

Synthèse de « Guide pour la constitution des jeux de données du SINP et le renseignement des métadonnées associées », Dupont, Touroult 2018 (INPN)

Synthèse du document : des définitions (cadre et jeux de données) ; des listes de mots clés standardisés pour les métadonnées ; en annexe : des définitions commentées puis des exemples de métadonnées.

Définitions

1. Métadonnées

sont des informations structurées servant à décrire une source de données (jeu ou série de données). Elles s'attachent à décrire aussi bien le contexte d'acquisition des données que les jeux de données.

Les métadonnées servent à « permettent une meilleure réutilisation des données, cohérente avec l'usage recherché. Elles permettent aussi d'augmenter la visibilité des structures impliquées dans l'acquisition des données.

« les objectifs et la manière dont les données ont été acquises »

« (1) les maîtrises d'ouvrage qui ont la responsabilité de décrire les métadonnées des cadres d'acquisition des données, (2) les producteurs de données qui ont la responsabilité de structurer des données acquises en jeux de données homogènes et de décrire les métadonnées associées »

2. Cadre d'acquisition

Contexte stratégique et opérationnel d'un projet ou d'un programme conduisant à la constitution d'un ou de plusieurs jeux de données.

Le CA est piloté et mis en place par : un maître d'ouvrage (administration et financement du projet) et un maître d'œuvre (acquisition des données).

Il correspond :

- au(x) objectif(s) à atteindre et aux besoins en termes d'acquisition de données ;
- au dispositif d'administration mis en place et piloté par une maîtrise d'ouvrage. Cette dernière élabore le montage financier du projet ou du programme et définit le calendrier des actions en partenariat avec une maîtrise d'œuvre (la maîtrise d'ouvrage « commande ») ;
- au(x) dispositif(s) de collecte associés aux observations et au(x) dispositif(s) d'informatisation et de gestion des données mis en place et pilotés par une maîtrise d'œuvre (la maîtrise d'œuvre « fait ou fait faire ») ;
- au(x) opérateur(s) qui sont responsables de la réalisation par délégation de la maîtrise d'œuvre (l'opérateur « fait »).

3. Jeu de données

Regroupement en unité homogène dictée par les besoins d'analyses et de réutilisations futures des données, ou fonction de recommandations spécifiques de la maîtrise d'ouvrage en lien avec les financeurs ou différents partenaires. Lié à la mise en place d'un protocole de collecte de donnée par exemple.

Lien direct entre JDD et l'usage que l'on peut faire des données.

Les JDDs sont liés à des processus réalisés ou décrits par un ou plusieurs maîtres d'œuvre ou opérateurs : collecte, production et gestion, validation diffusion dans le SINP des données

Les métadonnées d'un JDD permettent aussi de mettre en avant des personnes ou organismes ayant contribué à l'acquisition des données : on peut individualiser les jdd par financeurs.

Typologies

1. CA

Il est important d'associer une typologie à la démarche du CA (qui est la description de l'objectif du programme ou du projet). Cette typologie est la suivante :

1. Inventaire et cartographie :
2. Suivi/surveillance dans le temps
3. Expérimentation/recherche
4. Multiple ou autres

2. JDD

Paramètres en fonction du processus de collecte. Le cas échéant, il est préconisé de découper les JDDs en fonction de cette typologie :

- (1) Processus de collecte non intentionnel ou de réutilisation.
- (2) Processus de collecte opportuniste.
- (3) Processus de collecte étendu spatialement.
- (4) Processus de collecte restreint spatialement
- (5) Processus de collecte par piégeage ou échantillonnage indirect.
- (6) Processus de collecte compilatoire.

Conseils

- Grouper en CA puis en JDD
 - les données acquises dans le même objectif (même programme, projet, étude) : dans un même CA
 - les données financées / commandées par les mêmes structures : dans un même CA
 - les données collectées selon le même protocole : dans un même JDD
 - les données possédant le même profil de données : même JDD
- le groupe taxonomique, le périmètre géographique etc. sont des critères secondaires de groupement
- en cas de manque d'information (protocole inconnu, pas d'info sur le commanditaire, etc.) regrouper dans un JDD
- critères à éviter dans la création de JDD
 - séparation par année ou période de collecte si cette distinction n'est pas liée à un changement de protocole ou d'acteurs
 - séparation des taxons observés selon un même protocole (perte d'information de communauté d'espèce)
- L'usage des données est d'autant plus efficient si les métadonnées associées sont décrites de manière précise et détaillée
- Un JDD contenant des données homogènes en termes de collecte permettra un usage postérieur maîtrisé.

Commenté [RL1]: Ajout : groupement par zonage lorsqu'on peut faire une recherche de données sur cette zone (exemple : natura 2000). A confirmer

Définitions supplémentaires

- Rôle du ou des acteurs au niveau du CA
- **MOA** : Le ou les acteurs responsables de la maîtrise d'ouvrage. Cette dernière est constituée par une ou des personnes physiques ou morales pour laquelle un programme ou un projet est mis en œuvre ou réalisé (la maîtrise d'ouvrage « commande »).
- **MOE** : Le ou les acteurs responsables de la maîtrise d'œuvre. Cette dernière est constituée par une ou des personnes physiques ou morales qui a/ont la charge d'une ou de plusieurs phases opérationnelles d'un cadre d'acquisition (la maîtrise d'œuvre « fait ou fait faire », il coordonne, anime, pilote...).
 - L'opérateur responsable de la réalisation par délégation de la maîtrise d'œuvre (l'opérateur « fait »)
- **Financier** : Le ou les acteurs engagés dans le financement des actions (généralement, c'est le maître d'ouvrage qui finance directement ou indirectement).
- Rôle du ou des acteurs au niveau du JDD
- **Le producteur** du jeu de donnée. C'est la personne physique ou morale qui anime la saisie des données et des métadonnées associées dans la base de données dont il a la charge. Ce n'est pas nécessairement l'observateur. Il a autorité pour modifier les données contenues dans le jeu. Une personne physique ou morale fournissant un outil naturaliste de gestion de données, ne peut être considérée comme producteur, si celle-ci n'a pas autorité à modifier les données
- **Le fournisseur du jeu de données**. C'est la personne physique ou morale intermédiaire qui a récupéré et transmet le jeu de données et les métadonnées associées, au SINP. Dans la grande majorité des cas, le producteur est aussi le fournisseur.

Commenté [RL2]: NB : c'est le rôle qui définit la propriété de la donnée dans l'administration de notre base GeoNature 2

Commenté [RL3R2]: Non, maintenant c'est fournisseur...

Commenté [RL4]: Le fournisseur ça serait le PNR NM, ou Odin pour Normandie et Biodiv PdL pour Pays de la Loire (?) pas très important