



Coordination scientifique et technique
Système d'Information sur la Nature et les Paysages
Observatoire National de la Biodiversité



AVIS DE LA CST
(SUITE A UNE SAISINE)

Rappel des étapes :

- l'avis est pré-rempli par le secrétariat de la CST et envoyé aux rapporteurs avec copie aux contributeurs
- le rapporteur principal, en lien avec le co-rapporteur, rédige une première version de l'avis et recueil des remarques et compléments des contributeurs. Pour assurer un suivi, merci de mettre en copie le secrétariat de la CST lors des échanges de ce document entre rapporteurs et contributeurs.
- une fois consolidé, l'avis est transmis par le rapporteur principal au secrétariat de la CST :
Adresse : secrétariat_CST@mnhn.fr

Numéro de saisine	3 [19/09/2011]
Titre	Standard de métadonnées
Rappel des termes principaux de la saisine	<p>Question principale : Les résultats du groupe de travail « Standard de métadonnées » sont-ils cohérents, scientifiquement et techniquement, par rapport aux objectifs du SINP ?</p> <p>Sous-questions :</p> <ul style="list-style-type: none"> - La qualité scientifique du document final est-elle satisfaisante ? si non, quelles sont les améliorations à apporter ? - La qualité technique du document final est-elle satisfaisante ? Préciser les critères de satisfaction ? si non, quelles sont les améliorations à apporter ? - Quelles recommandations la CST peut-elle émettre en terme d'implémentation de ce modèle de métadonnées UML sur divers SI (outil naturaliste, Géosource, autres SI) ?
Rapporteurs	Michel DESHAYES
Contributeurs	Régine VIGNES, Patrice PRUVOST, Laurent COUDERCY

Historique du document

Version	Auteur	Date	Commentaires
V0	Secrétariat CST	04/01/2012	Envoi du document type
V1	Michel Deshayes	07/03/2012	Projet d'avis a destination des contributeurs
V2	Michel Deshayes, Régine Vignes	07/03/2012	Après prise en compte des remarques de Régine Vignes
V3	Michel Deshayes, Régine Vignes	05/07/2012	Après prise en compte des remarques de Régine Vignes

...

Préconisations à adapter selon la saisine

L'avis doit être synthétique (1 à 3 pages) avec des renvois en annexes si nécessaire (commentaires détaillés...). Les différents avis des rapporteurs et contributeurs doivent apparaître clairement et être synthétisés sous forme en faisant apparaître les points de consensus et les points de divergences.

On peut conseiller deux approches :

- *une analyse par document soumis*
- *une analyse par question et sous-question de la saisine*

Dans tous les cas, une conclusion est demandée afin de répondre aux interrogations de la saisine.

Les conseils et recommandations dépassant le cadre de la saisine doivent bien évidemment être conservés mais seront synthétisés dans un ou des paragraphes spécifiques.

Le Système d'Information sur la Nature et les Paysages (SINP) vise à faciliter l'accès à l'information aux données naturalistes, ainsi qu'à faciliter l'accès aux données mêmes, dans le respect des droits de propriété des producteurs. Pour avoir accès aux métadonnées, c'est-à-dire aux informations sur les contenus des bases de données des producteurs, il faut pouvoir échanger ces métadonnées et la définition de standards de métadonnées devient alors une pierre angulaire dans la construction du SINP.

La présente saisine concerne un avis à donner sur le rapport du groupe de travail « Standard de métadonnées », qui a travaillé d'octobre 2010 à juin 2011, et produit le document « Description du profil de métadonnées des ressources du Système d'Information sur la Nature et le Paysage (SINP) » (V1 du 19 octobre 2011).

Les contraintes de ce travail ne sont pas simples. Le SINP est un outil de travail collaboratif qui a vocation à fédérer tous les acteurs naturalistes. En même temps, si chacun comprend l'intérêt d'un catalogue de données, le remplissage d'un formulaire de métadonnées est parfois considéré comme une corvée (nécessaire). Un bon standard doit donc répondre aux contraintes suivantes :

- être suffisamment complet pour vraiment renseigner sur le contenu, mais aussi léger que possible pour ne pas être un obstacle lorsqu'il s'agit de décrire ces données,
- prendre en compte les besoins des différents acteurs, liés à la diversité des objets étudiés (phanérogame, mammifère, oiseau, insecte...) et aux actions envisagées (inventaire simple, observatoire, recherche...).

Le groupe de travail a réuni une diversité d'acteurs, producteurs de données, gestionnaires d'espaces naturels, scientifiques...

Après avoir rappelé l'objectif du système d'information sur la nature et les paysages, les concepts nécessaires à sa mise en œuvre, les dispositifs existants, le rapport traite successivement :

- de la définition du modèle de données UML applicable aux métadonnées du SINP
- des définitions des différents éléments de métadonnées et leur cardinalité,
- des référentiels associés à certains éléments et liste des valeurs possibles

La description des métadonnées concerne le profil « Dispositifs de collecte », le profil « Données », les services relatifs aux données INSPIRE, et les acteurs.

Question principale : Les résultats du groupe de travail « Standard de métadonnées » sont-ils cohérents, scientifiquement et techniquement, par rapport aux objectifs du SINP ?

Le rapport constitue le résultat du groupe de travail. Il décrit le modèle de données UML applicable aux métadonnées du SINP, puis donne la définition des éléments des métadonnées constitutifs du modèle de données. Le travail est cohérent avec les objectifs du SINP. Quelques remarques cependant :

- il manque une définition préalable de ce qu'est une métadonnée par rapport à des données.
- le rapport ne met pas en évidence la cohérence, ou la compatibilité entre les éléments de métadonnées des profils décrits dans les chapitres 4.2. à 4.5. avec les dispositifs existants (Ch. 3) : quels recouvrements ? quels éléments nouveaux ?

Sous-question 1. La qualité scientifique du document final est-elle satisfaisante ? si non, quelles sont les améliorations à apporter ?

La qualité scientifique du document final est satisfaisante. Les métadonnées proposées sont compatibles avec la directive INSPIRE (Directive 2007/2/EC) et avec les standards ISO 19115 (données) et 19119 (services).

Remarques :

- p.10, le schéma est intitulé « Schéma général des concepts du SINP ». Cependant il ne comprend ni le portail national, ni un portail régional (schéma = un seul acteur). Compléter le schéma, ou présenter auparavant un premier schéma présentant Portail national, Portail régional, et BD et données des acteurs.
- Les Bases de données sont un des concepts du SINP, représentés dans le schéma général des concepts (p.10). Or par la suite on ne trouve pas les métadonnées sur les bases de données. Ou cela entre-t-il dans les métadonnées du profil "données" ?
- L'élément 4.3.1.4 Type de ressource cite les termes Dataset et Series de la norme ISO 19115. On gagnerait à donner directement la définition, sans avoir besoin de consulter la norme. Il serait bon de donner des exemples, par ex. ensembles de données tels que les collections (herbiers etc.) ?
- La mise en place d'identifiant unique des données nécessite des identifiants contrôlés pour les bases de données. Métadonnées et données doivent donc suivre un processus commun.
- Préciser « info obligatoire/facultative » dans les différents champs de métadonnées. Et les métadonnées sont elles obligatoires pour tous les concepts SINP ?

Sous-question 2. La qualité technique du document final est-elle satisfaisante ? si non, quelles sont les améliorations à apporter ?

La qualité technique du document final est bonne. Le document est bien évidemment un document technique, mais il serait bon d'améliorer sa lisibilité pour des non-spécialistes.

Suggestions :

1. Les informations des Chap 4 et 5, le cœur du document, gagneraient en synthèse et lisibilité si elles étaient fusionnées sous forme d'un tableau à plusieurs colonnes : Référence, Élément de métadonnées, Définition, Exemple, Cardinalité, Condition), (avec présentation du tableau en format paysage)
2. La présentation est parfois sèche (ex. p.10, le schéma général des concepts du SINP n'est pas légendé, ni accompagné de texte. Les concepts sont cependant présentés dans la section qui suit)
3. Quelques coquilles Ex. modèle de données UML applicable aux métadonnées du SINP (dans résumé)..
4. Probablement erreur de cardinalité : Dispositifs de collecte (Tableau 1, p.32), élément « groupe taxonomique (au singulier ?), espèces/habitats (au pluriel ?) » : a priori 0..n (au lieu de 0..1).

Sous-question 3. Quelles recommandations la CST peut-elle émettre en terme d'implémentation de ce modèle de métadonnées UML sur divers SI (outil naturaliste, Géosource, autres SI) ?

Recommandation n°1. Le modèle de métadonnées proposé n'a pas été testé à ce jour, et il est bien évidemment souhaitable de le faire rapidement. Une fois ce test réalisé, et les conclusions tirées, il est recommandé d'implémenter ce modèle dans divers dispositifs actuellement en fonction (Géosource, Geonetwork, MDweb..) ou en cours de développement (outil naturaliste).

Recommandation n°2. Au début du document il est indiqué l'objectif de définir « un cœur commun de métadonnées » complété par des « données complémentaires ». Il est suggéré que le schéma UML mette en évidence ces deux niveaux, ce qui permettrait lors des implémentations de connaître clairement l'obligatoire commun, et ce qui peut ou non dépendre d'un domaine précis. Les outils doivent en effet alléger au mieux la saisie souvent fastidieuse, et donc s'adapter aux domaines. Il est aussi recommandé de lancer une réflexion destinée à évaluer la nécessité de développer des extensions aux profils de métadonnées, extensions de métadonnées nécessaires à un usage ou une catégorie d'acteurs

Recommandation n°3. Il est essentiel que les définitions des concepts SINP soient bien les mêmes entre les GT, et donc indiquer les interactions entre GT architecture - GT formats de données - GT juridique etc. En ce qui concerne les "données" on trouve dans le document des métadonnées :

"données élémentaires"
"données un peu plus élaborées" !!
"données utiles" pour le SINP
"données de référence"
"données de synthèse ou données élaborées"

Le GT architecture parle lui de "données sources", "données élémentaires d'échange" et "données de synthèse".

Remarques de détail

- Le GBIF ne collecte pas les métadonnées sur les institutions françaises mais sur les sources de données dites primaires cad sur les collections (herbier, collection zoologique, collection vivante etc.) et les bases de données d'observation.

- il est suggéré que le titre du document devienne « Description des profils de métadonnées.. » (au lieu de « Description du profil de métadonnées.. »)

- certains titres de chapitre (en quelque sorte les métadonnées du chapitre...) gagneraient à être plus descriptifs ex. Catalogue en ligne des sources d'information sur la forêt Ca-SIF (au lieu de Ca-SIF seul) ou encore « Résultats du projet européen NatureSDI+ » au lieu de « NatureSDI+ »

- la liste des participants au GT présentée dans le résumé (p.2) est assez différente de la liste présentée dans le corps du document (p.5) : 10 membres dans un cas et 14 membres dans l'autre, et des participants de chacune des listes non présents dans l'autre.

- Le résumé annonce que le rapport traite du périmètre d'application des métadonnées. Il est difficile de trouver la section du document où ce sujet est traité

- Dans le §4 Profils de métadonnées, harmoniser les titres des paragraphes. Privilégier la formulation « Description des éléments de métadonnées... »

- Dans le §4 Profils de métadonnées, harmoniser les titres des sous-paragraphes 4.2.22, 4.3.9, 4.4.15. Privilégier la rédaction « Métadonnées de la fiche Métadonnées du dispositif de collecte / des données / du service

- Si la rédaction du rapport a bénéficié de contributions significatives, ces contributeurs pourraient être cités

- pour la mise en page, le rapport fait appel à certains styles. Mais cette mise en page peut être améliorée. Exemple : les titres de niveau 4 sont marqués Paragraphes solidaires, pour ne pas apparaître seuls en bas de page, mais certains apparaissent cependant seuls en bas de page car ils sont suivis d'une ligne creuse. Il suffit de supprimer la ligne creuse, et d'utiliser, pour le paragraphe sous le titre, le style Corps de texte, avec un espace avant le paragraphe.