

Attribution d'un code de sensibilité à une données d'occurrence à partir d'un référentiel de sensibilité

Auteur : J. Panijel (MNHN), F. Vest (MNHN)

Date : 24 mai 2016

Contenu

1	Le référentiel de sensibilité	2
1.1	Définition des attributs du référentiel	2
1.2	Principes et règles métier contrôlés par le MNHN (BDC).....	2
1.3	Workflow GINCO pour le référentiel de sensibilité.....	3
2	Principes d'application d'un niveau de sensibilité sur une donnée	3
2.1	Règles d'application et d'héritage de la sensibilité sur les données.....	3
2.1.1	La sensibilité s'applique pour un taxon.	3
2.1.2	La sensibilité s'applique pour un taxon sur un département.....	5
2.2	Les restrictions d'application.....	5
3	Règles de calcul d'un niveau de sensibilité	6
3.1	Règle 1 : Période de sensibilité.....	6
3.2	Règle 2 : Sensibilité sans condition	6
3.3	Règle 3 : Sensibilité conditionnée au statut biologique du taxon observé	7
3.4	Règle 4 : Sensibilité conditionnée par un autre critère	8
3.5	Règle 5 : pas de sensibilité	8
3.6	Règle 6 : Pas de CD_NOM dans la donnée	9
4	Gestion de la sensibilité manuelle.....	9
4.1	Alerter l'administrateur à l'issue du processus automatique d'attribution d'un niveau de sensibilité.....	9
4.2	Attribution de la sensibilité manuelle à une donnée d'occurrence.....	9
5	Gestion de la fin de la période de sensibilité	10

1 Le référentiel de sensibilité

1.1 Définition des attributs du référentiel

Le référentiel de sensibilité constitué de deux tables comprenant les informations suivantes :

Table LISTE_SL

- CD_DEPT : code INSEE du département.
- CD_NOM : code du taxon valide.
- NOM_CITE : nom du taxon cité dans la liste
- DUREE : durée d'application de la sensibilité en année.
- CODAGE : code de sensibilité.
- AUTRE : commentaire sur l'application de la sensibilité.
- CD_SL : identifiant de la liste de sensibilité.
- CD_OCC_STATUT_BIOLOGIQUE : code du statut biologique pour lequel s'applique la sensibilité.

Table METADONNEES_LISTE_SL

- CD_SL : identifiant de la liste de sensibilité.
- LIBELLE_LISTE : Titre/ Nom de la liste de sensibilité
- DATE_CREATION : Date de première diffusion du référentiel
- DATE_MAJ : Date de dernière mise à jour du référentiel

1.2 Principes et règles métier contrôlés par le MNHN (BDC)

- Une liste de sensibilité ne peut se baser que sur une seule version du référentiel TaxRef
- Lors de l'import d'une liste régionale dans la BDC, celle-ci est mise en conformité avec la version de TaxRef utilisée dans l'INPN
- La sensibilité ne s'applique pas pour les données marines.
- Lorsqu'une liste régionale existe et est transmise au niveau national, cette liste remplace la liste nationale pour la région concernée. En l'absence de liste régionale ou de transmission au niveau national, la liste nationale sera appliquée.
- L'attribution de la sensibilité est effectuée sous la responsabilité de l'administrateur de la plateforme Régionale/Thématique.
- Un CD_NOM de rang taxonomique supérieur à l'espèce (genre, famille, etc.) ne peut être utilisé dans la définition des règles du référentiel de sensibilité.
- Une sous-espèce ne peut être moins sensible que son espèce dans le référentiel de sensibilité.
- Une règle du référentiel de sensibilité utilisant un CD_NOM supprimé dans la version publiée de l'INPN ne sera pas implémentée dans les calculs de sensibilité. Seuls les CD_NOM des listes de référence existant dans la version du référentiel Taxref en vigueur pourront être considérés pour l'application des règles de sensibilité.

- Une règle du référentiel de sensibilité utilisant un CD_NOM splitté de manière complexe fera l'objet d'un retour de la plate-forme nationale et nécessitera une révision de la liste régionale.

Commenté [JP1]: Principe susceptible d'être modifié et à vérifier.

1.3 Workflow GINCO pour le référentiel de sensibilité

- Le référentiel est fourni à GINCO par la plateforme nationale (INPN) dans la version de TaxRef publiée sur l'INPN.
- Lors de la publication d'une nouvelle version de TaxRef, la plateforme nationale met parallèlement à jour les listes de sensibilité régionales.

2 Principes d'application d'un niveau de sensibilité sur une donnée

2.1 Règles d'application et d'héritage de la sensibilité sur les données

2.1.1 La sensibilité s'applique pour un taxon.

Si un CD_NOM d'espèce est utilisé dans la définition des règles du référentiel de sensibilité

ALORS la règle s'applique au CD_NOM de l'espèce utilisé dans le référentiel de sensibilité

- Le CD_Nom renseigné pour l'observation "est égal à" un CD_NOM du référentiel de sensibilité (*ReferentielSensibilite.cdNom=Observation.cdNom*)

OU à un d'un CD_NOM "synonyme" du CD_NOM de l'espèce utilisé dans le référentiel de sensibilité

- le CD_REF renseigné pour l'observation "est égal à" un CD_NOM du référentiel de sensibilité (*ReferentielSensibilite.cdNom=Observation.cdRef*)

OU à un CD_NOM de sous-espèce "enfant" du CD_NOM de l'espèce utilisé dans le référentiel de sensibilité

- Si le CD_SUP renseigné pour l'observation "est égal à" un CD_NOM du référentiel de sensibilité (*ReferentielSensibilite.cdNom=Observation.cdSup* en récursif jusqu'au niveau du CD_NOM de la règle du référentiel)

OU à un CD_NOM "synonyme" du CD_NOM de la sous-espèce de l'espèce utilisé dans le référentiel de sensibilité

- Si le CD_SUP renseigné pour l'observation "est égal à" un CD_REF du CD_NOM du référentiel de sensibilité (*ReferentielSensibilite.cdSup=Observation.cdRef* en récursif jusqu'au niveau du CD_NOM de la règle du référentiel)

Commenté [JP2]: OUI ! Du coup on fait une boucle récursive qui remonte jusqu'à la racine ?

Commenté [SR3R2]: On a ce qu'il faut pour faire ça...on peut fournir la procédure ou la table pivot

Commenté [AM4]: Ok, pour faire le récursif je reprends le cdNom correspondant à Observation.cdSup et je regarde si le cdSup associé vaut ReferentielSensibilite.cdNom c'est ça ?

Commenté [SR5R4]: Perso, je ferai dans le sens inverse. On va chercher d'abord tous les enfants de la liste de référence et on croise ensuite avec les données.

2.1.1.1 Cas particuliers

1. Si les CD_NOM d'une espèce et d'une de ses sous-espèces sont utilisés dans deux règles distinctes du référentiel de sensibilité (code différent de sensibilité)

ALORS la règle de l'espèce s'applique au CD_NOM de l'espèce utilisé dans le référentiel de sensibilité

OU à un CD_NOM "synonyme" du CD_NOM de l'espèce utilisé dans le référentiel de sensibilité

OU à un CD_NOM de sous-espèce "enfant" du CD_NOM de l'espèce distinct du CD_NOM de la sous-espèce utilisé dans le référentiel de sensibilité

OU à un CD_NOM "synonyme" du CD_NOM de la sous-espèce de l'espèce utilisé dans le référentiel de sensibilité

ALORS la règle de la sous-espèce s'applique au CD_NOM de la sous-espèce utilisé dans le référentiel de sensibilité

OU à un CD_NOM "synonyme" du CD_NOM de la sous-espèce utilisé dans le référentiel de sensibilité

2. Si le CD_NOM d'une sous-espèce est utilisé dans une règle unique dans référentiel de sensibilité

ALORS la règle de la sous-espèce s'applique au CD_NOM de la sous-espèce utilisé dans le référentiel de sensibilité

OU à un CD_NOM "synonyme" du CD_NOM de la sous-espèce utilisé dans le référentiel de sensibilité

- La règle ne s'applique PAS aux données du CD_NOM de l'espèce
- Les données portant le CD_NOM de l'espèce ne sont pas sensibles

2.1.2 La sensibilité s'applique pour un taxon sur un département.

Un CD_DEPT est un attribut obligatoire dans la définition des règles de gestion du référentiel de sensibilité

La règle du référentiel s'applique au CD_DEPT à laquelle la donnée d'occurrence a été automatiquement rattachée

- *Le CD_DEPT renseigné pour l'observation "est égal à" un CD_DEPT du référentiel de sensibilité (ReferentielSensibilite.cdDept=Observation.cdDept)*

2.1.2.1 Cas particuliers

Si une donnée d'occurrence est fournie avec un objet géographique de type polygone ou multi-polygones rattaché automatiquement à plusieurs départements dont un est utilisé dans une définition des règles du référentiel de sensibilité pour le taxon cité.

ALORS la donnée est sensible sur l'ensemble des départements auxquels la donnée est rattachée, quel que soit le rang ou le pourcentage de recouvrement des départements.

2.2 Les restrictions d'application

L'attribution d'un code de sensibilité peut être restreinte :

- à un statut biologique donné (champ CD_OCC_STATUT_BIOLOGIQUE)
- et/ou une période donnée.

3 Règles de calcul d'un niveau de sensibilité

3.1 Règle 1 : Période de sensibilité

La sensibilité d'un taxon est définie sur une période. Cette période peut-être infinie (pas de limite dans le temps) ou bornée dans le temps.

Le calcul de cette période s'effectue à partir de la date ou période d'observation du taxon et du champ DUREE du référentiel de sensibilité.

Pour une observation d'un taxon, on considère que pour le/les département(s) où est effectuée l'observation, la Donnée élémentaire d'Echange (DEE) est considérée sensible jusqu'à la date calculée de la manière suivante :

dateFin de la DEE + valeur du champ DUREE (année) du référentiel de sensibilité.

Exemple : pour dateFin = 01/01/2012 et DUREE = 3, la DEE sera sensible jusqu'au 01/01/2015. Après cette date, la sensibilité ne s'applique plus.

3.2 Règle 2 : Sensibilité sans condition

Pour une observation d'un taxon, on considère que pour le/les département(s) où est effectuée l'observation :

Si le CD_NOM renseigné pour l'observation est :

- "est égal à" un CD_NOM du référentiel de sensibilité
- OU "est synonyme" d'un CD_NOM du référentiel de sensibilité
- OU "est enfant" d'un CD_NOM du référentiel de sensibilité

pour le département concerné

ET SI les champs du référentiel de sensibilité CD_OCC_STATUT_BIOLOGIQUE et DUREE du référentiel sont vides

ALORS la valeur du champ CODAGE s'applique pour l'observation concernée.

Les attributs de la Données Élémentaires d'Echange (DEE) prennent les valeurs suivantes :

- *sensible = OUI*
- *sensiNiveau = valeur du champ CODAGE*
- *sensiDateAttribution = date du jour de l'attribution de la sensibilité à la DSR ou la DEE*
- *sensiReferentiel = référentiel de sensibilité utilisé*
- *sensiVersionReferentiel = version du référentiel de sensibilité utilisé*

3.3 Règle 3 : Sensibilité conditionnée au statut biologique du taxon observé

Dans certain cas la sensibilité d'un taxon est conditionnée par le statut biologique du taxon observé.

Pour une observation d'un taxon, on considère que pour le/les département(s) où est effectuée l'observation :

Si le *CD_NOM* renseigné pour l'observation est :

- "est égal à" un *CD_NOM* du référentiel de sensibilité
- OU "est synonyme" d'un *CD_NOM* du référentiel de sensibilité
- OU "est enfant" d'un *CD_NOM* du référentiel de sensibilité

pour le département concerné

ET SI la valeur du champ *CD_OCC_STATUT_BIOLOGIQUE* du référentiel de sensibilité n'est pas vide ET SI elle est égale à la valeur de l'attribut *occStatutBiologique* de la DEE

ALORS la valeur du champ *CODAGE* s'applique pour l'observation concernée.

Les attributs de la Données Élémentaires d'Echange (DEE) prennent les valeurs suivantes :

- *sensible* = OUI
- *sensiNiveau* = valeur du champ *CODAGE*
- *sensiDateSensibilite* = date du jour de l'attribution de la sensibilité à la DEE
- *sensiReferentiel* = référentiel de sensibilité utilisé
- *sensiVersionReferentiel* = version du référentiel de sensibilité utilisé

3.3.1.1 Règle 3.1 : Traitement spécifique pour les valeurs vide ou 0 ou 1 ou 2

Si l'information liée au statut biologique de l'observation n'est pas fournie dans la DEE, ou si elle a des valeurs égales à 0 [Inconnu] OU 1 [Non renseigné] OU 2 [Non déterminé],

ALORS la règle de sensibilité du taxon ayant le *CODAGE* le plus élevé s'applique.

Les attributs de la Données Élémentaires d'Echange (DEE) prennent les valeurs suivantes :

- *sensiAlerte* = OUI
- *sensiManuelle* = NON
- *sensible* = OUI
- *sensiNiveau* = valeur du champ *CODAGE*
- *sensiDateSensibilite* = date du jour de l'attribution de la sensibilité à la DEE
- *sensiReferentiel* = référentiel de sensibilité utilisé
- *sensiVersionReferentiel* = version du référentiel de sensibilité utilisé

3.4 Règle 4 : Sensibilité conditionnée par un autre critère

La sensibilité d'une observation d'un taxon est conditionnée par un autre critère que le statut biologique. Le critère AUTRE peut compléter le statut biologique.

- Cette attribution ne pourra être effectuée qu'à l'appréciation de l'administrateur de la plateforme qui attribuera/validera alors la sensibilité « à la main ».

Pour une observation d'un taxon, on considère que pour le/les département(s) où est effectuée l'observation :

Si le CD_NOM renseigné pour l'observation est :

- "est égal à" un CD_NOM du référentiel de sensibilité
- OU "est synonyme" d'un CD_NOM du référentiel de sensibilité
- OU "est enfant" d'un CD_NOM du référentiel de sensibilité

pour le département concerné

ET si la valeur du champ AUTRE du référentiel de sensibilité n'est pas vide

ALORS les attributs de la Données Élémentaires d'Echange (DEE) prendront les valeurs suivantes :

- *sensible = OUI*
- *sensiNiveau = valeur du champ CODAGE*
- *sensiDateSensibilite = date du jour de l'attribution de la sensibilité à la DEE*
- *sensiReferentiel = référentiel de sensibilité utilisé*
- *sensiVersionReferentiel = version du référentiel de sensibilité utilisé*

ET DEUX CHAMPS DE GESTION PERMETTANT DE FLAGUER LA DONNEE SERONT MODIFIES

- *sensiManuelle = NON*
- *sensiAlerte = OUI*

3.5 Règle 5 : pas de sensibilité

L'observation n'est pas sensible

Si les conditions d'attribution du code de sensibilité définies par les règles 1, 2, 3 et 4 ne s'appliquent pas.

ALORS les attributs de la Données Élémentaires d'Echange (DEE) prennent les valeurs suivantes :

- *sensible = NON*
- *sensiNiveau = 0*
- *sensiDateSensibilite = date du jour de l'attribution de la sensibilité à la DEE*
- *sensiReferentiel = référentiel de sensibilité utilisé*

- *sensiVersionReferentiel* = version du référentiel de sensibilité utilisé

3.6 Règle 6 : Pas de CD_NOM dans la donnée

La sensibilité d'un taxon est définie en premier lieu par le champ CD_NOM. Si le Cd_Nom pour le taxon observé existe alors il doit être renseigné.

Dans le cas exceptionnel d'un taxon ne figurant pas encore dans le référentiel TaxRef, l'observation n'est pas sensible.

ALORS les attributs de la Données Élémentaires d'Echange (DEE) prennent les valeurs suivantes :

- *sensible* = NON
- *sensiNiveau* = 0
- *sensiDateSensibilite* = date du jour de l'attribution de la sensibilité à la DEE
- *sensiReferentiel* = référentiel de sensibilité utilisé
- *sensiVersionReferentiel* = version du référentiel de sensibilité utilisé

4 Gestion de la sensibilité manuelle.

4.1 Alerter l'administrateur à l'issue du processus automatique d'attribution d'un niveau de sensibilité

Si la valeur du champ *sensiAlerte* = OUI dans une ou plusieurs données d'occurrence du jeu de données traité

ALORS, la plateforme devra alerter l'administrateur sur le fait qu'il doit attribuer manuellement la sensibilité sur n observations du jeu de données traité.

4.2 Attribution de la sensibilité manuelle à une donnée d'occurrence

Si la valeur du champ *sensiAlerte* = OUI dans une donnée d'occurrence, les champs suivants de la donnée sont modifiables en édition

- *sensible* = OUI
- *sensiNiveau* = valeur du champ CODAGE

Lorsque la sensibilité manuelle est enregistrée les champs suivants sont automatiquement mis à jour avec les valeurs :

- *sensiDateSensibilite* = date du jour de l'attribution de la sensibilité manuelle à la DEE
- *sensiAlerte* = NON
- *sensiManuelle* = OUI

5 Gestion de la fin de la période de sensibilité.

A préciser plus tard

Le calcul et la mise à jour du champ Sensible devra être réalisée périodiquement.

Ce processus devra prendre en compte la périodicité des mises à jour des référentiels de sensibilité et TAXREF.

6 Gestion des données sensibles associées à un regroupement géolocalisé de données.

Lorsqu'une donnée appartient à un regroupement géolocalisé de donnée, il faut empêcher la corrélation des informations permettant de localiser la donnée sensible et indiquer le niveau de complétude du regroupement diffusé. Celle-ci est floutée et son lien avec le regroupement est masqué sur des niveaux de restitution inférieurs à sa sensibilité.

- *le champ "identifiant du regroupement" des données sensibles sera flouté à la diffusion.*
- *Les données non sensibles du regroupement sont diffusées mais l'entité regroupement porte un attribut booléen (d'information) sur sa complétude. S'il y a au moins 1 donnée sensible dans le regroupement l'attribut est "incomplet".*

7 Résumé de certains points discutés avec Julien Touroult

7.1 Prise en compte de la règle du guide méthodologique concernant l'ajout de taxon sensible pour des espèces dont les données sont déjà disponibles lors de l'élaboration des listes.

Problématique

Doit-on autoriser l'attribution d'une sensibilité plus forte lors du recalcul de la sensibilité pour des données déjà diffusées ?

Si l'on ne recalcule pas la sensibilité, cela veut dire que la sensibilité d'un groupe de données d'un même taxon sera hétérogène. Dans pas mal de cas la sensibilité d'un taxon deviendra ridicule si la majorité des données sont diffusées moins sensibles que ce qui est prévu dans la règle, voir diffusées non sensibles. La cohérence et la gestion de la sensibilité risque d'être aussi très difficile dans le temps.

Du coup, faut-il interdire que le niveau de sensibilité d'une règle devienne plus important et systématiquement appliquer la règle de sensibilité la plus faible en modifiant les listes de sensibilité ?

Autre cas d'usage à envisager : Si un CDNOM bouge dans TaxRef et devient le synonyme d'un CDREF qui correspond à une règle de sensibilité plus importante

Réponse

La règle de ne pas revenir sur la diffusion est une bonne règle, sur un principe de non régression mais certaines

espèces pourraient devenir sensibles (accroissement de la demande sur telle plante aromatique par exemple) et dans ce cas, si l'espèce apparaît dans une liste, il faudrait ré-appliquer la sensibilité aux données antérieures.

Conclusion

Il faut appliquer toutes les règles et recalculer toutes les données en fonction des listes de sensibilité

7.2 Calcul de la sensibilité de données rattachées à plusieurs entités administratives

Cas d'usage d'une donnée géolocalisée sur plusieurs départements

Si un producteur fournit une donnée avec un polygone. Par calcul, la donnée est rattachée au dpt 1 (40%) et dpt 2 (60%).

La liste régionale de sensibilité donne le taxon sensible sur un seul département.

La donnée est-elle sensible ou non ? Si oui, est-elle sensible sur les 2 départements ? sur un seul ? Faut-il prévoir une attribution manuelle de la sensibilité ?

Réponse

Le plus simple et prudent est de rendre sensible dans les deux départements. Autrement aller vers du manuel mais l'opérateur SINP n'aura pas forcément une idée claire de comment statuer...

Cas d'usage d'une donnée sur rattachée à plusieurs informations de géolocalisation

Si un producteur fournit une donnée avec un polygone + Dpt 1+ commune. Par calcul, la donnée est rattachée au dpt 1 (35%) et dpt 2 (65%) + commune A + commune B.

La liste régionale de sensibilité donne le taxon sensible sur le département 2 uniquement.

La données est-elle sensible ou non ? Si oui, est-elle sensible sur les 2 départements ? sur un seul ? Faut-il prévoir une attribution manuelle de la sensibilité ?

Réponse

Même réponse que le cas précédent.

7.3 Sensibilité des données en mer

Le groupe SINP interne a acté que la sensibilité ne s'applique pas pour les données marines. La sensibilité des données marines peut s'appliquer tant que les départements marins ou autres zones marines ne font l'objet d'aucun référentiel officiel.

Réponse

Dans le guide méthodo, on dit que les éventuelles listes marines doivent être établies par façade maritime (Méditerranée, Atlantique) et pas par approche régionale.