



Mapping OccTax v2/DwC

Version 1.0

Contexte

Dans le cadre de la livraison des données du SINP vers le GBIF et le Darwin Core, à l'avenir, une méthode doit être utilisée. 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

Table des matières

1. Organisation générale de la cible de l'échange	3
2. Détail des éléments cible	4
2.1.Event.....	5
2.1.1. <i>Cas du niveau correspondant au regroupement</i>	5
2.1.2. <i>Cas du niveau correspondant au sujet d'observation</i>	5
2.2.Measurement or Fact	6
2.3.Occurrence	6

1. Organisation générale de la cible de l'échange

Fichier Darwin Core Event
eventId
eventRemarks
samplingProtocol
sampleSizeValue
sampleSizeUnit
samplingEffort
fieldNumber
parentEventId
eventDate
eventTime

Champ obligatoire
Champ conseillé
Champ facultatif
Champ obligatoire conditionnel

Fichier Darwin Core Occurrence
eventId
occurrenceId
occurrenceStatus
originalNameUsage
minimumElevationInMeters
maximumElevationInMeters
minimumDepthInMeters
maximumDepthInMeters
recordedBy
scientificName
basisOfRecord
taxonId
nameAccordingTo
identifiedBy
dateIdentified
occurrenceRemarks
footprintWKT
Municipality
County
higherGeography
Habitat
waterbody
locationRemarks
coordinateUncertaintyInMeters

locality
country
countryCode
decimalLatitude
decimalLongitude
informationWithheld
basisOfRecord
institutionCode
modified
associatedReferences
dataGeneralization
identificationVerificationStatus
occurrenceRemarks
samplingProtocol
identificationRemarks
establishmentMeans
sex
lifeStage
reproductiveCondition
associatedMedia
behavior
Champ obligatoire
Champ obligatoire conditionnel
Champ conseillé
Champ facultatif

Fichier Darwin Core MeasurementOrFact
eventId
measurementType
measurementMethod
measurementValue
measurementUnit
Champ obligatoire
Champ conseillé
Champ facultatif

Les deux tables Darwin Core Occurrence et Darwin Core MeasurementOrFact sont rattachées à la table Event par le biais de l'eventId.

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.

2.1. Event

Description : Fichier DarwinCore contenant toutes les informations sur les événements. C'est à lui que se rattachera le fichier MeasurementOrFact. Un événement peut correspondre à un regroupement ou à une occurrence.

2.1.1. Cas du niveau correspondant au regroupement

- **eventId** (code unique de l'événement, sous forme d'UUID) -> Repris à l'identique de l'identifiant de regroupement
- **eventRemarks** (remarques sur l'événement) -> reprend la concaténation de methodeRegroupement et du libellé associé au code de typeRegroupement, avec pour séparateur « | »
- **samplingProtocol** (protocole d'échantillonnage) -> reprend la valeur du libellé associé au code présent dans techniqueEchantillonnage
- **sampleSizeValue** (valeur de la taille de l'échantillon, sans unité) -> reprend tailleEchantillon
- **samplesizeUnit** (unité de taille de l'échantillon) -> reprend uniteTailleEchantillonnage
- **samplingEffort** (effort d'échantillonnage) -> reprend effortEchantillonnage
- **fieldNumber** (numéro attribué à l'origine) -> reprend l'idOrigine du regroupement
- **parentEventId** (Identifiant d'événement parent) -> reprend l'idRegroupementParent
- **eventDate** (Date/heure d'observation. Peut contenir une période d'observation) -> On prend dateDebut de l'observation la plus ancienne et on concatène avec dateFin de l'observation la plus récente (maximum

de dateFin des observations du regroupement), avec pour séparateur « | ».

2.1.2. Cas du niveau correspondant au sujet d'observation

Utilisé dans le but de pouvoir embarquer les attributs additionnels du regroupement, du sujet d'observation, ou des éléments comme la profondeur moyenne de l'observation. En effet, l'extension MeasurementOrFact du DwC ne permet de rattacher des MeasurementOrFact qu'à un seul niveau : SOIT à l'événement, SOIT à l'occurrence.

- **eventId** (code unique de l'événement, sous forme d'UUID) -> Repris à l'identique de l'identifiant de regroupement
- **eventRemarks** (remarques sur l'événement) -> prend la valeur « Single occurrence event, to link MeasurementOrFact elements »
- **fieldNumber** (numéro attribué à l'origine) -> reprend l'idOrigine du sujet d'observation (concept Source)
- **parentEventId** (Identifiant d'événement parent) -> reprend l'idRegroupement pour rattachement au regroupement s'il y a lieu
- **eventDate** (date de l'événement) -> Si dateDebut/dateFin sont égales : on reprend dateFin. Sinon, on prend dateDebut, et on concatène avec dateFin, avec pour séparateur « | ».
- **eventTime** (horaire de l'événement) -> Si heureDebut/heureFin sont égales : on reprend heureFin. Sinon, on prend heureDebut, et on concatène avec heureFin, avec pour séparateur « | ».

2.2. Measurement or Fact

Description : Fichier DarwinCore contenant toutes les informations sur les mesures ou faits additionnels non prévus dans le standard DwC. Il sera rattaché à l'événement par le biais de l'eventId.

- **eventId** (code unique de l'événement, sous forme d'UUID) -> Repris à l'identique de l'identifiant de sujet d'observation (suivant à quoi l'attribut additionnel est rattaché)
- **measurementType** (type de mesure) -> reprend nomAttribut
- **measurementMethod** (méthode de mesure) -> reprend definitionAttribut
- **measurementValue** (valeur de la mesure) -> reprend valeurAttribut
- **measurementUnit** (unité de mesure) -> reprend uniteAttribut (quand disponible)

2.3. Occurrence

- **idSINPJdd** (Code unique de désignation d'un jeu de données : pas un attribut Darwin Core) -> repris à l'identique du même attribut de la source
- **eventId** (Identifiant de l'événement parent) -> Reprend la valeur de l'identifiant SINP idSINPOccTax.
- **occurrenceId** (Identifiant de l'occurrence) -> Reprend idSINPOccTax.
Cas particulier : si une donnée est rattachée à plusieurs mailles, plusieurs occurrences doivent être créées, chacune avec pour occurrenceId la concaténation de idSINPOccTax et du code de la maille considérée.

- **occurrenceStatus** (statut de l'occurrence) -> Reprend le libellé de la nomenclature associée (nomenclature n°18) au code présent dans le champ statutObservation
- **originalNameUsage** (nom cité) -> reprend la valeur de l'attribut nomCite
- **minimumElevationInMeters** (altitude minimale) -> reprend la valeur de altitudeMin
- **maximumElevationInMeters** (altitude maximale) -> reprend la valeur de altitudeMax
NB : dans le cas où on ne dispose que de l'altitude moyenne, on utilisera un MeasurementOrFact avec pour measurementType « Average altitude », measurement method « Average altitude in meters », la valeur d'altitude moyenne, et pour measurementUnit « m »
- **minimumDepthInMeters** (profondeur minimale) -> reprend la valeur de profondeurMin
- **maximumDepthInMeters** (profondeur maximale) -> reprend la valeur de profondeurMax
NB : dans le cas où on ne dispose que de la profondeur moyenne, on utilisera un MeasurementOrFact avec pour measurementType « Average depth », measurement method « Average depth in meters », la valeur de profondeur moyenne, et pour measurementUnit « m »
- **recordedBy** (Observateur) -> Reprend le nom de l'observateur d'OccTax. Séparer les observateurs (s'il y en a plusieurs) par le caractère « | »
- **scientificName** (Nom scientifique du taxon) -> Reprend le nom complet scientifique du taxon dans la version en cours de TAXREF
- **taxonId** (Code du taxon) -> Reprend l'attribut CD_NOM de TAXREF pour la version en cours

- **nameAccordingTo** (nom attribué en fonction d'un référentiel ou document de référence) -> Valeur « TAXREF v » et numéro de version TAXREF en cours par défaut (« TAXREF v12 » à la rédaction de ce document).
- **identifiedBy** (déterminateur) -> Reprend les valeurs de identite/organisme/mail pour le déterminateur.
- **dateIdentified** (date de détermination) -> **Reprend dateDetermination**
- **occurrenceRemarks** (remarques sur l'occurrence) -> reprend la concatenation du libellé associé au code présent dans occEtatBiologique (nomenclature n°7) et la valeur de obsDescription, et de obsContexte, ainsi que de commentaire.
- **footprintWKT** (contour d'objet géographique) -> Non utilisé, sauf dans le cas des données pour lesquelles le producteur a demandé le niveau de précision maximal à la diffusion (diffusionNiveauPrecision = 5) : on reprendra alors l'objet géographique
- **municipality** (municipalité) -> Reprend le codeCommune, concaténé avec le nomCommune. Séparateur : le caractère « | »
- **county** (comté, correspond à la notion de département en France) -> Reprend la valeur de codeDepartement.
- **higherGeography** (géographie de haut niveau) -> Reprend typeEN, codeEN, et la valeur de libellé associé à l'habitat dans la typologie idoine d'HABREF), le tout concaténé. Séparateur : le caractère « | »
- **habitat** (habitat) -> Reprend les informations du type habitat : concaténation du codeHabitat et de son libellé associé (cf. HABREF), et du code HabRef et de son libellé associé. Séparateur : le caractère « | »
- **waterBody** (masse d'eau) -> Reprend le code de la masse d'eau concaténé à son libellé. Voir Footprint WKT si la masse d'eau est la seule information géographique.
- **locationRemarks** (remarques sur la localisation) -> La donnée étant rattachée à la maille, valeur par défaut « Data isn't the original geo referenced one, but attached to the nearest 10x10 km grid cell »
- **coordinateUncertaintyInMeters** (Incertitude connue sur les coordonnées, en mètres) -> Valeur remplie, par défaut, avec le nombre « 5000 », soit 5000 mètres. Cas des données sensibles : pour le niveau 3, indiquer
- **locality** (localité) -> Reprend la valeur de nomLieu
- **country** (pays) -> Valeur remplie par défaut avec la valeur « France »
- **countryCode** (Code du pays) -> Valeur remplie, par défaut, avec le code « FR » signifiant « France » pour la France métropolitaine. Pour les DOM/TOM, utiliser les valeurs suivantes : Clipperton, TF, France, FR, Guadeloupe, GP, Guyane : GF, Martinique MQ, Mayotte YT, Nouvelle Calédonie NC, Polynésie Française, PF, Réunion, RE, Saint Barthélémy, BL, Saint Martin, MF, Saint Pierre et Miquelon PM, TAAF : TF, Wallis et Futuna WF
- **decimalLatitude** (Latitude de l'observation, en degrés décimaux) -> Reprend l'attribut X du centroïde de l'objet géographique ou de la maille considérée. **Est à considérer comme obligatoire. Dans le cas d'une occurrence rattachée à plusieurs mailles uniquement, tenter le croisement avec les communes et se rattacher à une concaténation des communes.**
- **decimalLongitude** (Longitude de l'observation, en degrés décimaux) -> Reprend l'attribut Y du centroïde de l'objet géographique ou de la maille considérée. **Est à considérer comme obligatoire. Dans le cas d'une**

occurrence rattachée à plusieurs mailles uniquement, tenter le croisement avec les communes et se rattacher à une concaténation des communes.

- **informationWithheld** (information non délivrée) -> Dans le cas d'informations sensibles (niveaux 1, 2, 3, ou 4), indiquer « Location information not given precisely : sensitive taxon ».
- **basisOfRecord** (Indique quel est le type de l'enregistrement) -> Est rempli par défaut avec la valeur « HumanObservation ». Pour les données issues de collection (statutSource = « Co »), on indiquera « Preserved Specimen »
- **institutionCode** (code ou nom de l'institution) -> Reprend les informations concernant organismeGestionnaire.
- **modified** (date de modification) -> reprend dEEDateDerniereModification
- **associatedReferences** (références bibliographiques associées) -> Reprend le contenu de referenceBiblio
- **dataGeneralization** (Informations sur une dégradation éventuelle des données à des fins de diffusion) -> Rempli par défaut avec la valeur avec le texte suivant : « Géographie transmise soumise à floutage (grille avec mailles de 10x10km) pour le grand public » en conformité avec les règles de diffusion du SINP | Geographic information generalized during aggregation (grid with 10x10km cells) for the general public, according to SINP communication rules ». **Cas des données sensibles :** « Geographic information generalized during aggregation due to sensitive information – endangered taxon for example – »
- **identificationVerificationStatus** (statut de validation de la détermination) -> Reprend les libellés suivants pour le niveau de validation nationale 1 -> « Certain, highly probable ». 2 -> « Probable, high degree of

reliability. ». 3-> « Doubtful : data is surprising, but there isn't enough information to invalidate it as such. ». 4 -> « Invalid ». 5 -> « Control could not be conclusive due to insufficient knowledge »

- **samplingProtocol** (technique d'observation) -> Reprend le libellé associé au code de nomenclature n°14 (<http://standards-sinp.mnhn.fr/nomenclature/14-methodes-dobservation-observationmethode-2018-05-14/> pour la version à jour) contenu dans l'attribut obsTechnique
- **identificationRemarks** (remarques sur l'identification) -> Reprend la valeur de occMethodeDetermination
- **establishmentMeans** (établissement de l'occurrence) -> Récupérer les valeurs de libellé correspondant aux codes trouvés dans occStatutBiogeographique (nomenclature 11) et occNaturalité (nomenclature 8). Concaténer les libellés, avec pour séparateur « | »
- **sex (sexe)** -> Reprend le libellé associé au code de nomenclature n°9 (<http://standards-sinp.mnhn.fr/nomenclature/9-sexe-occurrencesexe-23-06-2016/> pour la version à jour) contenu dans l'attribut sexe
- **lifeStage (stade de vie)** -> Reprend le libellé associé au code de nomenclature n°10 (<http://standards-sinp.mnhn.fr/nomenclature/10-stade-de-vie-stade-de-developpement-du-sujet-occurrencestadedevie-2018-05-14/> pour la version à jour) contenu dans l'attribut occStadeDeVie
- **reproductiveCondition** (état reproductif) -> UNIQUEMENT pour les valeurs 3 et 9 du champ occStatutBiologique : rempli avec les libellés correspondant aux codes de la nomenclature de statut biologique, n°13 (<http://standards-sinp.mnhn.fr/nomenclature/13-statut-biologique-occurrencestatutbiologique-2018-05-14/>)

- **associatedMedia** (médias associés) -> Reprend la valeur de uRLPreuveNumerique
- **behavior (comportement)** -> Reprend le libellé associé au code de nomenclature n°110 (<http://standards-sinp.mnhn.fr/nomenclature/110-comportement-des-occurrences-observees-2018-05-14/> pour la version à jour) contenu dans l'attribut occComportement

Remarque : dans le cas où plusieurs descriptifs du sujet seraient associés à une même occurrence, on concatènera dans les champs idoines.