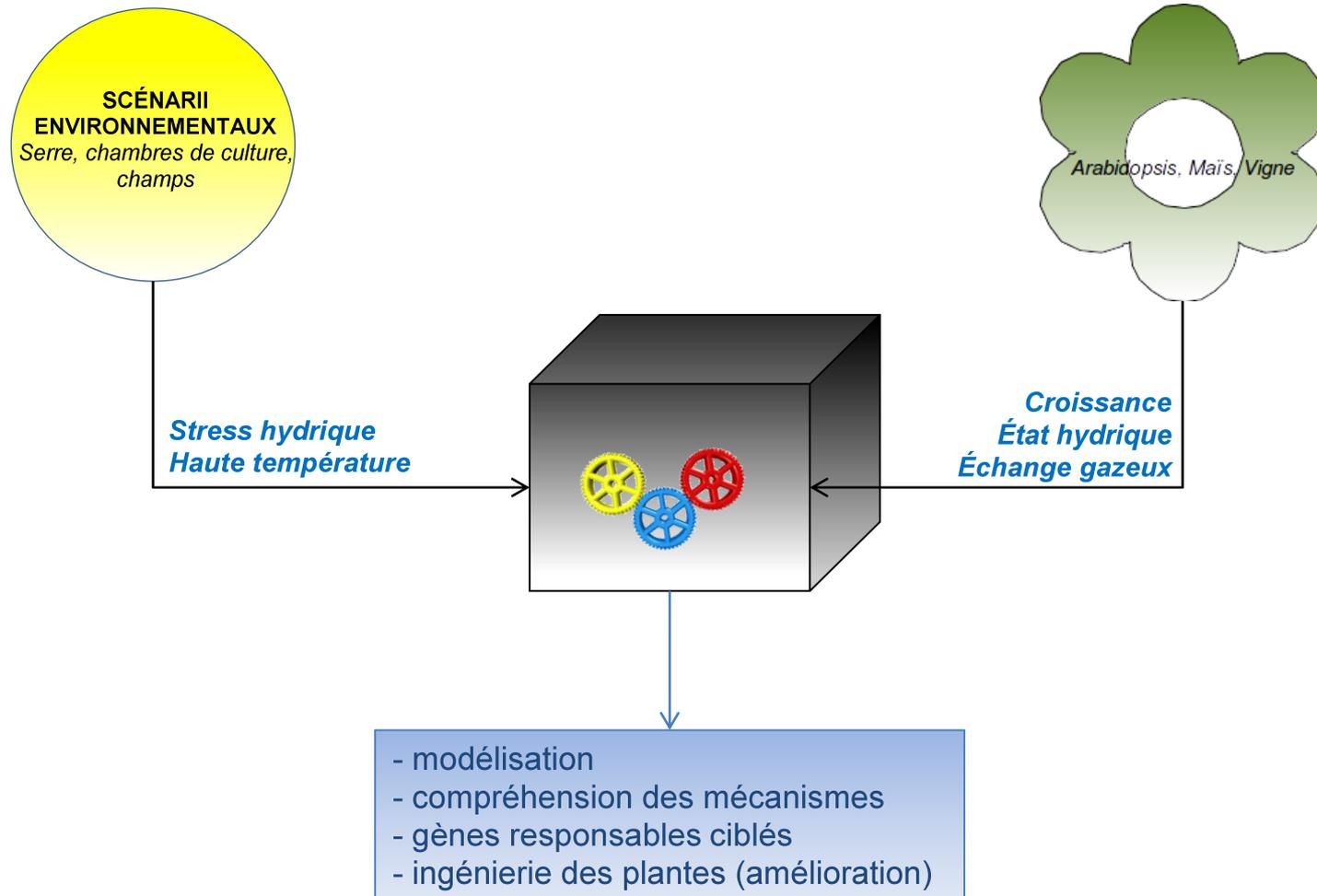


Phenotyping Hybrid Information System

Rencontres CATIs CODEX & SICPA – Décembre 2014

UMR LEPSE

Laboratoire d'Écophysiologie des *Plantes* sous *Stress Environnementaux*



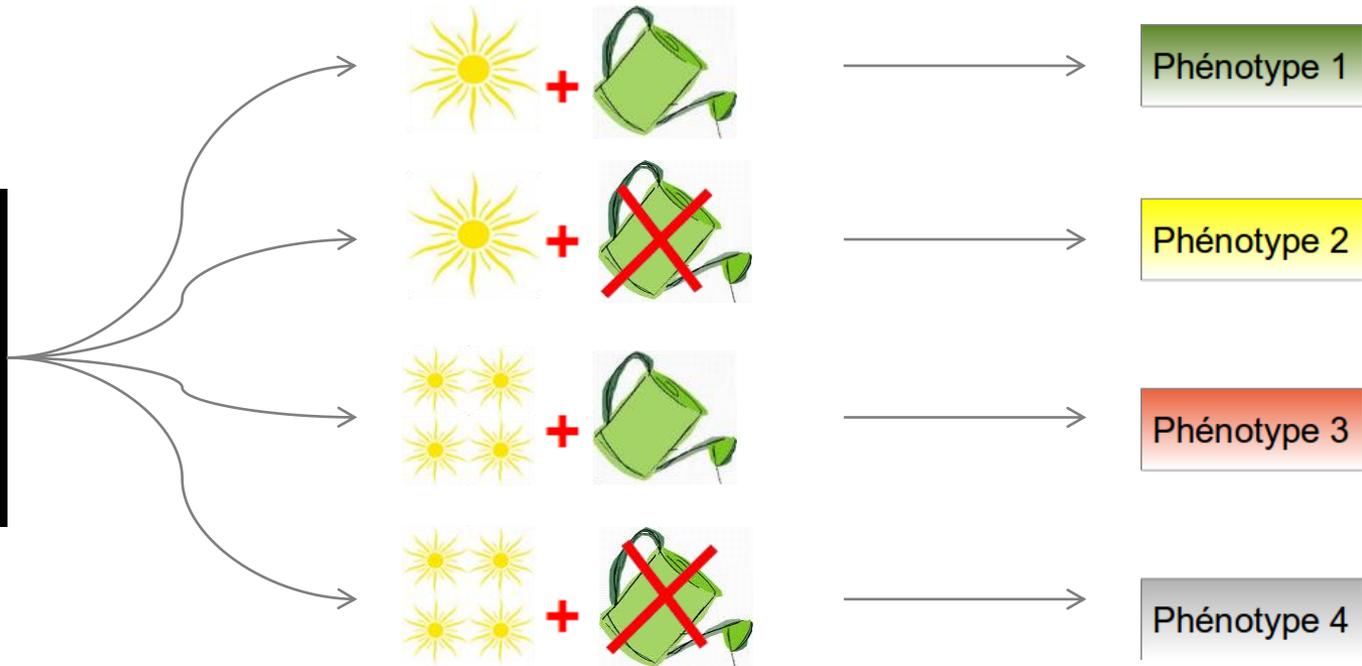
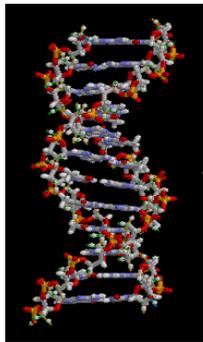
Approche intégrative "de la molécule à l'environnement"

« PHÉNOTYPAGE »

Une plante..

.. des scenarii ..

.. des phénotypes.



Collecte des caractéristiques structurales dont la variabilité est analysée en fonction du patrimoine génétique et des conditions environnementales contrôlées = **Haut débit, répétabilité, fiabilité**

Projet PHENOME



Construire
Développer
Diffuser

PCM1 : Capteurs et méthodes
PCM2 : Système d'information
PCM3 : Analyse de données

PCM2

« *Organiser les données phénotypiques générées par les plates-formes d'une manière coordonnée* »

1 : Installer et mettre en œuvre le système dans chacun des nœuds du projet (uniformisation)

2 : Développer un archivage « longue durée » en s'appuyant sur les datacenter INRA



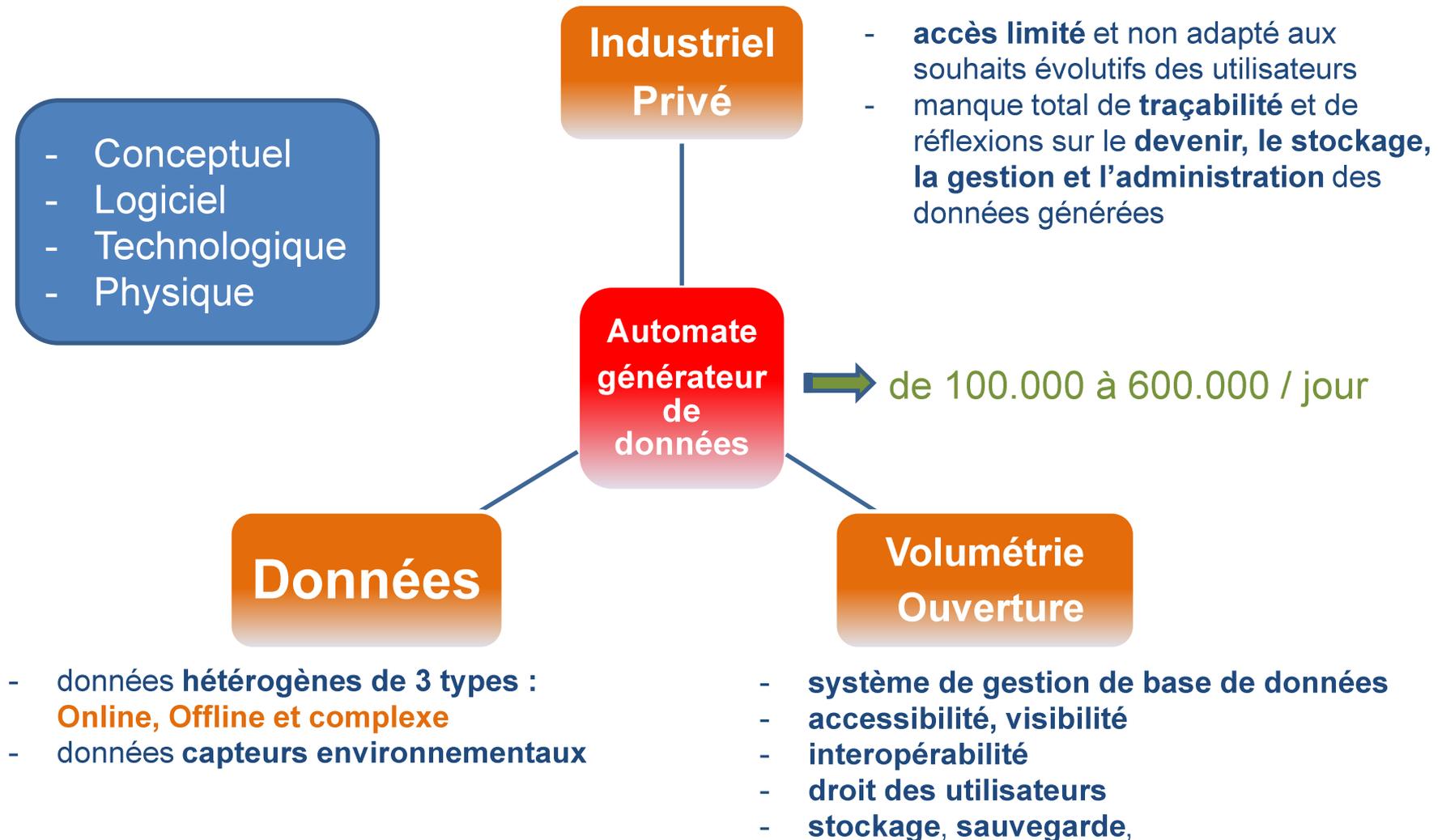
M3P

Montpellier Plant Phenotyping Platform

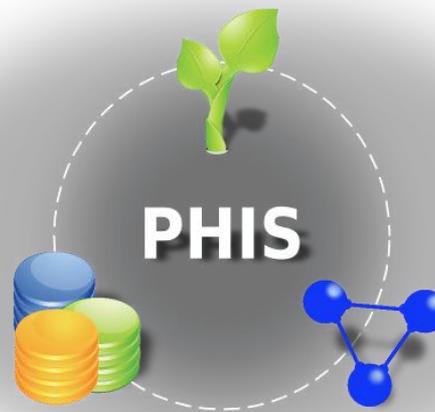


PhenoArch

Support d'étude : PHENOARCH



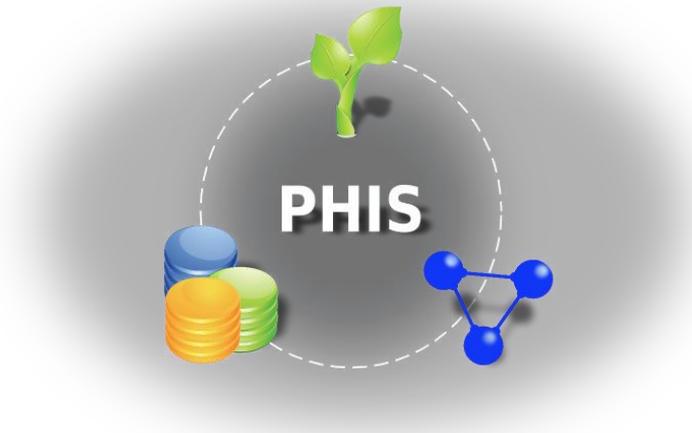
Approches collaboratives



CATI
CODEX



PHIS



**Base de données
relationnelle**



**Base de données
NoSql**



**Web
Sémantique**



PHIS

Base de données relationnelle

Propriétés

- ⊕ **Robuste** (Système bancaire et boursier)
- ⊕ **Mature** (1970)
- ⊕ Basé sur **principes ACID** :
 - Atomicité
 - Cohérence
 - Isolation
 - Durabilité

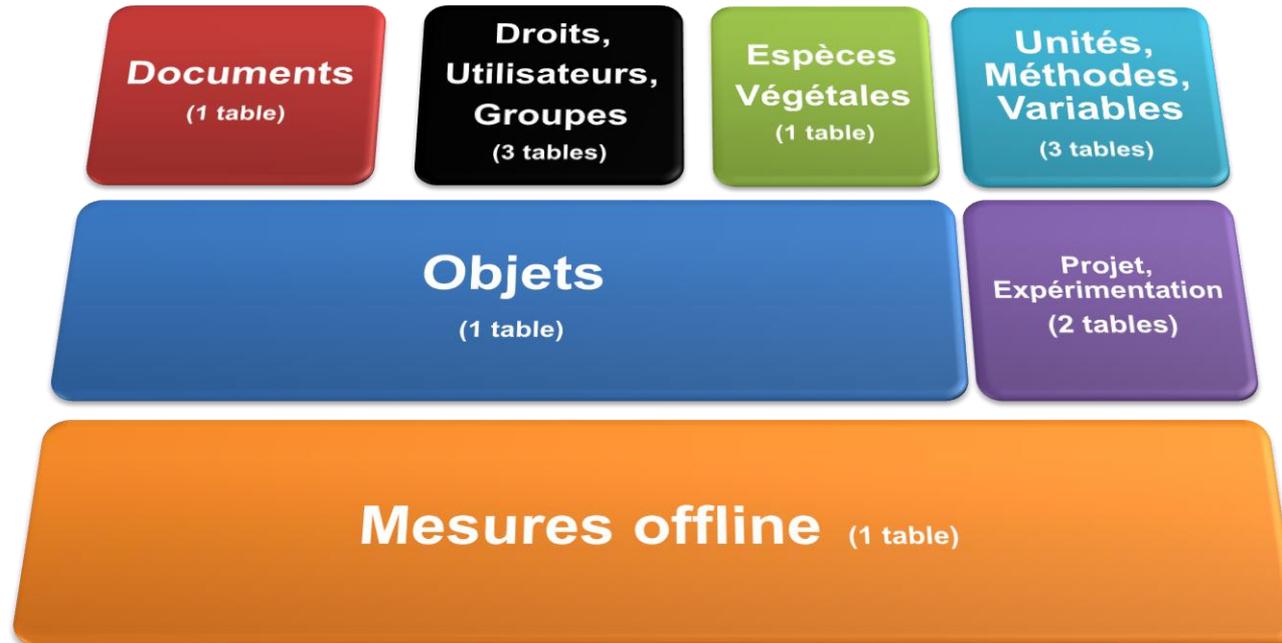
Intérêts dans le phénotypage

Volume des données
(issues de l'instrumentation et de l'observation)

Contrôle d'intégrité
(vérification de type, unicité, cohérence ..)

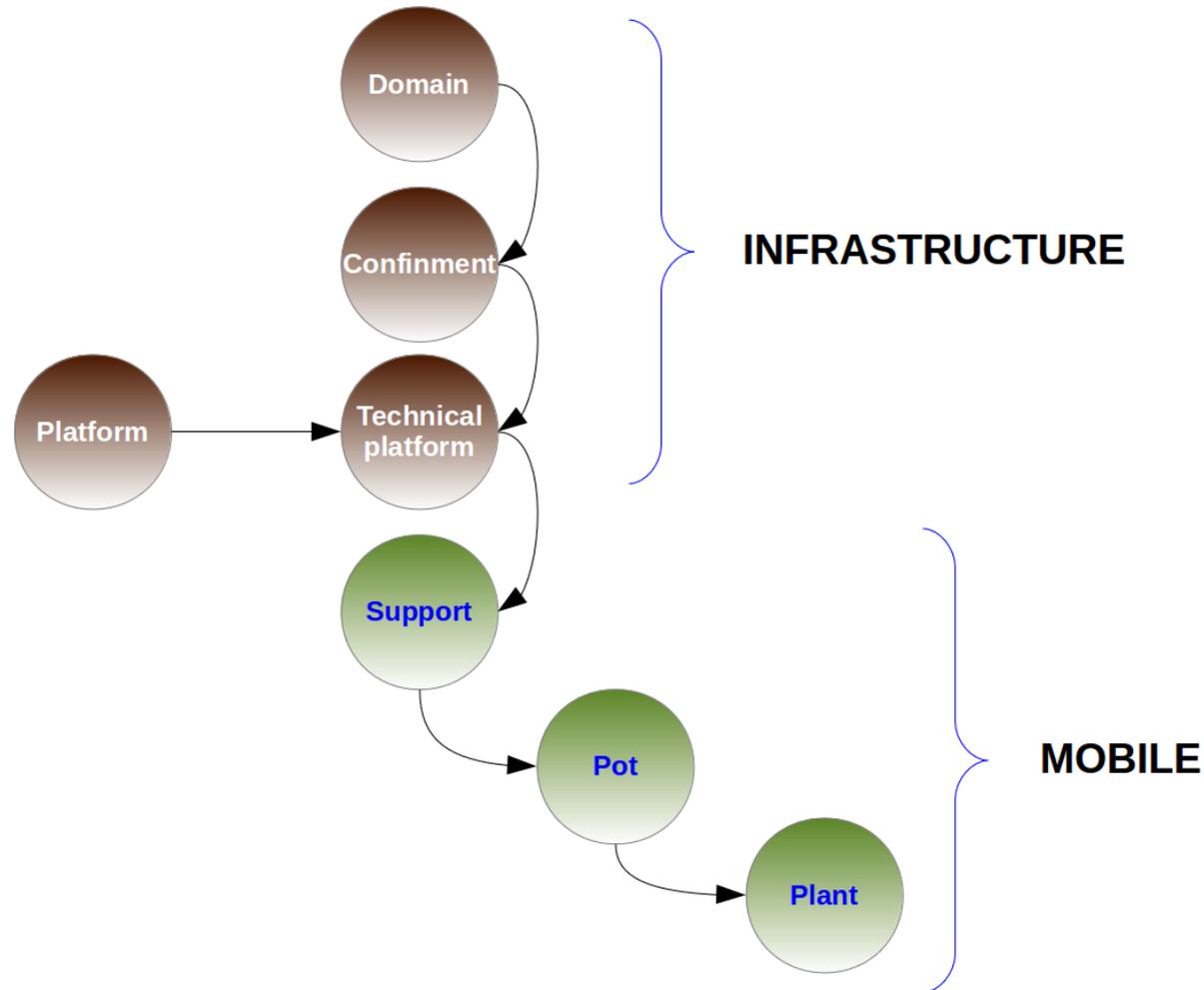
PHIS – Base de données

Package des concepts



PHIS – Base de données

Hiérarchie des objets



PHIS – Base de données

Les limites du relationnel



- Limitations des performances lors de la manipulation de **gros jeux de données**.
- Limitations des performances pour **la traçabilité** des plantes **dans l'espace et le temps** (complexité des requêtes). Certains concepts sont difficilement modélisables dans l'approche relationnelle (par exemple multi-parentalité). **Pas de raisonnement possible**

PHIS



Le NoSql : la gestion des gros volumes de données

IDÉE

- NoSQL pour “**Not Only SQL**”, littéralement “pas seulement SQL”. Représente l’ensemble des bases de données qui **proposent une alternative ou complètent** les fonctionnalités des SGBDR

BUT

- La **performance et résistance aux pannes**
- Les **Principes CAP** :
 - **C**ohérence
 - Haute disponibilité (**A**vailability)
 - Tolérance au **P**artitionnement

Particularités

- Ne répond plus au modèle relationnel : **perte du Contrôle d'intégrité** (vérification de type, unicité, cohérence ..).
- 4 grandes familles : orienté « **Clé/valeur** », orienté « **colonne** », orienté « **document** », orienté « **graphe** ».

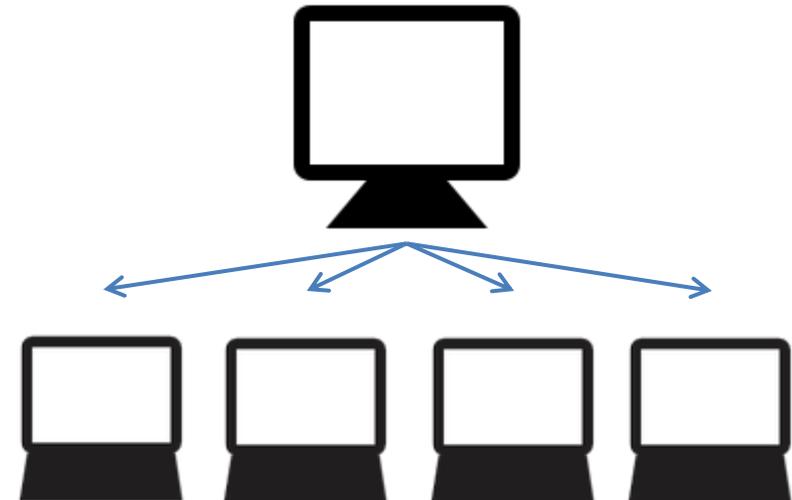


Données OnLine



MongoDB

Orienté Document
Format BSON
Structure Sharding



PHIS – Web Sémantique

IDÉE

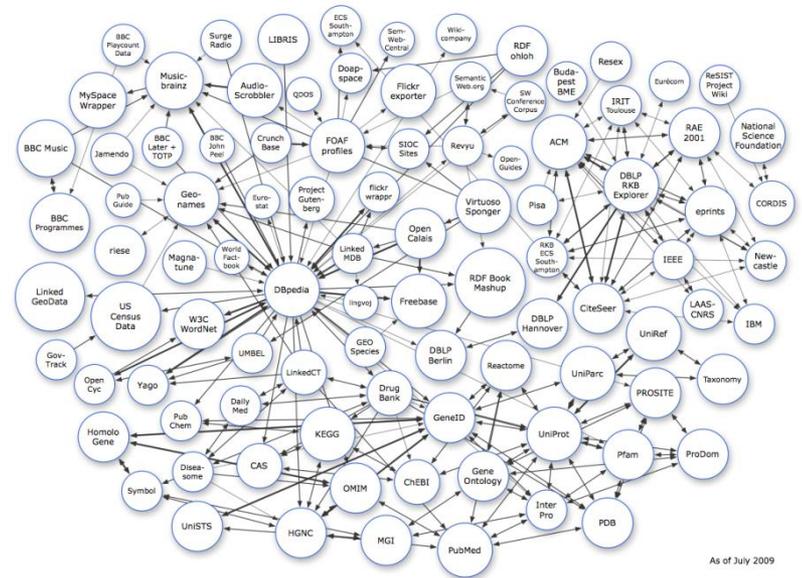
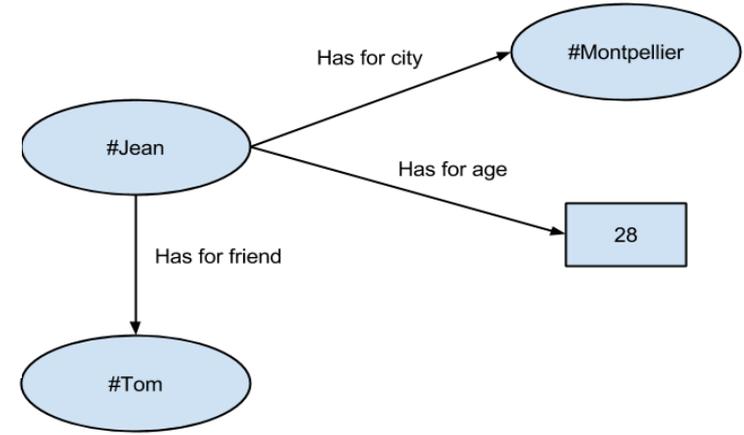
- Fournir aux machines des données qu'elles pourraient **comprendre** et à partir desquelles elles pourraient **tirer de nouvelles connaissances**.
- Les applications pourraient extraire des informations de **différentes sources et les combiner**.

BUT

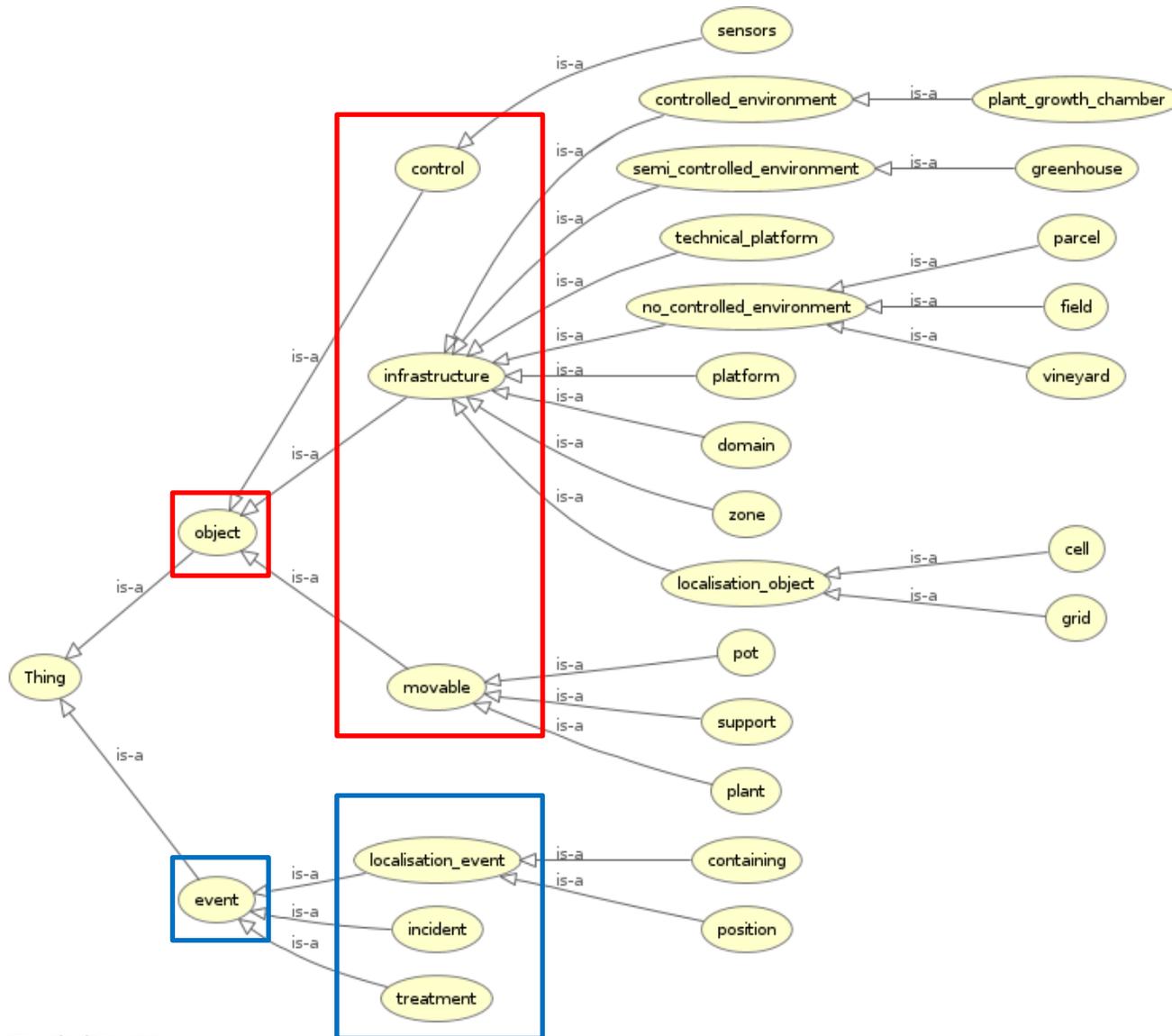
Fournir un **moyen standardisé d'utiliser et de publier des données** facilitant l'obtention et l'intégration de ces données.

INTÉRÊT

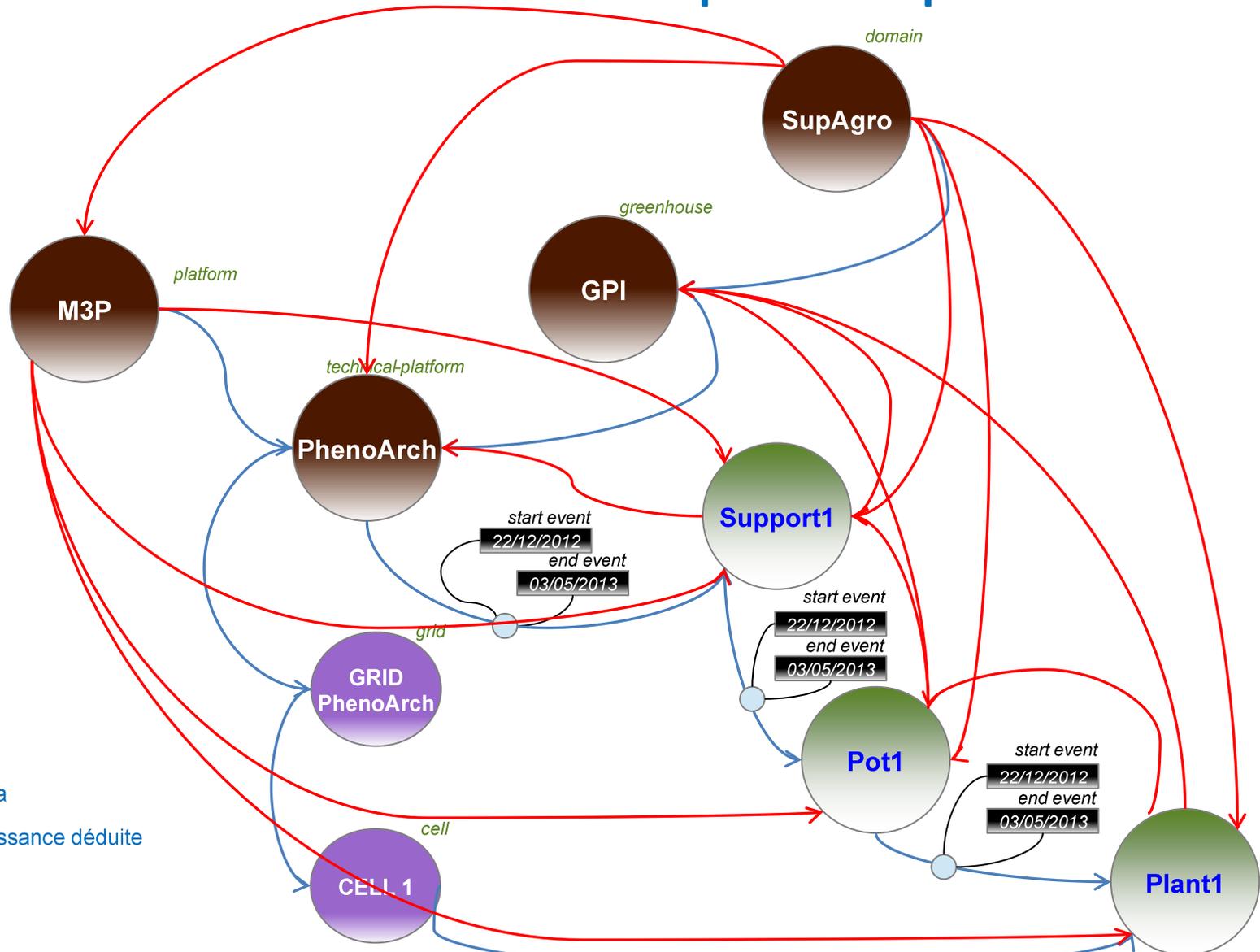
- Difficilement définissable en BDDR :
- Le **concept « objet »** (hiérarchiser et localiser les mesures)
 - **Suivi temporel**
 - **Facilité de requêtage** (parcours de graphe grâce à SPARQL).



PHIS – Web Sémantique : les concepts



PHIS – Web Sémantique : exemple



PHIS

