



Muséum  
national  
d'Histoire  
naturelle



# Standard Occurrence de taxon

Consultation intermédiaire en prévision de  
la saisine à la CST

*J. Chataigner (MNHN/SPN), Chef de projet Standardisation des données SINP*

*L. Poncet (MNHN/SPN), Coordination SINP/INPN au MNHN*



# Plan

- Historique et contexte
- Planning et méthodologie
- Présentation du standard
- Perspectives
- Questions



# Contexte et historique



# Contexte

- **Protocole SINP - Données élémentaires d'échange**

- « Ce sont des données standardisées inter-opérables. Elles sont élaborées à partir des données-source selon un format standard national propre à chaque thématique du SINP »
- « Le format standard des DEE comprend des informations obligatoires correspondant à des utilisations nationales strictement listées à l'article 10.3.6 du présent protocole et des informations facultatives »
- « Elles sont élaborées soit par les producteurs, soit par les plate-formes régionales ou thématiques et sont identifiées et qualifiées par les plate-formes régionales ou thématiques. »
- « Les DEE de tiers peuvent être géographiquement floutées c'est à dire :
  - pour les données terrestres, rattachées obligatoirement à une ou plusieurs commune(s) maille(s) terrestre(s) et selon les cas, une ou plusieurs masse(s) d'eau, zonage(s) de protection ou zones naturelles d'intérêt écologique, faunistique ou floristique (ZNIEFF);
  - pour les données marines, obligatoirement rattachées à une ou plusieurs maille(s) marine(s) et selon le cas à des zonage(s) de protection ou ZNIEFF. »
- « Les données élémentaires d'échange étant des données publiques... »

=> Les données « sources » ne sont pas dans le périmètre du SINP : uniquement Données Élémentaires d'Echange DEE



# Historique

- Saisine initiale CST n°7 : réunion plénière du **10 nov. 2011**
- Prise en compte des remarques / demande de modification
- Evolution du pilote du GT (FCBN vers MNHN)  
Création d'un groupe « pilote » au sein du MNHN :  
SPN/DSI(collection)/GBIF France/Recherche (taxonomie) – **dec. 2011**

- Nouveau mandat - **avril 2012**

Cercle 1 : =>

MNHN(SPN\*,collection, recherche, GBIF France), MEDDE,  
AAMP, ATEN, FCBN, IFREMER, IGN ,ONCFS, ONEMA,  
Niveau régional : Haute-Normandie et Franche-Comté

- Recrutement chef de projet : **aout 2012**

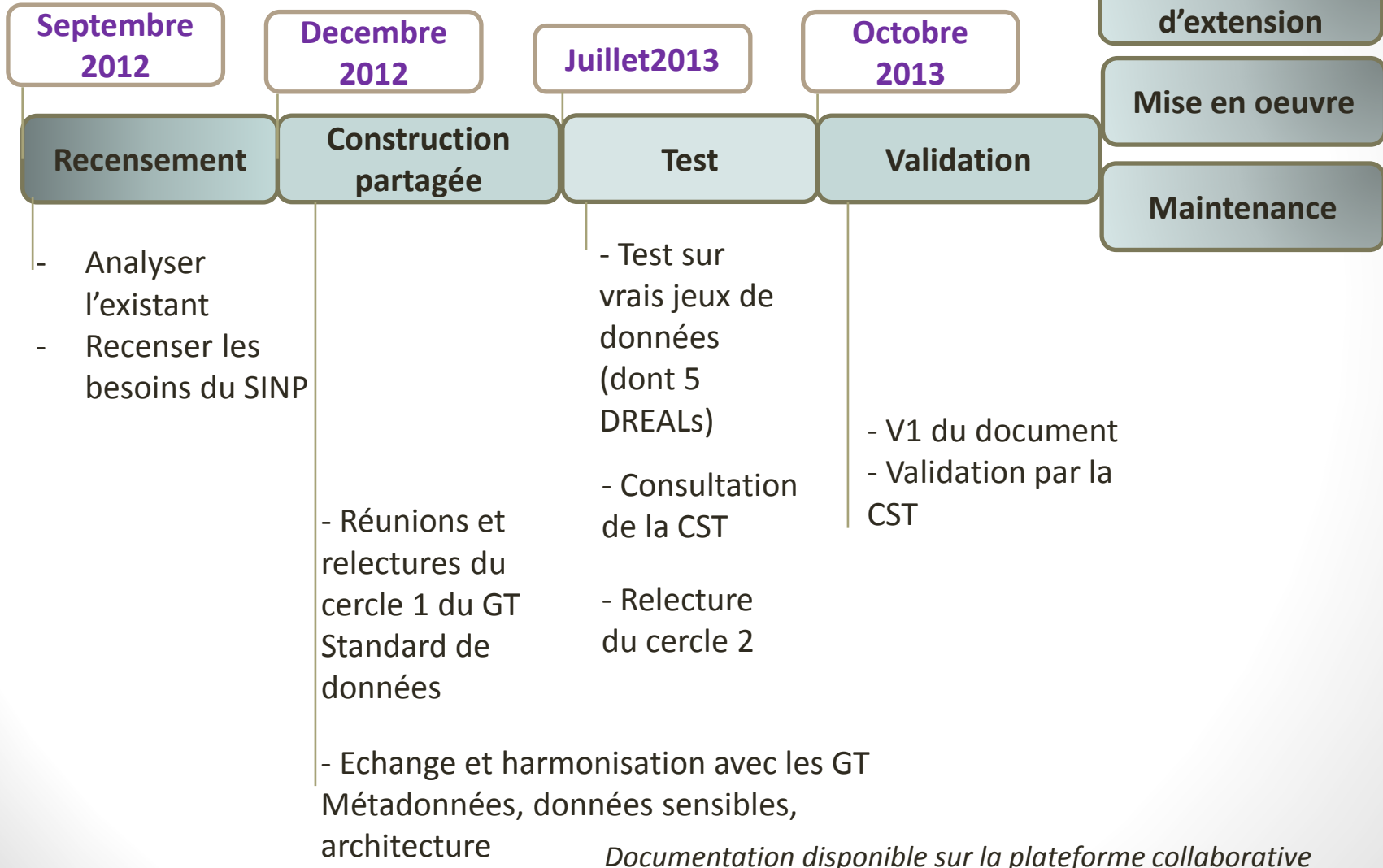


# Planning et méthodologie

# Planning et méthodologie

## La construction du cœur du standard

## Perspectives





## 1.2 Organisation : Livrables

- **Phase 1 & 2 : Recensement de l'existant et recensement des besoins**
  - Analyse de 14 standards Internationaux (darwincore, Obis, Manis...), nationaux (UK, Australie, Suisse...), régionaux (HauteNormandie, PACA, ... )
  - Les utilisations nationales du SINP
  - Retours des membres de cercle 1 et standards français**Livable** : Fiches de synthèse
- **Phase 3 : rédaction d'un document de cadrage méthodologique du format du standard**
  - Elaboration d'une ébauche de cœur de standard basé sur les phases 1 et 2 qui sera ensuite étendu via des extensions thématiques
  - Enrichissement par les réflexions et relecture du cercle 1 : 5 réunions de décembre 2012 à octobre 2013.
  - Consultation intermédiaire de la CST en aout-septembre 2013
  - Enrichissement par les relectures du cercle 2 : juillet- septembre 2013
  - Test du standard Cercle 1 &2 : juillet-septembre 2013 → Modification du standard
  - **Livable** : Dictionnaire du standard
- **Phase 4 : Définition des spécifications techniques**
  - Choix du format technique du standard (en cours)**Livable** : Description du standard technique
- **Phase 5 évaluation et veille**
  - Soumission du cœur du standards à la CST du **29 octobre 2013**.
  - Définition des extensions et de l'outillage du standard**Livable** : Planification 2014



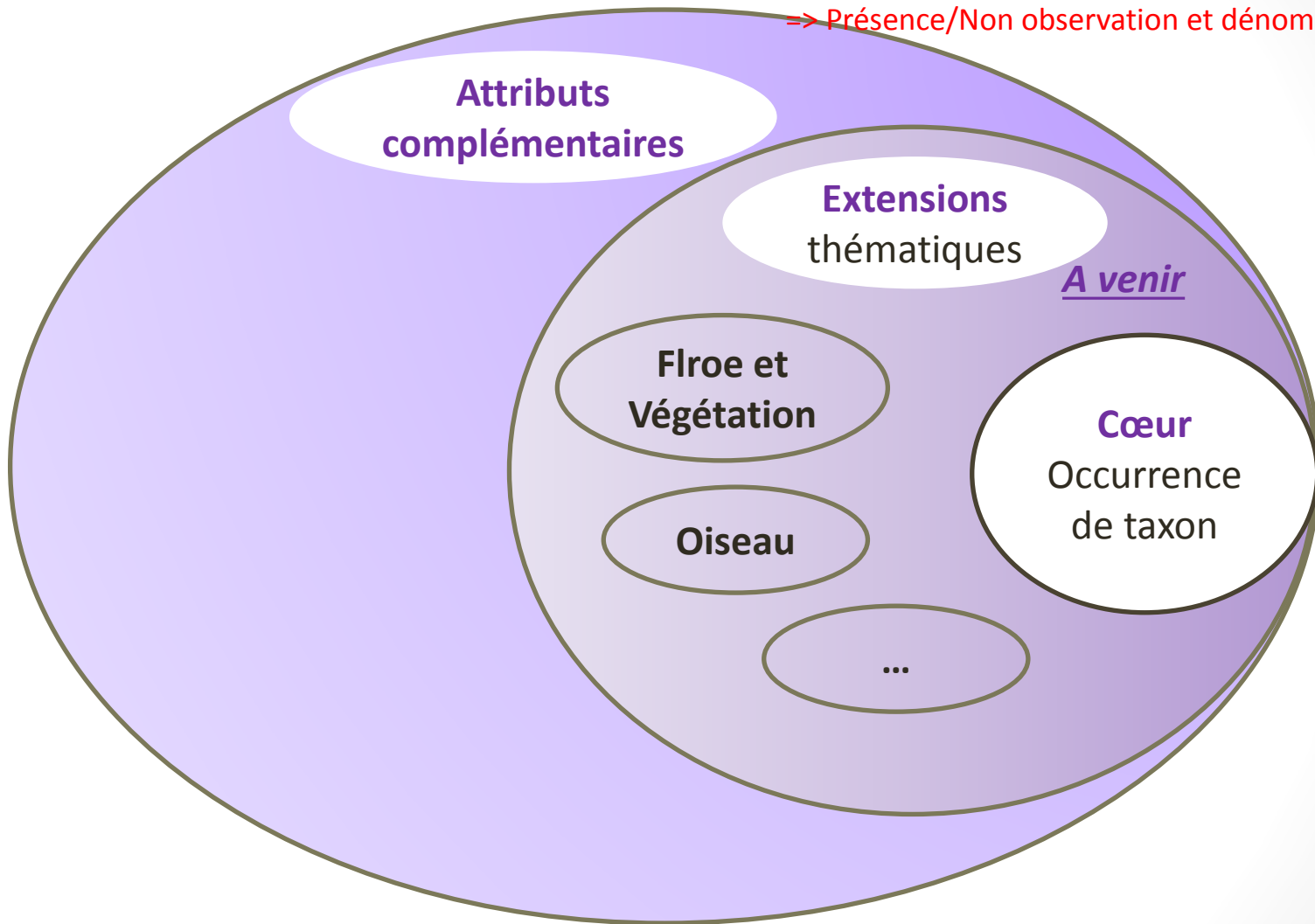


# Présentation du standard



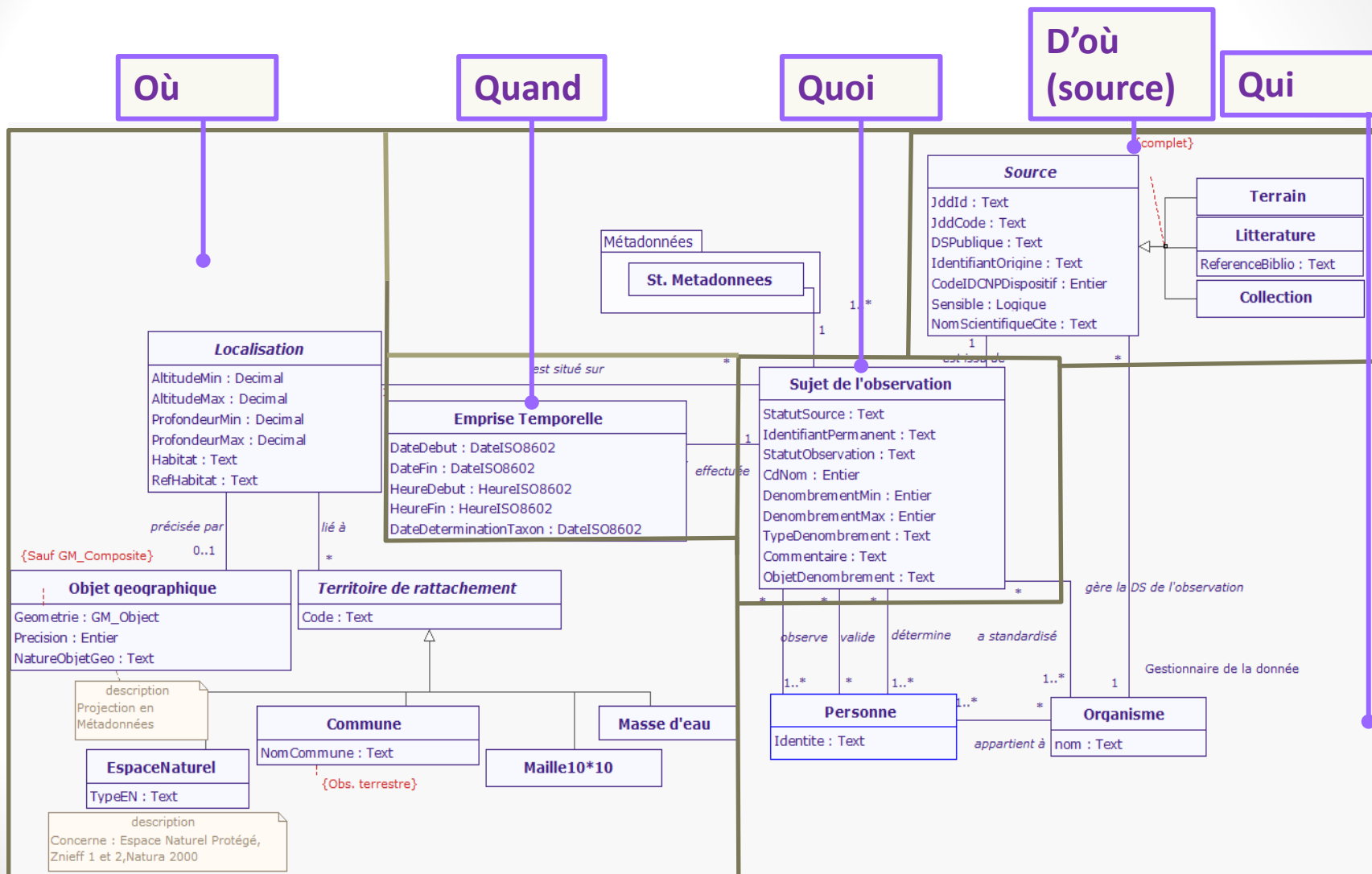
# Un standard modulaire

- Basé sur un cœur :  
=> Présence/Non observation et dénombrement



Les données complémentaires doivent être définies par le fournisseur de données : Thématique / Libellé/ Définition / Valeur /Unité

# Les informations portées par le standard



Un document Word complémentaire

# Les informations portées par le standard

Type

Vocabulaire  
Contrôlé

Exemple

StatutSource	Oui	Te
ReferenceBiblio		
JddId		00-18
JddCode		BDTerrain_32
IdentifiantOrigine		589636
IdentifiantPermanent		<i>En cours de définition – attribué par plateforme</i>
DSPublique	Oui	Pu
CodeIDCNPDDispositif	Oui	288
OrganismeStandard	Oui	
StatutObservation	Oui	Pr
NomCite		Bufo bufo (Linnaeus, 1758)
CdNom	Oui	259
Sensible	Oui	non
DenombrementMin		
DenombrementMax		
TypeDenombrement	Oui	
ObjetDenombrement	Oui	
IdentiteObservateur		
OrganismeObservateur	Oui	SPN MNHN
OrganismeGestionnaireDonnees	Oui	SPN MNHN
Determinateur		
Valideur		
Commentaire		
DateDebut		20/05/2012
DateFin		20/05/2012
HeureDebut		
HeureFin		
DateDeterminationObs		
AltitudeMin		
AltitudeMax		
ProfondeurMin		
ProfondeurMax		
Habitat	Oui	
RefHabitat	Oui	
<b>Localisation</b>		
Geometrie		45.256 -110.45 46.46 -109.48 43.84 -109.86 45.256 -110.45
Precision		25
NatureObjetGeo	Oui	NSP
CodeCommune	Oui	75056
NomCommune	Oui	Paris
CodeEP	Oui	
TypeEP	Oui	
CodeMaille	Oui	
CodeME	Oui	
<b>Données complémentaires</b>		
ThematiqueAttribut		
NomAttribut		
DefinitionAttribut		
ValeurAttribut		
UniteAttribut		

Version en cours



# Standard optimisé pour la diffusion ou la réutilisation des données : le compromis

- Moins le standard est contraignant et plus il est facile de diffuser des données.... Mais moins il est évident de les réutiliser ...

## D'où des contraintes :

- Données obligatoires (13 sur 43)
- Définitions, Règles et formats précis
- Liste de vocabulaire contrôlé :
  - Utilisations des référentiels existants gérés au niveau national (commune, TaxRef, Habitat...)
  - Création de liste de vocabulaire propre (concerne 4 attributs)
  - 2 listes en attentes : « organismes » et « référentiels bibliographiques »



## Concrètement

- Ces éléments de contraintes sémantiques sont définis dans un document :

<b>Nom de l'attribut</b>	<u>IdentiteObservateur</u>
<b>Définition</b>	Prénom, nom, alias ou surnom de la ou les personnes ayant réalisées l'observation
<b>Obligatoire</b>	Obligatoire
<b>Cardinalité</b>	1..N
<b>Format</b>	<u>Text</u>
<b>Règle</b>	Règle d'écriture : Nom Prénom Ou règle d'écriture : Surnom Si plusieurs personnes ont fait l'observation : concaténer les différentes identités séparées par des virgules « , » Exemple : Nom1 Prénom1, Nom2 Prénom2, Surnom3 Si l'observateur requiert l'anonymat, noter « Anonyme », si inconnu « NSP »
<b>Vocabulaire contrôlé</b>	non

**Vocabulaire contrôlé** Cd\_Nom du référentiel taxonomique TaxRef  
Lien :  
<http://inpn.mnhn.fr/telechargement/referentielEspece/referentielTaxo>



# Traçabilité de la donnée

Le cadre contraignant peut avoir simplifié la donnée

→ Importance de pouvoir identifier la DS d'où provient la DEE :

- Identification de la donnée source :
  - Informations sur l'identifiant IDCNP de la base source, l'identifiant de la DS dans la base ...
- Prise en compte de la chaîne des centralisateurs de données
  - Notion d'organismes impliqués dans la standardisation : non retrouvé dans les standards étudiés mais intéressants pour suivre où l'effort de standardisation a été produit
- Identifiant permanent
  - Tracer la donnée dans le cadre du SINP
  - UUID couplé à une URL



# Un standard en cohérence avec l'existant

- Au sein du SINP
  - Prise en compte d'éléments existants : GT Architecture Métadonnées, référentiels du SINP
  - Demande de mise en cohérence : GT Métadonnées : un standard de métadonnées est décliné pour le cas spécifique des DEE.
  - En attente des résultats de GT : donnée sensible, validation
- Au sein de la communauté internationale des données de la biodiversité
  - Prise en compte de l'existant dans l'analyse et la réflexion initiale
    - Livrable disponible « Draft »
  - On privilégie les mises en correspondance avec les standards existants (darwin core, standard du SIE, standards régionaux ...) plutôt que d'en reprendre des éléments afin de ne pas être tributaire des besoins des autres SI





# Un test avec les membres du GT

- Test très constructif et intéressant
  - => document de test rédigé
  - ONCFS : Castor
  - DREAL Bourgogne : BFF
  - DREAL Centre : Faune
  - IFREMER : Biocean
  - AAMP : Guide de référence technique N2000 DHFF
  - DWC : DWC
  - DREAL Pays de la Loire : Test non terminé
  - INPN : Document Standard
  - FCBN : Données FCBN.
  - SIE : Qesu v 3
  - OBHN : Données terrain STERF 2012 et autres données
- Montre clairement le besoin d'accompagnement dans la mise en œuvre du standard
  - Précision des définitions et concepts : itération
  - Problématique « politique » : interprétation du protocole : données publiques, floutage...



# Perspectives



# Perspectives

- Accompagnement de la mise en œuvre du standard « Occurrence de taxon »
  - Documenter et outiller
    - Standard en format « machine », forum, versionning des documents, outils de validation de fichiers standardisés pour vérifier le format
  - Promouvoir le standard produit
    - Communiquer / former, produire un document vulgarisé
    - Test sur identifiant permanent.
- Maintenir et faire évoluer le standard
  - Définir et mettre en place la gouvernance pour faire évoluer le standard de manière rationnelle et transparente
  - Etendre le standard aux thématiques les plus en demande
    - Flore et Végétation, géologie ...



# Questions



# CST - consultation intermédiaire :

## Questions/ Réponses

- Vous paraît-il judicieux d'intervenir dans les systèmes producteurs en les obligeant à attribuer et gérer un identifiant de données-source et à faire le lien entre leurs données-sources et les DEE ?
- Que pensez-vous de la proposition de supprimer la possibilité de créer des DEE composées ? Dans Quelles sont les conditions restreintes pour qu'une DEE composée soit possiblement assimilable à une DEE élémentaire ?
- Est-il pertinent de proposer deux standards de données occurrence de taxon qui véhiculeront les mêmes informations dans deux formats différents : un orienté « plate-forme informatique » de type GML, un orienté « utilisateur » basé sur plusieurs fichiers (fichiers csv, couche SIG) ?
- Ce choix d'organisation sur la validation vous paraît-il pertinent ? Avez-vous des remarques complémentaires ?
  - **2 réponses :**
    - **ne pas être intrusif, question sur l'intérêt de l'identifiant permanent, enjeu sur les doublons n'est pas un problème majeur**
    - **Identifiant à la source intéressant, ne pas permettre des DEE composées, proposer deux formats : un plus optimisé technique et un plus facile pour les utilisateurs, permettre de véhiculer des informations de validation existante dans les systèmes sources**



# CST intermédiaire : position du GT

- Identifiant permanent et impact dans les SI producteurs
  - Perte de traçabilité à la DS mais opérationnel
  - Solution pragmatique : on reste dans le cadre du suivi de la DEE, pas d'intrusion dans les SI sources
- Identifiant permanent et impact lien DEE- DSource
  - Pas d'identifiant permanent pour les DEE liées à plusieurs DS
- Optimisation informatique xml ou simple csv
  - Choix d'orienter XML – le standard ayant vocation d'être produit par des plateformes. On laisse aux plate-forme ce seront aux plateformes de proposer des formats pour les utilisateurs ?
- Valideur : gestion a minima
  - On garde ainsi, des données complémentaires de validation peuvent être ajoutées en attributs complémentaires