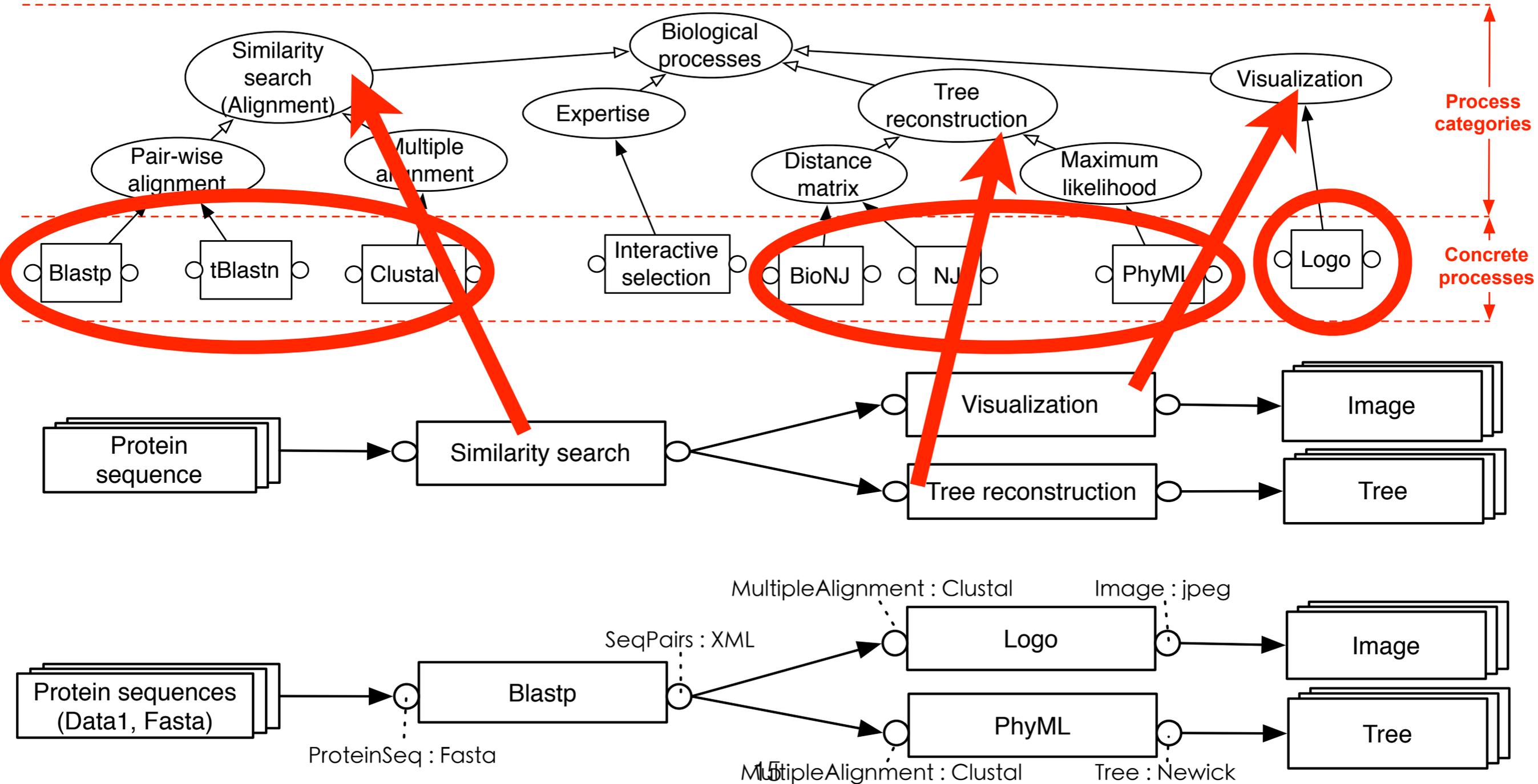
# Phase 2

## Abstrait métier —> Instancié



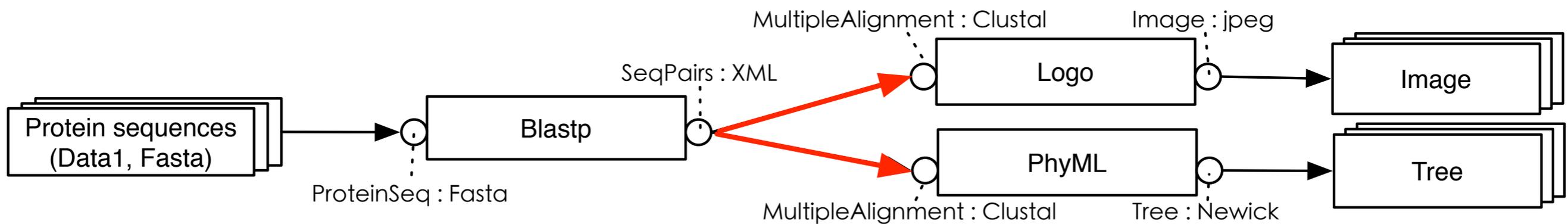
# Phase 2

## Abstrait métier —> Instancié



# Phase 2

Abstrait métier  $\longrightarrow$  Instancié



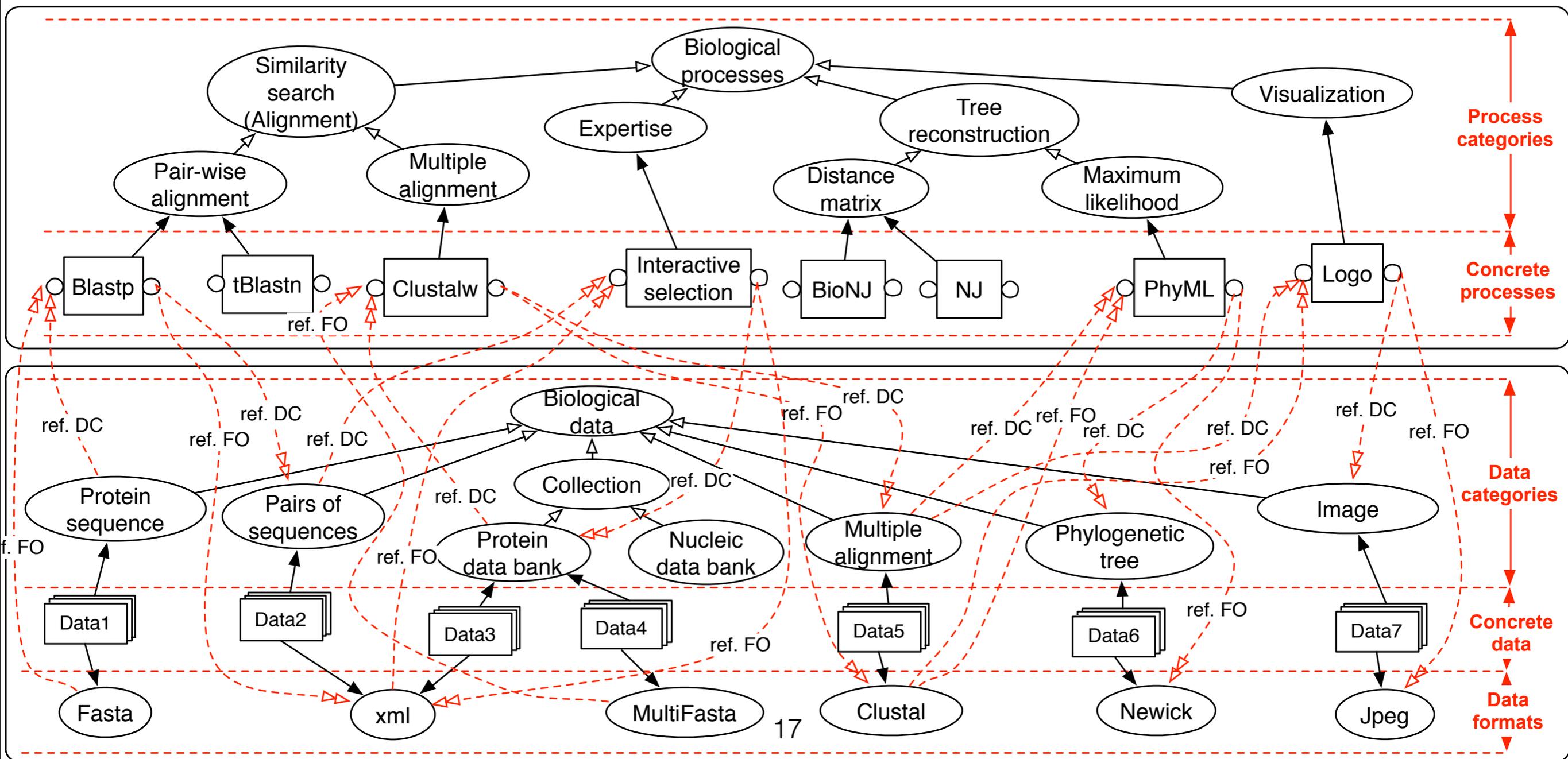
**Valide ?**

La validité des *compositions* entre :

- Blastp et Logo
- Blastp et PhyML

# Phase 2

## Abstrait métier $\rightarrow$ Instancié



# Perspectives

- Contexte de travail + Ontologie du domaine
- Apprentissage (Case-based reasoning)
  - amélioration des résultats de composition
  - extraction des fragments réutilisables

# Phase 3 (en cours)

## Instancié —> Exécutable

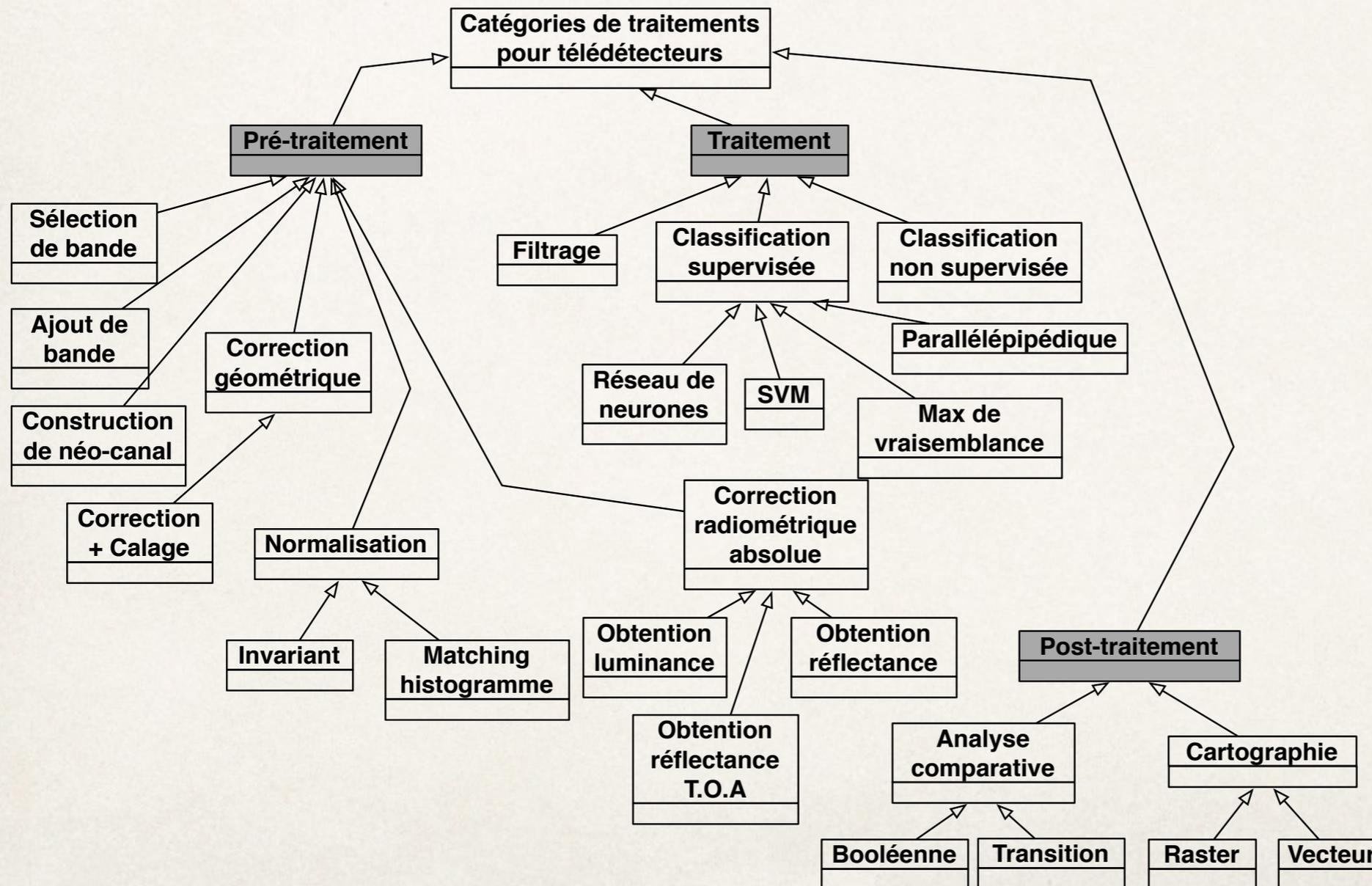
- Choix de paradigme d'exécution (locale/distribuée)
- Transformation des modèles
- Configuration / optimisation
- Traçabilité et reproductibilité (log, stockage/archivage)

# Conclusion

- Construction de workflows basée sur la réutilisation des ressources et des connaissances
- Validation dans divers domaines scientifiques (télétection, bio, hydrologie, astrophysique, climatologie)
- Assistance de construction de workflow

# Validation

- ❖ Les domaines concernés : biologie, risque, télédétection, etc.





Merci beaucoup

[yuan.lin@supagro.inra.fr](mailto:yuan.lin@supagro.inra.fr)