



SINP

Système d'Information
sur la Nature et le Paysage



Inventaire
National du
Patrimoine
Naturel



Mapping DwC/OccTax v2

Version 1.0

Contexte

Dans le cadre de la récupération des données du GBIF situées sur le territoire français, au format Darwin Core, à l'avenir, une méthode doit être utilisée pour intégrer ces données au format occurrences de taxons du SINP. La procédure sera effectuée suivant un certain nombre de règles de correspondance qui sont décrites dans ce document.

Objet du document

Ce document présente l'export semi-automatique.

Rédaction

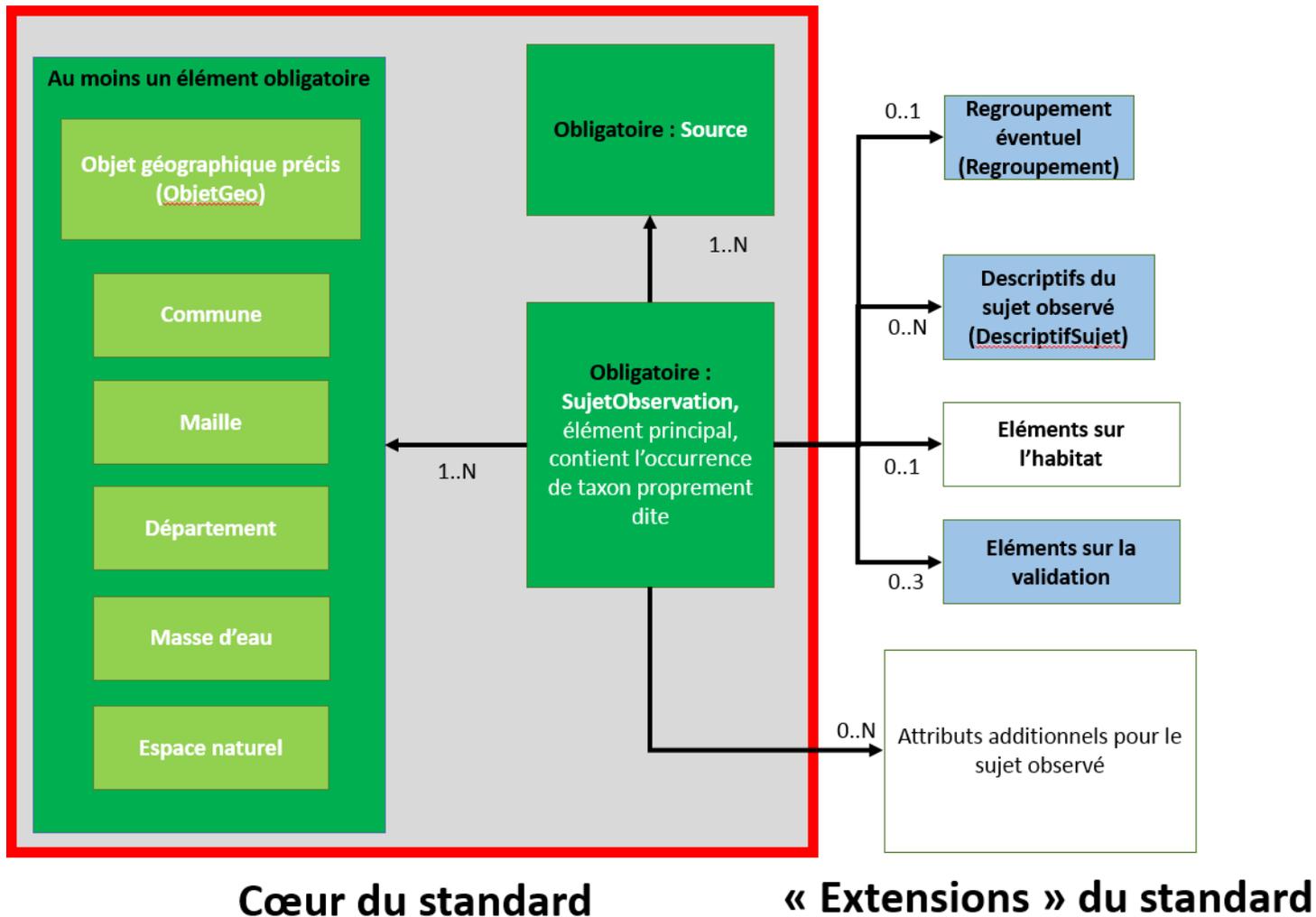
Relecture

Rémy JOMIER, Sophie PAMERLON

Solène ROBERT, Frédéric Vest

1. Organisation générale de la cible de l'échange

La cible de l'échange est le standard occurrences de taxons du SINP dans sa version 2.0.



2. Détail des éléments cible

Méthode : Dans les points qui suivent on distingue plusieurs types de champs pour lesquels la méthode de mapping varie :

- Les champs en vert sont obligatoires.
- Les champs en noir ne sont pas strictement nécessaires.

Un certain nombre de choix, devant la complexité de standardiser des informations en provenance de langues et de standards différents, sont nécessaires. En l'état, il est décidé de ne pas transférer toute information demandant un effort de standardisation trop important.

En conséquence, ne seront pas repris les éléments concernant :

- les attributs additionnels (hétérogénéité variable)
- les communes (hétérogénéité forte ne facilitant pas la réconciliation)
- les dénombrements (travail de réconciliation important entre organismQuantityType et objetDenombrement nécessaire)
- les départements
- le descriptif du sujet (beaucoup de champs du standard occurrences de taxons nécessitent l'usage d'une nomenclature, et il n'en existe pas actuellement pour le Darwin Core)
- l'habitat (pas de référentiel mondial existant)

- la maille 10x10 française (non utilisée au niveau mondial)
- la masse d'eau française (non utilisée au niveau mondial)

Ne seront donc repris que les concepts/objets suivants :

- Sujet d'observation
- Source de l'information
- Objet géographique (si présent)
- Regroupement
- Organismes et personnes quand nécessaire

2.1. Regroupement

Description : Concept du standard occurrences de taxons permettant de lier des occurrences entre elles. Correspond au fichier DarwinCore contenant toutes les informations sur les événements. Cet élément ne pourra être rempli que si le fichier EventCore est utilisé au sein du DarwinCore. Dans le cas où ce fichier ne serait pas présent, se reporter au paragraphe suivant.

- **idSINPRegroupement** (code unique du regroupement, sous forme d'UUID) -> Repris à l'identique de l'identifiant d'événement nommé eventID
- **methodeRegroupement** (méthode ayant présidé au regroupement) -> obligatoire : reprend le champ « eventRemark ». Si vide, prend la valeur « Inconnue »
- **typeRegroupement** (type de regroupement) -> prend la valeur « NSP »
- **techniqueEchantillonnage** (technique d'échantillonnage) -> reprend samplingProtocol
- **tailleEchantillon** (taille de l'échantillon) -> reprend sampleSizeValue

- **uniteTailleEchantillonnage** (unité d'échantillonnage) -> reprend sampleSizeUnit
- **effortEchantillonnage** (effort consenti pour l'échantillonnage ex : 2 jours) -> reprend samplingEffort
- **idOrigine** (Identifiant présent à l'origine chez le producteur) -> reprend fieldNumber

2.2. SujetObservation

- **idSINPOccTax** (Code unique de désignation d'une donnée au sein du SINP, sous la forme d'un UUID) -> repris à l'identique de OccurrenceId
- **statutObservation** (indique si on a observé, ou pas trouvé en le recherchant, le sujet d'observation) -> On affecte la valeur « Pr » et on considère que toutes les valeurs du DarwinCore sont des valeurs d'observation.
- **nomCite** (nom cité à l'origine par l'observateur) -> Reprend originalNameUsage. Si ce n'est pas rempli, alors, reprend ScientificName. **Si aucun des deux n'est rempli, alors, ce n'est pas une occurrence, pas d'intégration.**
- **objetGeo** (type objet géographique, obligatoire dans le cadre du transfert Darwin Core vers SINP, composé de 4 attributs) -> geometrie reprend la valeur convertie de footprintWKT. Si footprintWKT n'est pas rempli, alors, reprendre les valeurs X, Y les convertir en point et reprendre cela Si aucun des deux n'est rempli : pas d'intégration de la donnée qui ne possède pas d'information géographique précise. natureObjetGeo prend la valeur NSP. nomLieu prend la valeur de l'attribut locality.
- **dateDebut** (date de début d'observation) -> reprend la valeur de eventDate (présent dans le fichier event si ce sont des données utilisant Event Core, présent dans le fichier occurrence si ce n'est pas le cas). Attention, la date peut être une période de dates, auquel cas on reprendra la valeur la plus faible pour dateDebut
- **heureDebut** (heure de début d'observation) -> reprend la valeur de eventTime (présent dans le fichier event si ce sont des données utilisant Event Core, présent dans le fichier occurrence si ce n'est pas le cas). Attention, l'heure peut être une période de dates, auquel cas on reprendra la valeur la plus faible pour heureDebut. Si eventTime n'est pas rempli : prend la valeur 00 :00 :00
- **dateFin** (date de fin) -> comme pour dateDebut, reprend la valeur de eventDate. Si eventDate contient une période, reprend la plus grande valeur.
- **heureFin** (heure de fin) -> comme pour heureFin, reprend la valeur d'eventTime. Si eventTime contient une période, reprend la plus grande valeur. Si eventTime n'est pas rempli, prend la valeur 23 :59 :00
- **altitudeMin** (altitude minimale) -> reprend la valeur de minimumElevationInMeters. NB : attention au signe de décimale qui peut être ou un point ou une virgule.
- **altitudeMax** (altitude maximale) -> reprend la valeur de maximumElevationInMeters. NB : attention au signe de décimale qui peut être ou un point ou une virgule.
- **profondeurMin** (profondeur minimale) -> reprend la valeur de minimumDepthInMeters. NB : attention au signe de décimale qui peut être ou un point ou une virgule.

- **profondeurMax** (profondeur maximale) -> reprend la valeur de maximumDepthInMeters. NB : attention au signe de décimale qui peut être un point ou une virgule.
- **observateur** (type Observateur, contient 3 attributs, 2 obligatoires et 1 facultatif. Seuls les obligatoires peuvent être remplis) -> identite reprend recordedBy. Organisme prend la valeur « INCONNU ».
- **cdNom** (Code du taxon dans TAXREF) -> A attribuer par réconciliation sur la base de la valeur de l'attribut scientificName. Si ScientificName n'est pas rempli, alors, faire la réconciliation sur la base originalNameUsage
- **determinateur** (personne ayant déterminé l'espèce, obligatoire si un cdNom existe type composé de 3 attributs, dont 2 sont obligatoires. Seuls les obligatoires peuvent être remplis.) -> identite reprend la valeur de identifiedBy, organisme prend la valeur « INCONNU ».
- **dateDetermination** (date de détermination) -> Reprend la valeur de dateIdentified.
- **commentaire** (commentaire sur l'occurrence) -> reprend occurrenceRemarks
- **dEEFloutage** (indique si la donnée a été floutée) -> prend la valeur « OUI » si dataGeneralization contient une information
- **sensiNiveau** (niveau de sensibilité de la donnée) -> Sachant qu'une donnée déjà diffusée n'est pas sensible, sensiNiveau prendra la valeur 0.
- **statutSource** (origine de la donnée : terrain, collection, inconnue) -> si basisOfRecord prend la valeur PreservedSpecimen, alors, prend la valeur « Co » pour collection. Sinon, prend la valeur « NSP ».
- **idSINPJdd** (identifiant du jeu de données SINP) -> Toute donnée est rattachée à sa fiche de métadonnées au sein du GBIF. On importera au préalable les fiches de métadonnées du GBIF en reprenant leur UUID, puis on pointera vers leur identifiant.
- **orgTransformation** (organisme responsable de la transformation en DEE) -> prend la valeur « UMS Patrinat »
- **organismeGestionnaireDonnee** -> Reprend la valeur de InstitutionCode. Si non présent : OwnerInstitutionCode. Si non présent : Publisher. Si toujours pas présent, prendre la valeur définie dans les métadonnées.
- **dEEDateTransformation** (date de transformation en DEE) -> Prend la date du jour
- **dEEDateDerniereModification** (date de modification de la DEE) -> Prend la date du jour
- **referenceBiblio** (référence bibliographique) -> Reprend la valeur de associatedReferences

2.3. Source

Description : Concept SINP contenant toutes les informations sur l'origine de l'information.

- **idOrigine** (identifiant de la donnée chez le producteur) -> reprend l'UUID DarwinCore présent dans le fichier occurrence (occurrenceld)
- **dSPublique** (origine de la donnée, privée ou publique) -> prend la valeur « NSP »