

Intégration des données de terrain du projet DeepImpact

Avec l'ontologie sosa (« Semantic Sensor Network Ontology »)

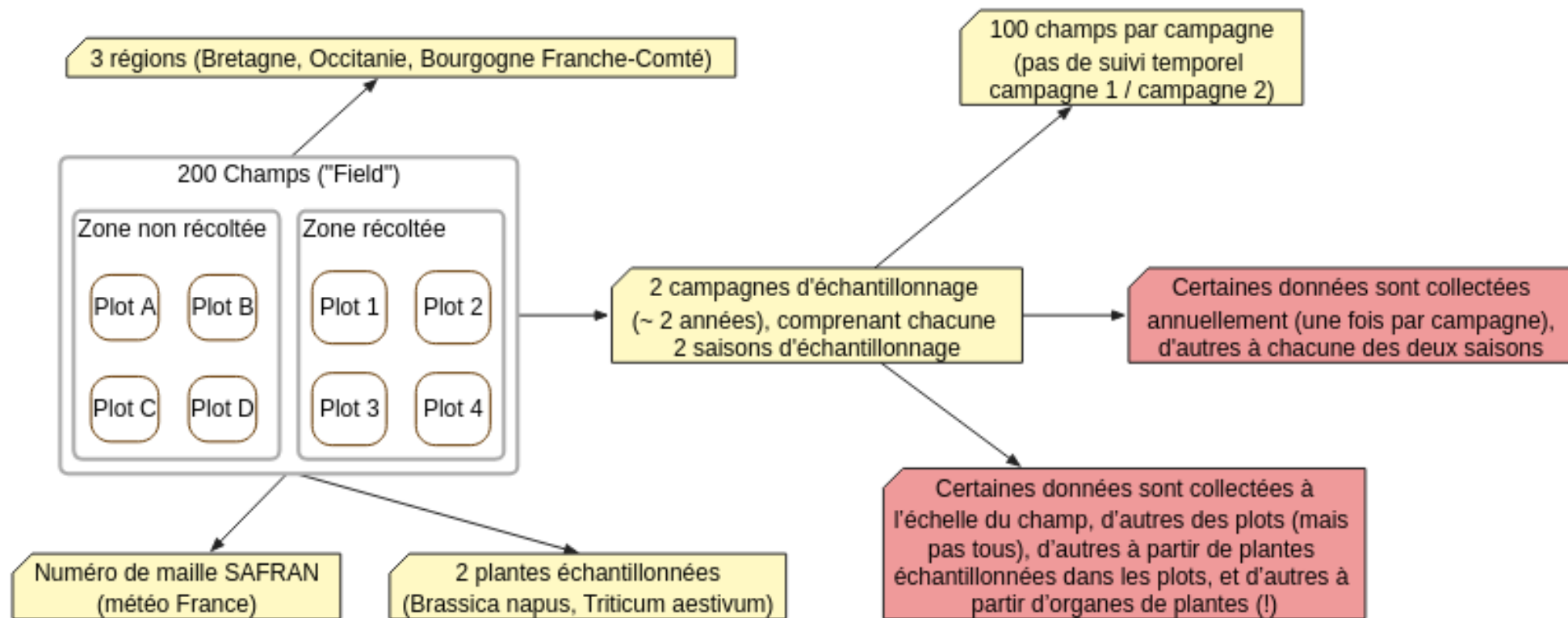
<https://w3c.github.io/sdw-sosa-ssn/ssn/>

<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

Protocole



Les données ne s'intègrent 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/>)

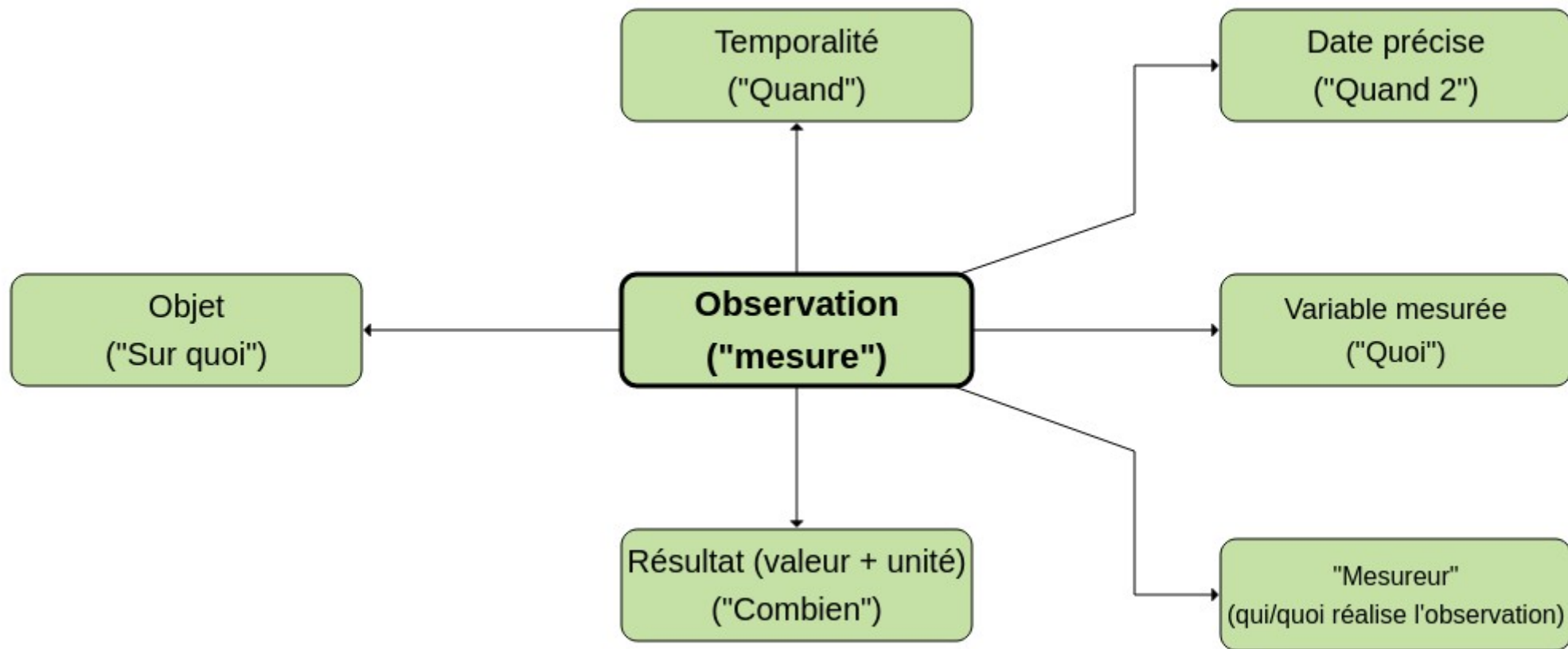
→ **Une ontologie servant à intégrer des données de mesures et d'observations**

DeepImpact et sosa

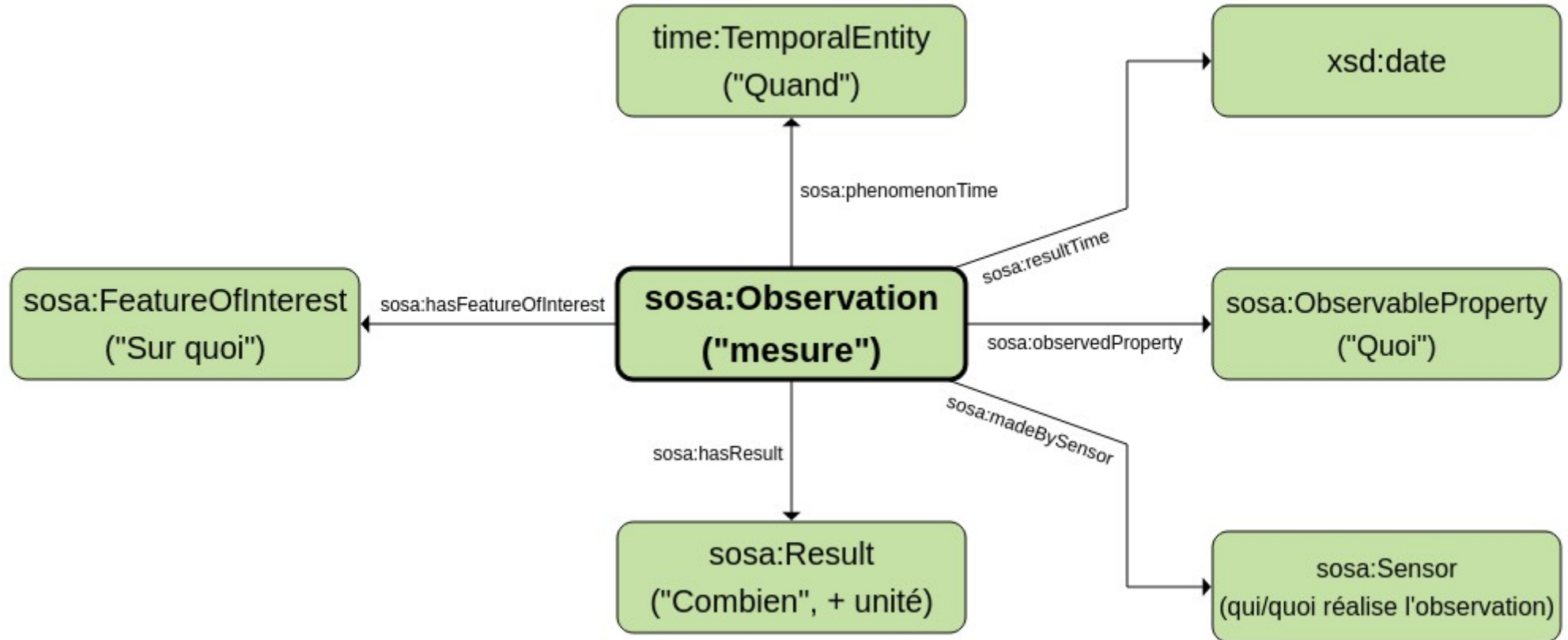
Objectif : intégrer les données de DeepImpact de façon la plus générique possible, c'est-à-dire avec le moins possible de classes/entités spécifiques à deepimpact et le plus possible d'utilisation de classes définies dans des ontologies déjà existantes

- 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)

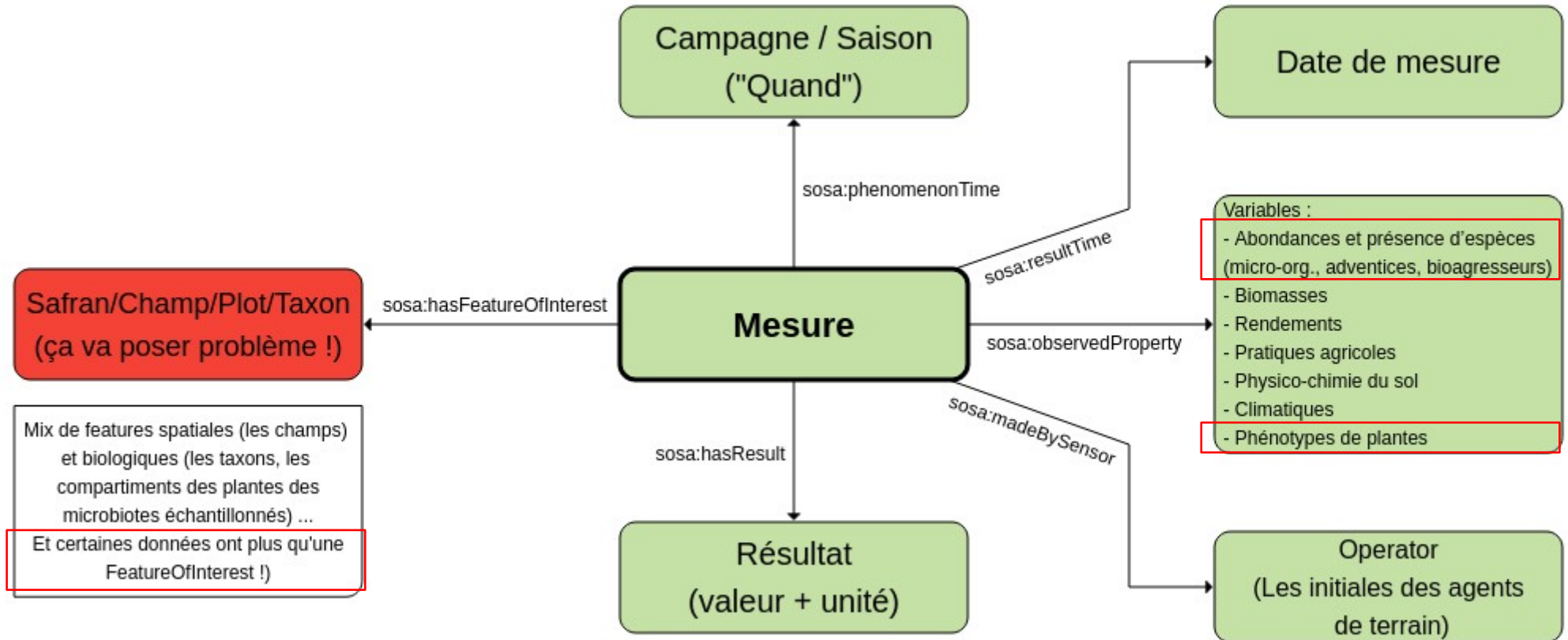
DeepImpact - De quoi avons nous besoin ?



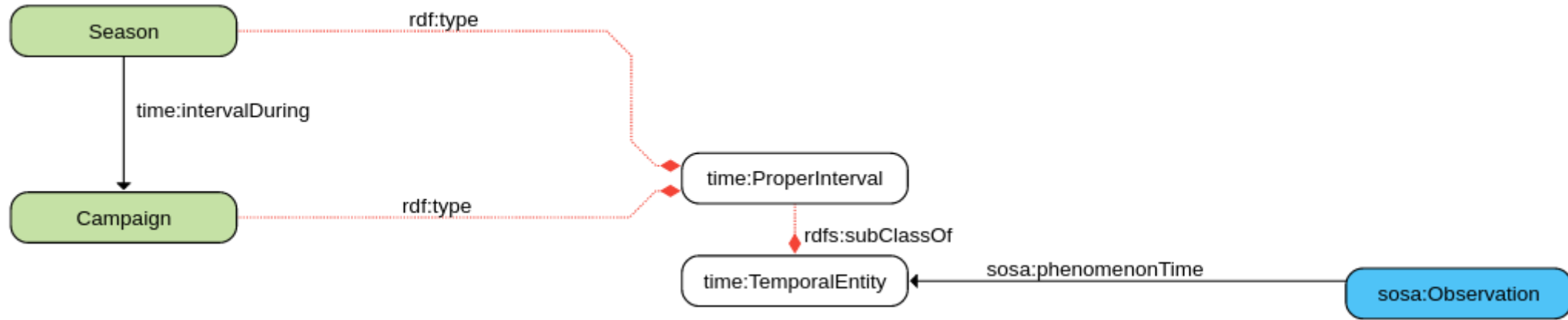
Base de SOSA



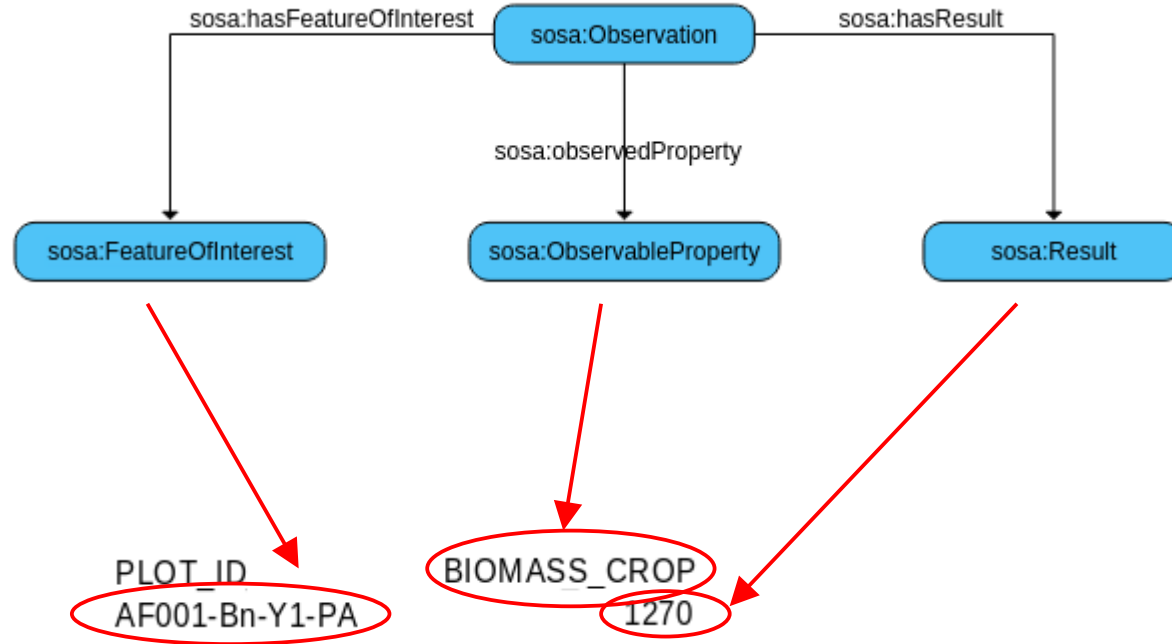
Matching DeepImpact - SOSA



Exemple - Intégration de la temporalité



Depuis un tableau de données



Données valides :

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

Mais...

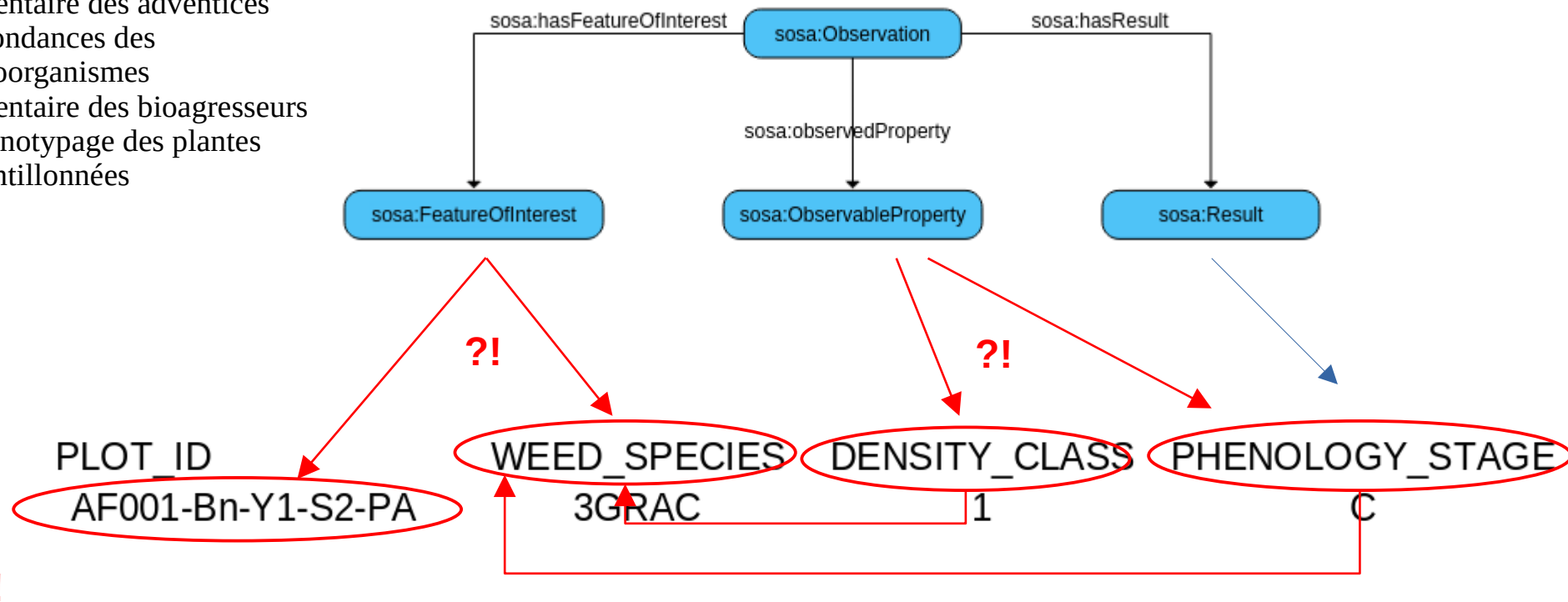
Deux problèmes :

- 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

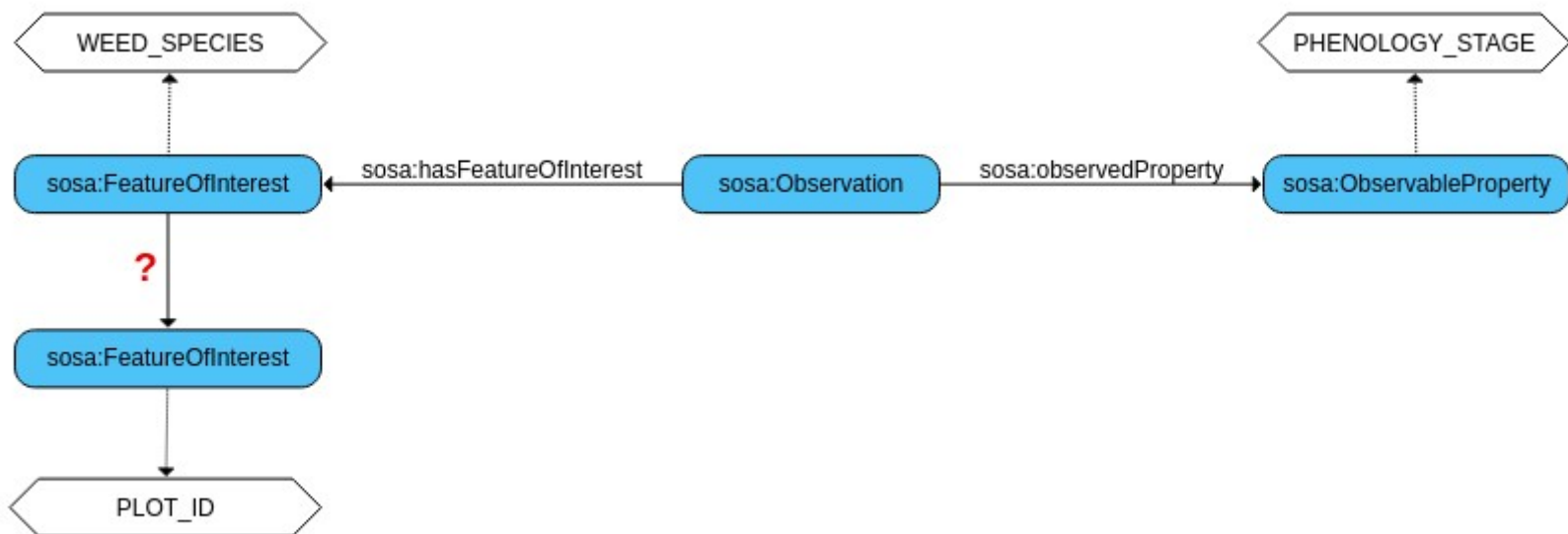
Données mal adaptées:

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



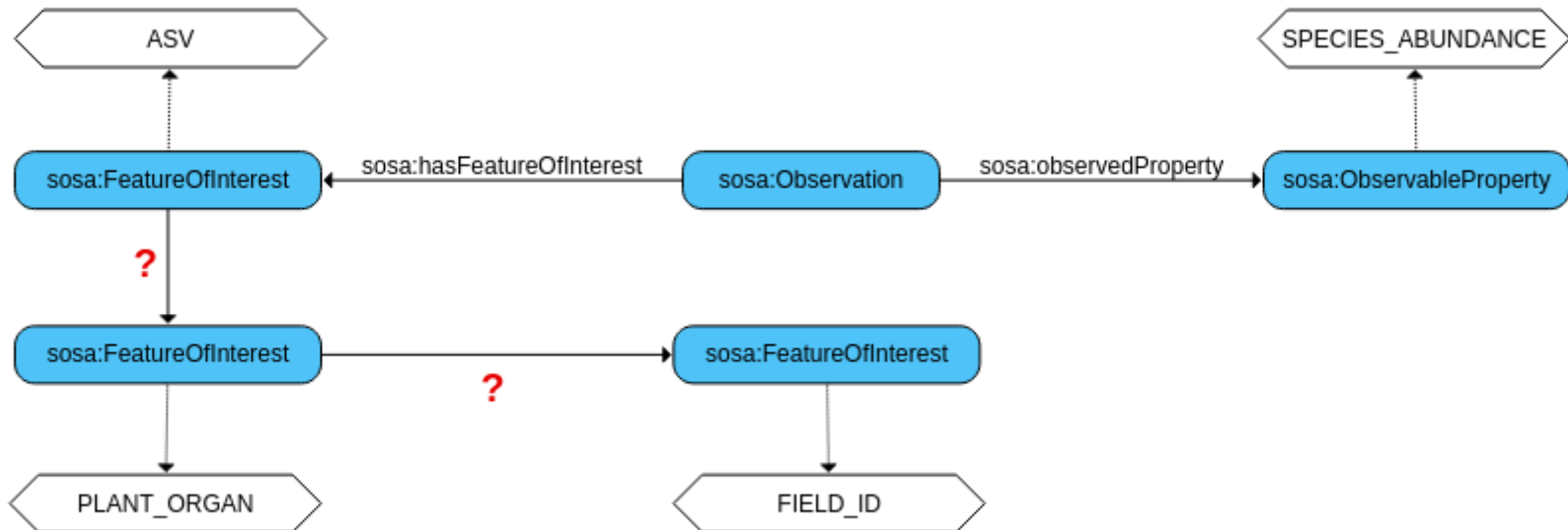
Exemples de données problématiques

- Adventices: des colonnes ne sont pas liées à l'ID de la ligne (le plot) mais à l'espèce identifiée : combinaison **Espèce + Plot**



Exemples de données problématiques

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



Les racines du problème

Selon les cas, champs et plots sont soit une donnée, soit une métadonnée (soit l'objet mesuré, soit la localisation de l'objet mesuré)

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



Double problème

- 1) Comment gérer l'imbrication des Safran/Champ/Plot/Plante prélevée/Organe de plante séquencé ?
- 2) Difficile d'incorporer l'espèce observée et d'autres `sosa:FeatureOfInterest` à une même `sosa:observation`



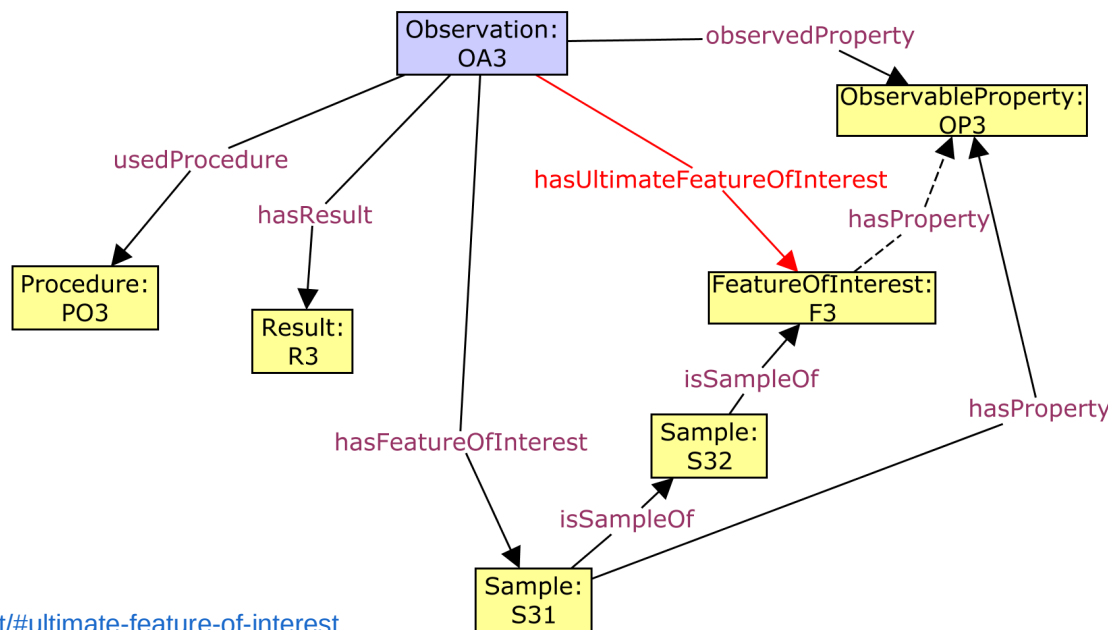
Comment faire ?

Plusieurs possibilités, par exemple :

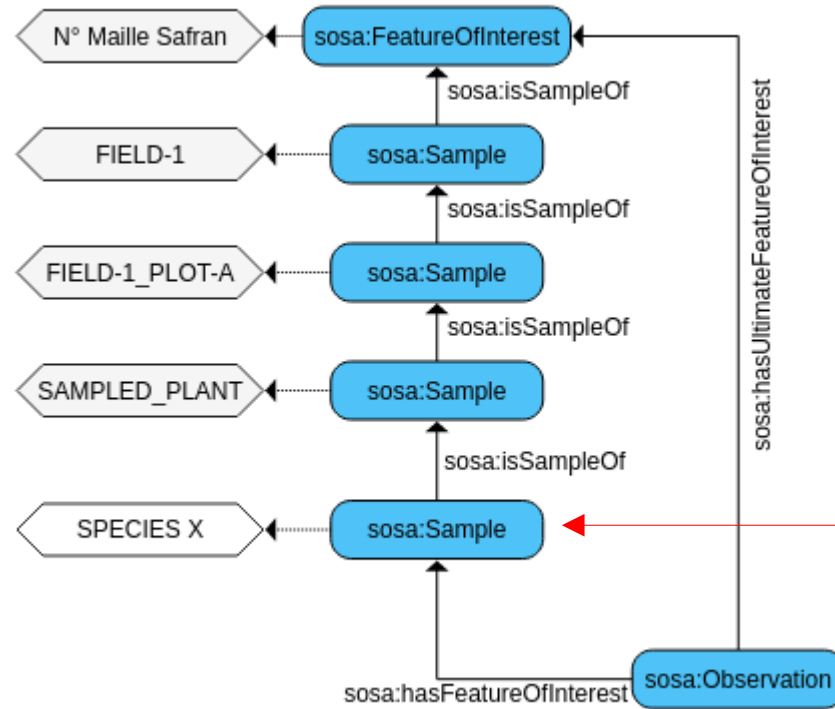
- A) Utiliser uniquement Sosa et la classe « *Sample* » et la propriété « *Ultimate FeatureOfInterest* »
- B) Rajouter une classe « *LivingOrganism* » et une propriété la reliant à *sosa:Observation* ?
- C) Déclarer la hiérarchie spatiale des zones d'échantillonnage dans un autre graphe

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



(Pas de restriction sur le nombre de propriétés « `sosa:isSampleOf` » pour une même instance de `sosa:Sample` :

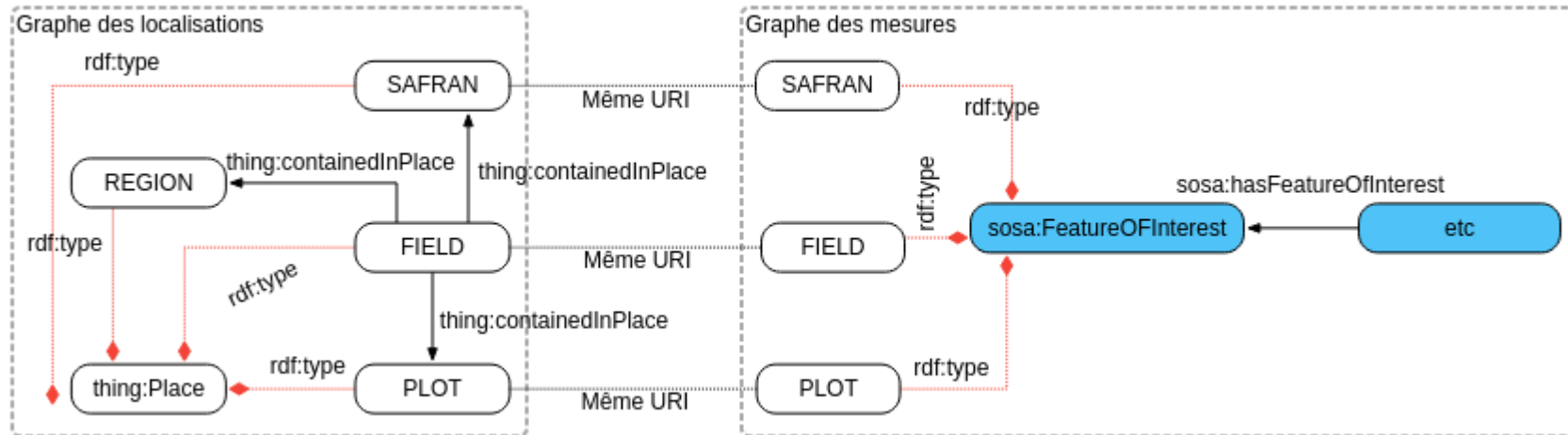
→ possible de relier plusieurs champs/plantes pour une même espèce)

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) Les samples imbriqués diffèrent selon le type de données (parfois pas de plot, parfois pas de plante échantillonnée, etc)
- 3) Le numéro de maille safran est obligatoirement la Ultimate Feature Of Interest quel que soit le type de donnée, ce qui n'a pas beaucoup d'intérêt pour les requêtes des biologistes (la grille safran n'est utilisée que pour les données climatiques)
- 4) *(Et c'est moche)*

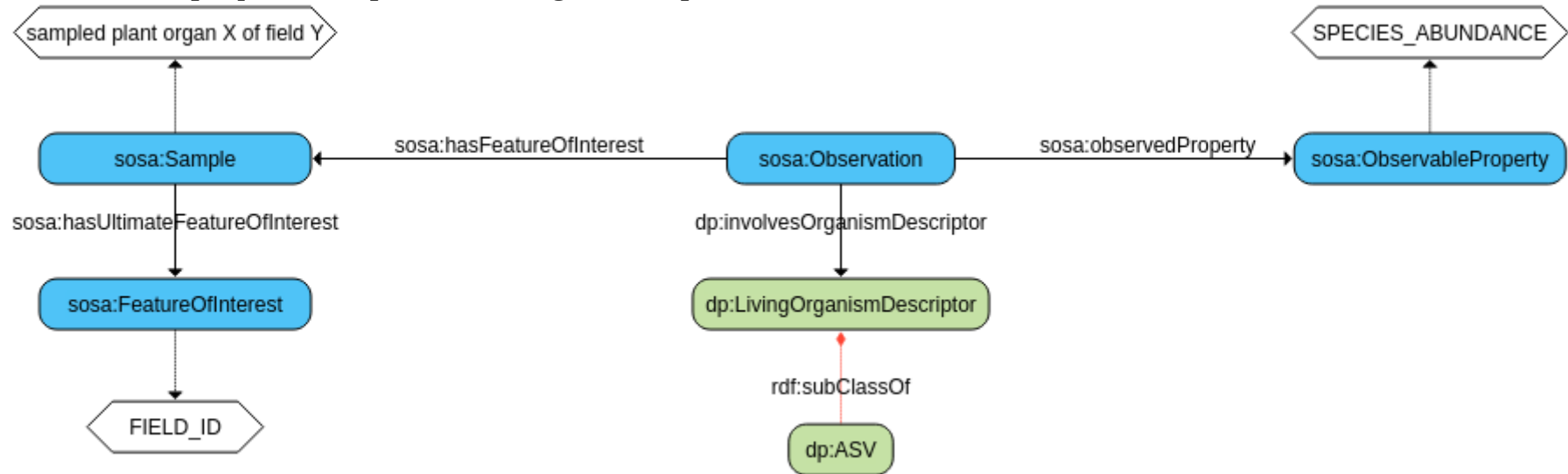
B-Préciser les liens avec plusieurs graphes ?

- 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*
- **Pas les mêmes types selon le graphe, mais les mêmes URIs → Q : Possible ou pas ?**



C-Une classe et une propriété supplémentaire

- Etendre le modèle en créant une classe « LivingOrganismDescriptor » 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 l'utilisation de `sosa:Sample` et `sosa:hasUltimateFeatureOfInterest` pour les données impliquant des plantes ou organes de plantes échantillonnés



La solution ?

