

Standard d'échange des données naturalistes

Modèle d'architecture de table de données naturalistes et de métadonnées associées

V 1 du 11/07/2016

Remerciement

Le présent document concrétise le travail d'une cinquantaine de personnes issues d'une grande diversité d'organismes, et notamment : Agence des aires marines protégées, Agence d'urbanisme, de développement économique et technopole du Pays de Lorient, Brest Métropole, Bretagne Grands migrants, Bretagne vivante – Sepnb, Collectif des associations gestionnaires d'espaces naturels, Communauté de communes du Val d'Ille, Conseils départementaux des Côtes d'Armor, Finistère, Ille-et-Vilaine et Morbihan, Eaux et rivières, Fédération régionale des chasseurs de Bretagne, , Conservatoire botanique national de Brest, Groupe de recherches et d'études des invertébrés armoricains, Groupe mammalogique breton, IGN DRNO, Institut national de recherche en sciences et technologies pour l'environnement et l'agriculture Montpellier, Lorient Agglomération, Lorient Trégor Communauté, Parcs naturels régionaux d'Armorique et du Golfe du Morbihan, Union régional des centres permanents d'initiatives pour l'environnement.

Qu'ils soient tous remerciés pour leur participation active à ce travail collectif.

Table des matières

1. Le contexte	3
2. Le projet.....	3
3. Référentiels communs.....	4
4. Un modèle pour une architecture de table de données naturalistes	5
5. Un modèle pour une architecture des métadonnées associées à un lot de données naturalistes	8
6. Fiabilité des données.....	10
7. Sensibilité des données	10

Ressources

Contact :

GIP Bretagne environnement - Observatoire du patrimoine naturel et de la biodiversité en Bretagne

La feuille de route du pôle métier biodiversité ainsi que les documents relatifs aux diverses étapes d'élaboration sont accessibles sur <http://cms.geobretagne.fr/biodiversite>.

La fiche de métadonnée de GéoBretagne fait l'objet d'explicitations détaillées de ses champs, compléments utiles au présent modèle, et accessibles sur <http://kartenn.region-bretagne.fr/wiki/doku.php>.

Le standard d'occurrence des taxons du MNHN est accessible via <https://inpn.mnhn.fr/telechargement/standard-occurrence-taxon>

Un exemple de définition de données « sensibles » au niveau national est accessible sur : http://www.naturefrance.fr/sites/default/files/fichiers/ressources/pdf/sinp_guide_technique_donnees_sensible_v1_avril_2014.pdf

1. Le contexte

L'élaboration de ce modèle s'inscrit dans la dynamique du pôle métier biodiversité de GéoBretagne.

Le pôle métier biodiversité accueille tous les acteurs régionaux qui le souhaitent autour de trois organismes animateurs : Dreal, Conseil régional, GIP Bretagne environnement.

Au sein du pôle métier, les groupes de travail thématiques sont les lieux des réflexions et productions. La dynamique est collégiale et participative ; chaque groupe est animé par un ou des acteurs volontaires.

Validée en juin 2014, la feuille de route du pôle métier fixe des axes de travail (*cf.* Ressources page 2).

L'un des axes de travail du pôle métier biodiversité est de proposer et de promouvoir des bonnes pratiques facilitant les échanges de données naturalistes entre partenaires et permettant leur mise à disposition, sous conditions éventuelles, auprès des acteurs, notamment des collectivités.

En cours de développement, cette dynamique régionale prône ainsi l'établissement d'un standard d'échange.

Le but est de mutualiser les données naturalistes. Au-delà de la donnée au sens le plus simple du terme (un taxon, un lieu, une date), c'est plus globalement l'information naturaliste qui est visée et qui nécessite une réflexion partagée sur les besoins en matière d'échange et les meilleurs moyens d'y répondre.

Une des attentes concerne le fait qu'une donnée doit être parfois interprétée, vulgarisée et accompagnée sur le terrain au plus près des acteurs. Ainsi, un annuaire régional des compétences et de référents est en cours de réflexion. Ce sera une ressource pour tout producteur ou utilisateur de données qui aurait des questionnements.

2. Le projet

Une des bonnes pratiques en matière d'échange concerne la structuration de la table de données qui est échangée, ainsi que celle de la métadonnée associée. Le projet a pour ambition d'identifier et de décrire les champs nécessaires et suffisants qui composeraient une table de données naturalistes échangée entre partenaires ou entre plateformes. La métadonnée accompagnant la table est aussi un objet du projet.

Les données concernées en priorité portent sur les taxons (espèces, sous espèces), les syntaxons (groupements végétaux) et les habitats (« milieux » au sens de Corine Biotope, Eunis...) dont l'occurrence (absence / présence) est spatialement connue (géoréférencée). Le cas des données synthétiques est aussi abordé.

Le modèle a fait l'objet d'une première réflexion globale et d'orientations générales lors d'une réunion du groupe de travail en décembre 2015 à Brest, animée par le GIP Bretagne environnement et le Conservatoire botanique national de Brest. Une quarantaine d'acteurs y participait. Puis, sous forme de questionnaire électronique, une formalisation du modèle a été proposée en mars 2016 à plus de deux cents acteurs régionaux de la biodiversité. En avril 2016 à Loudéac, le groupe de travail a examiné les retours et stabilisé le modèle. Le présent document est issu de cette dernière étape (*cf.* Ressources page 2).

3. Référentiels communs

La notion de référentiel est au cœur des problématiques d'échange et se trouve donc abordée dès la phase de structuration des tables. Il s'agit de coder de manière univoque et intemporelle un objet et d'explicitier d'éventuelles relations entre les objets (genre – espèce – sous espèce, cours d'eau – tronçon, etc.). C'est un langage commun qui facilite les échanges et le croisement de sources diverses. Il y a un intérêt majeur à utiliser ces référentiels afin de constituer une communauté de partage de données.

Deux types de référentiels sont généralement usités en biodiversité : géographique et nomenclatural. Les référentiels disponibles sont nombreux et couvrent de larges sujets : descripteurs administratifs (canton, communes...), descripteurs géographiques (lieux-dits, cours d'eau...), divers découpages géographiques (mailles d'atlas, secteurs régionaux de prospections...), faune, flore, habitats...

L'utilisation de certains référentiels peut s'avérer délicate : espèces non encore référencées, précision de certaines entités géographiques ambiguës, etc. Comme peut s'avérer complexe le croisement de deux référentiels sur une même thématique (ex. la flore). Le futur annuaire régional des compétences et de référents, en cours de réflexion, permettra de mobiliser les savoir-faire.

Le modèle d'architecture standard proposé ci-après sous-entend plusieurs points cruciaux.

Concernant les référentiels nomenclaturaux :

- chaque producteur reste libre d'utiliser son référentiel propre (le référentiel « producteur ») ;
- par contre, lors de la mise à disposition d'une table, un champ (CodeNomTaxRef) doit permettre de faire le lien entre le référentiel « producteur » et le référentiel national TaxRef du MNHN ;
- cette correspondance recherchée n'est pas sans posée certains problèmes, notamment pour les taxons de flore et les taxons d'invertébrés marins et terrestres ; un travail régional collaboratif est à mener afin de consolider ces correspondances, voire être force de proposition pour abonder le niveau national.

Concernant les référentiels géographiques :

- dans le cas de limites administratives ou d'entités liées à l'eau, il est impératif d'utiliser les référentiels institutionnels : IGN Geofla (commune, départements...), BD Carto ou BD Carthage (cours d'eau, bassins versants...), etc.
- dans le cas de découpage de type atlas, l'utilisation d'un même système de maille ou de secteur par tous les acteurs est nécessaire ; un travail régional d'homogénéisation est finalisé pour les objets de type d'oiseaux d'eau hivernants et oiseaux marins nicheurs.

4. Un modèle pour une architecture de table de données naturalistes

Ce modèle a vocation à satisfaire à la fois des besoins régionaux et nationaux en termes d'information naturaliste. Il vise par conséquent à être en adéquation avec les modèles développés, d'une part, au niveau national (SINP, système d'information sur la nature et les paysages) et, d'autre part, au niveau régional par les structures naturalistes en place.

Il s'appuie donc sur le standard de données Occurrence des taxons du SINP (cf. Ressources page 2) en tenant compte des bases de données et autres standards développés en région, notamment ceux du Conservatoire botanique national de Brest (ex. base de données eCalluna ou Référentiel des Noms d'usage de la Flore de l'Ouest de la France), des associations naturalistes bretonnes BV-Sepnb, Geoca, GMB, Gretia, LPO, VivArmor Nature (ex. bases de données Faune-Bretagne ou Observatoire régional des oiseaux marins) ou encore de divers organismes produisant ou manipulant des bases de données naturalistes (Onema, fédérations de pêche ou de chasse, GIP Bretagne environnement...).

Le nombre de champs dans le modèle a été réduit au strict minimum dans un premier temps. Ils concernent essentiellement les données d'occurrence.

Certains champs sont obligatoires : c'est l'information minimale qui explicite la sémantique du lot de données.

Les autres champs sont optionnels et permettent de s'adapter aux différents cas rencontrés, notamment, ceux concernant des données synthétiques et non plus une simple occurrence de taxons.

Quelques champs sont dits « optionnel conditionné » : si la valeur de ces champs est unique pour toute la table, cette valeur peut être positionnée dans la métadonnée.

Si le modèle est fondé sur un nombre minimal de champs, libre à chaque producteur de le compléter et d'enrichir l'information en ajoutant des champs, pour autant que la structure minimale soit préservée.

Quelques points clefs généraux à respecter sous peine de rendre obsolète une table de données :

- Toutes les cellules d'une même colonne contiennent exclusivement l'information d'un seul champ.
Contre-exemple : les colonnes « Sensibilité » et « Fiabilité » sont interverties sur certaines lignes.
- Les cellules d'une ligne contiennent les informations relatives à une seule donnée.
- Toutes les informations relatives à une donnée se trouvent dans les cellules d'une seule ligne.
- Chaque ligne porte une information qui est unique à l'échelle de la table.
Contre-exemple : présence de lignes en doublons.

Quelques points clefs généraux améliorant la qualité structurelle d'une table de données :

- Toutes les cellules d'une même colonne ont le même format d'écriture.
- Un seul type d'information est présent dans les cellules d'une même colonne.
Contre-exemple : regrouper « Observateur » et « Producteur » dans la même colonne limite les capacités d'exploitation.
- Lorsqu'un seul champ descripteur ne suffit pas à caractériser l'unicité de la ligne dans la table, il est recommandé d'ajouter un champ (numérique entier) faisant office de clé primaire.

	nom du champ	format	contenu	commentaire
obligatoire	CodeNom	entier	code du taxon* selon le référentiel utilisé par le producteur de la donnée	*une donnée du type nom de genre (ex. « puffin sp. ») est gérée par tout « bon » référentiel
obligatoire	NomScientifique	texte	nom scientifique du taxon selon le référentiel utilisé par le producteur de la donnée	un nom scientifique sans ambiguïté cite le(s) nom(s) d'auteur(s)
optionnel*	NomFrancais	texte	nom français du taxon	*le nom français n'existe pas toujours ; d'où le caractère optionnel, mais fortement recommandé pour la lisibilité de la table par les non spécialistes
obligatoire	ReferentielNom	texte	citation du référentiel nomenclatural utilisé par le producteur de la donnée	l'utilisation d'un référentiel est très fortement recommandé et si possible un référentiel déjà existant
obligatoire	CodeNomTaxRef	texte	code du taxon selon le référentiel national TaxRef du MNHN en utilisant le champ CD_NOM de TaxRef	si le producteur utilise TaxRef pour son référencement, alors CodeNom = CodeNomTaxRef ; ce champ permet d'agréger des tables qui utiliseraient des ReferentielNom différents ;
obligatoire	Presence	texte	seules 2 valeurs possibles : oui / non	Valeur « non » = non observé ; cf. le cas échéant DenombComplement pour des précisions
optionnel	Denombrement	texte*	la quantité dénombrée	*valeurs possibles : valeur entière, valeur décimale, fourchette de valeur... d'où le format texte
optionnel	DenombComplement	texte	toutes spécifications nécessaires à la compréhension de Denombrement	valeurs possibles : grandeur mesurée (la métrique), ordre de grandeur, niveau de précision, niveau d'estimation...
obligatoire	DateDebut	entier	valeur ISO8601 de la date d'observation soit année mois jour	ex. : 20160530 pour 30 mai 2016
obligatoire*	DateFin	entier	valeur ISO8601 de la date d'observation soit année mois jour	*si la donnée concerne une date unique DateFin = DateDebut

	nom du champ	format	contenu	commentaire
obligatoire	EntiteGeographique	texte*	code spécifique à l'entité selon le référentiel utilisé ; si aucun référentiel n'est utilisé, c'est un numéro classant	si le fichier est au format SIG, ce champ est aussi dans la table attributaire ; *texte : le codage peut être de nature textuelle d'où le format générique texte
obligatoire	TypeGeographique	texte	type d'entité codée dans le référentiel : maille, commune... ou secteurproducteur si le type n'est pas référencé	
obligatoire	ReferentielGeographique	texte	citation du référentiel géographique utilisé ; valeur « néant » si aucun référentiel n'est utilisé	le référentiel peut être institutionnel (commune, cours d'eau...) ou propre au producteur
obligatoire*	X	décimal	coordonnées métriques en Lambert93 du point	*obligatoire si la géométrie est ponctuelle et que le fichier n'est pas au format SIG
obligatoire*	Y			
obligatoire	Sensibilite	texte	seules 2 valeurs possibles : oui / non	se référer au référentiel régional « Sensibilité » (en cours de réflexion) pour catégoriser la donnée
optionnel	Fiabilite	entier	seules 3 valeurs possibles : 1 / 2 / 3	hiérarchie : 3 > 2 > 1 ; se référer au référentiel régional « Fiabilité » (en cours de réflexion) pour catégoriser la donnée
optionnel	TypeObservation	texte	seules 3 valeurs possibles : terrain / littérature / collection	
obligatoire	Observateur	texte	la personne (ou l'organisme) à créditer de l'observation ; peut être complété avec l'auteur de l'identification du specimen	valeurs possibles : une personne, un organisme...anonyme, inconnu...
optionnel* conditionné	Producteur	texte	organisme ayant collecté l'observation auprès de l'observateur	ces trois champs permettent de gérer les multiples cas ayant conduit à l'élaboration de la table ; la notion de maître d'ouvrage permet de gérer une compilation de données issues de divers producteurs optionnel* conditionné: si ils ne sont pas cités dans la table, ils sont cités dans la métadonnée champ Responsable
optionnel* conditionné	Maitredouvrage	texte	organisme ayant compilé l'ensemble des données de la présente table	
optionnel* conditionné	Commanditaire	texte	organisme ayant commandité l'élaboration de la table	

Explicitation des termes Observateur Producteur Maitre d'ouvrage Commanditaire

OBSERVATEUR observation terrain	PRODUCTEUR organisme qui compile les données des observateurs.	MAITRE D'OUVRAGE organisme qui élabore la table à partir des données des producteurs.	COMMANDITAIRE organisme ayant commandité l'élaboration de la table.
OBSERVATEUR observation terrain			
OBSERVATEUR observation terrain	PRODUCTEUR organisme qui compile les données des observateurs.	Si c'est un producteur qui élabore la table, et sur ses propres données, alors maitre d'ouvrage = producteur.	Permet d'afficher qui est à l'origine de la demande d'élaboration.
OBSERVATEUR observation terrain			

5. Un modèle pour une architecture des métadonnées associées à un lot de données naturalistes

Ce modèle s'enracine dans la fiche de métadonnée telle que développée par GéoBretagne (cf. Ressources page 2). Tous les champs du modèle se retrouvent dans cette fiche et en respectent les normes d'écriture. Ainsi, le modèle est conforme aux obligations réglementaires imposées par la directive INSPIRE et assure l'interopérabilité entre logiciels.

	nom du champ	format	contenu	commentaire
obligatoire	TitreLotDonnee	texte	nom donnée au lot de donnée	
obligatoire	DescriptionLotDonnee	texte	description sémantique du contenu du lot de donnée	
obligatoire	IdentifiantLotDonnee	texte	code identifiant de manière unique* le lot de donnée	*une procédure possible pour assurer l'unicité : FR + n° Siren + nom du fichier (cf. wiki GéoBretagne)
obligatoire	ThemeISO	texte	autant de valeurs que nécessaire parmi le lot de mots clés fermés dans le référentiel ThemeISO de GéoBretagne	
obligatoire	ThemeInspire	texte	autant de valeurs que nécessaire parmi le lot de mots clés fermés dans le référentiel ThemeINSPIRE de GéoBretagne	
optionnel	MotClef	texte	valeur(s) au choix du producteur	
optionnel	ExtensionGeographique	texte	liste des limites administratives correspondant à l'extension géographique des données, ainsi que le référentiel administratif utilisé (ex. communes Geofla2015)	ne s'utilise que si l'extension correspond à une limite administrative : « Bretagne », « Départements 22 et 29 »...; seules les limites de la plus grande dimension sont citées (ex. les départements mais pas les communes contenues)
obligatoire	LatitudeN	décimal	les coordonnées métriques des 4 coins du rectangle de l'emprise maximale du lot de données, en Lambert93	l'automatisation du calcul est possible quand la fiche de métadonnée est directement remplie dans GéoNetWork
obligatoire	LatitudeS			
obligatoire	LongitudeE			
obligatoire	LongitudeO			
obligatoire	DateCreation	entier	valeur ISO8601 annéemoisjour de la date de création du lot de données	ex. : 20160530
obligatoire	DatePublication	entier	valeur ISO8601 annéemoisjour de la date de publication du lot de données	
obligatoire	DateRevision	entier	valeur ISO8601 annéemoisjour de la date de révision du lot de données	révision : correction apportée a posteriori, ajout de champs complémentaire, ajout de lignes de données, etc.
obligatoire	DateDebut	entier	valeur ISO8601 annéemoisjour de la première date de données	
obligatoire	DateFin	entier	valeur ISO8601 annéemoisjour de la dernière date de données	

	nom du champ	format	contenu	commentaire
obligatoire	Methode	texte	concerne les aspects méthodologiques d'acquisition sur le terrain ainsi que de traitement et d'analyse de la donnée	il est possible de citer un lien vers une documentation externe
optionnel	LimiteUtilisation	texte	citer les limites méthodologiques éventuelles liées à l'utilisation du lot de données	ne concerne que les aspects méthodologiques : « données non pertinentes dans telles conditions », etc.
obligatoire	EchelleUtilisation	texte	citer la gamme d'échelle pour laquelle le lot de donnée reste pertinent	
optionnel	ContrainteUtilisation	texte	citer les contraintes éventuelles (autres que méthodologiques) liées à l'utilisation du lot de donnée	ex. : « usage libre sous réserve des mentions obligatoires sur tout document de diffusion... », « ne pas diffuser ce lot de donnée en l'état car contient des données sensibles précises », etc.
optionnel	AccesDonnes	texte	lien pour accéder à la donnée	cas où la donnée peut être téléchargée via un site, une plateforme
obligatoire	Contact	texte	personne à contacter pour tout renseignement sur le lot de donnée ou sur son accès	l'information doit permettre d'accéder le plus directement possible à la personne ressource
obligatoire	Responsable	texte	organisation(s) ou personne(s) responsable(s) de la création, gestion, maintenance et diffusion du lot de données	ce champ permet d'exprimer aussi selon les cas : le/les producteurs et/ou le maître d'ouvrage et/ou le commanditaire

6. Fiabilité des données

Le producteur de la donnée peut vouloir exprimer une information sur la fiabilité de la donnée qu'il met à disposition.

Ainsi, dans un même lot de données, certaines peuvent être entachées d'un biais qui porte atteinte à leur qualité informationnelle (ex. compilation de données d'effectif quantifié exactement avec des données à créneau minimal et maximal).

La fiabilité peut concerner différents niveaux : taxonomique, géographique, méthodologique, etc.

Il ne faut pas confondre la notion de fiabilité telle qu'utilisée dans ce modèle avec celle de validation de l'information. En toute rigueur, une information non validée n'a pas à être échangée. Toutes les données concernées par le présent modèle sont considérées *a priori* comme validées.

La fiabilité s'exprime au travers de trois valeurs : « 3 » qualifiant une donnée fiable, qui est supérieure à « 2 », elle-même supérieure à « 1 ».

Un travail régional collaboratif reste à mener pour encadrer les critères déterminant ces trois valeurs.

7. Sensibilité des données

Certaines données peuvent être qualifiées de sensibles, dès lors qu'elles sont destinées à être diffusées de manière précise, suivant en cela la problématique portée par la directive Inspire ou par les chartes déontologiques des producteurs de données.

La communication du positionnement géographique précis de ces données dites sensibles peut éventuellement générer des nuisances à l'objet observé ou constituer une indiscretion sur la localisation de l'observateur.

Lorsque les données sont diffusées à une échelle communale ou encore plus large, elles ne sont généralement pas sensibles.

Les trois conséquences usuelles en matière d'échange de données sensibles sont :

- le producteur ne communique pas les données ;
- le producteur communique les données en l'état de l'observation, donc précises, mais elles sont marquées comme « sensibles » afin d'imposer des réutilisations spécifiques (ex. : gestion automatique du flou par des plateformes de diffusion ou non rediffusion par l'utilisateur primaire) ;
- les données sont communiquées avec un flou sur le positionnement géographique et marquées comme « sensibles » pour signifier cette spécificité géographique.

Un travail régional collaboratif concernant la sensibilité (définition, critères, espèces...) est en cours d'élaboration.