



Feuille de route du GT Architecture

Historique du document

Version	Date	Commentaire
0.1	20/04/17	Version initiale
0.2	16/06/17	Révision
0.3	19/06/17	Révision
0.4	27/06/17	Révision suite à la réunion du GT
1.0	19/07/17	Validation de la feuille de route

Rédacteurs

Thomas Milon (UMS 2006 PatriNat)

Relecteurs

Frédéric Vest, Solène Robert (UMS 2006 PatriNat), Grégory CAZE (Représentant région Nouvelle Aquitaine), Florent CLET (DREAL Normandie), Laurent COUDERCY (AFB), Laetitia DE MONICAULT (MEEM)
Julien DEFENOUILLE (DREAL Normandie), Stéphane GUILLIN (DREAL Grand-Est), Rémy JOMIER (UMS PATRINAT), Valentin LE TELLIER (DEAL La Réunion), Dorothée MEYER (DREAL Provence-Alpes-Côte-d'Azur), Camille MONCHICOURT (PN des Écrins), Léa VERDIER (DREAL Grand-Est), Mathieu WILLMES (DREAL Centre).

Diffusion

Participants du groupe de travail (cercle 1, 2 et 3)

Objectifs du document

Ce document a pour objectif de cadrer et d'organiser les actions à mener sur les dossiers « Architecture »



Table des matières

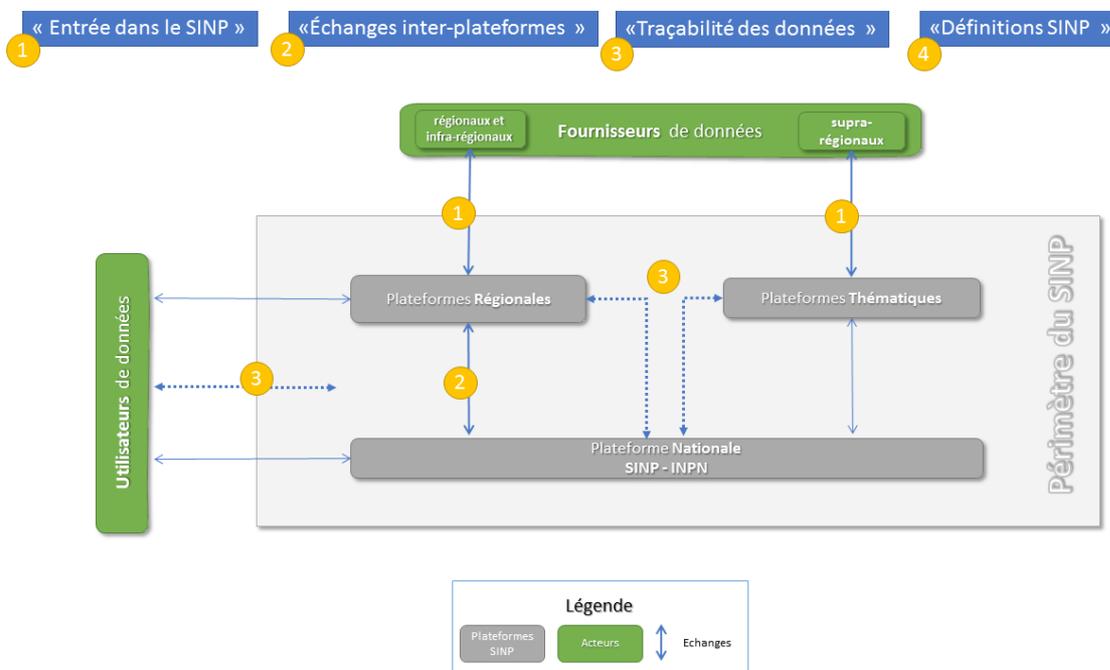
Introduction	3
1 Dossier « Entrée des données dans le SINP »	4
1.1 Problématiques.....	4
1.2 Actions.....	4
1.2.1 Définir les bonnes pratiques concernant l'entrée des données dans le SINP.....	4
1.2.2 Gérer le cas des « données brutes de biodiversité » (téléservice IPN - loi biodiversité).....	5
1.3 Éléments de priorisation.....	5
2 Dossier « Échanges inter-plateformes »	6
2.1 Problématiques.....	6
2.2 Actions.....	6
2.2.1 Analyser l'organisation des échanges inter-plateformes.....	6
2.2.2 Analyser le volet technique des échanges inter-plateformes.....	7
2.2.3 Proposer des bonnes pratiques d'échanges inter-plateformes.....	7
2.2.4 Améliorer l'intégration et l'accès aux données au niveau de la plateforme nationale.....	7
2.2.5 Accompagner les acteurs régionaux.....	7
2.3 Éléments de priorisation.....	7
3 Dossier « Traçabilité des données »	8
3.1 Problématiques.....	8
3.2 Actions.....	9
3.2.1 Consolider le système d'identification des données et des métadonnées (identifiant permanent / unique).....	9
3.2.2 Mettre en place un outil facilitant la traçabilité des données et métadonnées.....	9
3.2.3 Limiter la présence de doublon dans le SINP.....	10
3.3 Éléments de priorisation.....	10
4 Dossier « Définitions SINP »	11
4.1 Problématiques.....	11
4.2 Actions.....	11
4.2.1 Clarifier les termes structurant le SINP.....	11
4.3 Éléments de priorisation.....	11
5 Autres dossiers	11



Introduction

Suite à la mise en place de l'enquête sur l'architecture du SINP et suite à la réunion du GT Architecture du 30 mars 2017, les dossiers suivants ont été choisis pour être traités de manière prioritaire :

- « Entrée des données dans le SINP » ;
- « Échanges inter-plateformes » ;
- « Traçabilité des données » ;
- « Définitions SINP ».



Cette feuille de route décrit :

- les problématiques auxquelles répondre ;
- les actions à entreprendre ;
- les éléments pouvant conditionner la priorisation de ces actions.

Les produits de ces actions pourront être validés au besoin par le COPIL s'il implique des sujets qui ont un lien avec le pilotage.

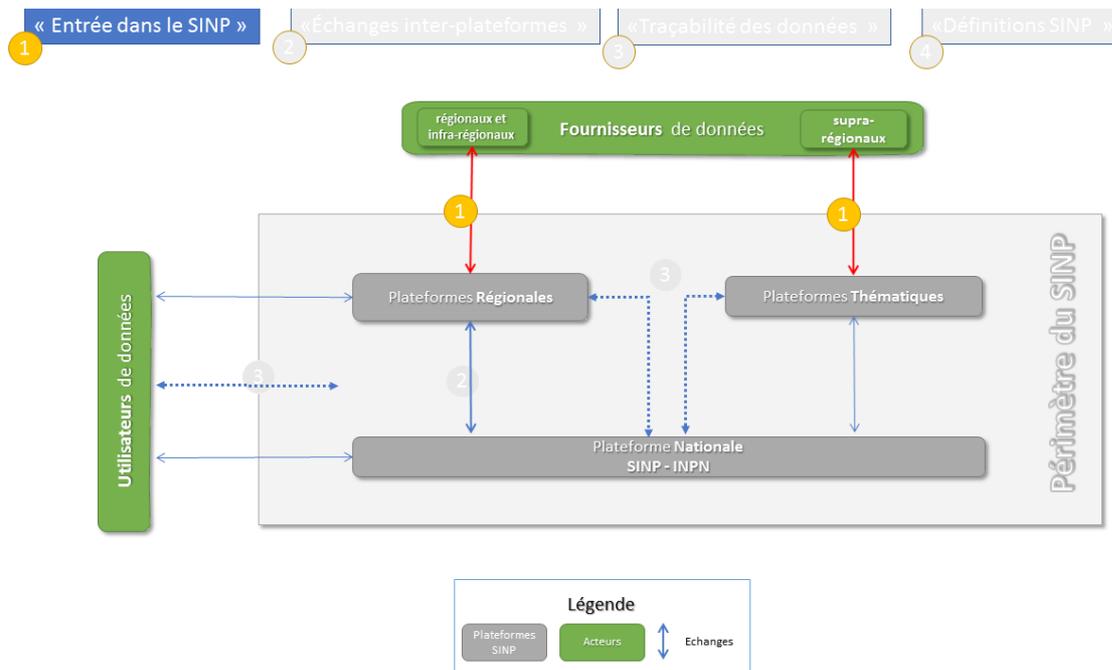
Pour rappel, le périmètre prioritaire de travail donné par le mandat au groupe de travail englobe :

- les occurrences de taxons ;
- les relevés phytosociologiques ;
- les occurrences d'habitats et de végétations ;

A ce jour, seule la thématique « occurrences de taxon » connaît une plateforme thématique.



1 Dossier « Entrée des données dans le SINP »



1.1 Problématiques

- Comment intégrer des données dans le SINP au niveau des plateformes R/T tout en :
 - préservant le principe de subsidiarité du SINP ;
 - assurant une gouvernance partagée, solide et simple ;
 - limitant les doublons dans le système ;
 - s'articulant avec les autres SI nationaux et internationaux (SIE, SIMM, GBIF...)?
- Comment gérer le cas des « données brutes de biodiversité » (téléservice IPN – loi biodiversité) ?

1.2 Actions

1.2.1 Définir les bonnes pratiques concernant l'entrée des données dans le SINP

- Note technique : « Analyse des pratiques d'intégration des données dans le SINP » ;
cible : Ptf et producteurs
 - Cartographie de l'organisation des entrées de données et métadonnées au niveau des plateformes régionales et thématiques (organismes dont pôles thématiques régionaux, jeux de données) ;
 - Identification et analyse de cas concernant les sources pouvant constituer des entrées doubles au niveau régional et thématique (ex : SERENA et RNF, ONF, ONCFS, AFB-ex ONEMA, Vigie Nature, bureaux d'étude, données de collection, Parcs Nationaux, autres SI nationaux) ;
 - Description de l'articulation avec la validation (conformité, cohérence, doublons, validation scientifique) ;



- Proposition de scénarios de gestion des cas généraux et des cas particuliers ;
- Note méthodologique : « Bonnes pratiques d'intégration de données dans le SINP – cas généraux » ;
cible : Plateformes et producteurs
- Préconisation sur les pratiques d'intégration en fonction du contexte – prise en compte de la question de l'auteur de la décision (« qui décide ? ») ;
- NB : la question de la gestion systématique des doublons (analyse lors de l'intégration) sera gérée dans le dossier « Traçabilité des données » ;
- Identification des moyens d'accompagnement pour favoriser la mise en place des bonnes pratiques ;
cible : Plateformes

1.2.2 Gérer le cas des « données brutes de biodiversité » (téléservice IPN - loi biodiversité)

- Bilans réguliers sur les avancés du projet Téléservice IPN – Loi Biodiversité (mail au GT) ;
cible : Cercle 1 du GT
- Note technique : « Scénarios d'entrée des données brutes de biodiversité dans le SINP » ;
 - Contexte lié au téléversement des données brutes de biodiversité, en particulier le devenir des données avant entrée dans le SINP ;
 - Proposition de scénarios d'entrée de ces données dans le SINP ;*cible : Plateformes*
- Note méthodologique : « Bonnes pratiques d'intégration de données dans le SINP – cas des données brutes de biodiversité ».
cible : Plateformes

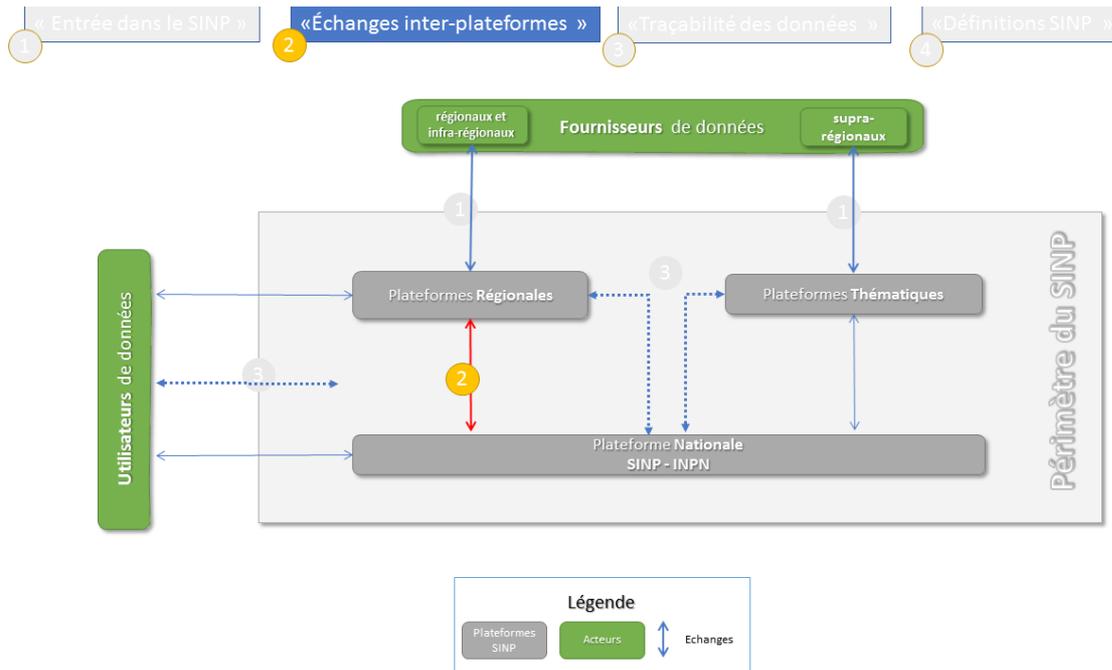
1.3 Éléments de priorisation

Les données brutes de biodiversité (Téléversement IPN - loi biodiversité) doivent être déposées à partir du 1^{er} janvier 2018. Ces données devront être gérées par le SINP. La description du contexte pour comprendre les enjeux d'intégration et impacts potentiels de ces données sur le SINP doit être gérée de manière prioritaire.

D'autre part, le dossier « traçabilité » aura un impact sur ce dossier (notamment la question de l'identification des données dans le SINP).



2 Dossier « Échanges inter-plateformes »



2.1 Problématiques

- Comment assurer une bonne visibilité et une bonne réactivité des échanges de données et métadonnées dans les 2 sens (plateforme régionale ↔ plateforme nationale) ?
- Comment limiter les problèmes techniques dans les échanges de données et métadonnées (respect du format standard d'un point de vue sémantique, format technique des données...) ?
- Quelles sont les informations nécessaires à échanger pour assurer une utilisation optimale de la donnée (question de la compatibilité des chartes régionales et le protocole national) ?
- Comment positionner les pôles thématiques régionaux dans les échanges de données et métadonnées ?
- Comment mettre à jour facilement et selon les besoins les données et métadonnées ?

2.2 Actions

2.2.1 Analyser l'organisation des échanges inter-plateformes

- Note technique : « Organisation des échanges de données et métadonnées entre les plateformes régionales et nationale – cartographie et préconisation » ;
cible : Plateformes
 - Analyse de l'organisation des échanges - cartographie des acteurs, des jeux de données, et des flux de données et métadonnées (schéma général) ;
 - Analyse des informations nécessaires à partager pour assurer une utilisation optimale des données (question de la compatibilité des chartes régionales avec le protocole) ;



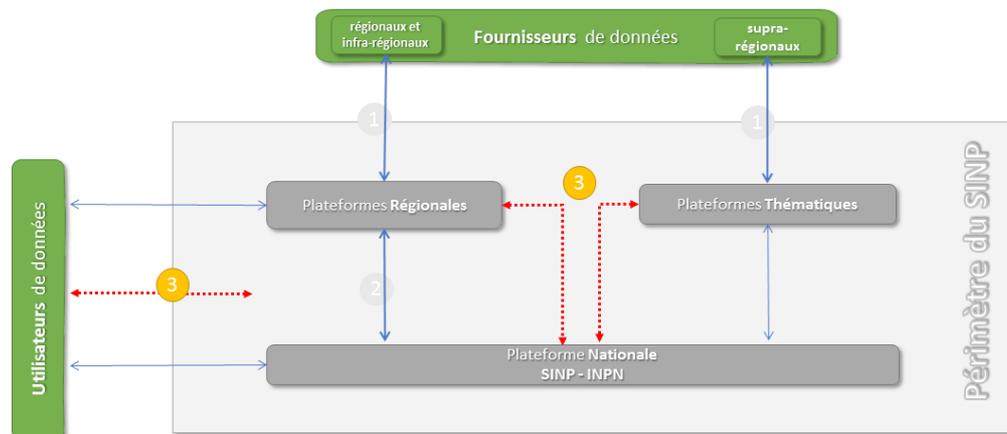
- Formalisation des grandes tendances sous forme de cas d'organisation ;
- 2.2.2 Analyser le volet technique des échanges inter-plateformes
- Note technique: « Analyse technique des échanges inter-plateformes » ;
cible : Plateformes
 - Cartographie des techniques, outils et formats utilisés ;
 - Formalisation des grandes tendances sous forme de cas d'utilisation ;
 - État de l'art des méthodes de partage de données, prenant en compte les questions d'ajout, mise à jour et suppression de données ;
- 2.2.3 Proposer des bonnes pratiques d'échanges inter-plateformes
- Note méthodologique : « Bonnes pratiques d'échanges de données et métadonnées inter-plateformes » ;
cible : Plateformes
 - Préconisations organisationnelles et techniques concernant les échanges pour ajouter, mettre à jour et supprimer des données du système (notamment sur les fréquences et/ou délais souhaitables d'échanges) ;
- 2.2.4 Améliorer l'intégration et l'accès aux données au niveau de la plateforme nationale
- Cahier des charges fonctionnel : « évolution fonctionnelle de la plateforme nationale pour faciliter le partage de données » ;
cible : Plateformes
 - Dépôt de lots de données et métadonnées ;
 - Service de vérification des doublons, de la conformité et cohérence des données et métadonnées (multi-format) ;
 - Traçabilité et notification des échanges ;
 - Mise à jour ponctuelle de données ;
 - Log des échanges ;
 - Cahier des charges techniques ;
cible : Plateforme nationale
 - Prototype à partir du cahier des charges ;
cible : Plateforme nationale
- 2.2.5 Accompagner les acteurs régionaux
- Identification des moyens d'accompagnement pour la mise en place des bonnes pratiques.
cible : Plateformes

2.3 Éléments de priorisation

Les notes techniques 2.2.2 et 2.2.3 alimenteront la cartographie générale du SINP. Elles sont donc à faire en priorité. Le reste des actions peuvent être réalisées dans un second temps.



3 Dossier « Traçabilité des données »



Traçabilité = suivi du cycle de vie de la données dans le SINP



3.1 Problématiques

- Où se trouve la donnée ? La donnée est-elle présente dans la plateforme régionale concernée ainsi que dans la plateforme nationale ?
- Quel traitement a-t-elle subi ?
 - Est-elle identique / à jour dans la totalité du système ?
 - Quel est son niveau de validation scientifique ?
- Comment limiter le nombre de doublons dans le système ?
- Quel suivi de l'utilisation des données ?



3.2 Actions

3.2.1 Consolider le système d'identification des données et des métadonnées (identifiant permanent / unique)

- Note technique : « Système d'identification des données et métadonnées du SINP - Analyse de l'existant et nouvelles propositions » ;

cible : Plateformes et producteurs

- Reprise des travaux sur les identifiants et analyse des problèmes rencontrés ;
 - Proposition d'évolutions du système d'identification ;
- Note technique: « Système d'identification des données et métadonnées du SINP – spécifications » ;

cible : Plateformes et producteurs

- Description des spécifications techniques du système d'identification ;
- Identifier les moyens d'accompagnement de l'implémentation de ce système ;

cible : Plateformes et producteurs

- au niveau de la plateforme nationale, et thématique ;
- au niveau des plateformes régionales ;

3.2.2 Mettre en place un outil facilitant la traçabilité des données et métadonnées

- Cahier des charges fonctionnel : « Système de traçabilité des données et métadonnées du SINP » ;

cible : Plateformes

- « Où se trouve la donnée » ;
 - Ajout / import des identifiants nouvellement ajoutés dans le système ;
 - Visualisation des identifiants présents dans les différentes plateformes / Interrogation du système pour vérifier l'existence d'un identifiant / Téléchargement d'identifiants existants ;
 - Mise à jour des informations relatives aux identifiants (date de mise à jour, plateforme, signature numérique) ;
 - Identification des données ayant été modifiées au niveau d'une plateforme (basée sur une signature numérique) ;
 - Création et export de nouveaux identifiants permanents par lot (avec vérification de l'unicité) ;
 - Lien URL vers la donnée / métadonnée complète sur la plateforme nationale (et vers la plateforme régionale quand cela est possible) ;
- « Quels traitements a-t-elle subis » - cas de la validation ;
 - Ajout de l'information concernant la validation (producteur, nationale ou régionale) ;
 - Visualisation / téléchargement de l'information concernant la validation ;



- Mise à jour de l'information concernant la validation ;
 - « Quels traitements a-t-elle subis » - cas de mises à jour ;
 - modification de cd_ref suite à évolution de Taxref ;
 - modification de cd_insee suite à évolutions des périmètres de communes ;
 - modification des cd_sensibilité suite à évolution des référentiels de sensibilité ;
 - « Quel suivi de l'utilisation des données ? » ;
 - Solution pour améliorer la visibilité de l'utilisation nationale et régionale ;
- Cahier des charges technique ;
cible : Plateforme nationale
- Prototype : Mise en place d'un prototype basé sur le cahier des charges – travail avec 2 régions pilotes, la plateforme nationale et thématique ;
cible : Plateforme nationale
- Analyse : Différentiels entre plateforme nationale et régionale (présence/absence de données, mise à jour de données) ;
cible : Plateformes (dont 2 régionales)
- Industrialisation du prototype ;
cible : Plateforme nationale

3.2.3 Limiter la présence de doublon dans le SINP

- Limiter les doublons à travers l'organisation des flux d'entrée (cf. dossier « Entrée des entrées dans le SINP ») ;
- Note technique : « Scénarios de gestion des doublons de données dans le SINP » ;
cible : Plateformes
 - Méthodes d'analyse et d'évaluation systématique des doublons au sein du SINP en lien avec la procédure de contrôle de conformité ;
 - Proposition de méthode de gestion des doublons ;
- Analyse : application des méthodes d'analyse et de gestion des doublons avec 2 régions pilotes ;
cible : Plateformes (dont 2 régionales)
- Note technique : « Procédure de gestion des doublons de données dans le SINP » ;
cible : Plateformes
 - Comment traiter les doublons ?

3.3 Éléments de priorisation

Le système d'identification des données aura un impact sur beaucoup de dossiers. Il est à gérer en priorité.



4 Dossier « Définitions SINP »

4.1 Problématiques

- Comment faciliter la communication tout en utilisant les termes parfois complexes du SINP ? Dans le SINP, un même terme peut être utilisé avec des niveaux de discours différents (vulgarisation vs communication technique vs note juridique et texte réglementaire) ou avoir des significations multiples en fonction du contexte (ex : données sources, données SINP, métadonnées, plateformes, plateformes thématiques, pôle thématique régional, jeu de données).

4.2 Actions

4.2.1 Clarifier les termes structurant le SINP

- Note technique : « Polysémie des concepts clés du SINP – Analyse et proposition de gestion » ;
cible : Plateformes et producteurs
 - Analyse des concepts clés basés sur les livrables de GT et le glossaire (définitions et contextes) ;
 - Priorisation des concepts à étudier ;
 - Proposition d'un système de gestion de la polysémie ;
- Prototype d'un système appliqué sur les termes prioritaires ;
cible : Plateformes
 - Note méthodologique : « Bonne pratique pour la création et modification de concept clé dans le SINP ».

4.3 Éléments de priorisation

Ce dossier n'est pas prioritaire. Cependant, il est proposé qu'il constitue un fil rouge tout au long du GT. Il pourra être ouvert, en fonction du besoin, pour répondre à un besoin spécifique.

5 Autres dossiers

Les dossiers suivants ont une priorité moins importante que ceux décrits précédemment.

- Habilitation des plateformes ;
 - critères d'habilitation ;
 - Procédure d'habilitation ;
- Animation / Accompagnement ;



- Modèle d'animation ;
- Outils d'animation ;
- Accompagnement ;
- Validation de données ;
 - retour vers les fournisseurs => procédure en cours, GINCO ;
 - Qui valide quoi et où ? Positionnement de la validation dans l'architecture (données non validées, données erronées invalidées...) ;
- Production de métadonnées ;
 - outil métadonnées proposé par le MNHN ;
- Implémentation de la sensibilité ;
 - mise en place des listes, implémentation régionale et nationale en cours ;
- Juridique ;
 - GT Juridique?