



Atelier technique : schéma de données pour les itinéraires

Compte-rendu - 07/06/2021

Cet atelier et cette démarche existe grâce à l'implication technique, financière et logistique du Parc national des Écrins et bénéficie également de l'aide des fonds européens via le programme inter-régional ALCOTRA et son PITEM MITO, PS 2 Outdoor DATA.



Rédaction :

Lise Le Lann

Participants :

- Apidae Tourisme : Serge Bregliano et Pierre Granger
- Conseil départemental des Côtes d'Armor : Eric Ravenet
- Conseil départemental des Alpes-de-Haute-Provence : Nadège Marboutin
- Etalab / DINUM : Geoffrey Aldebert
- IGN : Stephanie Barriere-Riffort, Delphine Boulogne et Herver Quinquel
- Parc national des Écrins : Camille Monchicourt, Lise Le Lann, Cendrine Hoarau
- Parc national des Cévennes : Amandine Sahl et Idrissa Djepa Creutz

Objet :

Définir le format du schéma de données.

Supports de travail :

- https://github.com/PnX-SI/schema_randonnee
- https://github.com/PnX-SI/schema_randonnee/blob/master/schema.json

Ordre du jour :

- 1 Avancement du projet depuis le précédent atelier (07/05/2021)
- 2 Acter sur le format du schéma
- 3 Acter sur les champs complexes

1 Avancement du schéma depuis le précédent atelier

Des tests d'automatisation des exports de données Goetrek conformes au schéma de données ont été réalisés.

Le Parc national des Cévennes a testé différents formats de schéma proposés par Etalab pour la publication de schémas.

Les formats sur lesquels le Parc s'est attardé sont les *tables schema* et les *Json schema*.

- o *table schema* : ce schéma a été testé du au fait que le support de travail du Groupe de travail était un tableur partagé. Les deux principaux freins à ce format sont la valeur des champs (qui peut être différente d'une structure à l'autre) et la gestion des objets complexes qui n'est pas permise par ce format.
- o *Json schema* : ce format est intéressant pour traiter des données géographiques et permettre de hiérarchiser les objets.

Un premier squelette de schéma est déjà présent dans le dépôt GitHub. Afin de pouvoir publier le schéma il faut travailler sur les contraintes de certains champs.

2 Validation du format du schéma

Entre les formats tabulaires et Json, lequel est le plus pertinent pour publier un schéma de données d'itinéraire ?

Discussions :

- Le format Json
 - o Il est plus adapté que le format csv pour les données géographiques.
 - o Ce format n'est pas un frein pour l'export de données issues de divers outils.
 - o Ce format peut être un frein pour des utilisateurs non techniques

- o Il permet de gérer des objets et donc les champs hiérarchiques.
- o Ce format peut être généré par SIG même si QGIS ne permet pas la hiérarchisation des champs
- o Etalab conseille ce format pour des données géographiques. IGN et APIDAE sont aussi favorables à ce format.
- Format csv :
 - o Les champs et leurs hiérarchies définis lors du précédent atelier rendent l'écriture complexe pour ce format. Si nous le sélectionnons une mise à plat des champs est nécessaire afin de réduire au maximum les objets.
 - o Il est plus facile de faire une interface de saisie pour ce format que pour le Json
- Pour nous aider à choisir le format du schéma il faut se poser la question suivante : Qui va utiliser la donnée ?
 - o Les techniciens de SI ont l'habitude d'utiliser ce format mais si c'est une personne moins technique le format va être compliqué à utiliser.
- Peut-on faire un mix des deux formats ? C'est-à-dire dans le Json mettre à plat les champs pour éviter l'arborescence ?
 - o C'est possible mais ça retire la force du Json, autant utiliser un csv.
- On peut créer deux jeux de données. Le premier lié au schéma (Json) et un autre non lié au schéma (csv avec une colonne coordonnée).
 - o Les coordonnées sont difficiles à utiliser pour une personne moins techniques.
 - o Le csv peut être rajouté en fichier complémentaire, mais pas rattaché au schéma afin d'éviter les confusions

Format retenu : **Json**

3 Présentation de la trame du schéma et validation du type de certains champs.

Le début de schéma concerne les métadonnées et le corps du schéma comprend les champs.

- Les contraintes présentes ici sont celles identifiées dans la V1
- Le champ **structure** est obligatoire mais son champ qualificatif ne l'est pas, ce qui génère une arborescence. Le problème est que l'on double l'information en créant deux champs.
 - o Pour le niveau 1 il est possible de faire une table des structures
 - o Le champ contact correspond à celui qui produit la donnée ? (voir CR atelier 2)
- Le champ **nom itinéraire** est obligatoire
- Il faut que les champs complexes ne soient pas des champs obligatoires sinon cela va freiner les créateurs de données.
- Le champ **pratique** doit avoir une liste afin de limiter les variations d'écriture (ex : à pied, pied, pédestre....)
 - o Il est difficile d'imposer un vocabulaire alors qu'il n'existe pas de référentiel.
 - o Peut-on faire des recommandations de vocabulaire sur le format Json ?
 - Peut être décrit dans la description du champ et en recommandation

- Actuellement il n'est pas possible de mettre des recommandations dans un schéma.
 - Si une liste n'est pas proposée elle peut l'être par la suite car les schémas évoluent
 - Les listes sont intéressantes pour faire des mappings (faciliter les imports et exports de données) et pour faire des analyses de données. Il peut y avoir une valeur autre afin de faire évoluer la liste en fonction des besoins.
- o Travail de création de liste de pratiques compatibles avec le contexte d'itinéraire (liste actée) :
 - Pédestre
 - Ne pas mettre vélo mais :
 - VTT
 - Cyclo
 - Gravel
 - Descente
 - Trail
 - Équestre
 - Ski de randonnée
 - Ski de fond
 - Ski orientation
 - Raquette
 - Autre
- Le champ **type d'itinéraire** :
 - o Les valeurs de ce champ doivent être en lien avec l'ontologie de DataTourisme
 - o valeur de la liste actée
 - Aller simple
 - Aller / retour
 - Boucle
 - Itinéraire
 - Etape
- Le champ **commune**, la valeur et le type de ce champs varient selon les sources étudiées (nom de la commune et code INSEE)
 - o Le code INSEE n'est reconnu uniquement en France et ne prend pas en compte le nom non officiel comme c'est le cas en station
 - o Ce champ est un champ composé actuellement, si le créateur de données code mal ce champ, il risque d'y avoir un décalage des valeurs dans le fichier. Ce champ est compliqué à gérer mais si on le scinde en deux, on efface l'atout du format Json
 - On le sépare en deux champs avec le **code INSEE** comme champ obligatoire.
- Le champ **média** est un champ complexe mais il est facultatif. Il comprend une arborescence dont le format et le type
- Le champ **niveau** reste un champ libre non standardisé en attendant qu'un référentiel soit construit.

4 Poursuite du travail

- Reprendre le schéma avec les éléments actés lors du 2nd atelier
- L'hébergement du schéma reste identifié Parcs nationaux le temps de finaliser cette version après migration chez Etalabe