

INRAE

# > CATI SysMics

Sandra Dérozier

# Présentation générale du CATI

## Un CATI de 3e génération orienté “recherche”

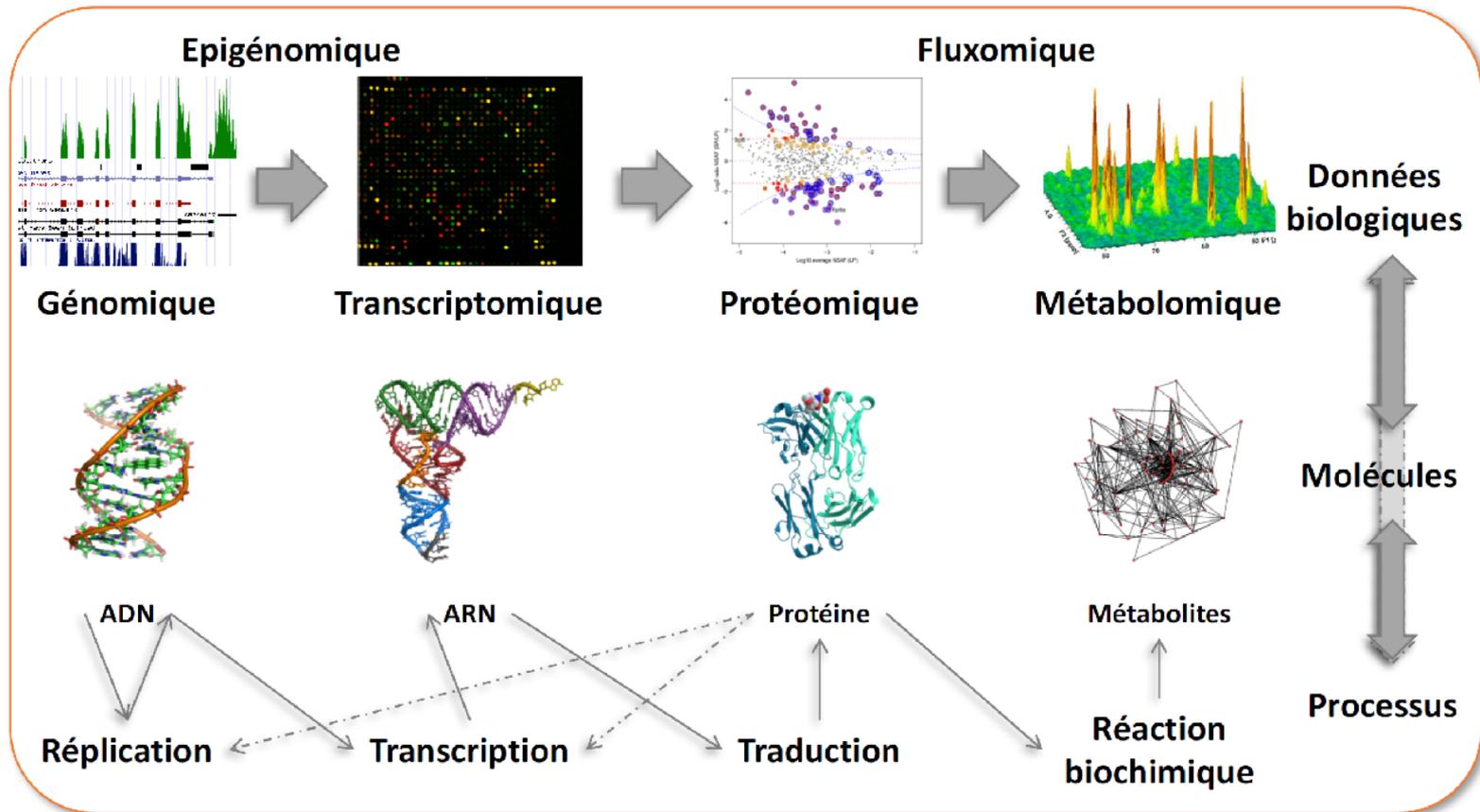
### Objectifs

- **développer** un **environnement informatique** intégrant la **notion de processus** (dynamiques) afin de **représenter** un **organisme vivant**,
- **déployer** des **méthodes statistiques** exploitant à la fois des **données “omiques” hétérogènes** (transcriptome, protéome, fluxome, métabolome, phénomène, imageries, ...) et les **processus**.

### Missions

- **consolider** et/ou **établir** une preuve de concept de l'**environnement informatique** pour des organismes modèles
- **élaborer** et **déployer** une 1<sup>ère</sup> **stratégie de transfert** pour un ou des organismes d'intérêt agronomique
- **constituer** une **communauté interdisciplinaire** autour de l'**intégration systémique** des données et des connaissances biologiques

# Données hétérogènes et processus biologiques

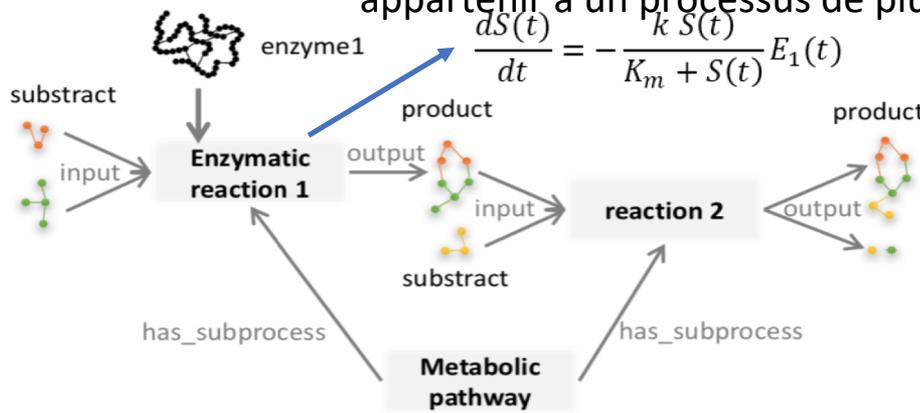


**Un processus biologique : des entités biologiques + un modèle mathématique**

# Biologie des systèmes

La Biologie des Systèmes s'appuie sur une description « systémique » des objets biologiques manipulés, basée sur les approches systèmes en Science de l'Ingénieur

- un système (au sens des SI) est formé de constituants (objets) en interaction. Un modèle mathématique est associé au système et décrit son comportement dynamique.
- un processus biologique est constitué d'une ou plusieurs réactions biochimiques impliquant des entités biologiques, et a un comportement dynamique.
  - ◆ peut être considéré comme un système au sens des SI
  - ◆ peut être décomposé en sous-systèmes de granularité plus fine ou appartenir à un processus de plus haut niveau.



Exemple : voie métabolique

zoom-in

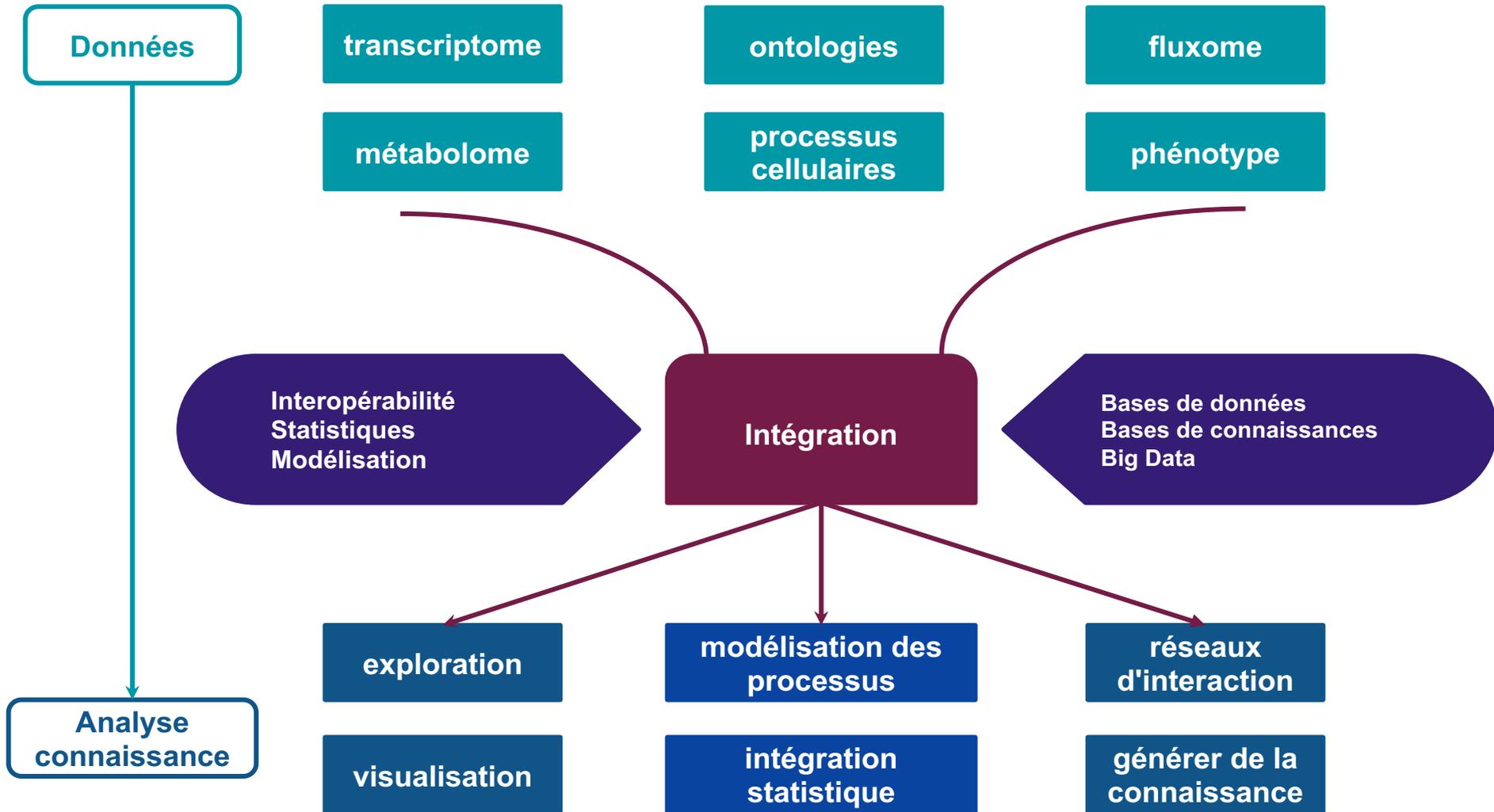
Granularité

zoom-out

## Multi-échelle

- moléculaire
- cellulaire
- organisme
- population

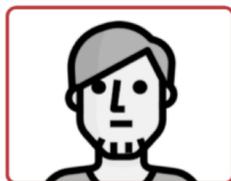
# Contexte



# Les membres

<https://sysmics.cati.inrae.fr/qui-sommes-nous>

Création en 2019 et piloté par **MathNum**, **BAP** et **MICA**



Harry Belcram  
GQE



Nadia Bessoltane  
IJPB



Eric Biot  
IJPB



Véronique Brunaud  
IPS2



Guillaume Cerutti  
ENS Lyon



Delphine Charif  
IJPB



Sandra Dérozier  
MaAGE



Marc Dinh  
MaAGE



Thomas Duigou  
MICALIS



Cyprien Guérin  
MaAGE



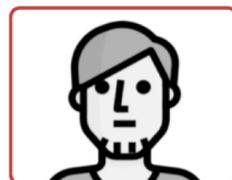
Cécile Guichard  
IPS2



Anne Goelzer  
MaAGE



Olivier Inizan  
MaAGE



Johann Joets  
GQE



Sandrine Lefranc  
IJPB



Christine Paysant-Le Roux  
IPS2



Franck Samson  
LAMME

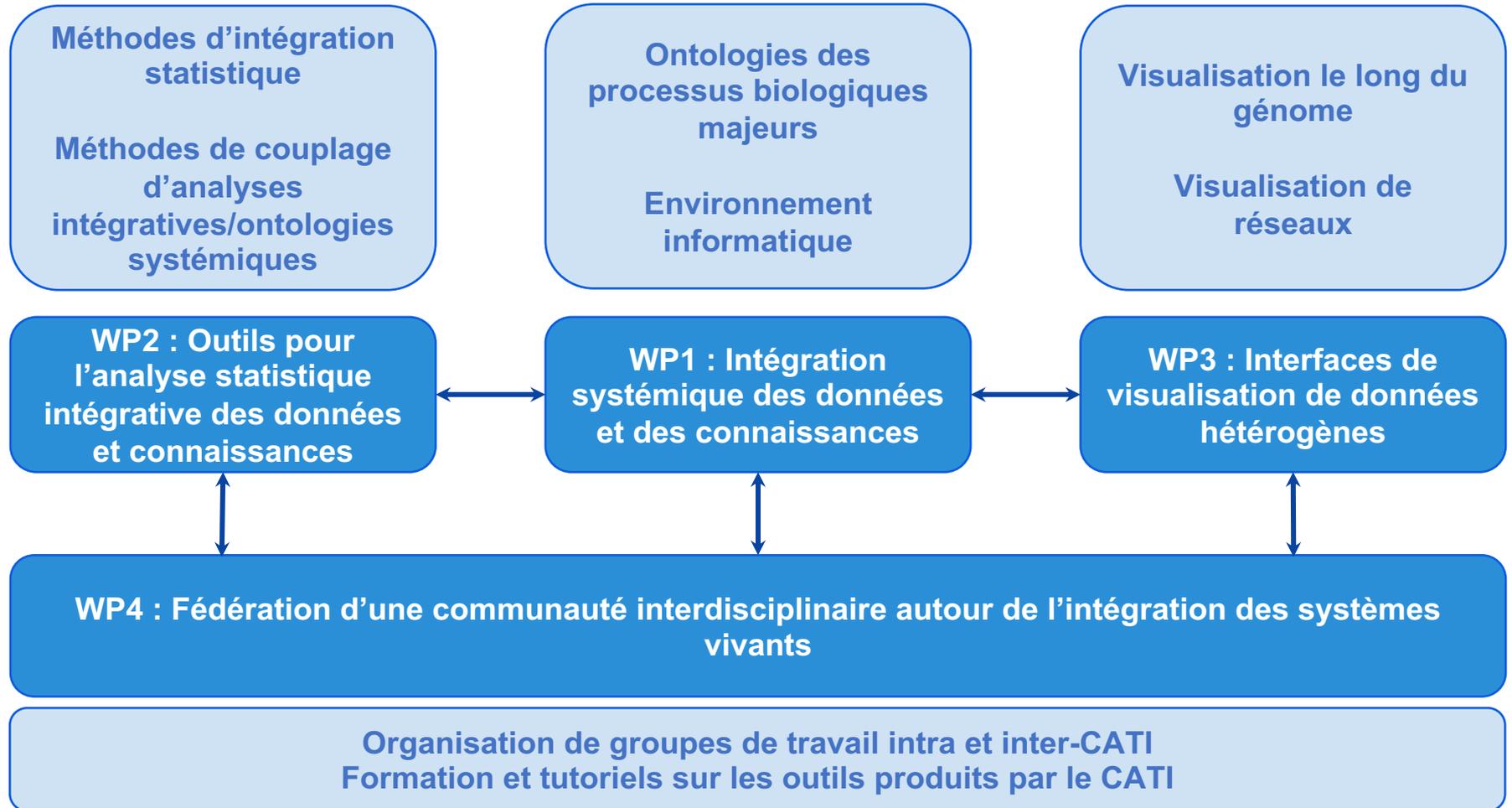


Alain Trubuil  
MaAGE

# Structure du CATI SysMics

<https://sysmics.cati.inrae.fr/workpackages>

Structuration en 4 *workpackages*



# Focus sur le WP 3

<https://sysmics.cati.inrae.fr/workpackage3>

## Interface de visualisation de données hétérogènes

### Deux volets

- interface de visualisation le long du génome
- interface de visualisation de réseaux

# Focus sur le WP 3

<https://sysmics.cati.inrae.fr/workpackage3>

## Interface de visualisation de données hétérogènes

### Deux volets

- **interface de visualisation le long du génome**
- interface de visualisation de réseaux

### Objectif

- enrichir et adapter les outils existants afin de visualiser les données “omiques” mises à disposition au sein du système d’informations proposés par le WP 1

# Focus sur le WP 3

<https://sysmics.cati.inrae.fr/workpackage3>

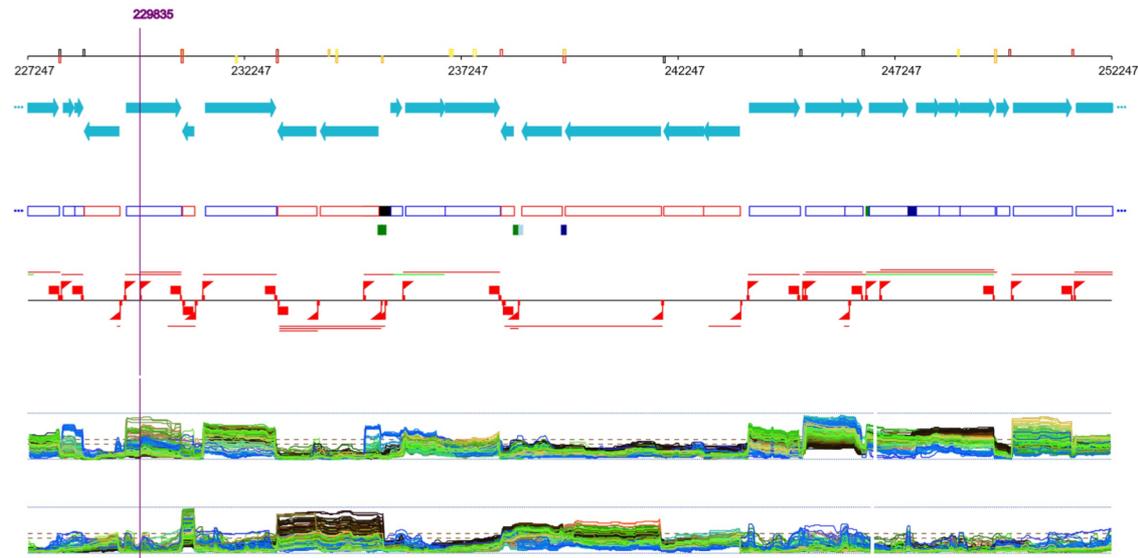
## Interface de visualisation de données hétérogènes

### Deux volets

- interface de visualisation le long du génome
- interface de visualisation de réseaux

### Objectif

- enrichir et adapter les outils existants afin de visualiser les données “omiques” mises à disposition au sein du système d’informations proposés par le WP 1



# Focus sur le WP 3

<https://sysmics.cati.inrae.fr/workpackage3>

## Interface de visualisation de données hétérogènes

### Deux volets

- interface de visualisation le long du génome
- **interface de visualisation de réseaux**

### Objectif

- proposer un outil permettant de visualiser les processus biologiques (réseaux métaboliques et/ou de régulation) à partir des données extraites du système d'information structuré proposé par le WP 1

# Focus sur le WP 3

## Interface de visualisation de données hétérogènes

### Deux volets

- interface de visualisation le long du génome
- **interface de visualisation de réseaux**

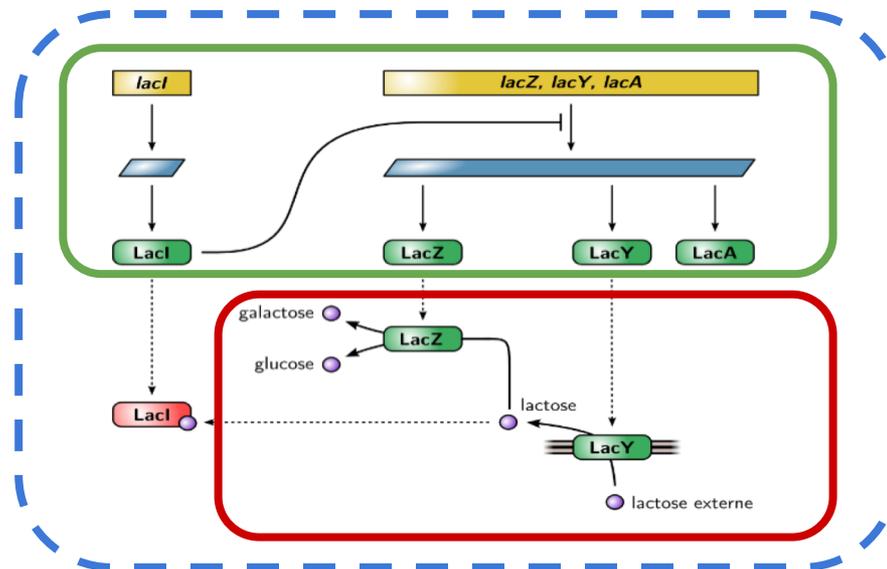
### Objectif

- proposer un outil permettant de visualiser les processus biologiques (réseaux métaboliques et/ou de régulation) à partir des données extraites du système d'information structuré proposé par le WP 1

réseaux métaboliques

réseaux de régulation

modules fonctionnels



# Focus sur le WP 3

<https://sysmics.cati.inrae.fr/workpackage3>

## Interface de visualisation de données hétérogènes

### Deux volets

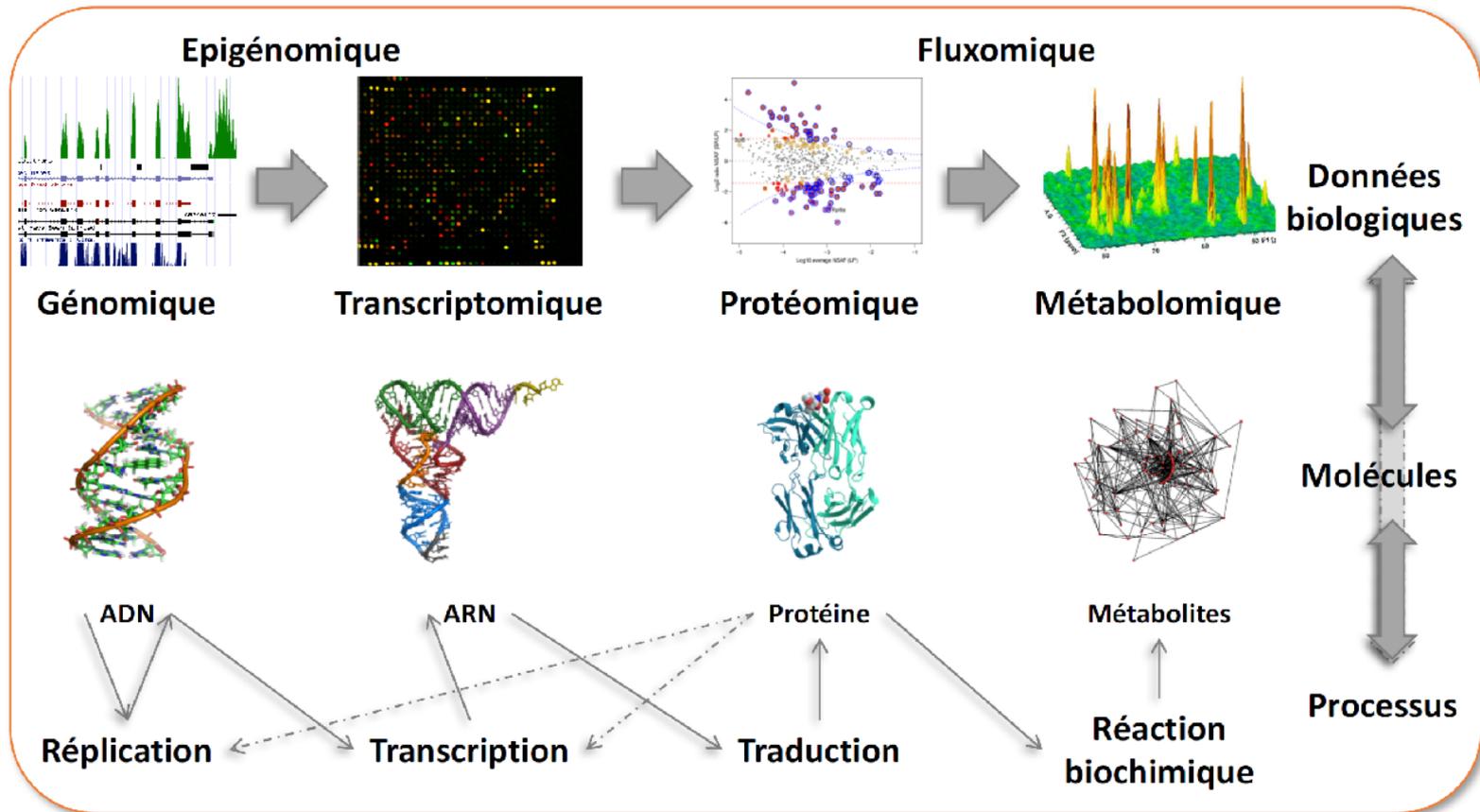
- interface de visualisation le long du génome
- **interface de visualisation de réseaux**

### Plusieurs tâches

- veille sur les outils de visualisation de réseaux existants afin d'en évaluer les limites
- définition des besoins de visualisation de réseaux au regard des données à visualiser et des limitations évaluées précédemment
- évaluation des solutions envisagées sur des jeux de données définis sur les organismes modèles, *B. subtilis* et *A. thaliana*
- implémentation de fonctionnalités supplémentaires si besoin

# Panorama des données manipulées

# Données hétérogènes et processus biologiques



**Un processus biologique : des entités biologiques + un modèle mathématique**

# Espèces d'intérêt



*Arabidopsis thaliana*  
(crédit : Wikipedia)



*Bacillus subtilis*  
(crédit : Cambridge commodities)

# Espèces d'intérêt



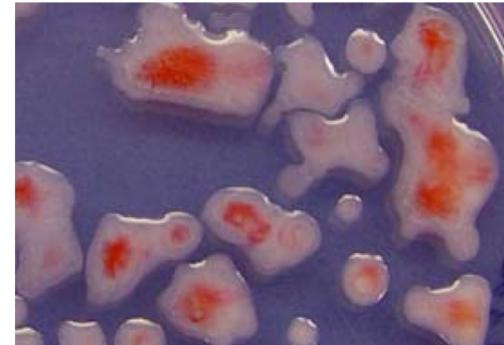
*Arabidopsis thaliana*  
(crédit : Wikipedia)



*Bacillus subtilis*  
(crédit : Cambridge commodities)



*Zea mays*  
(crédit : Efor)



*Ralstonia solanacearum*  
(crédit : T. Momol & S.M. Ilson, University of Florida)

# Ontologies

Principe : mettre les processus biologiques au coeur de la représentation.

En partant de ce principe :

- décrire le fonctionnement d'une cellule → création de deux ontologies :
  - ◆ BiPON pour les processus "itératifs" (régulation génétique)
  - ◆ BiPOM pour le métabolisme
- affiner cette description au niveau des modèles mathématiques avec comme objectif de représenter les aspects modulaires du vivant : une nouvelle ontologie en cours de construction.

Liens avec quelques lexiques / ontologies existants : [Gene Ontology](#), [ChEBI](#), [KEGG](#).

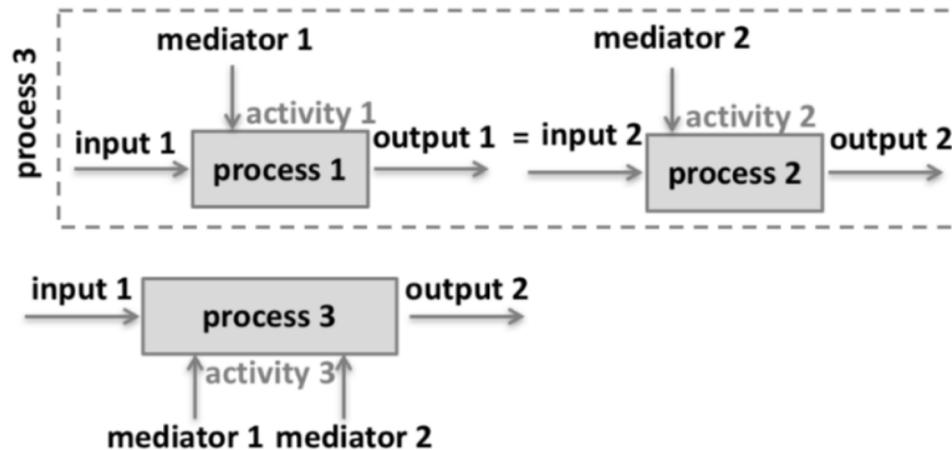


# Ontologies

Principe : mettre les processus biologiques au coeur de la représentation.

Enjeux pour la visualisation :

- travailler à l'échelle de la cellule entière *B.subtilis*,
- être en mesure de superposer plusieurs échelles (de la réaction biochimique au module fonctionnel).



# Outils utilisés

# Veille - Outils pour la visualisation de réseaux

## Grille de lecture

- Utilisation en tant que service web possible ?
- Open Source ?
- Code maintenu et documenté ?
- Facilité à utiliser ? À adapter ?

# Veille - Outils pour la visualisation de réseaux

## Grille de lecture

- Utilisation en tant que service web possible ?
- Open Source ?
- Code maintenu et documenté ?
- Facilité à utiliser ? À adapter ?

## Document sur la veille

- [https://sysmics.cati.inrae.fr/sites/default/files/wp3/201906\\_outil\\_visualisation\\_processus.pdf](https://sysmics.cati.inrae.fr/sites/default/files/wp3/201906_outil_visualisation_processus.pdf)
- liste non exhaustive de 20 outils ou librairies



# Veille - Outils pour la visualisation de réseaux

## Quelques points saillants

- Beaucoup d'outils ne sont plus maintenus
- Le niveau d'abstraction est hétérogène, de la librairie “bas niveau” (ex. : D3.js) à la visualisation de cartes spécifiques de KEGG

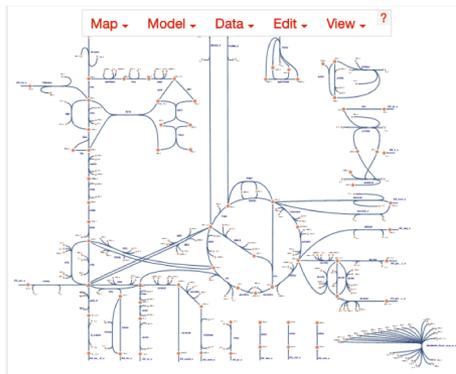
# Veille - Outils pour la visualisation de réseaux

## Quelques points saillants

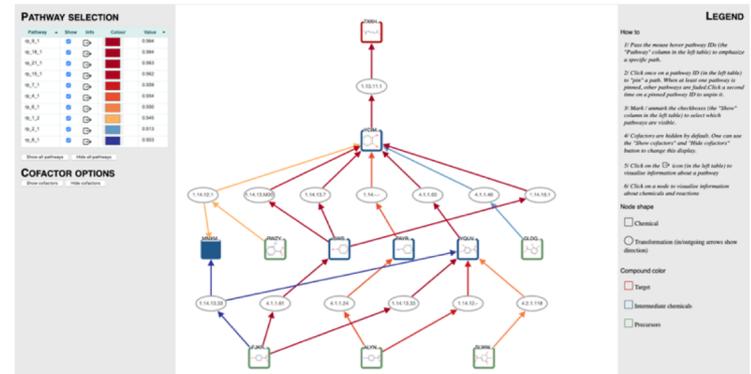
- Beaucoup d'outils ne sont plus maintenus
- Le niveau d'abstraction est hétérogène, de la librairie "bas niveau" (ex. : D3.js) à la visualisation de cartes spécifiques de KEGG

## Deux outils investigués en particulier

- **rpViz** (<https://github.com/brsynth/rpVisualiser>) : outil développé à Micalis pour visualiser des voies métaboliques prédites par rétrosynthèse
- **Escher** (<https://escher.github.io/>) : outil que Thomas va vous présenter dans quelques minutes



Escher



rpViz

# Organisation des travaux réalisés

# Organisation

## Le collectif en interne

- 6 personnes
- réunion mensuelle

## Un collectif plus large

- volonté de partager avec d'autres CATI, notamment BARIC, GREP et EmPrEInTE
- co-organisation du *workshop* work4Graph

## Partage et Gestion des productions

- Site web du CATI : <https://sysmics.cati.inrae.fr/productions>
- ForgeMIA



# Questions / cas d'utilisation pertinents

# Cas d'utilisation

## Visualisation des processus biologiques (connaissance et données)

**Granularité fine** : visualiser un processus biologique (défini en espace et en temps), ses régulations connues, données *in silico*, et pouvoir projeter des données hétérogènes dessus, par exemple :

- voie métabolique + ses régulations enzymatiques et génétiques + combinaison de données (métabolome, fluxome, protéome, interactome si complexe enzymatique)
- transcription du gène *X* : *in silico* (promoteur (seq prédites), TSS) + métabolome, complexe RNAPolymérase (protéomique, transcriptomique), fixation de la RNAPolymérase et TF, (Chip-Seq, Chip-Chip), transcriptome du gène en aval

**Granularité moyenne** : en se basant sur la systémique, définir une stratégie pour passer des niveaux fins aux niveaux plus élevés.

# Questions

## Visualisation des processus biologiques (connaissance et données)

- Faut-il générer une représentation générique pour l'ensemble des processus ou plutôt par famille de processus visualisés ?
- Faut-il préférer une représentation générée automatiquement ou s'appuyer sur une collection de cartes agencées "à la main" par des experts ?
- En pratique, quel traitement pour les données pour la visualisation ? Faut-il normaliser entre données d'un même type ?
- Au delà de la faisabilité, comment enchaîner les processus ? Avec quels outils ?

exemple : voie métabolique A → voie métabolique B

exemple : transcription → traduction d'un gène X

