



Automatisation des flux entre instances GeoNature

Cahier des clauses techniques particulières

Maître d'ouvrage : Parc National des Écrins

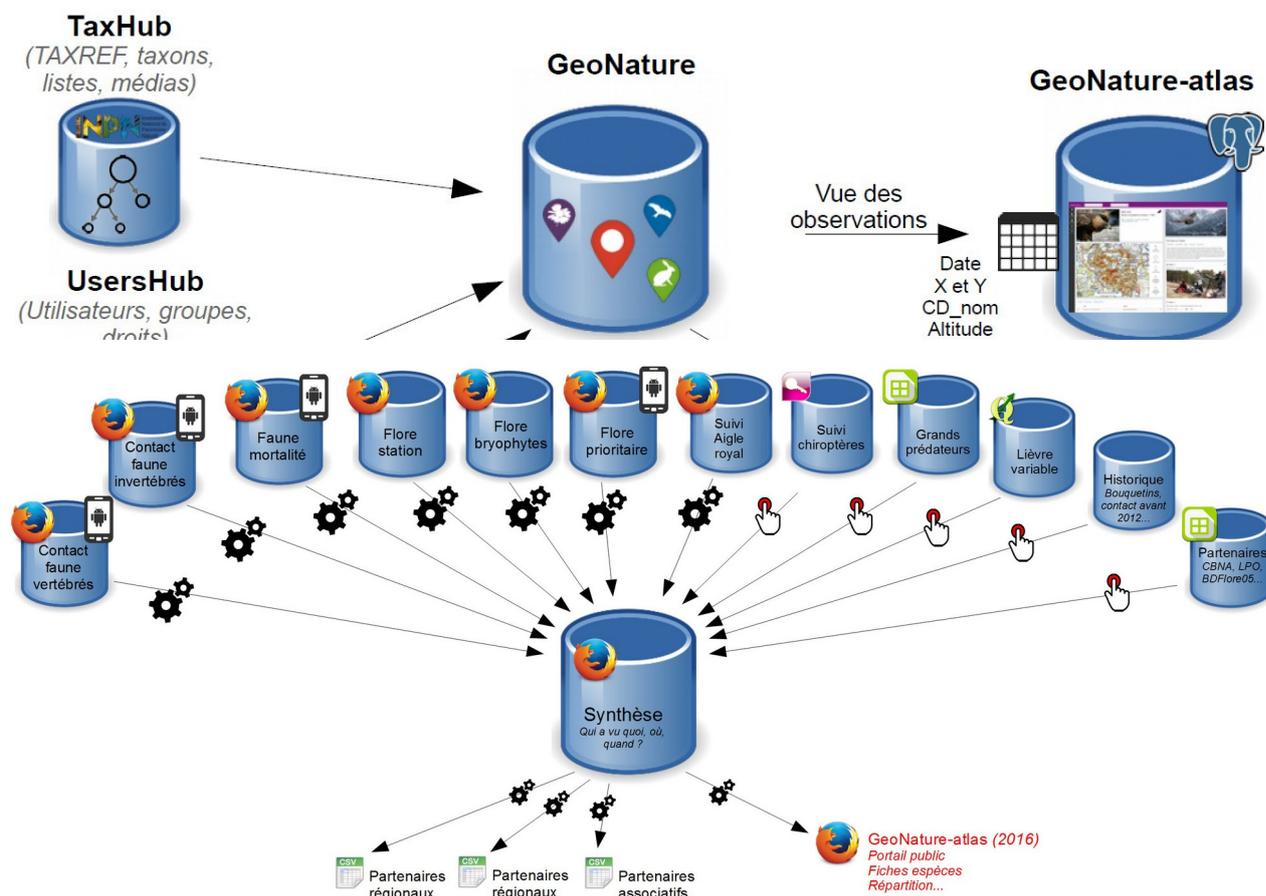
Sommaire

1. Préambule.....	3
1.1. GeoNature.....	3
1.2. gn_module_export + GN2PG pour une amélioration des échanges entre bases de données de biodiversité en lien avec les SINP.....	4
1.2.1. gn_module_export.....	4
1.2.2. GN2PG.....	4
2. Objet de la commande.....	4
2.1. Phase 1 : Amélioration des performances et des fonctionnalités du module d'export de GeoNature.....	4
2.2. Phase 2 : Amélioration de l'outil GN2PG.....	5
2.3. Phase 3 : Intégration de GN2PG en tant qu'outil satellite de l'écosystème GeoNature et développement d'une interface associée.....	5
3. Modalités de mise en œuvre.....	6
3.1. Pré-requis.....	6
3.2. Une solutions open source.....	6
3.3. Déroulement de la prestation.....	6
4. Livrables.....	7
5. Offre et éléments financiers.....	7
6. Critères de sélection des offres.....	7

1. Préambule

1.1. GeoNature

GeoNature est un ensemble d'applications WEB et mobile pour saisir, gérer, synthétiser et diffuser les données faune et flore multi-protocoles d'un territoire (<https://geonature.fr>). Elles sont publiées sous licence libre et sont maintenues par le collectif de développeurs du réseau des Parcs Nationaux.



GeoNature permet ainsi de déployer un système d'informations complet pour la gestion des données Faune/Flore d'une structure, allant de :

- la gestion des référentiels (taxonomiques et utilisateurs),
- à la saisie web et mobile dans différents protocoles,
- à la gestion de leurs métadonnées,
- à l'intégration de données de partenaires,
- à l'export des données selon les formats attendus par chaque partenaires
- à la synthétisation des données des différents protocoles sous forme de DEE
- à la diffusion des données sur un portail web grand public

GeoNature peut être testé sur <https://demo.geonature.fr> (admin / admin)

GeoNature (et ses dépendances) peut être installé de manière automatisée à l'aide de scripts de déploiement : <https://docs.geonature.fr>

GeoNature regroupe plus de 90 structures utilisatrices et chaque structure dispose de sa propre instance de GeoNature. Dans un contexte de partage de la donnée entre ces structures et de manière générale, la fluidification et l'automatisation des échanges de données est un enjeu majeur.

Aujourd'hui deux outils permettent une automatisation de ces échanges, mais ceux-ci doivent être améliorés pour être rendus simples d'utilisation.

1.2. gn_module_export + GN2PG pour une amélioration des échanges entre bases de données de biodiversité en lien avec les SINP

Les échanges de données sont un enjeu clé pour la connaissance de la biodiversité à des échelles spatiales et temporelles diverses. Avec la création des SINP, les données publiques doivent être remontées régulièrement aux organismes régionaux identifiés. Aujourd'hui, ces échanges de données sont le plus souvent effectués manuellement avec la chaîne d'actions suivante : export des données au format tableur ou dump de la base de données par la structure source, transmission via mail à la structure puits et import par la structure puits.

Ce processus est souvent long, chronophage et fait face à de nombreuses problématiques et notamment au fait que les données exportées par la ou les bases sources ne sont pas homogènes et ne correspondent pas au format de la base puits. Il est donc primordial d'automatiser les exports de données, le pont entre les bases de données et l'import de celles-ci. De plus, avec la structuration des SINP régionaux, ces exports seront de plus en plus réguliers et feront face à de nouvelles problématiques comme la quantité et le type de données échangées.

Dans le cadre de ce cahier des charges nous nous intéresserons à deux outils déjà existants qui permettent l'export de données de biodiversité d'une part et le pont entre une base de données GeoNature et une base PostgreSQL d'autre part : gn_export et GN2PG.

1.2.1. gn_module_export

Le module d'Export de GeoNature, développé par la société Natural Solutions puis repris par les parcs nationaux des Cévennes et des Écrins permet, à partir de vues dans la base de données GeoNature, de mettre à disposition des exports sous forme de fichiers ou d'API JSON.

Une description complète de l'outil peut être trouvée ici :

https://github.com/PnX-SI/gn_module_export

1.2.2. GN2PG

GN2PG est une API, développée par la LPO Auvergne-Rhône-Alpes, qui permet d'interroger une API JSON mise à disposition par le module d'export afin de récupérer les données et de les insérer dans une base de données PostgreSQL.

Dépôt de l'outil : <https://github.com/lpoaura/GN2PG>

Un document de vulgarisation du fonctionnement de l'outil peut être trouvé à l'adresse suivante :

<https://geonature.fr/documents/2021-10-GN2PG.pdf>.

Ces deux outils sont complémentaires et permettent de simplifier le travail d'intégration des données d'une base GeoNature, base de données très largement utilisée par les acteurs de la biodiversité français. Cependant, ces deux outils doivent être améliorés pour répondre aux exigences grandissantes des utilisateurs.

2. Objet de la commande

2.1. Phase 1 : Amélioration des performances et des fonctionnalités du module d'export de GeoNature

- Proposer un retour technique sur le fonctionnement actuel de l'export et des pistes d'amélioration (l'ouverture d'un ticket complet sur le dépôt est attendue)
- Ordonner le résultat des exports : https://github.com/PnX-SI/gn_module_export/issues/101
- Permettre la gestion de gros jeux de données (plusieurs millions de données) de manière performante : https://github.com/PnX-SI/gn_module_export/issues/95
Les SINP, gérant souvent plusieurs millions de données, doivent être équipés d'outils performants qui répondent aux besoins d'échanges de données et de métadonnées.
- Pouvoir supprimer un export : https://github.com/PnX-SI/gn_module_export/issues/93
- Pouvoir mettre à disposition des exports publics sans authentification

2.2. Phase 2 : Amélioration de l'outil GN2PG

L'outil GN2PG : <https://github.com/lpoaura/GN2PG> récupère au sein d'une base de données PostgreSQL cible, les données mises à disposition via un export d'un GeoNature source. Fonctionnel, le téléchargement des données s'effectue aisément. Plusieurs problématiques demandent encore à être traitées :

- Proposer un retour technique sur le fonctionnement actuel de GN2PG et des pistes d'amélioration (l'ouverture d'un ticket complet sur le dépôt est attendue)
- **Les performances de l'outil** sont encore une fois un enjeu de taille pour les SINP qui s'échangent plusieurs milliers/millions de données. L'outil doit être amélioré pour absorber des transferts importants.
- **Automatiser les actions de GN2PG.** Mise en place d'un CRON permettant de programmer de manière régulière les imports de données.
- **Les métadonnées** doivent être le cœur des échanges de données entre les acteurs. Aujourd'hui, l'outil récupère les données et les métadonnées telles que formatées dans le GeoNature source. Il est souhaité de pouvoir indiquer un CA de destination différent des CA de la source.

2.3. Phase 3 : Intégration de GN2PG en tant qu'outil satellite de l'écosystème GeoNature et développement d'une interface associée

Actuellement l'outil GN2PG est une application Python téléchargeable via l'installer PIP de Python. L'objectif est de le rendre générique et accessible aux gestionnaires des bases de données des acteurs de biodiversité. Ainsi l'objectif est de proposer :

- **Une installation packagée simple pour un utilisateur de GeoNature**
L'outil doit cependant pouvoir rester indépendant pour ceux qui l'utiliseraient sans disposer de GeoNature. Possibilité d'utiliser Flask-admin à étudier
- **Une interface de l'outil permettant la gestion et le suivi des flux de données avec un système d'authentification**
 - Système de configuration des connexions entre outils
 - Système de suivi des téléchargements/imports :
 - Système d'alerte

Cette interface pourra fonctionnellement ressembler à celle ci-présentée :

Tableaux de bords de vos échanges de données

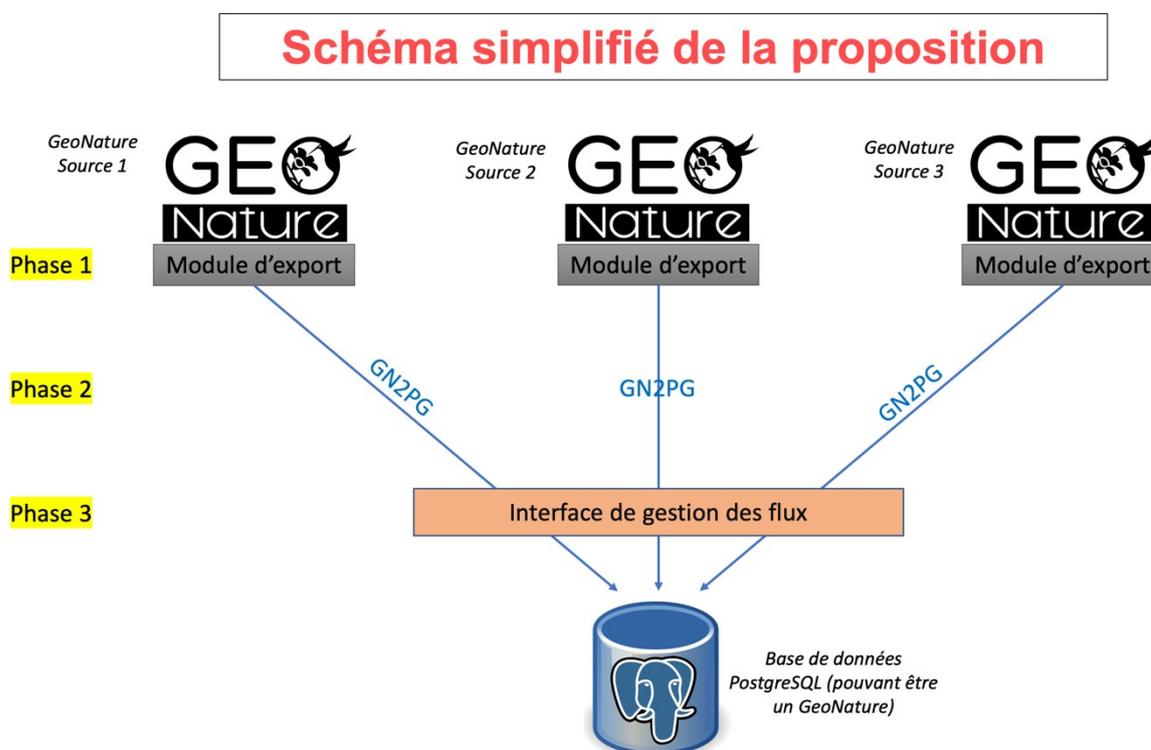


L'organisme sera associé à son identifiant du référentiel national des organismes. A terme l'interface pourra être liée à l'API mise à disposition par le SINP.

** À tout moment, il pourra être possible de lancer un moissonnage manuel des données d'une source déjà connectée. Par exemple, le flux de l'association 3, planifié de manière hebdomadaire pourra être lancé à n'importe quel moment par un utilisateur.

*** Faut-il une historisation des imports pour une traçabilité ? Stockage en JSON par source ?

Voici un schéma simplifié des grandes phases de notre proposition :



3. Modalités de mise en œuvre

3.1. Pré-requis

Le prestataire installera les applications sur un serveur de démonstration, accessible au maître d'ouvrage. Il devra respecter les prérequis techniques existants pour faire fonctionner les applications GeoNature.

L'ensemble des codes sources de GeoNature sont consultables sur les forges aux adresses indiquées plus haut. Ainsi que sa documentation sur <https://docs.geonature.fr>.

Le prestataire fournira une procédure d'installation de l'ensemble des nouveaux développements.

3.2. Une solutions open source

Les applications préexistantes ont été développées avec des outils libres et font l'objet d'une libération sur une forge. La présente commande s'inscrit dans la continuité et il est attendu du prestataire la proposition de solutions open source, réutilisables par d'autres opérateurs sous licences libres et qui ne généreront pas l'achat de licences d'utilisation récurrents.

Les développements doivent être conduits de manière générique et documentés de manière à pouvoir être utilisés par les différents utilisateurs de GeoNature. De la même manière, toutes améliorations des bibliothèques libres utilisées devront être publiées sur une forge publique, de préférence GitHub.

A l'issue du marché, les applicatifs développés devront être publiés sur une forge et être réutilisables par d'autres opérateurs sous licences libres.

3.3. Déroulement de la prestation

Le lancement de la prestation sera réalisé par une réunion de cadrage fixant les modalités d'échanges et de suivi ainsi que le planning opérationnel. Cette réunion sera, soit physique, soit par

conférence téléphonique associée à du partage d'écrans.

A l'issue de l'intégration des données de la base actuelle, une phase de test sera mise en œuvre pour vérifier le bon fonctionnement de l'application et son ergonomie.

A l'issue de la livraison, le prestataire proposera un délai de garantie pour répondre aux corrections éventuelles de dysfonctionnements survenant lors de la phase d'exploitation de l'application.

4. Livrables

Le prestataire fournira :

- Les développements de la branche améliorée du module d'export sur le dépôt Github PNX-SI/gn_module_export en vue d'une nouvelle version
- Les développements de la branche améliorée de GN2PG sur le dépôt Github lpoaura/GN2PG en vue d'une nouvelle version
- Les développements réalisés devront être accompagnés de tests automatisés sur les outils où ils sont déjà en place pour garantir leur bon fonctionnement et leur maintenance

5. Offre et éléments financiers

- Proposition méthodologique et technique précisant l'approche envisagée
- Moyens humains (composition de l'équipe et équipe dédiée au projet)
- Coût de la prestation

Cette proposition devra s'intégrer dans l'enveloppe maximale destinée au projet, soit 17500 euros TTC.

6. Critères de sélection des offres

Le Parc National des Écrins se réserve le droit de demander des compléments d'informations relatifs aux propositions des candidats.

L'offre de chaque candidat sera évaluée sur 20 :

- Proposition méthodologique et technique : 1 à 5
- Références techniques : 1 à 5
- Moyens humains (composition de l'équipe dédiée au projet) : 1 à 5
- Coût de la prestation : 1 à 5