

# Pour une gestion des données de recherche en environnement : l'ontologie méta-observatoire

MARIE-LAURE BETBEDER , SYLVIE DAMY, BÉNÉDICTE HERRMANN



# Plan

---

- Les données de recherche au laboratoire Chrono-environnement
- Constats
- Notre proposition le méta-observatoire
- L'existant
- Notre modèle

# Gestion des données de recherche d'un laboratoire

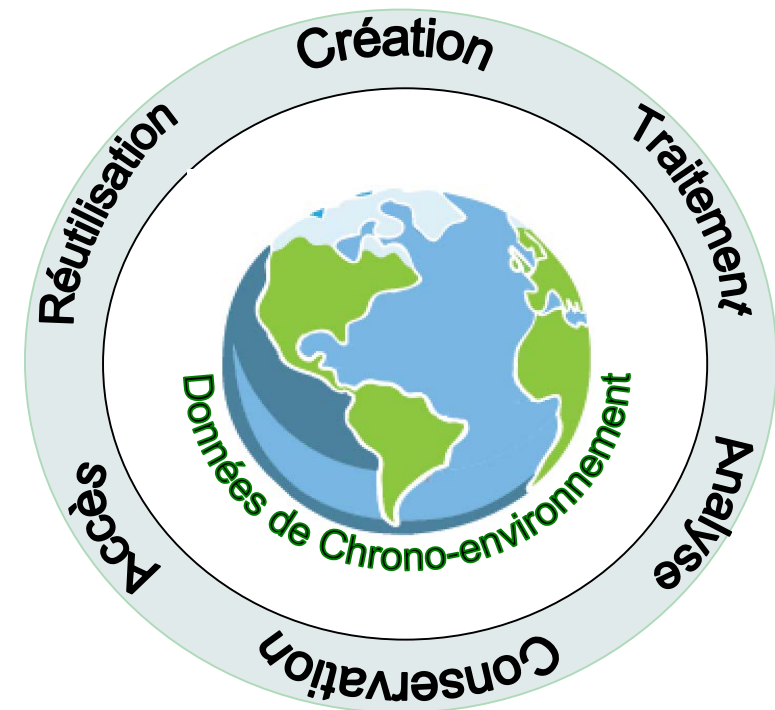
---

Gérer les données tout au long de leur cycle de vie

Plusieurs étapes

Axe transversal Bases de données du laboratoire Chrono-environnement

- valorisation des données
  - Traitement
  - Accès



# Valorisation des données

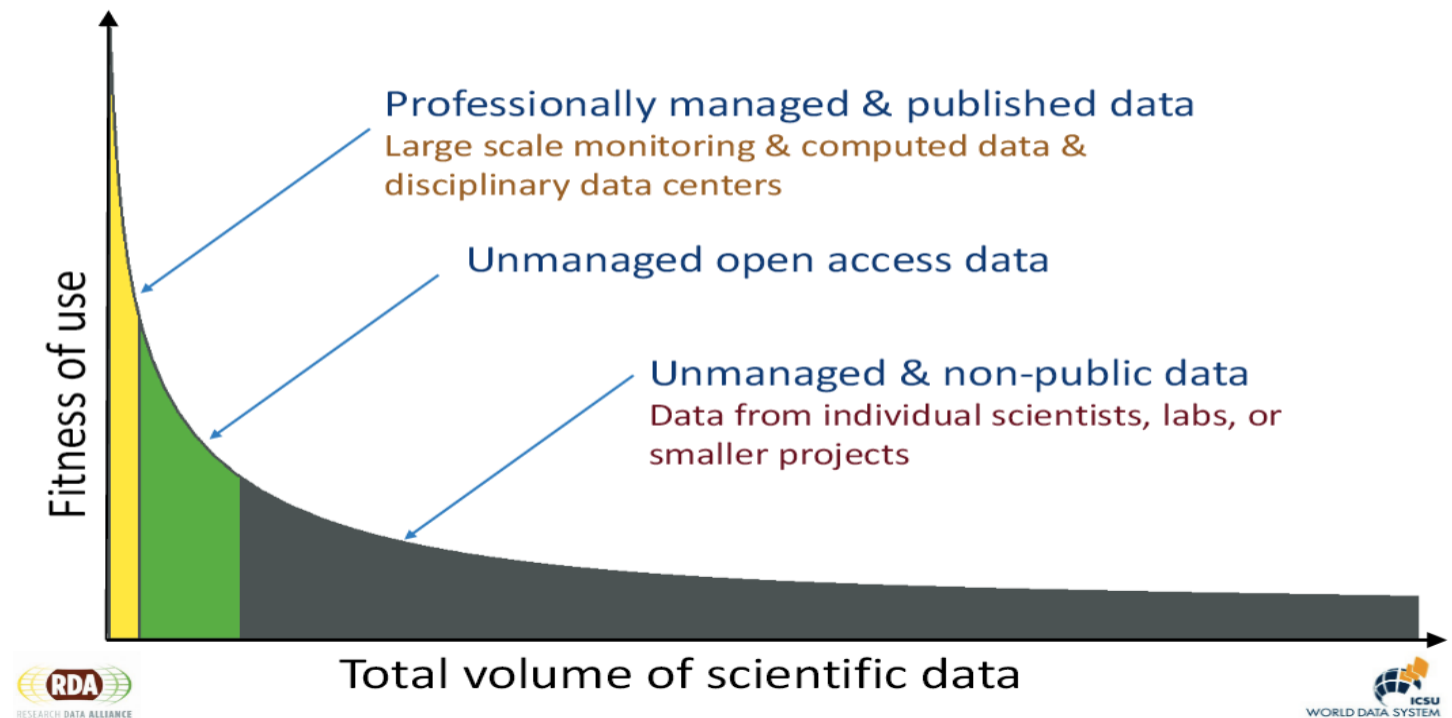
Etude de 2013 (de Waard A. Making Research Data Discoverable and Usable, 2013) :

90% des données de la recherche seraient sur les machines des chercheurs

-> Portail de référencement des données

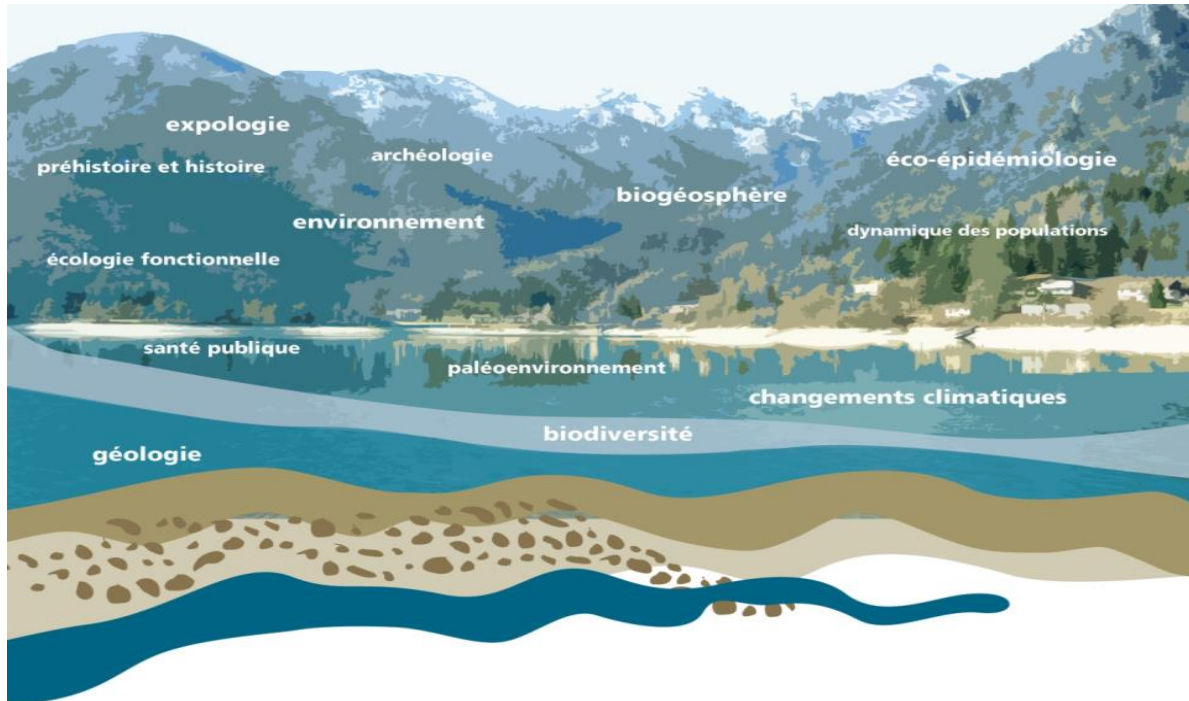
dat@osu

## The Long Tail of Data



# Les données de recherche de Chrono-environnement

---



Données d'observation

Disciplines très diverses

Peu de moyens/ressources pour assurer une gestion systématique des données

Données pas toujours bien organisées (feuille de calcul, ...)

# Les données de recherche de Chrono-environnement

|    | A                 | B     | C                        | D             | E       | F         |          |      |
|----|-------------------|-------|--------------------------|---------------|---------|-----------|----------|------|
| 1  | Date              | Heure | TOC                      | NO3-N         | DOC     | Turbidity |          |      |
| 2  | 2015-06-25        | 16:30 | 1.78                     | 1.49          | 1.28    | 1.46      |          |      |
| 3  | 2015-06-25        | 17:00 | 1.66                     | 1.48          | 1.32    | 0.52      |          |      |
| 4  | 2015-06-25        | 17:30 | 1.72                     | 1.47          | 1.31    | 0.80      |          |      |
| 5  | 2015-06-25        | 18:00 | 1.71                     | 1.49          | 1.35    | 0.93      |          |      |
| 6  | 2015-06-25        | 18:30 | 1.67                     | 1.50          | 1.35    | 0.66      |          |      |
| 7  | 2015-06-25        | 19:00 | 1.62                     | 1.48          | 1.32    | 0.19      |          |      |
| 8  | 2015-06-25        | 19:30 | 1.70                     | 1.49          | 1.32    | 0.86      |          |      |
| 9  | 2015-06-25        | 20:00 | 1.67                     | 1.48          | 1.33    | 0.61      |          |      |
| 10 | 2015-06-25        | 20:30 | 1.67                     | 1.49          | 1.30    | 0.60      |          |      |
| 11 | 2015-06-25        | 21:00 | 1.69                     | 1.50          | 1.32    | 0.93      |          |      |
| 12 | 2015-06-25        | 21:30 | 1.69                     | 1.49          | 1.31    | 0.91      |          |      |
| 13 | 2015-06-25        | 22:00 | 1.62                     | 1.48          | 1.34    | 0.60      |          |      |
| 14 | 2015-06-25        | 22:30 | 1.61                     | 1.47          | 1.27    | 0.46      |          |      |
|    |                   |       | GGUN-FL Fluorometer #396 | Signals in mV |         | 0.15      |          |      |
|    |                   |       |                          |               |         | 0.69      |          |      |
| #  | Time              | R     | Tracer1                  | Tracer2       | Tracer3 | Turbidity | Baseline | Batt |
| 1  | 13/11/27-13:50:09 | 0     | 7.32                     | 1.12          | 4.33    | 29.25     | 13.83    | 0.55 |
| 2  | 13/11/27-14:05:09 | 0     | 7.10                     | 1.09          | 4.30    | 14.71     | 13.93    |      |
| 3  | 13/11/27-14:20:08 | 0     | 7.18                     | 1.10          | 4.25    | 14.50     | 13.78    |      |
| 4  | 13/11/27-14:35:08 | 0     | 7.07                     | 1.09          | 4.14    | 14.53     | 13.82    |      |
| 5  | 13/11/27-14:50:08 | 0     | 7.09                     | 1.09          | 4.22    | 14.57     | 13.82    |      |
| 6  | 13/11/27-15:05:08 | 0     | 7.03                     | 1.10          | 4.16    | 14.51     | 13.80    |      |
| 7  | 13/11/27-15:20:08 | 0     | 7.08                     | 1.08          | 4.18    | 14.59     | 13.82    |      |
| 8  | 13/11/27-15:35:08 | 0     | 7.00                     | 1.07          | 4.11    | 14.54     | 13.82    |      |
| 9  | 13/11/27-15:50:08 | 0     | 7.15                     | 1.08          | 4.23    | 14.60     | 13.81    |      |

Données issues de capteurs

Des millions de mesures dans des milliers de fichiers

## Jeu de données : Jurassic Karst

### Métadonnées générales

**Identifiant :** local : FR-18008901306731-2015-03-08-02

**Description :** Hydrogeochemical monitoring of karst springs.

**Disciplines :** géologie (sciences de l'univers), géosciences (sciences de l'univers), ressources en eau (sciences de l'univers), sciences de l'environnement (sciences de l'univers)

**Mots clés :** altération continentale, géochimie, hydrochimie, hydrologie, karst, Montagnes du Jura, physicochimie, surveillance, zone critique

**Dates :** Obtention des données : depuis 2009 et en cours

Fiche de métadonnées : Création : 01/01/2015  
Mise à jour : 09/02/2016 Publication : 25 sept. 2014

**Rythme de mise à jour :** quand nécessaire **Langue :** Anglais (eng)

Accès aux données

Droits



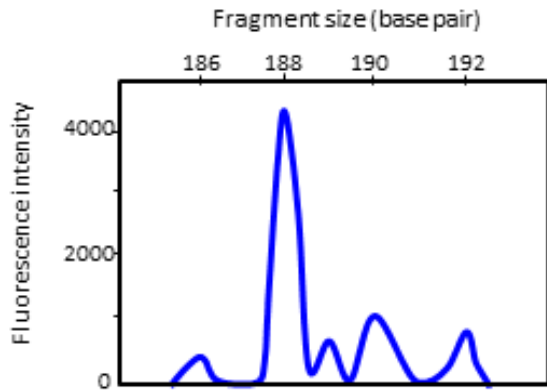
Citation

Chrono-environnement, RBV, SNO Karst, ZAAJ, Catherine Bertrand, Hélène Celle-Jeanton, Jean-Baptiste Charlier, Cybèle Cholet, Sophie Denimal, Marc Steinmann (2014): *Jurassic Karst*. Chrono-

8 sites d'observations  
 Nombreuses sondes sur chaque site  
 Mesures tous les quart d'heure

# Les données de recherche de Chrono-environnement

1  
 GTGTGGATGAGTGTGCCATCCATCACGCAACGTCCATCAATGCACTCTGCTCCA  
 CGCGCACCTGCTCAGCACAGGCCAGTGACACTAACTCGACTCACTCACGCTC  
 ACACACACACACACAGAGAGAGAGGGTGAGAGAGAGAGGGGTGAGAGGA  
 TGACGGCGGTCATGATTGCAGTAGGGAAGGTGGCGTGGTGCAGTTGGAGCCA  
 GTGCCATCCATCACGCAACGTCCATCAATGCACTCTGCTCC



Données résultant de calculs effectués à partir de prélèvements

| 1   | EWET-code1 | EWET-code2 | EWET-code3 | EWET code4 | EV |
|-----|------------|------------|------------|------------|----|
| 574 | Fra        | Dou        | 573        | fox        | 57 |
| 575 | Fra        | Dou        | 574        | fox        | 57 |
| 576 | Fra        | Dou        | 575        | fox        | 57 |
| 577 | Fra        | Dou        | 576        | fox        | 57 |
| 578 | Fra        | Dou        | 577        | fox        | 57 |
| 579 | Fra        | Dou        | 578        | fox        | 57 |
| 580 | Fra        | Dou        | 579        | fox        | 57 |

Données provenant de centaines d'utilisateurs différents (au niveau mondial)

# Axe transversal Bases de données

---

Aide à la gestion des données

A la demande des chercheurs

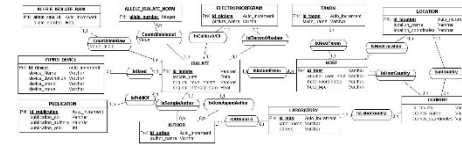


Conception de bases de données



# Conception de bases de données

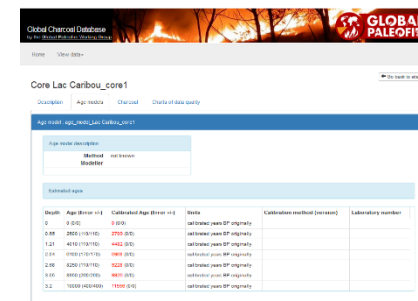
- Modélisation



- Intégration des données

| Nom de la tâche | Charge initiale | Temps passé depuis le début du projet à n-1 | Reste à faire à n-1 |
|-----------------|-----------------|---|---------------------|
| Parking         | 8,00            | 0,00  | 8,00                |
| PRET            | 3,00            | 0,00  | 3,00                |
| PPPGM           | 30,00           | 0,00  | 30,00               |
| PaTEST          | 2,00            | 0,00  | 2,00                |
| VRET            | 2,00            | 0,00  | 2,00                |
| VPPGM           | 15,00           | 0,00  | 15,00               |
| VWTEST          | 1,5             | 0,00  | 1,5                 |
| ACRE            | 8,00            | 0,00  | 8,00                |
| ACTEST          | 2,00            | 0,00  | 2,00                |
| ACPPGM          | 25,00           | 0,00  | 25,00               |
| ACTEST          | 2,5             | 0,00  | 2,5                 |
| ESRE            | 3,00            | 0,00  | 3,00                |
| ESRE            | 1,5             | 0,00  | 1,5                 |
| ESPPGM          | 15,00           | 0,00  | 15,00               |
| ESTEST          | 1,5             | 0,00  | 1,5                 |
| In              | 10,00           | 0,00  | 10,00               |

- Développement des interfaces de manipulation des données



# Plan

---

- Les données de recherche au laboratoire Chrono-environnement
- **Constats**
- Notre proposition le méta-observatoire
- L'existant
- Notre modèle

# Constats : après la conception de nombreux modèles

---

**Grande hétérogénéité** des données

Mais présence **d'éléments communs** (nommés différemment)

Toutes les BD créées sont des **observatoires** qui ont des objets d'observation différents

→ Capitalisation de l'expérience acquise

# Éléments communs des Bases de données

---

Dans chaque base de données on décrit les **contextes d'observation** :

- Qui : personne, structure
- Quoi : observation
- Quand : dates
- Où : géolocalisation, site, ...
- Comment : méthode d'obtention des données
- Pourquoi : objectifs, problématique

# Observatoire / Observation

---

Une base de données **observatoire** regroupe un ensemble d'observations et un contexte d'observation dans le cadre de projets de recherche

Une **observation** consiste en la mesure d'une caractéristique sur un élément. L'observation est décrite par :

- Une mesure
- Un acteur
- Une date
- Des méthodes ou protocoles
- Des outils

# Plan

---

- Les données de recherche au laboratoire Chrono-environnement
- Constats
- **Notre proposition le méta-observatoire**
- L'existant
- Notre modèle

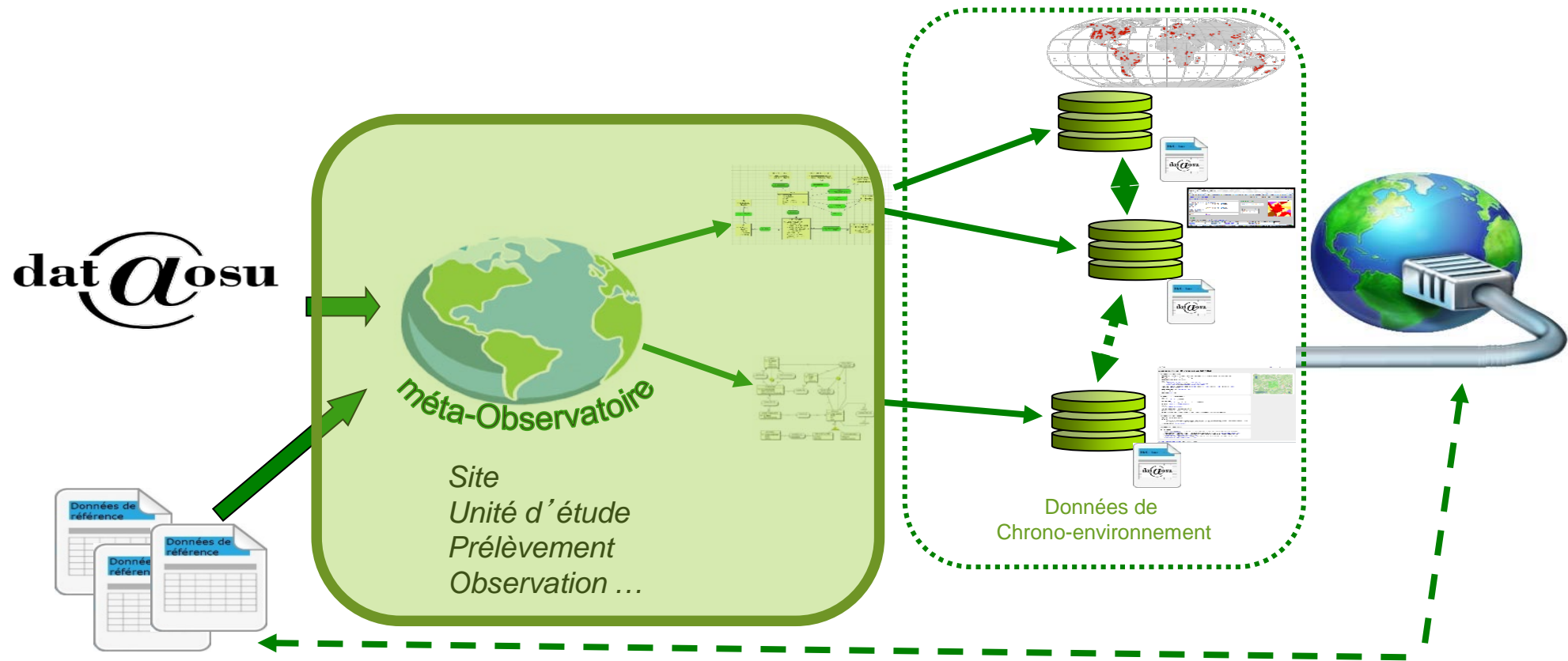
# La proposition

---

Concevoir un environnement, qui s'appuie sur un méta-modèle, pour guider les chercheurs pour la création de bases de données

- qualité (traçabilité, validation, contrôle, explicitation, ...)
- pérennes (serveur de base données, sauvegarde, ...)
- interopérables avec d'autres bases de données

# La proposition





# Méta-observatoire

---

Plate-forme permettant aux chercheurs de construire leurs observatoires avec l'aide d'un informaticien

Première étape : Construction d'un modèle décrivant les concepts d'un observatoire le plus générique possible tout en étant adaptable

Modélisation par une ontologie

# Plan

---

- Les données de recherche au laboratoire Chrono-environnement
- Constats
- Notre proposition le méta-observatoire
- **L'existant**
- Notre modèle

# Travaux existants

---

OBOE : Extensible Observation Ontology

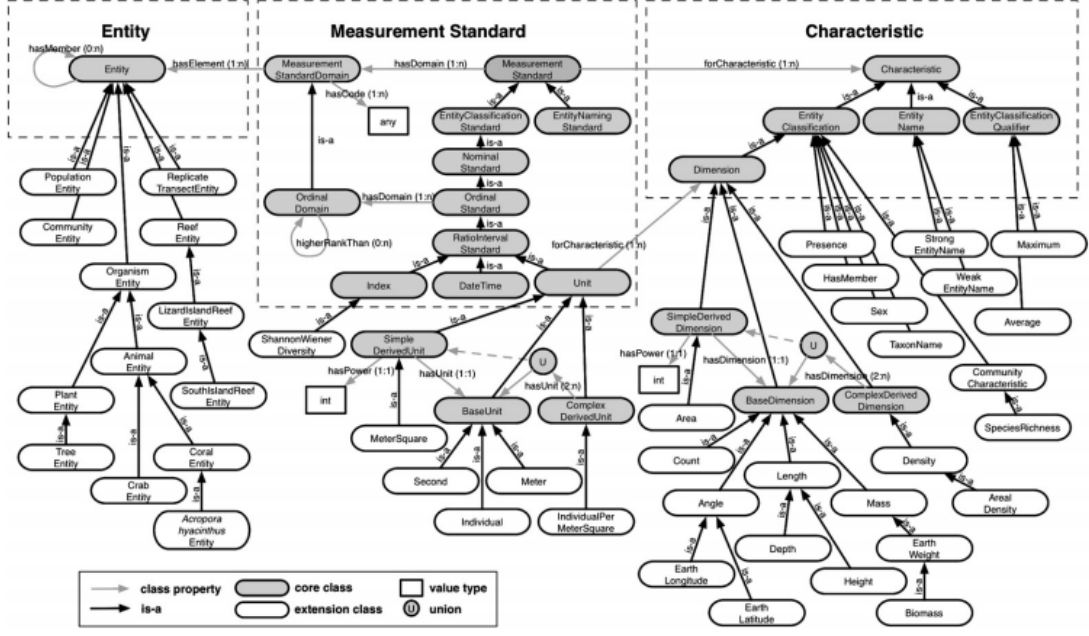
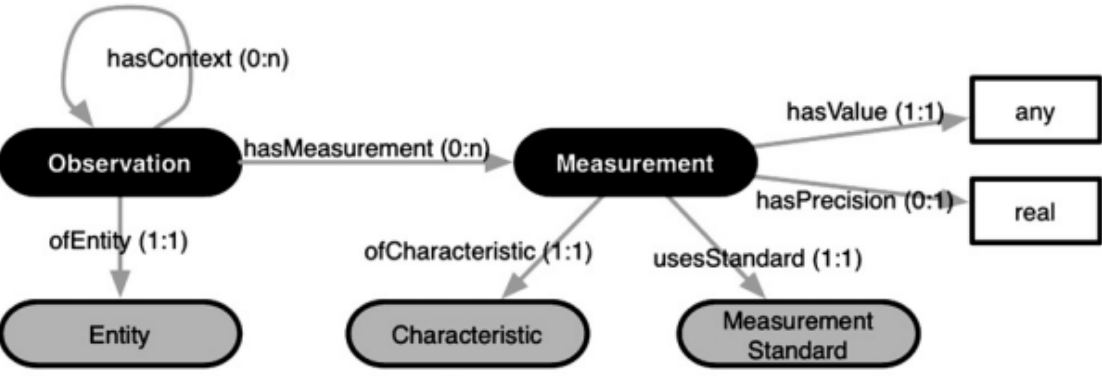
SERONTO

ODM

EML

...

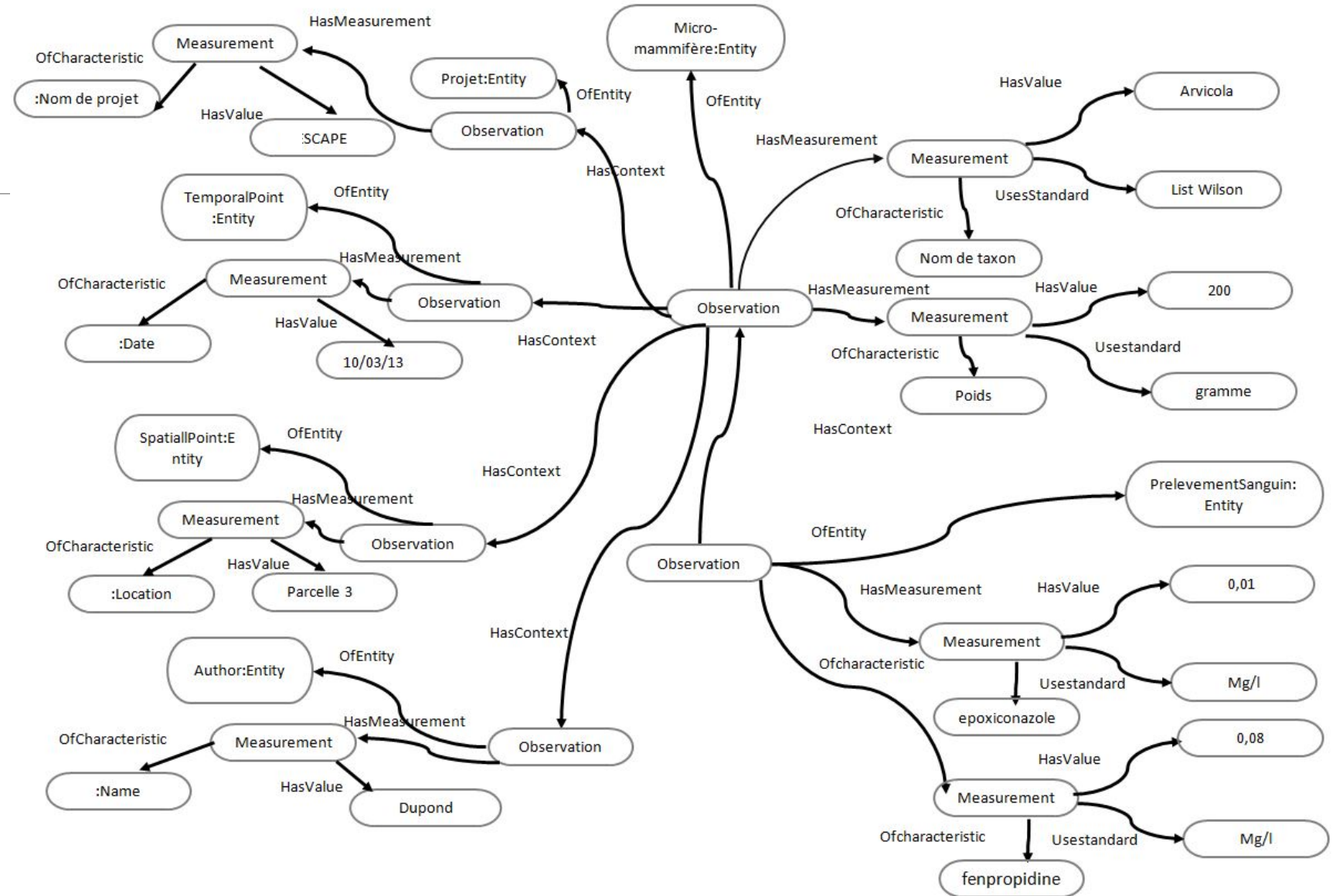
# OBOE (Extensible Observation Ontology)



Tout est observation

## Ensemble concret d'observations simplifié

Micromammifère de taxon "arvicola", capturé dans le cadre du projet SCAPE, le 10 mars 2013 sur la parcelle 3 par M. Dupont. On mesure son poids et on effectue un prélèvement sanguin pour réaliser des analyses permettant de doser de l'époxiconazole et de la fenpropidine.



# Constat de l'instanciation

---

Tout est au même niveau

Difficulté pour les chercheurs de représenter leurs données avec ce modèle

- Peu lisible
- Pas de hiérarchie
- Le contexte est une observation
- Pas de regroupement d'observations

Mais concept d'observation très précis

# Plan

---

- Les données de recherche au laboratoire Chrono-environnement
- Constats
- Notre proposition le méta-observatoire
- L'existant
- **Notre modèle**

# L'ontologie Méta-Observatoire

---

Notion d'observatoire

Contexte clairement défini

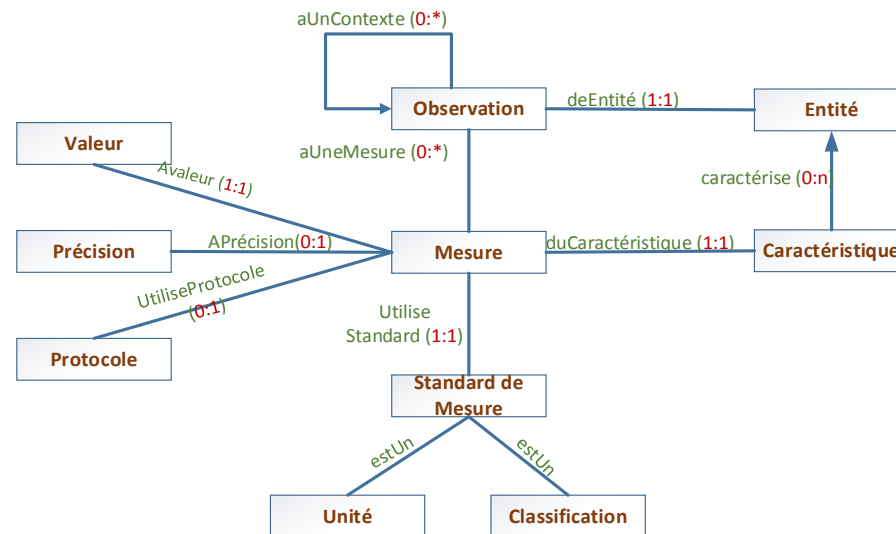
Définir un ensemble de concepts obligatoires

Différents niveaux d'observation : unité d'étude et prélèvement

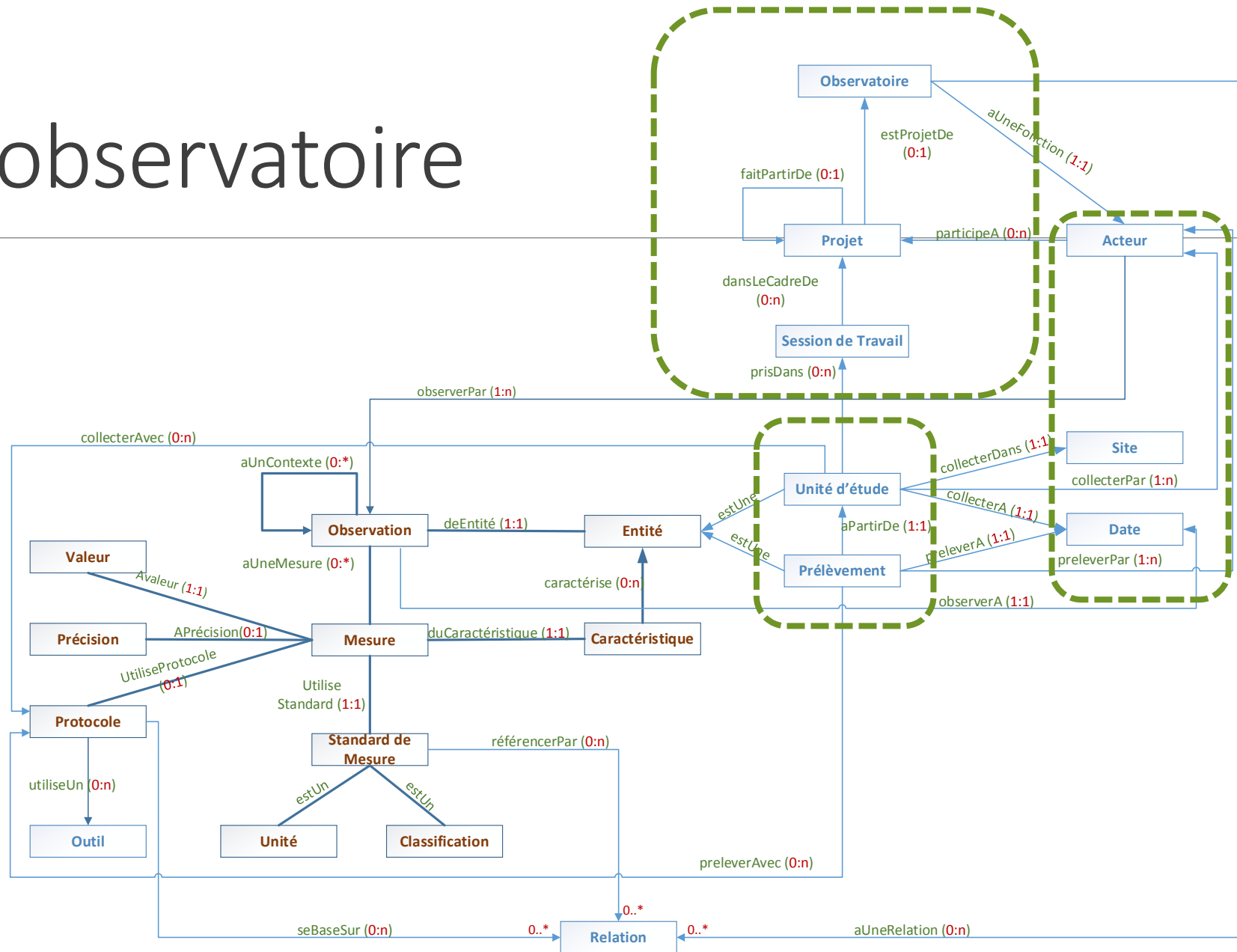


# Méta-observatoire

## Ontologie basée sur OBOE

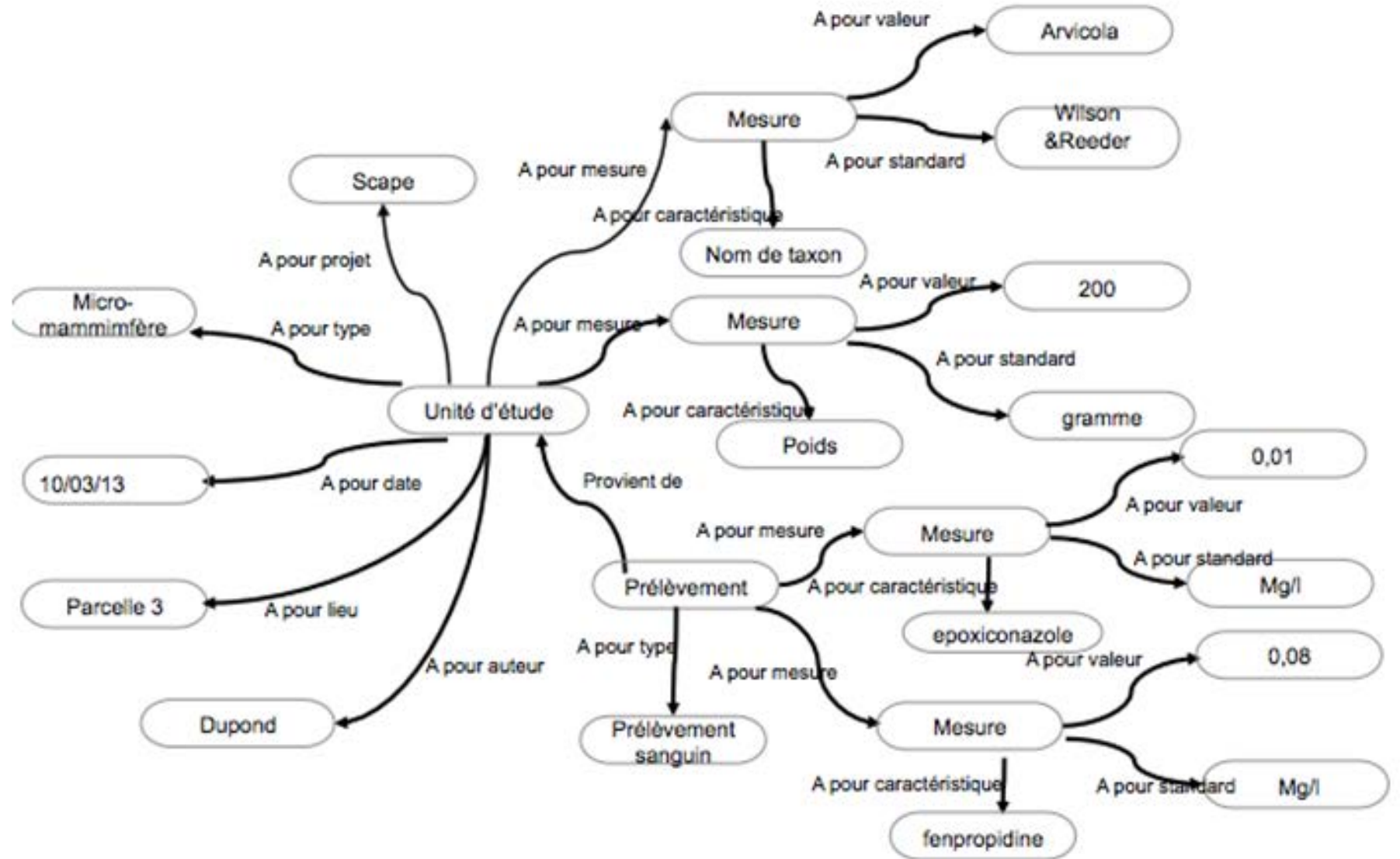


# Méta-observatoire



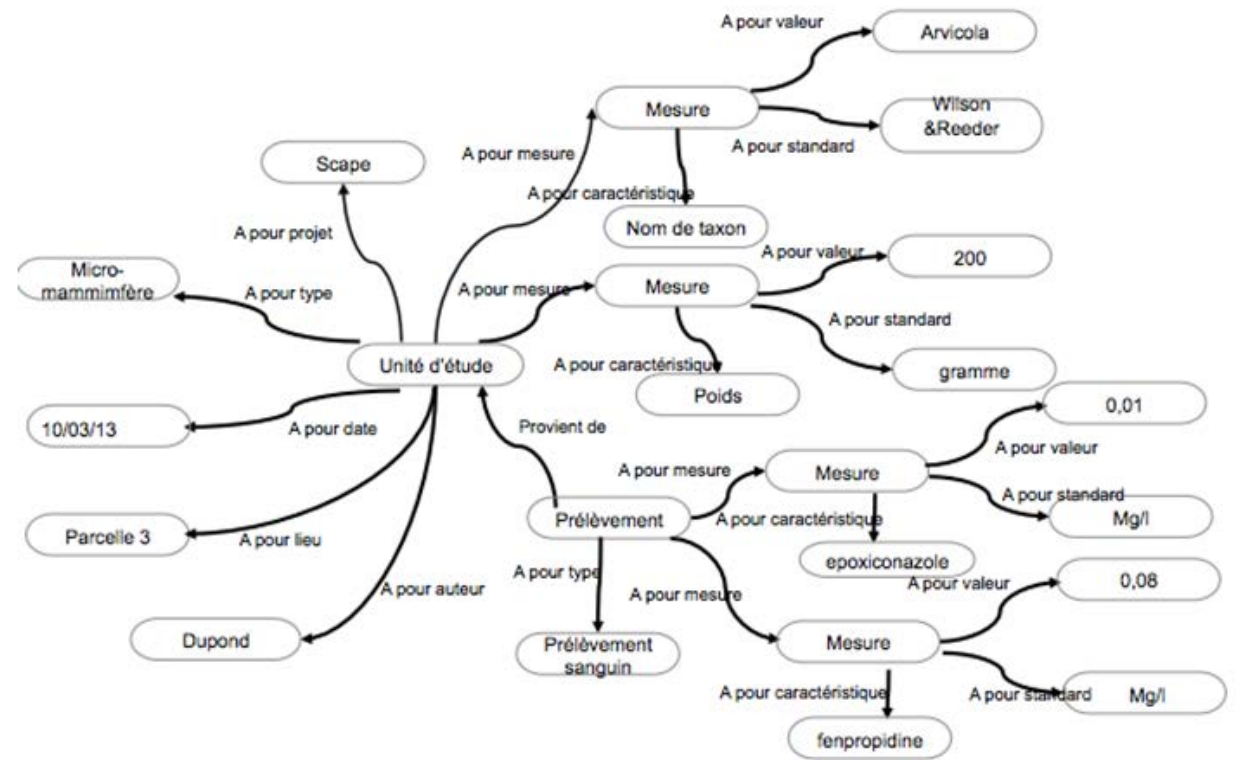
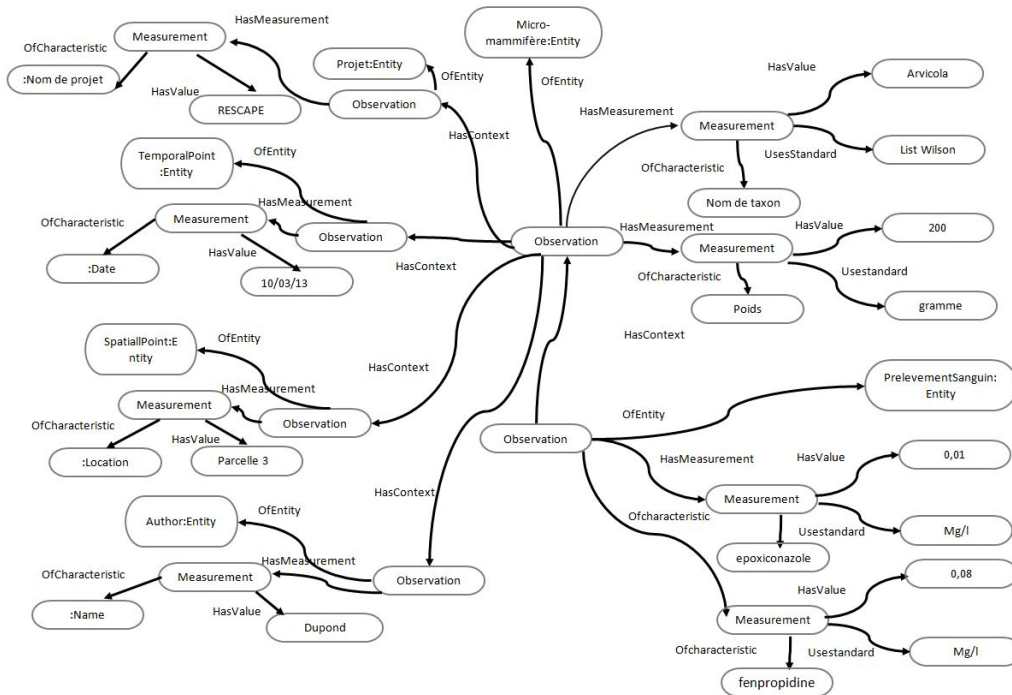
## Même exemple

Micromammifère de taxon “arvicola”, capturé dans le cadre du projet SCAPE, le 10 mars 2013 sur la parcelle 3 par M. Dupont. On mesure son poids et on effectue un prélèvement sanguin pour réaliser des analyses permettant de doser de l'époxiconazole et de la fenpropidine.



# OBOE

# Méta-Observatoire



# Conclusion

---

Représentation d'un jeu de données moins dense

Plus lisible et explicite les contextes

Met en évidence l'unité d'étude

1ère version de l'ontologie modèle Meta-Observatoire

# Perspectives

---

Mise en situation complète de l'ontologie sur des observatoires

## Développement de la plate-forme

- Intégration de l'ontologie et définition des points d'extension
- Création d'un modèle spécifique :
  - Spécialisation et Sélection
  - Génération de l'ontologie spécifique à un projet
  - Génération du modèle de base de données

# Merci de votre attention

---