

# Intégration des données de terrain du projet DeepImpact

Mise en application de l'ontologie sosa (« Semantic Sensor Network Ontology »)

<https://w3c.github.io/sdw-sosa-ssn/ssn/>

<https://www.w3.org/TR/2020/WD-vocab-ssn-ext-20200116/>

# DeepImpact

- Projet de recherche visant à identifier et valider des consortia de microorganismes favorables à la santé des plantes ou à la performance des cultures
- Divisé en plusieurs volets, dont le premier est dédié à l'acquisition de données environnementales d'un ensemble de parcelles agricoles :
  - Pratiques agricoles (bio/conventionnel, élevage, rotation ...)
  - Physico-chimie des sols
  - Biomasses et rendements des plantes cultivées
  - Inventaire des plantes adventices et des bioagresseurs (champignons pathogènes, insectes phytophages...)
  - Abondances de microbiennes (bactéries+champignons) par métagénomique
  - Variables climatiques

# Différents niveaux d'échantillonnage

- 200 champs provenant de trois régions de France (Bretagne Occitanie, Bourgogne-Franche-Comté)
- Deux campagnes d'échantillonnage (~Deux années distinctes), chacune divisées en deux saisons d'échantillonnage
  - (Pas de suivi temporel entre campagne 1 et campagne (une parcelle = une seule campagne))
- Chaque champ agricole est divisée en plusieurs plots de prélèvement
  - + Les données climatiques sont acquises via Météo France selon les mailles de la grille Safran : chaque champ est relié à un numéro de maille)
- Certaines données sont collectés annuellement (une fois par campagne, sans considération pour les saisons)
- Certaines données sont collectées à l'échelle du champ, d'autres des plots, d'autres à partir de plantes échantillonnées dans les plots, et d'autres à partir d'organes de plantes (!)

**Les données ne s'intègrent donc pas toutes au mêmes niveaux spatiaux et temporels, même si ces niveaux sont tous imbriqués**

# SSN / SOSA

« The Semantic Sensor Network (SSN) ontology is an ontology for describing sensors and their observations, the involved procedures, the studied features of interest, the samples used to do so, and the observed properties, as well as actuators. SSN follows a horizontal and vertical modularization architecture, with the core classes and properties defined using minimal axiomatization in a graph called SOSA (Sensor, Observation, Sample, and Actuator) supplemented with additional axiomatization and terms in further graphs. »

(<https://www.w3.org/TR/2017/REC-vocab-ssn-20171019/>)

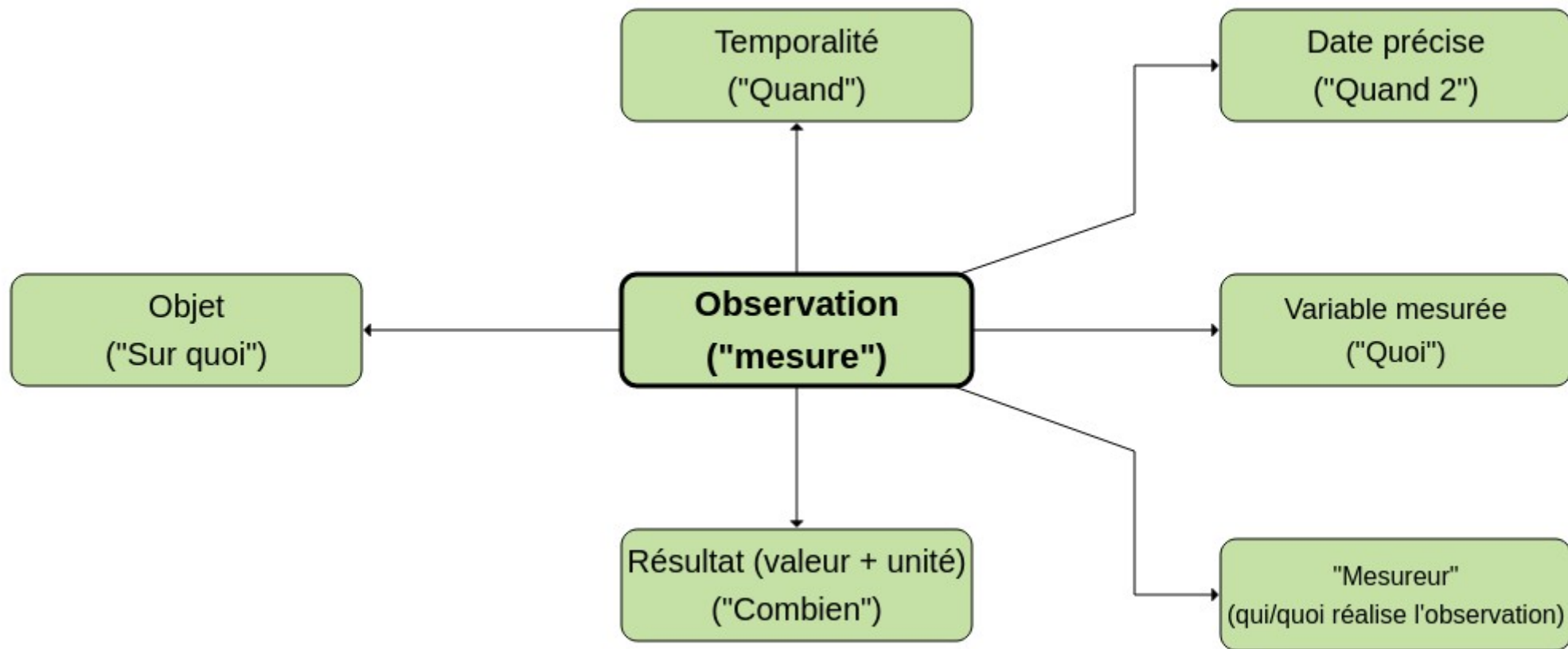
→ **Une ontologie servant à intégrer des données de mesures et d'observations**

# DeepImpact et sosa

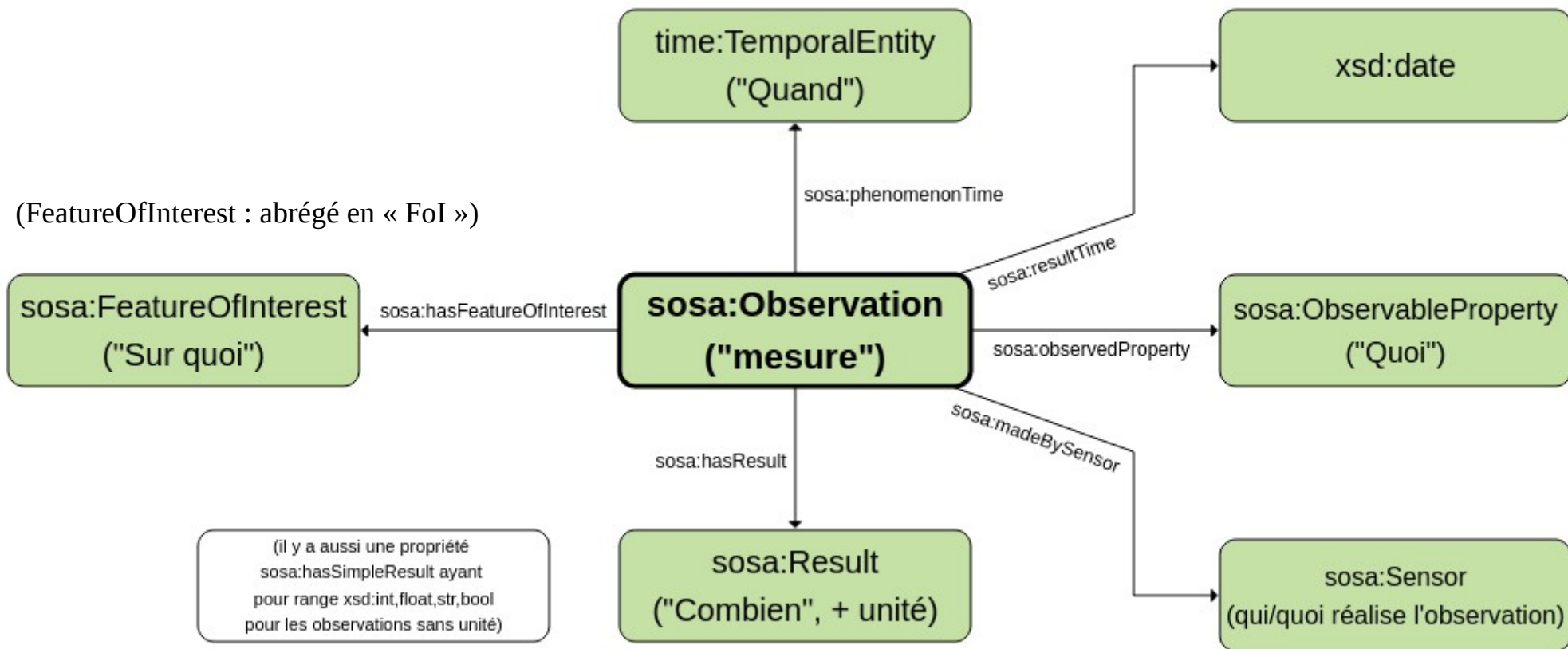
Objectif : intégrer les données de DeepImpact de façon la plus générique possible, c'est-à-dire avec le moins possible de classes/entités spécifiques à deepimpact et le plus possible d'utilisation de classes définies dans des ontologies déjà existantes

- Avec SSN/SOSA pour l'intégration des mesures elle-même
- En faisant le lien avec d'autres ontologies pour décrire les parcelles agricoles, régions, espèces (taxons)

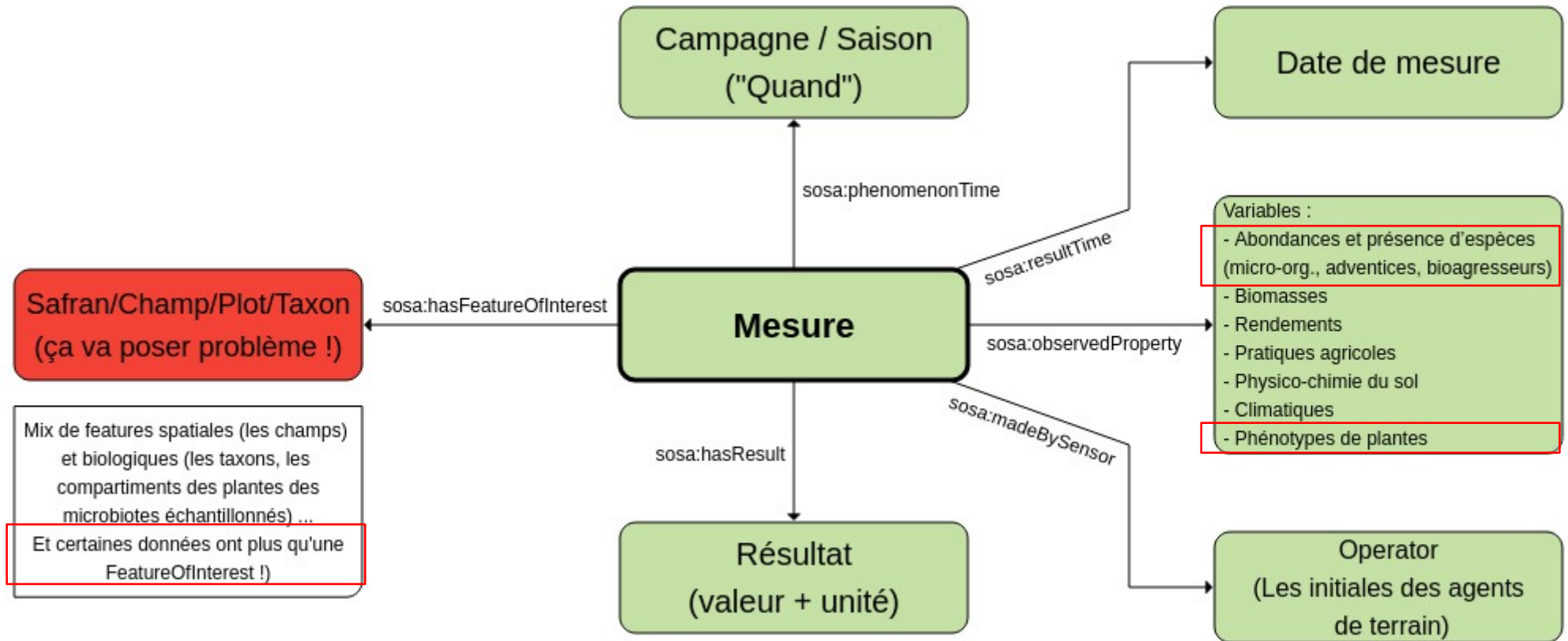
# DeepImpact - De quoi avons nous besoin ?



# Base de SOSA

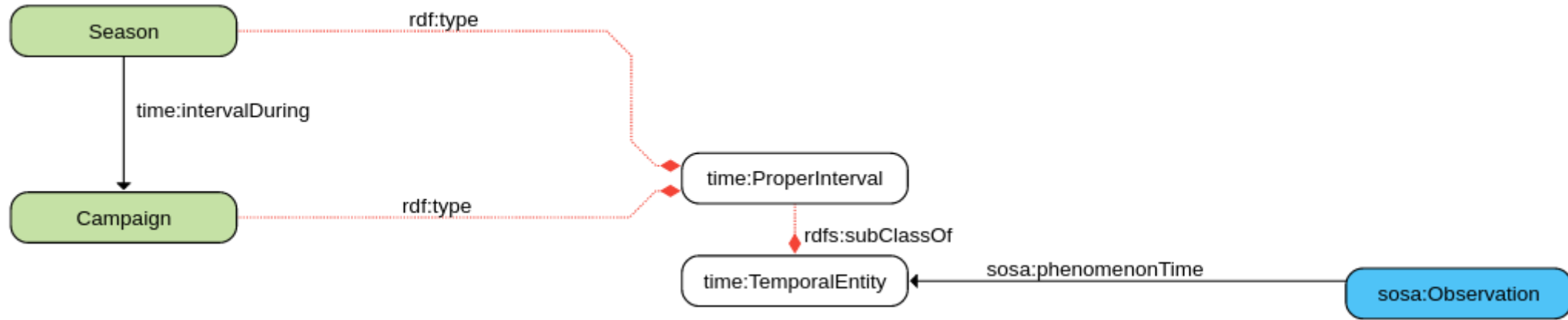


# Matching DeepImpact - SOSA



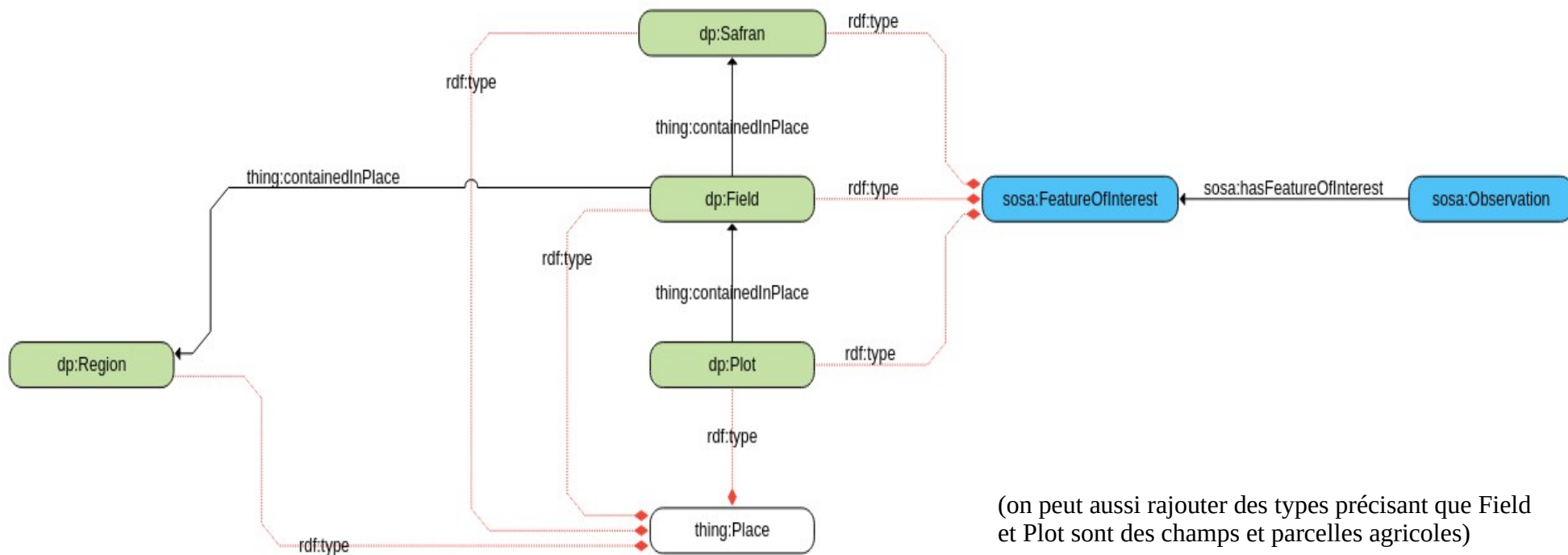


# Sosa / DeepImpact - Temporalité



# Sosa / DeepImpact – Hiérarchie des zones d'échantillonnage

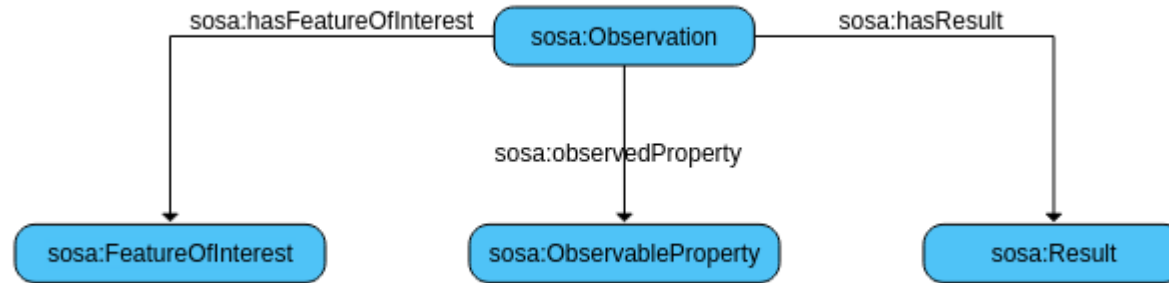
- Établi avec la classe `thing:place` et la propriété `thing:containedInPlace`



# Depuis un tableau de données

Données valides :

- Biomasses
- Rendement
- NIRS (graines)
- Climatiques
- Physico-chimie des sols



PLOT_ID_ANNUAL	ROW_NUMBER	PLANT_DENSITY	BIOMASS_CROP	BIOMASS_WEED	OPERATOR	DATE
AF001-Bn-Y1-PA	3	28	1270	10,75	PLG,CM,LL	20/05/2022
AF001-Bn-Y1-PB	3	23	815	0,6	PLG,CM,LL	20/05/2022
AF001-Bn-Y1-PC	3	17	1225	3,15	PLG,CM,LL	20/05/2022
AF001-Bn-Y1-PD	3	23	1160	0,3	PLG,CM,LL	20/05/2022

# Mais...

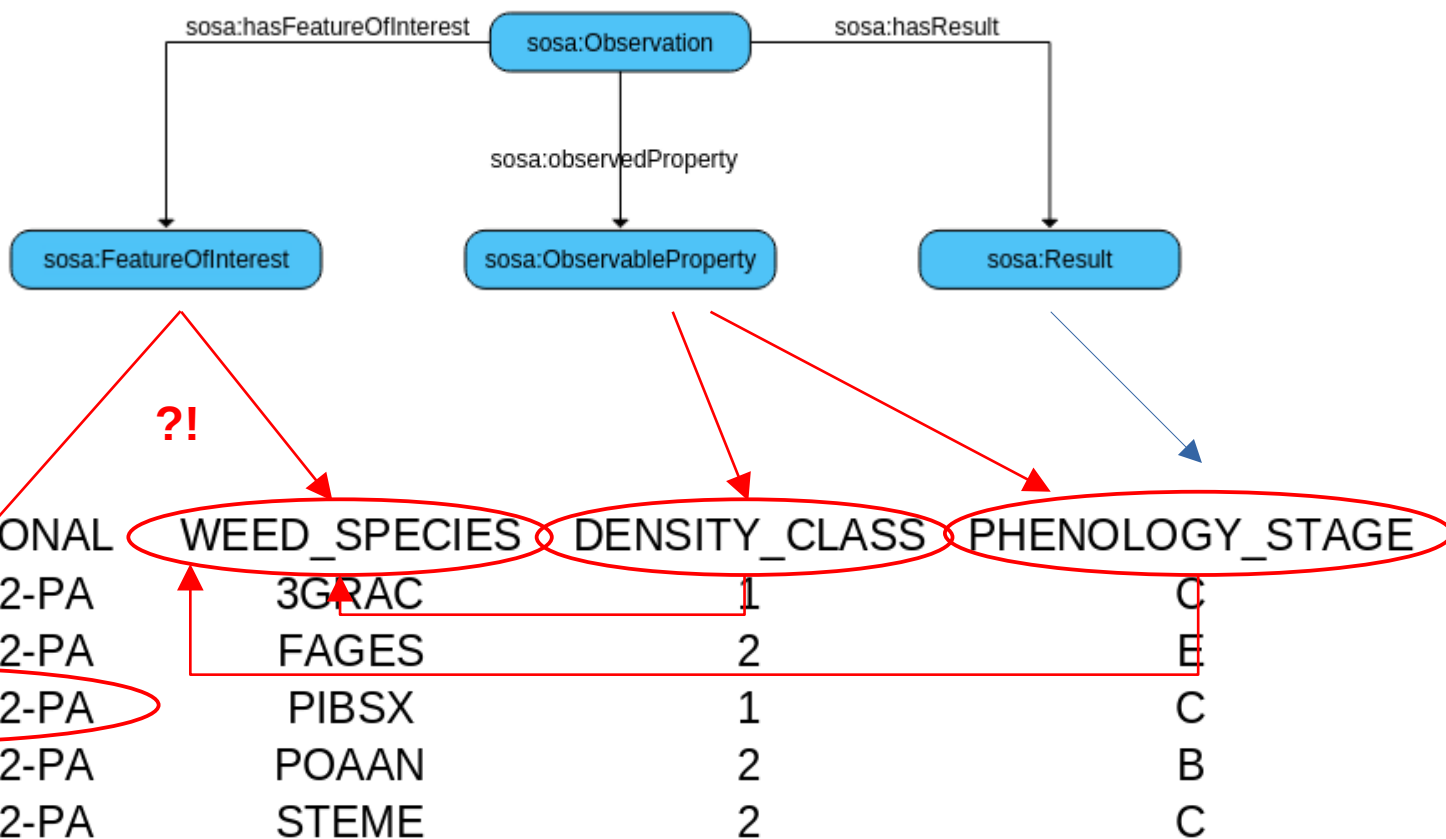
Ce schéma pose deux problèmes majeurs :

- Avec les données qui impliquent des observations à l'échelle des organismes vivants au lieu de l'échelle des champs / plots
- Avec l'imbrication des champs / plots (...) en tant que FeatureOfInterest

# Exemples de données problématiques

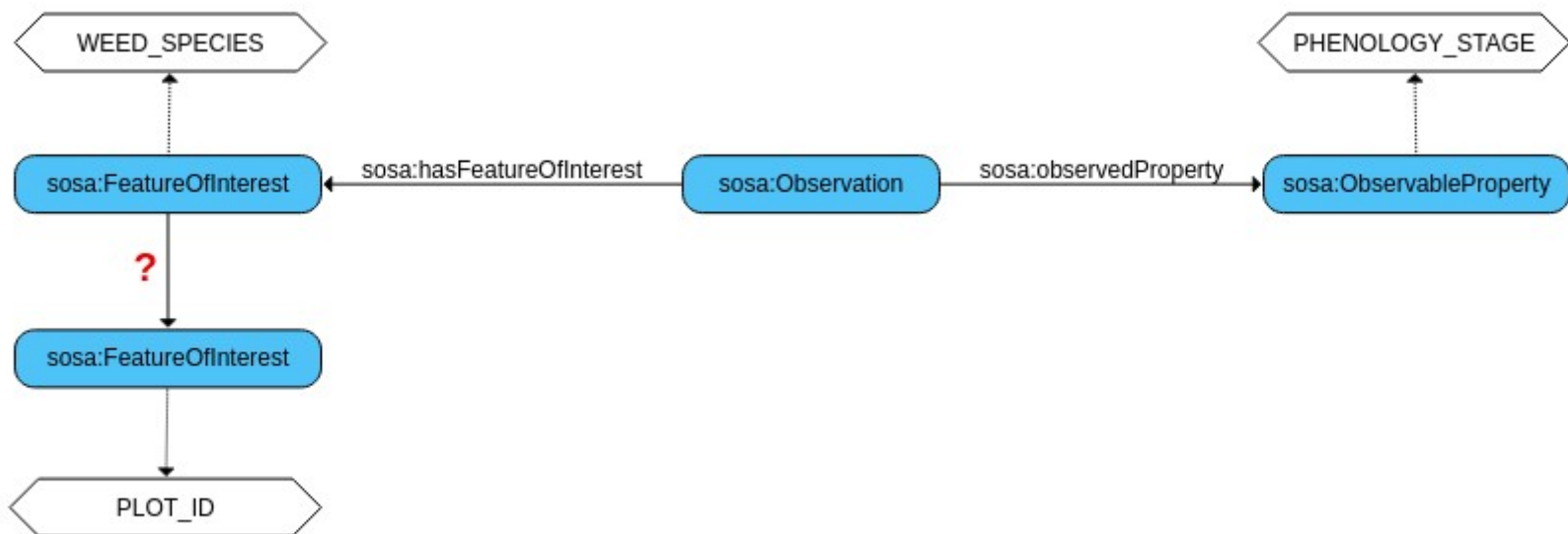
Données mal adaptées:

- Inventaire des adventices
- Abondances des microorganismes
- Inventaire des bioagresseurs
- Phénotypage des plantes échantillonnées



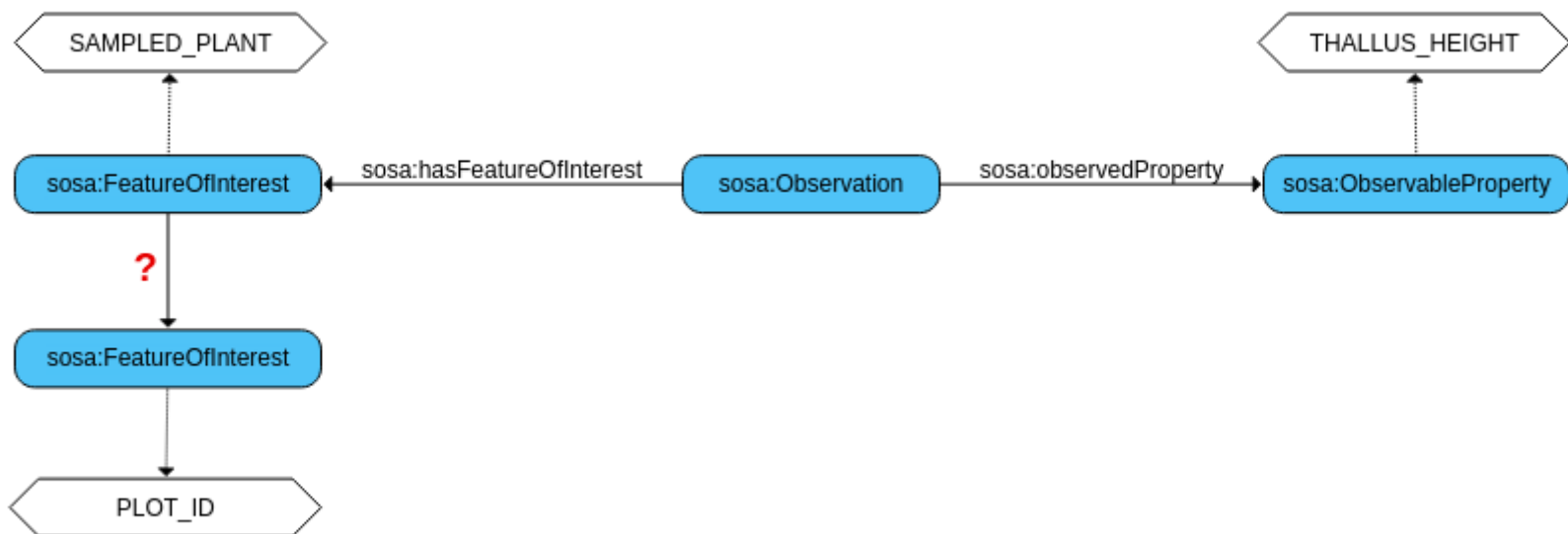
# Exemples de données problématiques

- Weeds : des colonnes ne sont pas liées à l'ID de la ligne (la feature of interest « habituelle ») mais à l'espèce identifiée : combinaison **Espèce** + **Plot**



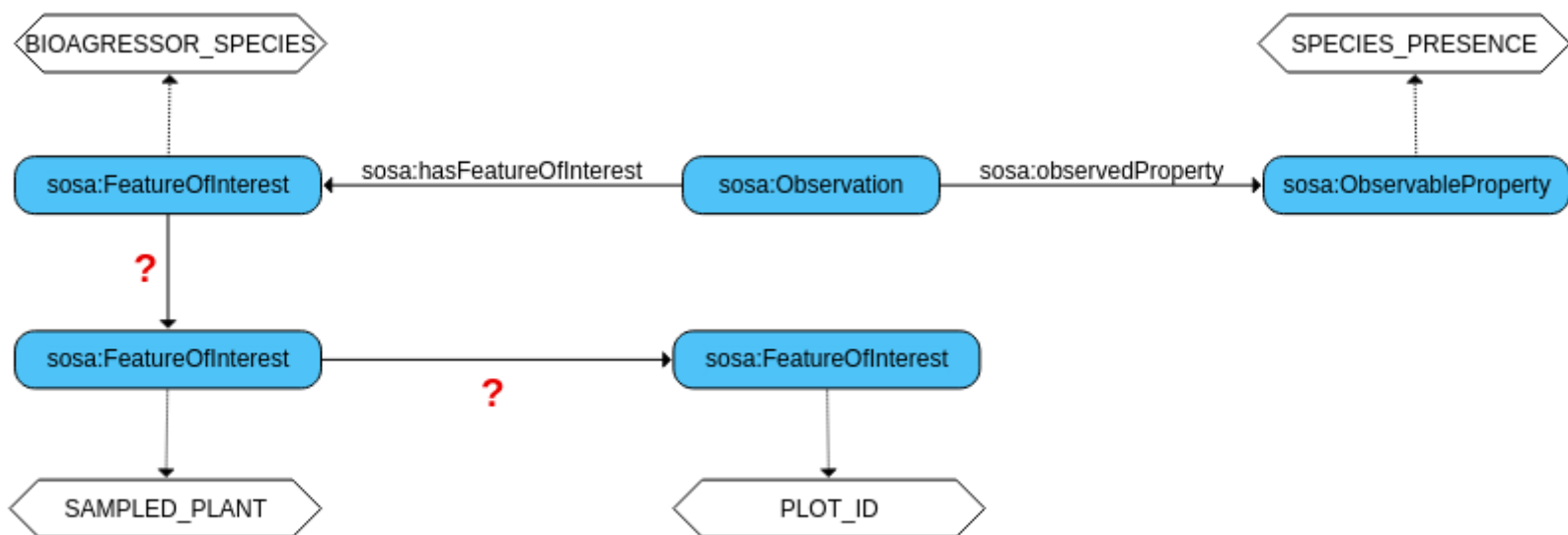
# Exemples de données problématiques

- Bioagresseurs / phénotypage de plantes: combinaison **Plot** + **Plante** prélevée



# Exemples de données problématiques

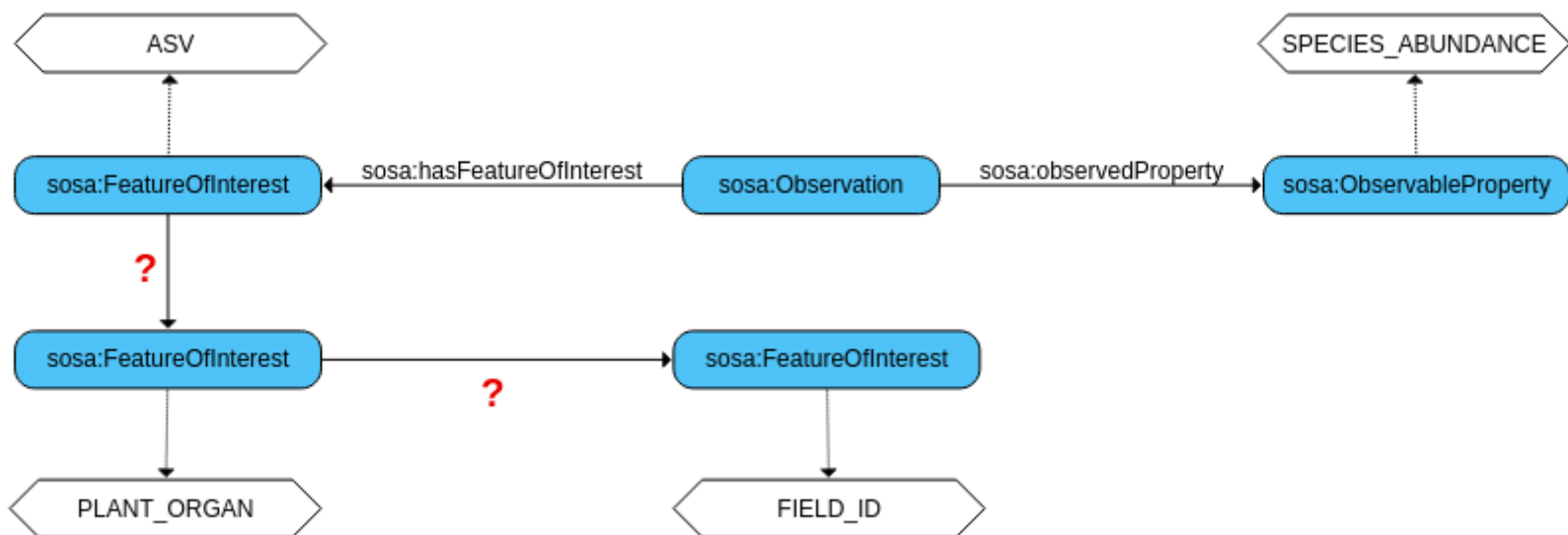
- Bioagresseurs / phénotypage de plantes: combinaison **Plot** + **Plante** prélevée





# Exemples de données problématiques

- Séquençage métagénomique: combinaison abondance de **l'ASV séquencé** + **Organe de plante échantillonné** + **champ échantillonné**



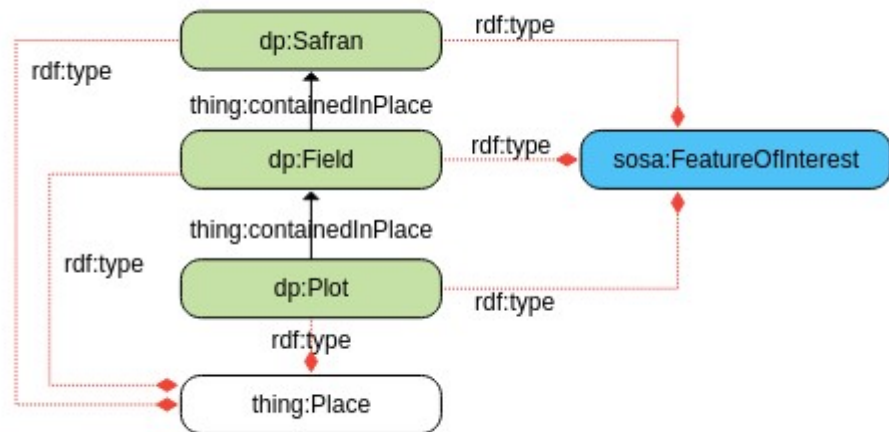
# Les racines du problème

- Sosa ne permet **pas** l'imbrication de plusieurs `sosa:FeatureOfInterest`
- Une `sosa:Observation` ne peut avoir **qu'une seule** `sosa:FeatureOfInterest`



# Double problème

- 1) Le double type `thing:Place` et `sosa:FeatureOfInterest` des champs, plots, et mailles safran : l'un autorise les imbrications, l'autre pas
  - **Q : est-ce que c'est autorisé/habile, ou est-ce que ça pose problème ?**
- 2) Difficile d'incorporer l'espèce observée et d'autres `sosa:FeatureOfInterest` à une même `sosa:observation`



# Comment faire ?

Plusieurs pistes, pas de certitudes sur la faisabilité / la pertinence de chacune, par exemple :

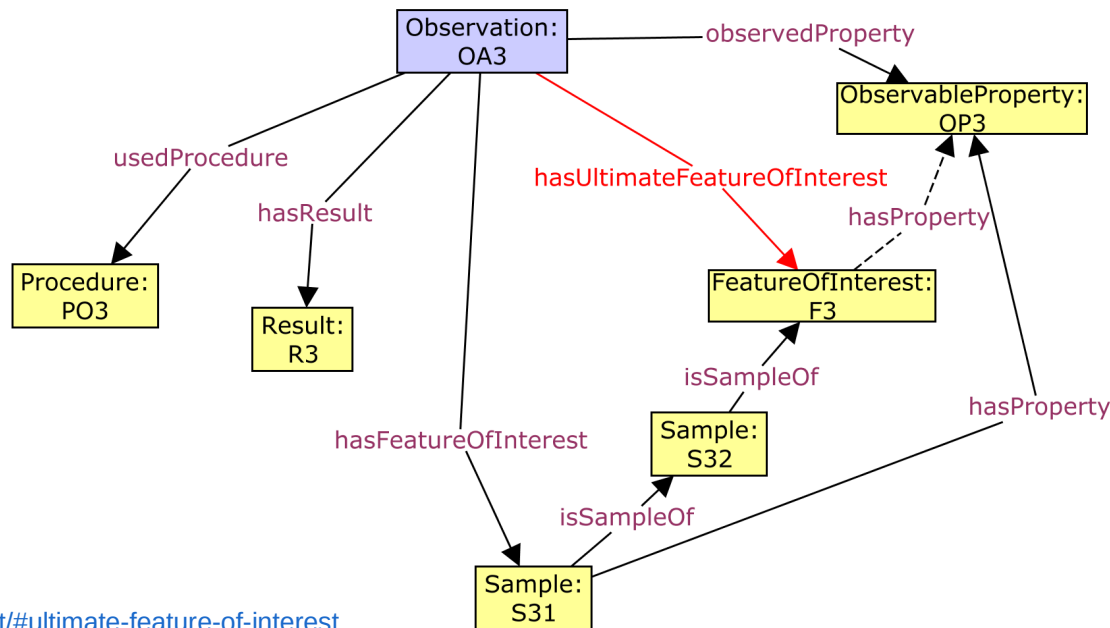
A) Problèmes 1 et 2 : les entités « Ultimate FeatureOfInterest » et « Sample »

B) Problème 1: séparer en plusieurs graphes ?

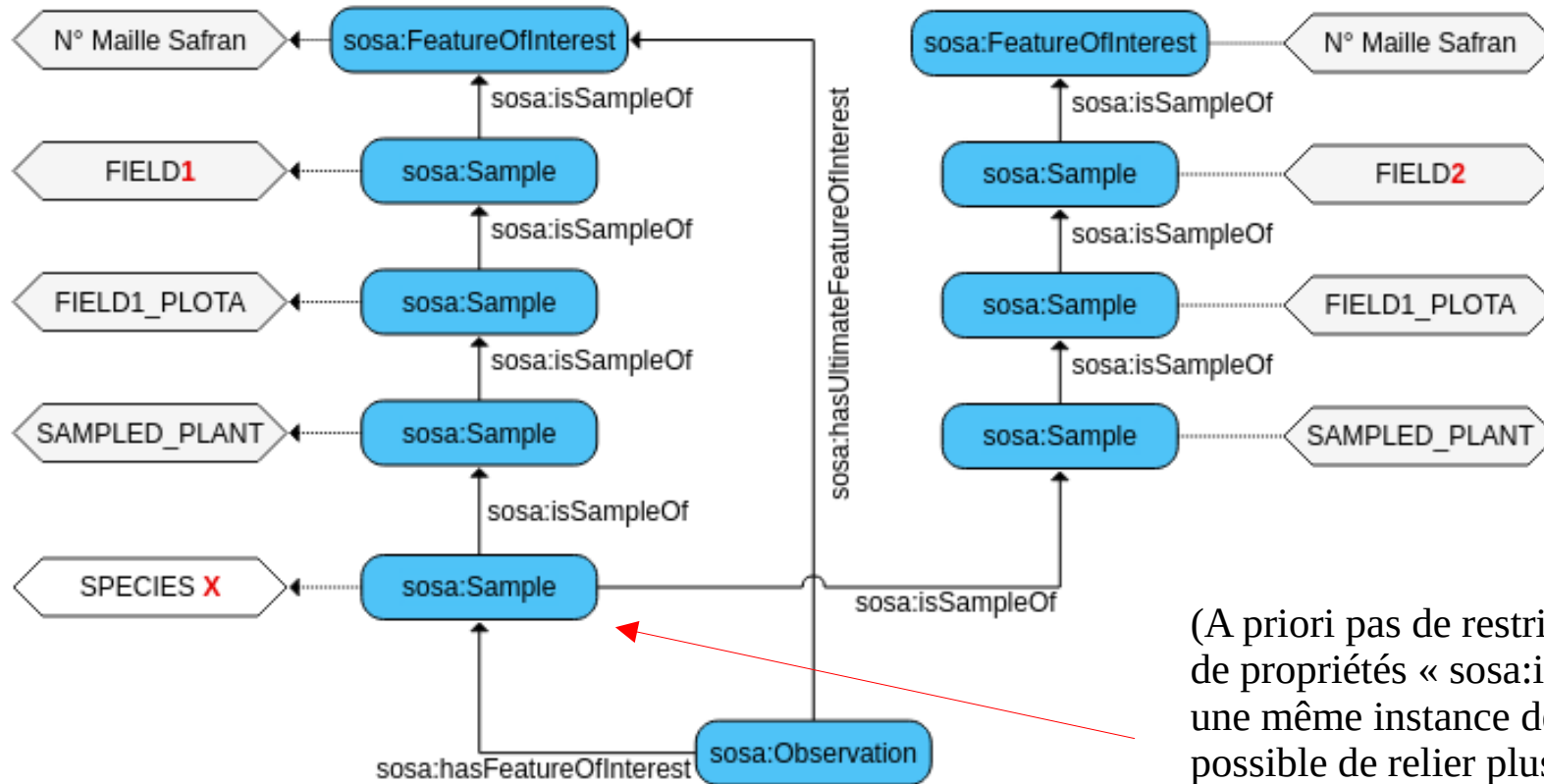
C) Problème 2: rajouter une classe « LivingOrganism » et une propriété la reliant à sosa:Observation ?

# A-Sosa:UltimateFeatureOfInterest

Ainsi que l'entité `sosa:Sample`, une sous-classe de `sosa:FeatureOfInterest` permettent d'imbriquer des éléments de l'observation jusqu'à l'objet d'intérêt de plus haut niveau (données → plante prélevée → plot → champ ...)



# A-Sosa:UltimateFeatureOfInterest



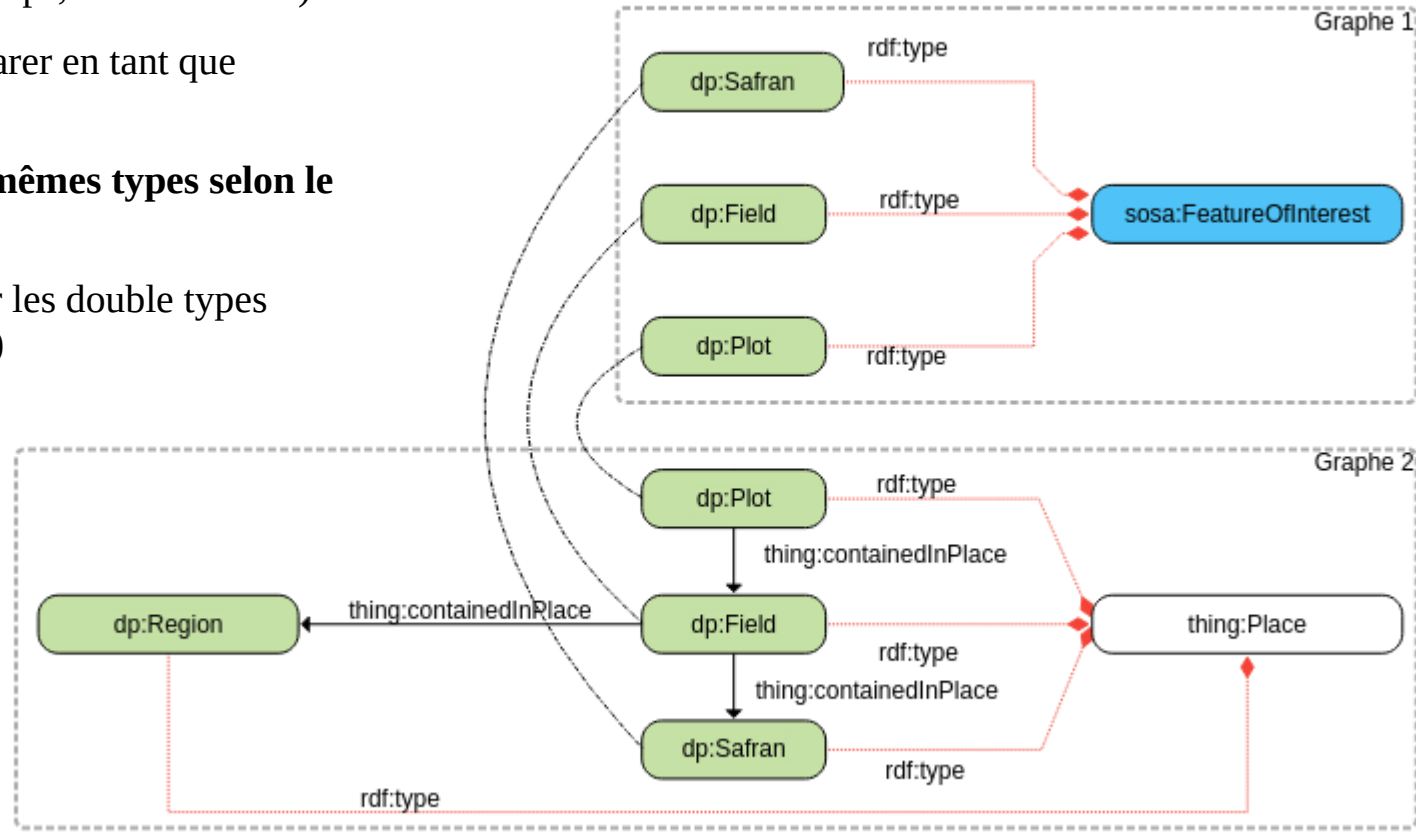
(A priori pas de restriction sur le nombre de propriétés « **sosa:isSampleOf** » pour une même instance de **sosa:Sample** : possible de relier plusieurs champs pour une même espèce)

# A-Mais :

- 1) On utilise le type « Sample » pour des éléments qui ne sont pas réellement des échantillons (plutôt des entités échantillonnées) : les plots, les champs ...
- 2) Le numéro de maille safran est obligatoirement la Ultimate Feature Of Interest, ce qui n'a pas beaucoup d'intérêt pour les requêtes des biologistes (ils travaillent à l'échelle du champ)
- 3) *(Et c'est moche)*

# B-Préciser les liens avec plusieurs graphes ?

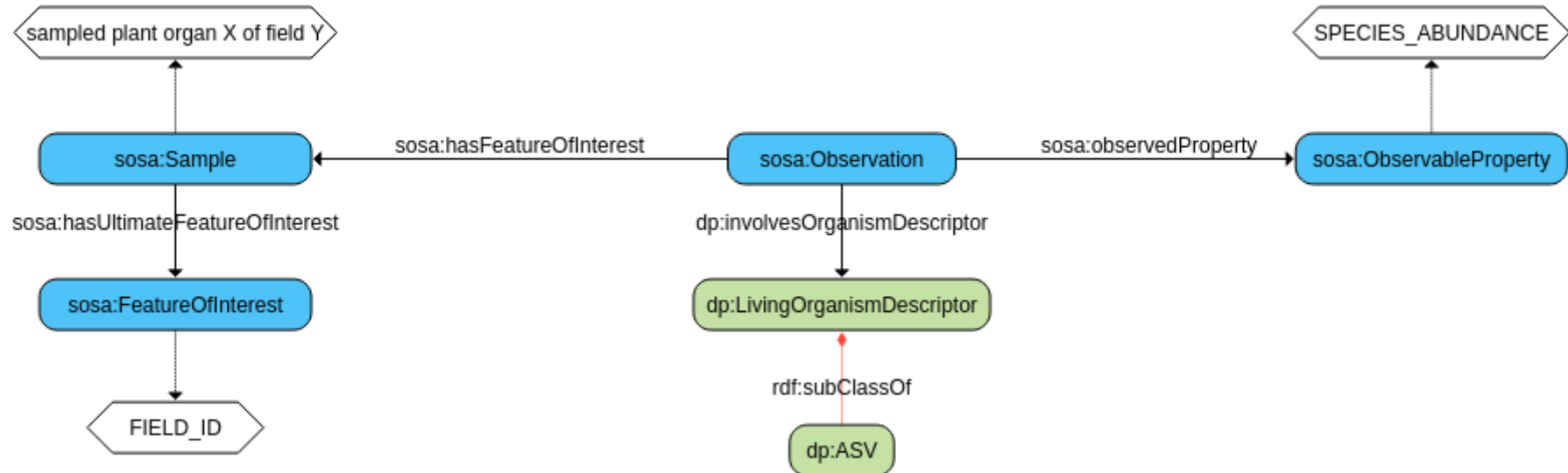
- Un premier graphe utilisant thing:Place pour hiérarchiser les features spatiales (plots, champs, mailles safran)
- Un second graphe pour les déclarer en tant que `sosa:FeatureOfInterest`
- **Q : Possible ou pas ? (pas les mêmes types selon le graphe...)**
- (Pas besoin si c'est ok de garder les double types `thing:Place` et `sosa:FeatureOfInterest`)



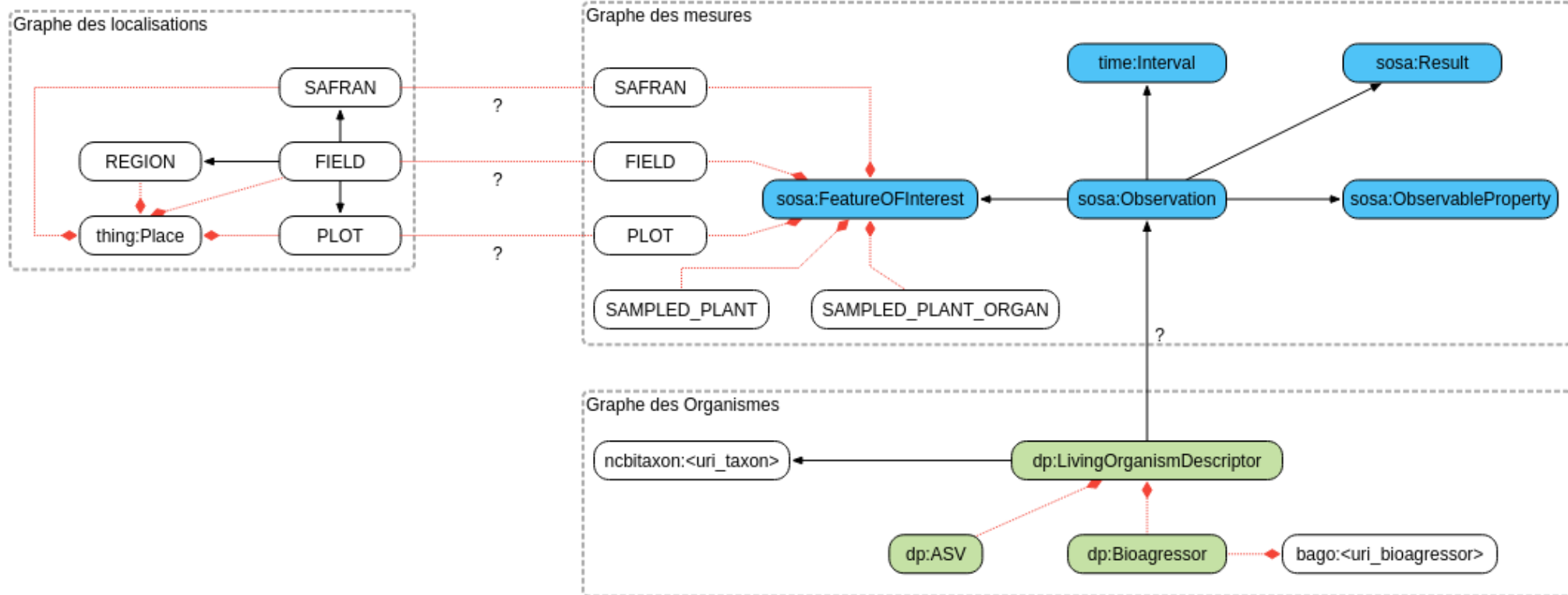


# C-Une classe et une propriété supplémentaire

- Créer une entité « LivingOrganismDescriptor » (qui n'appartient pas à sosa) pour regrouper tout ce qui sert à identifier un organisme vivant, et une propriété la reliant à sosa:Observation quand c'est nécessaire → **Q : Est-ce que c'est autorisé ?**
- L'organe de plante séquencé peut être déclaré comme un sosa:Sample, permettant l'utilisation de sosa:hasUltimateFeatureOfInterest
- Solution à combiner avec la solution pour l'imbrication des plots/champs/mailles Safran



# La solution idéale



# Bonus : bidouiller

- Par exemple : Combiner les valeurs espèces de plante adventice / stade phénologique en une seule instance de `sosa:ObservableProperty`
- Ok mais pas très élégant pour les variables catégorielles, pas ok pour les variables numériques
- A combiner avec les autres solutions pour la hiérarchisation field / plot

