

MISE EN PLACE D'UNE BASE DE DONNÉES NATURALISTES

AU SERVICE ENVIRONNEMENT ET SAUVEGARDE SOL DU CNES-CSG

STAGE MASTER 2 GÉOMATIQUE

Soutenance du stage effectué entre le 15 février et 15 août 2023

Nathan FOLMER

Plan



1. Présentation du STAGE

Présentation du CNES au Centre Spatial Guyanais et ses missions
Objectifs du stage
Problématique et démarche méthodologique



2. Enjeux autour de la gestion des données naturalistes au CSG

Une production diversifiée
Enjeux régionaux avec le SINP
Conclusion: les attentes du service



3. Présentation du processus de gestion des données

Structures et contenus des jeux de données naturalistes
Standardisation des données
Compilation de l'information naturaliste dans un SGBD
Valorisation des données



4. Discussion des résultats et évolution du projet

Défis et enjeux du stage
Limites des outils employés
Maintenance et évolution du projet

Centre National d'Etudes Spatiales – Centre Spatial Guyanais



Missions du CNES:

- Coordonner les opérations de lancement
- Piloter la préparation des satellites
- Assurer le suivi du lanceur
- Concevoir les installations
- Protéger les biens, la population et l'environnement

Plan de Gestion de la Biodiversité du Centre Spatial Guyanais

2021-2030

Tome 1

Etat des lieux et diagnostic



1. PRÉSENTATION DU STAGE

Objectif Global: Maîtriser les activités industrielles pour protéger l'environnement, les biens et le personnel

Afin de réduire l'empreinte environnementale du CNES, un plan de gestion de la biodiversité a été mis en place afin de mieux **connaître, protéger et valoriser la biodiversité** du CSG.

Mise en place d'une base de données naturalistes pour **capitaliser l'ensemble des informations environnementales (faune/flore)** produite sur le CSG.

Tome 1 : Etat des lieux et diagnostic

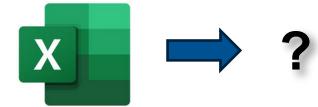
Tome 2 : enjeu, tableau de bord et plan d'actions

Tome 3 : fiches d'action

Création du sujet de mon stage

Mes objectifs initiaux de stage:

- Étudier les **potentialités des outils** existants pour gérer efficacement la base de données des espèces du CSG, en prenant en compte les **besoins** (localisation, description et statut de l'espèce observée) , les **habitudes** et les **contraintes** existantes.
- Convertir « **la base de données excel existante** » dans l'outil choisi en concertation avec mon tuteur de stage.
- Compléter la nouvelle base de données pour y intégrer les **observations plus récentes**.
- Rédiger un **manuel d'utilisation** ou une procédure expliquant comment mettre à jour la base de données.
- Proposer une **formation** (avec support à l'appui) pour l'utilisation et le **maintien** de la base de données.



Quels sont les enjeux autour de la conception d'une base de données naturalistes au service Environnement du CNES?



Grande diversité dans les projets:

- **Conservation et restauration:** Etudes et recensements de la faune et de la flore
- **Recherche scientifique:** Via des partenariats, les chercheurs collectent des données naturalistes (interactions entre espèces et leurs environnement, suivie de population...)
- **Projet d'Aménagement:** Evaluation environnementale d'un site avant la réalisation d'un projet d'urbanisation
-
-
-
-
-
-
-

Diversité des acteurs

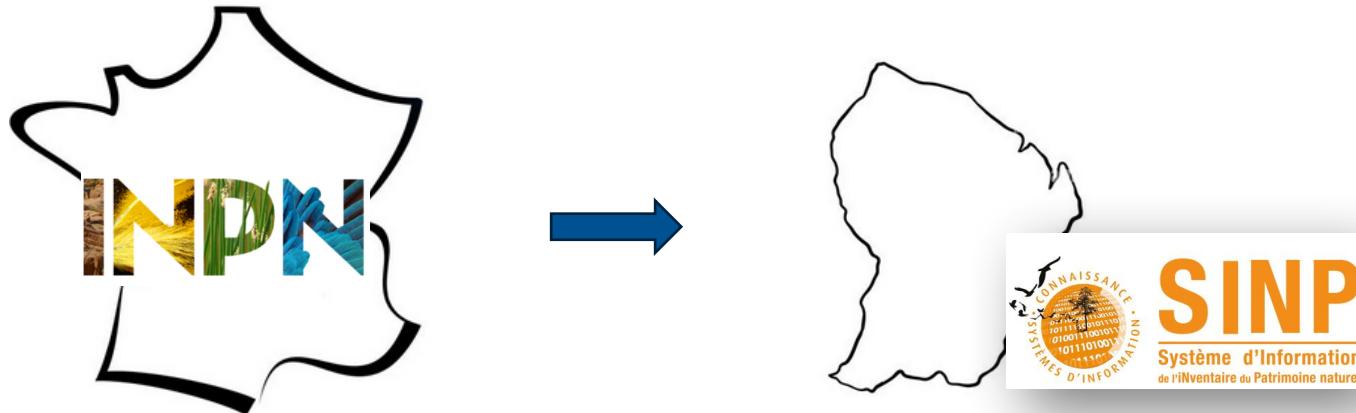
- **Biotope**: Bureau d'études
 - **AGE**: Bureau d'études
 - **GEPOG**: Groupe d'étude et de protection des oiseaux en Guyane.
Mène des projets de recherche visant à la protection des écosystèmes et biodiversité locale (abatage des niaoulis/accacias)
 - **OFB** : Office Français de la Biodiversité. Mène différents projets de recherche sur le CSG (Hocco, Tapir, Biches des Palétuviers...)
 - **ONF**: Office National des Forêts: Gestion du patrimoine naturel du CSG
 - Autres: **Cerato** , Association herpétologique

Diversité des méthodes, protocoles, outils...

- Exemples d'outils de collecte: les applications mobiles
 - Pour les observations faunistiques:
Inaturalist/ Faune Guyane /CardObs/Shuriken
 - Pour les observations floristiques: /Inaturalist/CardObs/
Shuriken



Enjeux nationaux et régionaux: le SINP de Guyane, quelques précisions:



L'objectif est de fournir des éléments précis sur les écosystèmes afin de prendre des décisions éclairées en matière de **conservation** et de **gestion de la biodiversité**.

Cette infrastructure de données (SINP) a pour objectif de collecter, gérer et diffuser les informations sur la biodiversité des régions.

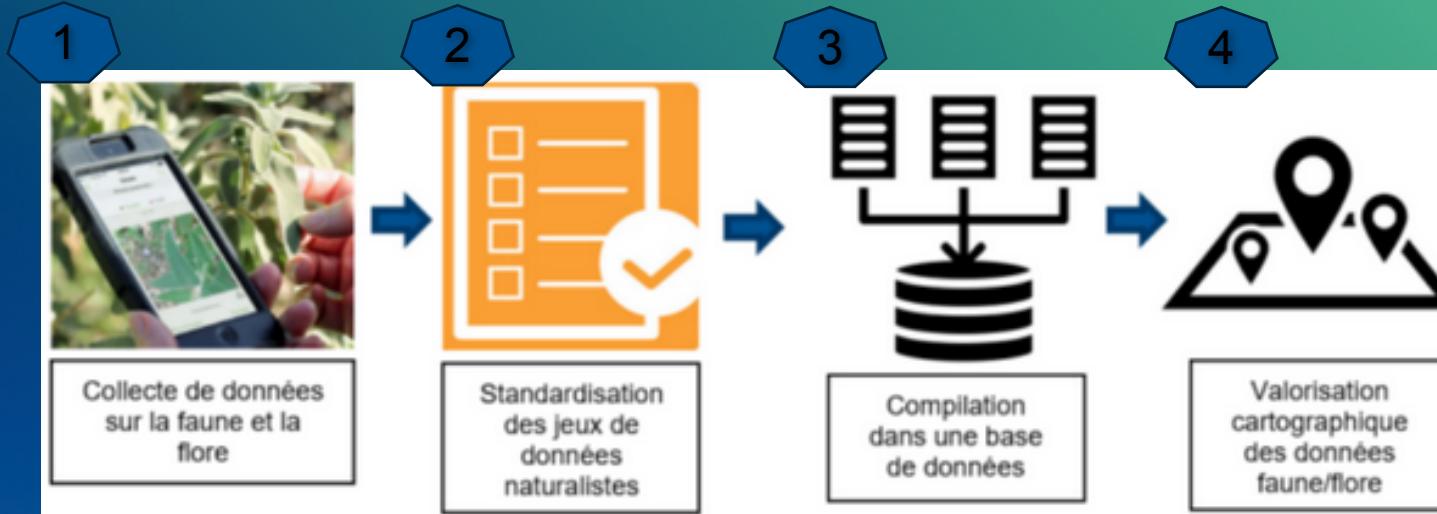
- **Améliorer les connaissances:** état de la biodiversité et les menaces
 - Prendre des décisions éclairées: **conservation/gestion des espaces naturels/ aménagement du territoire**
 - **Partage et coopération:** transfert des données d'observations naturalistes entre différents acteurs de la biodiversité (chercheurs, gestionnaires d'espaces naturels, décideurs)
- nécessite le **respect de certains standards** pour la transmission des données

Quels sont les besoins et attentes du service?

Développer un outil qui:

- capitalise l'ensemble des observations naturalistes produites ces 20 dernières années (données d'archives)
- intègre les observations depuis 2020
- fournit une information claire et structuré intégrant des informations sur la localisation, et les statuts des espèces observées
- permet une valorisation des données géographiques sous QGIS
- est facile à prendre en main par le personnel avec l'aide d'une documentation et/ou d'une formation.
- répond aux exigences du SINP de Guyane pour la transmission de l'information naturaliste du CSG
- Si possible gratuit

Démarche méthodologique



- 1 L'étape de collecte des données est gérée directement par les partenaires et prestataires
- 2 Structures et contenus des jeux de données multi-sources → vers un besoin de **standardisation**
- 3 Définir un SGBD capable de capitaliser l'information environnementale
- 4 Valoriser les données naturalistes de la base de données

Des jeux de données hétérogènes: comment définir des normes de standardisation ?

Analyse multi-critères des **jeux de données de référence** par rapport aux **standards** du SINP et les **attentes** du service.

Catégorie	Sujet de l'observation (Quoi?)					
	Standard régional de données du SINP	nomCite	nomvalide ^(R)	cdNom	versionTAX REF	CommentSujet ^(R)
AGE - Faune guyane - data Faune	x	x				
CERATO - GPS - data Faune		x				
OFB - GPS - data Faune	x	x				
Gepog - Faune guyane - data Faune	x	x				
Biotope - Shuriken - data Flore		x				
AGE - Inaturalist - data Flore	x	x				
AGE - CardObs - data Flore	x	x				
Attentes du service Environnement		x				

Comparaison en fonction:

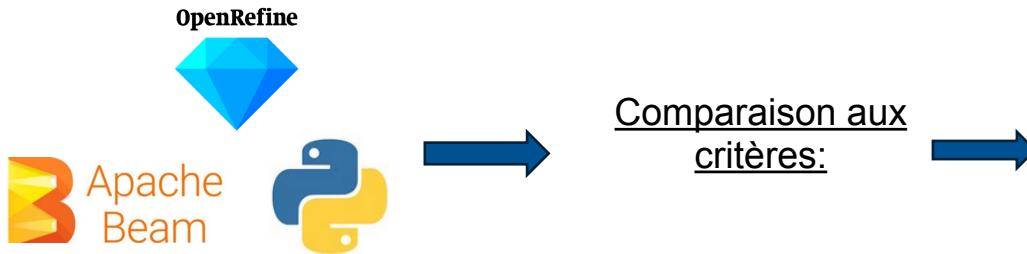
- Sujet de l'observation
- Le nombre d'entités composant l'observation
- La description fine de l'observation
- Période d'observation
- Observateur
- Cadre acquisition



Définition de **30 champs** qui forment les standards à respecter pour homogénéiser les jeux de données multi-sources



OpenRefine: l'outil de standardisation



- Adaptation aux contraintes de l'environnement numérique
- Coût
- Facilité d'utilisation
- Scalabilité
- Automatisation
- Formats pris en compte
- Fonctionnalités de nettoyage

Gratuit, facile à prendre en main

Ne demande pas de compétences poussées en programmation

Accepte les formats des jeux données transmis par les prestataires du CNES

Automatisation de la restructuration des jeux de données similaires via des instructions JSON

Mal adapté pour les grands jeux de données



PostgreSQL, le SGBD adapté pour l'organisation et la structuration des données naturalistes

3

Définir l'outil de gestion des données :



selon les critères:

- Fonctionnalités
 - Type de données stockées
 - Compatibilité avec QGIS
 - Conformité au SINP
- Coûts
 - Licence
 - Mise en production, exploitation, maintenance, formation...
- Facilité d'utilisation
 - Interrogation des données
 - Efficacité/stabilité
- Sécurité
 - Insertion dans l'environnement numérique du CNES
 - Gestion des droits utilisateurs
- Evolutivité
 - Local ou Serveur?
 - Répondre au besoin d'amélioration de la gestion des couches géographiques au service environnement

Extension Postgis

Gratuit, performant, stable

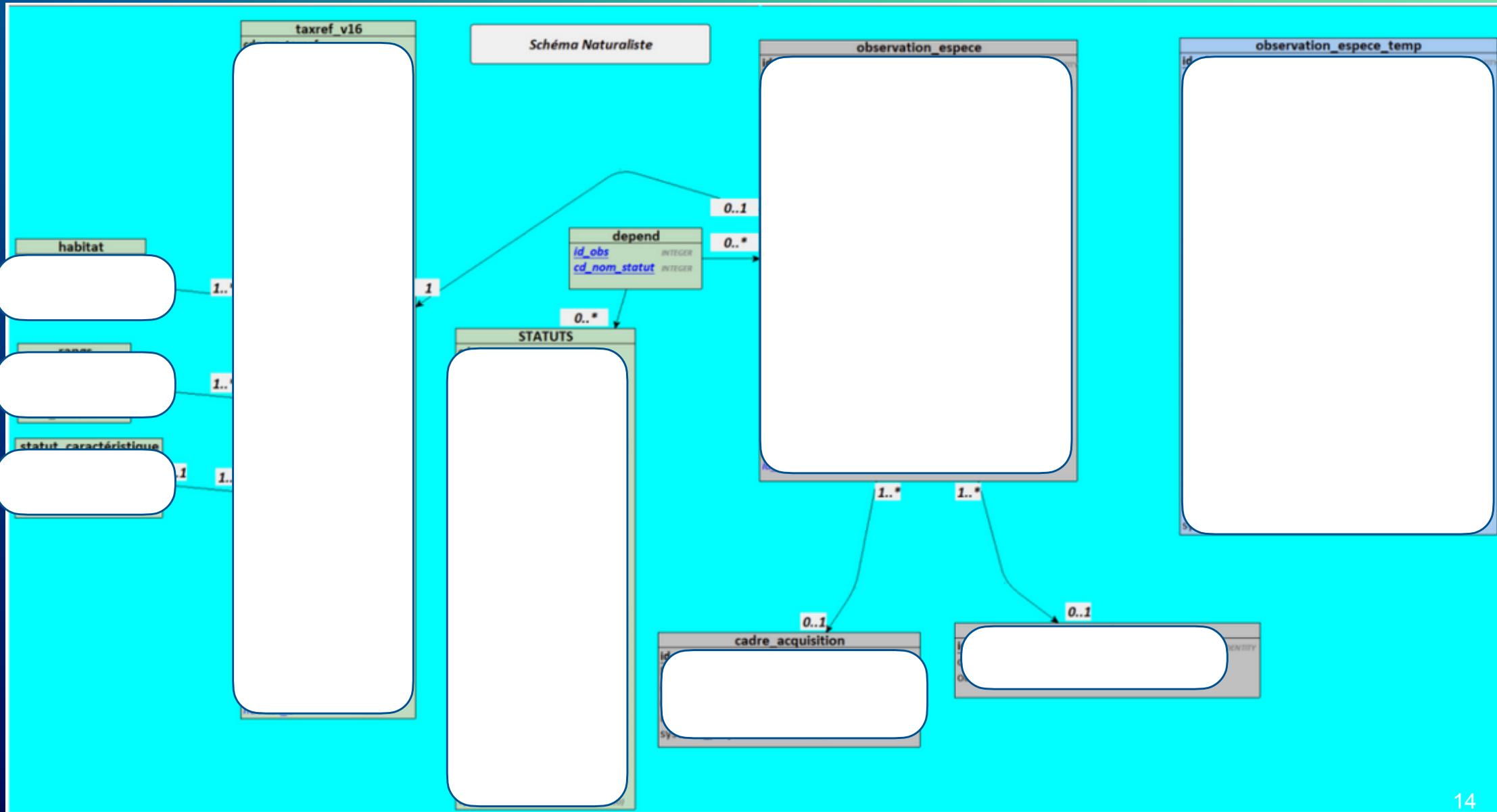
Modulable, sécurisé

PostgreSQL est le meilleur outil car :

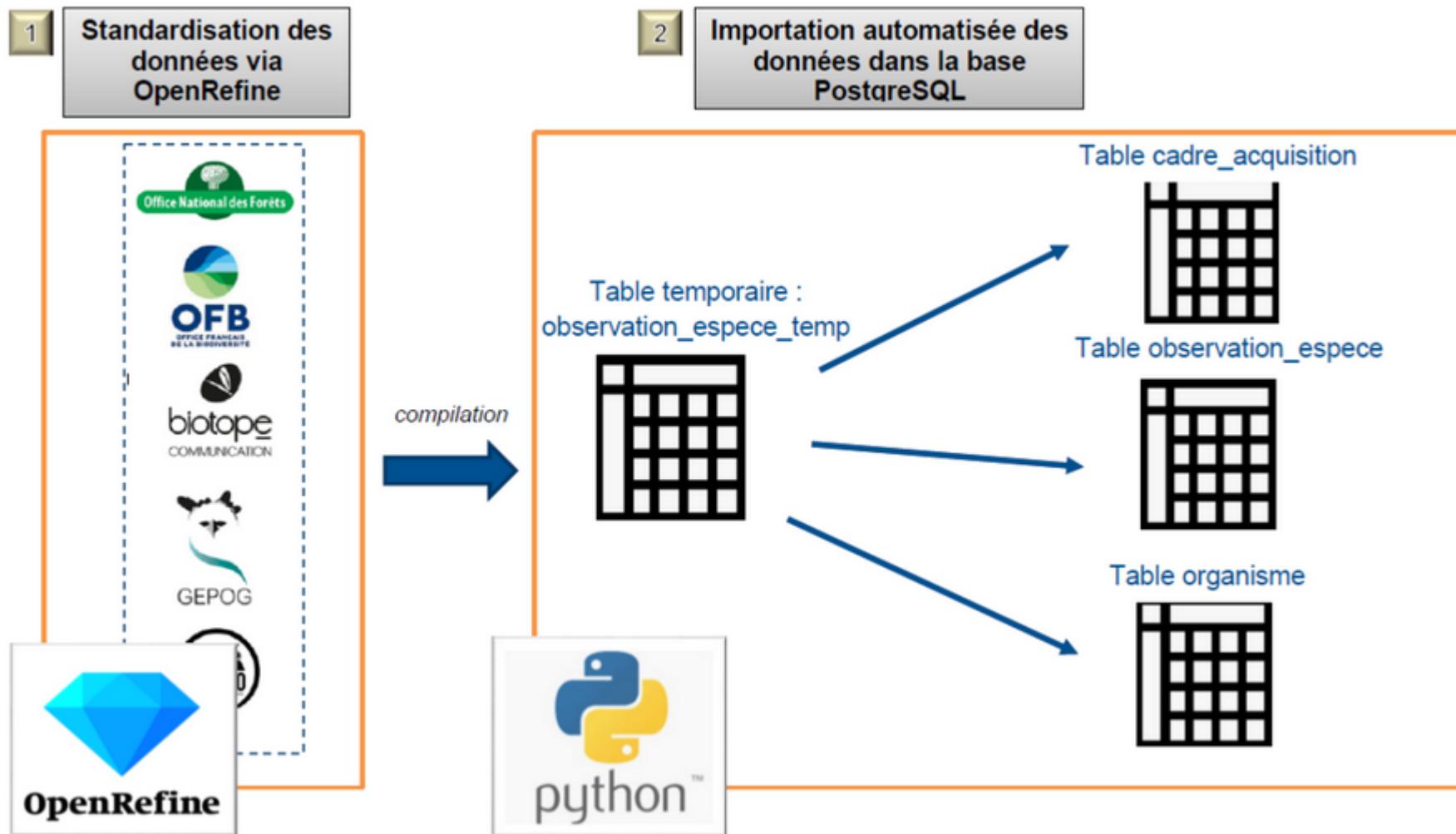
- Compétences en administration des données
- Requiert un poste linux déconnecté du réseau



Modèle logique de la base de données



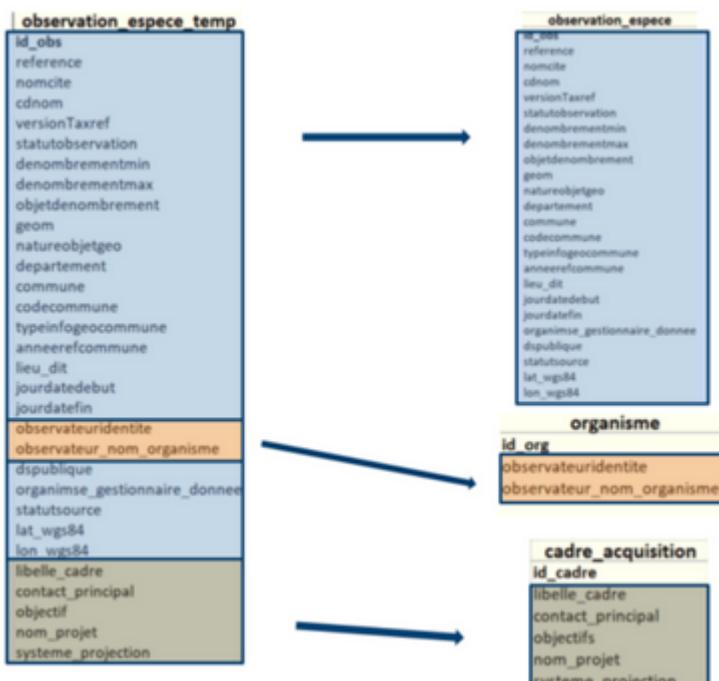
Automatisation de l'importation des données dans la base



Logique du script python

1) Création de la fonction « parcours »: qui vient récupérer dans un dossier donné, tous les jeux de données standardisées.

Puis grâce à la fonction « insert_data », ces derniers sont importés dans la table temporaire « observation_espece_temp »



2) Mise à jour de certains champs:

```

# création de la géométrie pour les coordonnées WGS 84 dans la table observation_espece_temp
update_geom_query_wgs84 = """
    UPDATE observation_espece_temp
    SET geometry = ST_Transform(ST_SetSRID(ST_MakePoint(lon_wgs84, lat_wgs84), 4326), 2972)
    WHERE systeme_projection LIKE 'WGS%'
"""
cur.execute(update_geom_query_wgs84)
conn.commit()
  
```

Alimentation des tables du schéma:

```

# execution des requêtes sur les différentes tables
def execute_query (conn) :
    CUR = conn.cursor()
    # Importer les données dans la table organisme
    requete_organisme = """
        INSERT INTO organisme (observateuridentite, observateur_nom_organisme)
        SELECT DISTINCT observateuridentite, observateur_nom_organisme
        FROM observation_espece_temp
        WHERE NOT EXISTS (
            SELECT 1
            FROM organisme
            WHERE organisme.observateuridentite = observation_espece_temp.observateuridentite
            AND organisme.observateur_nom_organisme = observation_espece_temp.observateur_nom_organisme
        );
    """
    cur.execute(requete_organisme)
    conn.commit()
  
```

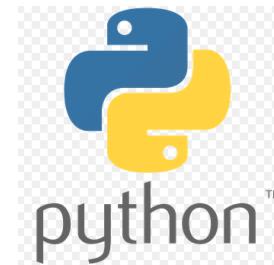
Récapitatif: Méthode pour l'implémentation de données dans la base

Option Retenue:

OpenRefine



Standardisation
des jeux de
données



Récupération des jeux de
données standardisés,
automatisation de l'envoie
des données dans la base



Compilation de l'ensemble
des données faune/flore.
Visualisation des données
possible dans QGIS via
l'extension PostGis

Valorisation des données naturalistes de la base de données PostgreSQL

Via QGIS: → répondre aux besoins internes de consultation des données

- Création d'un second schéma dans ma base
 - Création de vues SQL:

```
CREATE OR REPLACE VIEW public.v_test
AS
SELECT t1.nomcite,
t3.regne,
t3.classe,
t2.lb_type_statut_1,
t2.label_statut,
t2.code_statut,
t1.lieu_dit,
t4.nom_projet,
t5.observateuridentite,
t5.observateur_nom_organisme,
t1.geom
FROM observation_espece t1
JOIN statuts t2 ON t1.cdnom_taxref = t2.cd_nom
JOIN taxref_v16 t3 ON t1.cdnom_taxref = t3.cdnom_taxref
JOIN cadre_acquisition t4 ON t1.id_cadre = t4.id_cadre
JOIN organisme t5 ON t1.id_org = t5.id_org;
```

Via Ginco: → répondre aux enjeux régionaux en transmettant l'information naturaliste au SINP



Play ✓ Validati

- • • • • • • •

→ Possibilité de téléverser les données grâce aux respects des standards



Les défis et enjeux du stage:

- Comment les données naturalistes sont produites sur le CSG ?
 - Comprendre l'environnement numérique du CNES
 - Comprendre le rôle des SIG, des utilisations cartographiques au service environnement ?
 - → quelles compétences en informatiques et en géomatiques ?
 - → comment assurer la pérennité de mon travail ?

Limites des outils:

- Une standardisation des jeux de données qui reste complexe:
 - Des données très hétérogènes dans leurs structures et dans leurs contenus liés à la diversité des acteurs, des projets et des outils utilisés.
 - Besoin de définir un standard dans la structure des jeux de données par producteur.
 - Jointure attributaire qui possède des limites car le champ de jointure est de type texte.

Maintenance et Formation

Objectif: garantir la pérennité de mon travail:

- **Procédure:** document qui détaille le cycle de vie de la donnée au service:
 - Homogénéisation des jeux de données multi-sources
 - Script python qui envoie les données dans la base PostgreSQL
 - Possibilités de valorisation sous QGIS
 - Envoie des données sur la plateforme Ginco
- **Formation et tutoriel:** Mise en pratique de certains processus (standardisation et transfert des données dans la base PostgreSQL)
 - Réalisation d'un exercice par le personnel

Quelles évolutions ?

- Objectif: **Déployer la base de données postgresql sur serveur** (dépasser les contraintes des systèmes de sécurité de l'ASSI)
- - Un projet qui peut être utile au-delà du service environnement
-
-
- - Organisation de réunions: pour **consultation** les besoins d'autres services du CNES (Patrimoine, Gestion de crise)
- - étudier les **synergies possibles** autour de la mise en place d'une base de données: ambition **d'améliorer la gestion des projets SIG et des informations géographiques...**
-
-
-
- Vers des solutions Web SIG pour la consultation de données ?
- Enjeux de sécurité en temps réel sous forme de cartes et dashboard?
-

Conclusion

Un projet de stage qui me permet de découvrir l'ensemble des étapes dans la conception d'une base de données.

Mise en valeur d'enjeux:

- Besoins internes, contexte information que l'ASSI
- Homogénéisation, réglementations régionales
- Enjeux autour de la protection de l'environnement

