

Les standards de données

Présentation

Les standards de données proposés par la FCBN sont des outils communs permettant de faciliter l'échange de données au sein du réseau.

Historique

Le travail sur les formats standards de données est un travail de longue haleine qui a débuté en 2009. Ce travail, décomposé en 4 domaines - [DATA](#), [TAXA](#), [SYNDATA](#), [SYNTAXA](#) - s'est concentré dans un premier temps sur le DATA. Une première version du FSD (Format Standard de Données) DATA a été proposé en juillet 2010 :

- FSD DATA v1.4

Ce format standard a été utilisé pour réaliser la première agrégation nationale de données d'observation. Cette agrégation s'est finalisé en février 2013 par la mise en place de l'Atlas de la flore et l'affichage de résultat de cette agrégation (<http://siflore.fcbn.fr>).

Dans un deuxième temps, le FSD TAXA a été élaboré à partir de 2013. Ce travail s'est fait sur la base des travaux réalisés en 2010 (analyse des besoins), des catalogues de la flore existant dans les CBN et la tenu de Groupe de Travail SI FLORE. Le travail d'agrégation du TAXA a été validé par le CODIR et le le FSD TAXA a été validé par le GT SI FLORE en avril 2014.

- FSD TAXA

et

cadrage de l'exercice

Dans un troisième temps, le FSD SYNDATA a été élaborer pour partager les premières données de végétation. Ce travail a été réalisé en collaboration avec le CBN MCE (contact : Gilles Bailly). Ce FSD a été largement inspiré par ce qui avait été proposé pour le DATA, adapté pour les relevé de végétation. Il a été utilisé par un seul CBN, dans l'objectif de réaliser un test et montrer la capacité du réseau à répondre à cette demande de partage.

- FSD SYNDATA

et

cadrage de l'exercice

La perspective de travailler sur la partie SYNDATA et SYNTAXA, ainsi que la mise à jour des données DATA et TAXA ont ouvert la réflexion pour une mise en cohérence des FSD. De plus, la compatibilité avec le FSD SINP devait être recherché, FSD qui lui aussi était en cours d'évolution.

Il a été proposé alors une révision des FSD pour que ceux ci soit i) plus cohérent entre eux, ii) plus durable et iii) plus flexible afin de permettre des évolutions sans chambouler la totalité des modèles. Une première version d'un FSD TOTAL a été proposé :

- première version du FSD TOTAL

Une procédure un peu plus carrée a alors été réalisée concernant les “Standards” de données : l'idée a été de définir un modèle conceptuel commun à tous les domaines d'échanges (data, taxa, syndata, syntaxa), un dictionnaire de données commun avec des champs transversaux qui puissent être utilisé dans les différents FSD, un modèle conceptuel pouvant être directement utilisé pour générer une base de données - permettant ainsi l'accueil des données au format standard, des FSD qui découle de tout cela, accompagné de vocabulaires contrôlé, discuté au sein du réseau. Tous ces éléments sont présentés sur ce wiki et ont vocation à évoluer.

La partie SYNTAXA a été développée par Anaïs de manière un peu indépendante à cette réflexion. Cependant, les éléments apportés sur ce sujet ont aujourd'hui trouvé leur place sur le wiki.

Les différents chapitres

Les éléments composant ces standards sont :

- le [modèle conceptuel de données](#)

Le modèle conceptuel de données (MCD) aide à analyser la structure conceptuelle du système d'information global, afin d'identifier les principales entités à représenter, leurs attributs et les relations entre ces attributs. Ce schéma ne représente que les “grands concepts” que nous souhaitons partager ainsi que les relations entre ces concepts.

- Le [dictionnaire de données](#)

Le dictionnaire de données (DDD) définit de manière plus précise les champs issus des concepts définis dans le MCD. C'est une liste de champs qui sera utilisé pour créer les formats standards d'échanges et le modèle physique de données.

- Le [modèle physique de données](#)

Le modèle physique de données (MPD) est une représentation plus concrète du Modèle conceptuelle, applicable en base de données. Il donne une vision globale avec un certain degré de précision des éléments à partager. Ce MPD sera utilisé tel quel dans le hub.

- Les [vocabulaires contrôlés](#)

Les vocabulaires contrôlés (ou listes de valeur) définissent les valeurs utilisables pour certains champs. Si ces champs ne sont pas renseignés avec une valeur issue de cette liste, le jeu de données sera considéré comme non conforme. ces vocabulaires contrôlés ont vocation à évoluer pour permettre au réseau de partager les données qu'ils souhaitent. Ces évolutions n'impacte pas les MCD et MPD (modèle conceptuel et modèle physiques)

- Les [formats standards de données](#)

Les formats standards de données décrivent la manière de partager les données selon un format spécifique. Il s'appuie sur les dictionnaire de données et définit un ordre particulier et des règles de renseignement spécifiques. Pour le moment, 3 formats sont disponibles : [FSD META](#), [FSD DATA](#), [FSD TAXA](#)

Évolutions à prévoir

- homogénéisation avec le SYNTAXA

Une homogénéisation de la conceptualisation des standards serait intéressant entre le SYNTAXA et le reste des standards (gestions des référentiels en dehors du MCD et MPD, homogénéisation de nomenclature des champs et passage par un dictionnaire de données).

- Procédure d'évolution des standards

Aujourd'hui, la procédure d'évolution des standards proposé pour l'évolution des standard n'a pas été validé par le réseau. Il reste à l'état de proposition et doit être discuté pour que le réseau se retrouve là dedans.

Procédure d'évolution des standards

Aucune procédure n'est encore en place mais il est temps d'en proposer une. En effet, afin d'éviter de gagner de temps en évitant de repartir à zero, de permettre de faire évoluer les outils et script de manière itérative et non brutale, il est nécessaire de partager une méthode de mise à jour des Standard avec les utilisateurs (= le réseau).

La procédure proposée est la suivante :

1. Les standards sont versionnés avec une fréquence de 6 mois (2 par ans),
2. Les standards sont "gelés" entre 2 versions c'est à dire qu'ils ne connaissent aucune modification entre ces 2 versions, à l'exception des vocabulaires contrôlés dont l'évolution n'impacte que peu,
3. Les demandes de modifications et points de discussion concernant le standard sont recensés sur le wiki sur la page [Propositions/évolutions standard](#). Toutes personne peut proposer des évolutions, à condition qu'elles soient i) personnalisées (qui fait la proposition) et ii) argumentées (texte qui accompagne la proposition d'évolution). Il est également possible de proposer une remise en question sans pour autant proposer une évolution concrète des standards.
4. A l'occasion d'une GT SI FLORE/VEG, tous les 6 mois, les propositions d'évolution et questions sont débattus (les éléments concernant ces propositions sont figés 2 semaine avant le GT pour préparer les réflexions). Toutes propositions sont votés (acceptées, repoussées ou refusées) et le résultat de ces votes sont consignés dans un rapport d'évolutions
5. Le rapport d'évolution est par le CODIR suivant le GT. Si certains éléments de ce rapport ne sont pas validé par le CODIR, il sont remis à l'ordre du jour du GT suivant.

Suivi des modification

- [Changement entre la version 3.1 et 3.2](#)
- [Changement entre la version 3.2 et la 3.3](#)

Note admin sur le lien entre le Hub et le Codex : les standards (ddd,fsd,voca_ctrl) sont géré à travers le codex, rubrique "FSD". La fonction `export_fsd(version varchar)` permet l'export des standards sur le serveur (dossier /home/hub/00_ref et /home/hub/01_fsd) afin qu'ils soient utilisé par le hub.

From:

<https://wiki.fcbn.fr/> - **Wiki - FCBN**

Permanent link:

https://wiki.fcbn.fr/doku.php?id=les_standards_de_donnees&rev=1475065894

Last update: **2022/03/07 12:00**

