



Séminaire TAXREF – 22 janvier 2016
Muséum national d'Histoire naturelle
Table ronde des utilisateurs : quels usages, quelles perspectives ?

➤ **Simon Barreau (Office international de l'eau, ST Sandre)**

Le Service d'administration nationale des données et référentiels sur l'eau (Sandre) élabore le langage commun des données sur l'eau. Il est chargé d'établir la « normalisation » de ces données afin de rendre compatibles et homogènes la définition et l'échange des données entre les producteurs, les utilisateurs et les banques de données. D'un point de vue informatique, le Sandre garantit l'interopérabilité des systèmes d'information relatifs à l'eau.

Dans le cadre du Système d'information sur l'eau (SIE), le Sandre diffuse le référentiel des données sur l'eau. Ce référentiel, composé de spécifications techniques et de listes de codes libres d'utilisation, décrit les modalités d'échange des données sur l'eau à l'échelle de la France.

Ainsi, le Sandre diffuse et/ou administre certains référentiels. Sur le site du Sandre, on trouve des référentiels géographiques (exemple la BD Carthage ®) et des référentiels non géographiques (paramètres, méthodes, unités de mesure, appellations de taxons...).

Pour la taxonomie (référentiel nommé « appellations de taxons ») le ST Sandre administre et diffuse ce référentiel en s'appuyant notamment sur l'expertise du MNHN et sur l'utilisation de TAXREF. Le Sandre dispose de son propre référentiel taxonomique car les besoins des utilisateurs des données sur l'eau sont parfois non compatibles avec les règles de TAXREF. Un exemple parmi d'autres : TAXREF est un référentiel taxonomique où chaque taxon doit faire l'objet d'une publication et dont la présence en France doit être avérée. Cela n'est pas toujours possible pour le Sandre. Dans le référentiel Sandre on trouvera des « taxons » qui ne peuvent appartenir à TAXREF comme [34786] *Geissleria bourbonensis* f. anormale.

- Pourquoi utilisez-vous TAXREF ?

Parfois, certains experts du domaine de l'eau ne sont pas d'accord sur la taxonomie ou sur le libellé à utiliser pour décrire une espèce observée. Dans ces cas problématiques, le ST Sandre s'appuiera sur TAXREF pour prendre une décision sur le choix à effectuer pour la codification de son référentiel des appellations de taxons.

- Comment utilisez-vous TAXREF ?

Le ST Sandre suit les recommandations de TAXREF en manière de synonymie, d'auteurs de libellés à utiliser... De plus, un croisement est réalisé chaque année entre les taxons présents à la fois dans TAXREF et dans le référentiel des appellations de taxons du Sandre. Le CD_NOM est ainsi notifié dans le référentiel Sandre dès que cela est possible.

- Quels seraient vos besoins que ne rempliraient pas pour l'instant TAXREF ? / Quelles perspectives ?

Il serait intéressant de pouvoir suivre les évolutions pour chaque taxon entre les différentes versions de TAXREF. Par exemple, si un taxon référent devient synonyme entre deux versions de TAXREF, cela serait un plus de pouvoir repérer rapidement quels sont les taxons qui sont concernés par ces changements. Certains, taxons ne sont pas diffusés (certains niveaux taxonomiques, taxons dont la présence en France n'est pas avérée...) mais disposent néanmoins d'un CD_NOM. Pour le ST Sandre, avoir une version « étendue » de TAXREF où ces taxons seraient présents peut être profitable.

Pour finir, en plus du référentiel TAXREF en lui-même, le ST Sandre s'appuie régulièrement sur les administrateurs de TAXREF pour répondre à certaines questions taxonomiques pour la gestion de son référentiel. Cet accompagnement à l'utilisation de TAXREF nous apparaît comme essentiel.

➤ Mathieu Bossaert (CEN Languedoc-Roussillon)

- Pourquoi utilisez-vous TAXREF ?

Tout travail de consolidation et de comparaison de données nécessite l'utilisation de référentiels communs. TAXREF a été pour nous très tôt la référence évidente, permettant de mettre fin aux discussions sans fins relatives à la qualité de tel ou tel référentiel. Chacun pouvant y contribuer et faire remonter d'éventuelles erreurs. Son intégration en base de données est simple et son utilisation aisée (http://si.cenlr.org/postgresql_fdw_taxref & http://si.cenlr.org/integration_des_statuts_de_protection)

- Comment utilisez-vous TAXREF ?

Comme référentiel de base pour l'ensemble de nos outils de collecte de données (web ou mobile) et comme référence centrale au sein de notre SI, auquel nous rattachons de nombreuses données. C'est aussi le référentiel que nous utilisons pour exporter et envoyer nos données aux partenaires.

Il est stocké au sein de notre base de données PostgreSQL au sein d'un schéma propre «l'INPN». Nous utilisons aussi les données de réglementation diffusées par l'INPN.

A la manière de ce que nous trouvons sur l'INPN, le CD_NOM et constitutif de l'url d'accès aux fiches de synthèse relatives à un taxon (cela facilite aussi le lien vers des références extérieures) :

- <http://atlas.libellules-et-papillons-lr.org/papillons/monographie/53576>

De par sa structure arborescente, facile à exploiter de manière récursive, nous avons travaillé en 2015 à la mise en place d'une API d'accès aux données d'observations rattachées à ce référentiel (<https://github.com/CENSIG/api>). Si nous nous représentons TAXREF comme un arbre inversé. Nous pouvons, en tout nœud de l'arbre explorer les nœuds descendants, jusqu'aux feuilles, de manière très performante : ceci est une maquette (<https://github.com/CENSIG/client-interface>), qui affiche les données issues de l'API. En cliquant sur le bouton bleu à droite de l'écran, on peut descendre dans l'arborescence jusqu'aux taxons terminaux.

- Quels seraient vos besoins que ne rempliraient pas pour l'instant TAXREF ? / Quelles perspectives ?

L'intégration à TAXREF, ou plutôt la fourniture en référentiels joints, du statut des taxons dans les listes rouges « labellisées » IUCN est très attendue par mes collègues des CENs. Ce paramètre est de plus en plus utilisé pour définir les enjeux écologiques de nos sites.

➤ Emilie Gauthier (IFREMER)

L'Ifremer gère le système d'information Quadrigé, qui permet la bancarisation et la diffusion des données nationales de surveillance environnementale et sanitaire du milieu marin. Il est le SI de référence pour la gestion des données acquises pour la DCE dans les eaux littorales. A ce titre, Quadrigé fait partie du SIE et participe aux travaux du SANDRE. La bancarisation des données s'appuie sur des référentiels spécifiques à ce système d'information car créés dans les années 80 à la mise en place de la surveillance nationale opérée par l'Ifremer. Aujourd'hui, ces référentiels ont été rendus interopérables avec le SANDRE pour permettre l'échange et le partage des données.

Parmi ces référentiels, on trouve le référentiel taxinomique, composé :

- Des taxons : liste interne Ifremer de 1996, actualisée avec l'ERMS en 2006, puis avec le WoRMS depuis 2009
- Des regroupements taxinomiques correspondant à des unités taxinomiques identifiables sur le terrain (ex : algues filamenteuses) ou pratiques pour le traitement des données (ex : espèces carnivores).

Depuis 2015, une nouvelle application de gestion des données de suivi des récifs coralliens, nommée BD Récif, a été développée. Partageant le modèle de données et les référentiels Quadrigé, BD Récif utilise des listes taxinomiques établies via TAXREF. Cela implique :

- La gestion des CD_NOM de chaque taxon
- L'utilisation des informations biogéographiques de TAXREF pour établir les listes régionales.

Le travail de mise en correspondance du référentiel taxinomique Quadrigé (dont BD Récif) avec TAXREF a fait l'objet d'un effort important en 2015 pour les équipes du MNHN et de l'Ifremer.

Le bilan est le suivant :

- Nb de taxons Q² codifiés TAXREF : 28 553 / 34 884 taxons soit 82% des taxons sont codifiés
- Dont 812 taxons à codifier sont associés à des données (515 d'entre eux sont référents, les autres sont des synonymies).
- Pour les suivis des récifs coralliens (outil BD Récif), des filtres taxons régionaux ont été constitués. L'utilisation des statuts biogéographiques de TAXREF n'était pas suffisante pour le faire (ex : 27 taxons identifiés dans les jeux de données réunionnais n'avaient aucun statut biogéographique dans TAXREF pour La Réunion).

Outils utilisés :

- Chargement de la table TAXREF v8.0 format .txt dans notre base de données pour requêtage (mapping sur les libellés de taxons et éventuellement les auteurs)
- Si un taxon n'est pas trouvé par ce biais-là : utilisation de l'outil de mapping http://taxref.mnhn.fr/taxref-match/spring_security_login
- Tentative de complément via le Sandre (taxon Quadrigé → taxon Sandre → Taxon TAXREF) : ajout de 1418 codes. Il reste 1537 taxons codifiés Sandre et dont le CD_NOM n'est pas renseigné (ni dans Quadrigé, ni dans le Sandre), et dont 1480 sont des référents.

Quels seraient vos besoins que ne rempliraient pas pour l'instant TAXREF ? / Quelles perspectives ?

Tout comme le SANDRE, l'Ifremer aurait besoin d'accéder à l'ensemble des taxons codifiés dans TAXREF, pour faciliter les expertises.

L'Ifremer gère à la fois les équivalences avec le SANDRE, et avec le WoRMS. Dans la mesure où TAXREF contient des équivalences avec le WoRMS, et que le SANDRE gère les équivalences avec TAXREF, il est indispensable d'organiser le circuit de gestion de ces tables de transcodage pour éviter le doublonnage des tâches, et mutualiser les moyens mis en œuvre pour compléter les équivalences.

L'Ifremer a également besoin de ressources humaines pour rechercher les références bibliographiques permettant d'ajouter les taxons observés dans les eaux littorales françaises dans TAXREF, et de fournir les informations utiles à l'ajout des statuts biogéographiques des taxons dans TAXREF.

➤ Johan Gourvil (FCBN)

- Pourquoi utilisez-vous TAXREF ?

Car pour la flore vasculaire de métropole, il est basé sur l'Index de Kerguelen (Index synonymique de la flore de France), le référentiel historique utilisé par de nombreux CBN. Parce qu'il a été retenu comme référentiel d'échange de données commun aux CBN. Car il constitue le Référentiel officiel du SINP.

- Comment utilisez-vous TAXREF ?

Comme référentiel d'échanges de données au sein du réseau des CBN et avec nos partenaires (SPN, Tela Botanica...) à partir des noms valides (CD_REF).

Comme référentiel de gestion des données (observations).

- Quels seraient vos besoins que ne rempliraient pas pour l'instant TAXREF ?

Nécessité d'avoir un référentiel qui ne change pas tous les 6 mois. La stabilité d'une version est un gage de fiabilité. Eviter de suivre systématiquement la dernière publication au risque de faire constamment des allers-retours.

Assurer une meilleure traçabilité des modifications d'une version à l'autre. TAXREF_CHANGES ne remplissant pas totalement ce besoin pour les rangs supérieurs à l'espèce. Il semble que le fichier TAXREF_CHANGES n'indique que les changements effectués pour des éléments du rang espèce ou d'un rang inférieur.

TAXREF n'est pas suffisamment pensé pour être un outil synonymique.

TAXREF n'est pas fait pour gérer de la donnée au regret des utilisateurs de ce référentiel qui eux gèrent de la donnée et doivent donc créer des groupes artificiels sous peine de perdre des données.

- Quelles perspectives ?

Interactions avec les référentiels internationaux (Euro+Med, The Plant List...).

- Pourquoi utilisez-vous TAXREF ?

En tant que producteur du référentiel source de TAXREF sur la flore vasculaire de France métropolitaine (BDTFX), l'utilisation de TAXREF dans nos outils et nos plates-formes internes est assez limitée. Nous utilisons principalement TAXREF pour échanger des données d'observations avec des partenaires et plates-formes SINP, notamment les CBN. TAXREF permet alors d'interagir avec d'autres structures sur des bases de connaissances utilisant les mêmes index taxonomiques. C'est un point très positif.

- Comment utilisez-vous TAXREF ?

Nous utilisons principalement le CD_NOM de TAXREF pour permettre une correspondance entre TAXREF et les autres référentiels.

- Quels seraient vos besoins que ne rempliraient pas pour l'instant TAXREF ? / Quelles perspectives ?

TAXREF est un référentiel intégratif qui se veut unique sur l'ensemble des régions couvertes. De ce fait, il implique une fusion de données taxonomiques et nomenclaturales en provenance de différentes sources géographiques. Cette fusion est une opération difficile qui nécessiterait des moyens relativement importants pour être pleinement satisfaisante sur le plan scientifique, ce qui ne peut être réalisé actuellement compte tenu des moyens alloués à la constitution des référentiels.

Afin de contourner cet écueil et pour travailler au plus près des acteurs de terrain nous privilégions les référentiels sources proposés par les structures locales, avant fusion dans TAXREF.

Tela Botanica travaillant à l'échelle de la francophonie, il n'est de toute façon pas envisageable actuellement de fusionner tous les référentiels que nous utilisons (Afrique du Nord, Liban, Afrique de l'Ouest, Afrique Centrale...)

Parallèlement à cela nous travaillons à la mise en correspondance des référentiels locaux avec des référentiels de plus large couverture géographique. En 2014-2015, nous avons travaillé à la mise en correspondance de la BDTFX et avec *The Plant List*. De cette étude ressort un besoin d'échange entre référentiels locaux et référentiels plus généraux qui bénéficie aux deux parties, sans pour autant fusionner ces référentiels (détections des erreurs, problématiques soulevées par des synonymies distinctes ou des noms retenus différents, etc.) qui sont traitées au cas par cas.

En résumé, TAXREF est une nécessité impérieuse à l'échelle nationale pour permettre les échanges des données entre différentes bases et participer à l'élaboration du SINP. En revanche, en ce qui concerne la botanique, domaine dans lequel les pratiques et les spécificités locales en matière de conservation sont très différentes d'une région naturelle à l'autre, et où les travaux de systématique en sont à des stades d'avancement très inégaux selon les groupes et les régions, la fusion des référentiels nous semble poser plus de problèmes qu'elle n'en résout. Chaque référentiel territorial étant utilisé isolément, sa fusion dans un référentiel unique n'apporte pas de valeur ajoutée supplémentaire aux utilisateurs mais introduit nécessairement des problèmes difficiles à résoudre. Un référentiel propre à chaque grand groupe biologique (botanique, zoologique...) et à chaque territoire biogéographiquement distinct et distant (Métropole + Corse, Réunion + Mayotte, Nouvelle-Calédonie, Les Antilles, la Guyane, etc.), exploitant correctement les sources d'information disponibles, nous aurait ainsi semblé plus pertinent.