



Les défis qui attendent le GEOCATALOGUE

**Atelier technique N° 1 : Articulation entre niveau local et national :
comment optimiser l'accessibilité aux données et services dans
l'infrastructure de données géographiques?**

Journée INSPIRE, Paris le 16/11/2015

Maëlle Decherf et Etienne Taffoureau, BRGM

Le Géocatalogue en 2015 – contexte général

- > Depuis plusieurs années, le Géocatalogue apporte une réponse efficace à la mise en œuvre de la directive INSPIRE
 - Avec une orientation très « réglementaire » (INSPIRE) et peu orientée « usagers »

- > Constats :
 - 1/ Les PRODUCTEURS rencontrent des difficultés pour moissonner leurs métadonnées et les rendre conformes avec les règles nationales et européennes
 - 2/ Les USAGERS trouvent difficilement les données qu'ils recherchent ...
 - pertinence des résultats de recherche
 - multitude des sources
 - qualité des métadonnées
 - 3/ ... et lorsqu'ils les trouvent ne peuvent pas y accéder
 - 4/ Les efforts des PRODUCTEURS de données ne sont pas suffisamment valorisés à l'échelon national

- > Les actions menées en 2015 constituent un début de réponse à ces questions

Grandes axes de travail

- > **améliorations du référencement des données**
 - articulation des différents niveaux (local / régional / thématique <> national) = aspects organisationnels
 - liste des catalogues moissonnés = aspects quantitatifs
- > **amélioration des résultats de recherche**
 - qualité des métadonnées
 - présence de métadonnées de services
 - doublons
 - harmonisation et catégorisation des noms des producteurs
 - conformité INSPIRE / outils de validation
 - qualité des données
 - identificateur unique de ressource (URI)
 - listes de code
- > **valorisation des contenus et acteurs**
- > **implication des utilisateurs**
 - retours utilisateurs
 - implication des producteurs dans la démarche qualité

Positionnement du Géocatalogue dans l'IDG nationale

principes et recommandations (volet organisationnel)

> Orientations portées par le MEDDE

- Le Géocatalogue est le nœud privilégié de l'infrastructure de données géographiques (IDG) nationale pour la recherche de données et la consultation des métadonnées au niveau national (entrepôt de moissonnage)
- Il s'appuie sur des catalogues fédérateurs au niveau régional ou thématique pour la collecte des informations

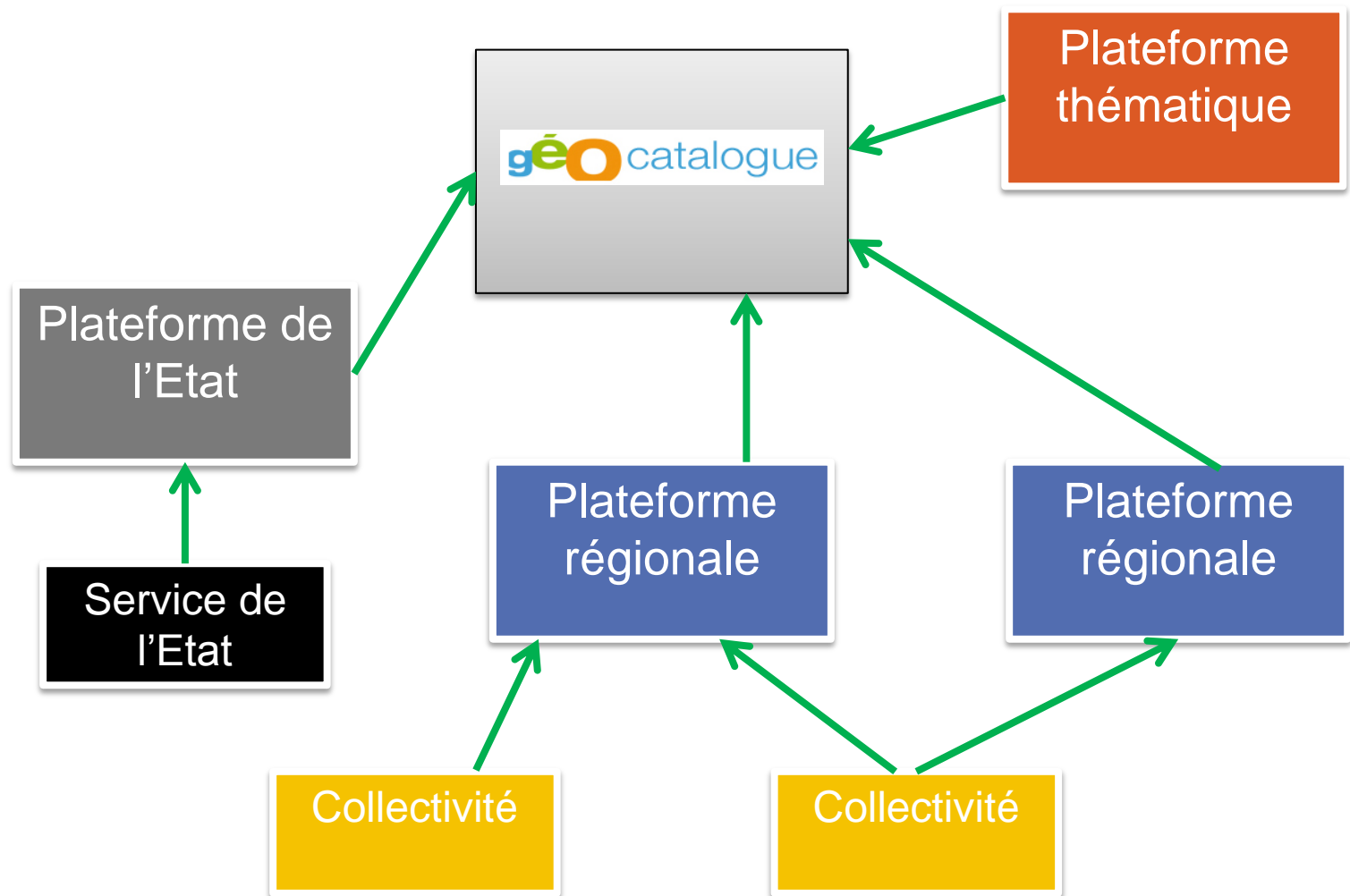
> Recommandations nationales

- la métadonnée doit être produite et diffusée au plus près du producteur
- chaque administrateur doit s'assurer que *a minima* les métadonnées concernées par INSPIRE dont il a la responsabilité sont remontées au Géocatalogue
- seule l'organisation qui produit une ressource ou l'une des organisations responsable de sa diffusion, alimente le catalogue fédérateur des métadonnées correspondantes, de façon à éviter les duplications

> *cf. guide de gestion des catalogues*

Positionnement du Géocatalogue dans l'IDG nationale

illustration des principes et recommandations



Le moissonnage dans le Géocatalogue

- > Dans le Géocatalogue, la création des comptes et des points de moissonnage est basée sur une démarche déclarative
- > Actuellement, plus d'une centaine de catalogues sont moissonnés dans le Géocatalogue (cf. liste publiée sur geocatalogue.fr)
- > Constats :
 - manque d'exhaustivité dans le moissonnage des catalogues existants
 - configuration et règles du moissonnage insuffisamment claires
- > Ce qui est prévu (début 2016) :
 - réalisation d'un inventaire et d'une cartographie des catalogues existants (« catalogue de catalogues »)
 - diffusion d'un tutoriel expliquant la démarche
- > Ce qui est envisagé :
 - moissonnage automatique des catalogues non connus par le Géocatalogue

Les métadonnées de services de découverte

> Contexte :

- un service de découverte peut être assimilé à un catalogue ou un sous-ensemble de catalogues
- sa fourniture est rendue obligatoire par le règlement métadonnées
- intérêt : facilite le moissonnage par le Géocatalogue mais également par les autres plateformes
- constat : il est compliqué pour les gestionnaires de catalogue de renseigner ces métadonnées
- seuls 10% des catalogues moissonnés sont décrits par une fiche de service

> Ce qui a été réalisé :

- création et publication automatique des fiches de services par le Géocatalogue à partir des capacités des services
- diffusion d'un mode opératoire pour l'élaboration des fiches (http://www.geocatalogue.fr/www/Geocatalogue_accessibilite_services.pdf)
- mise en place d'une procédure pour récupérer, compléter et publier les fiches (par les producteurs)

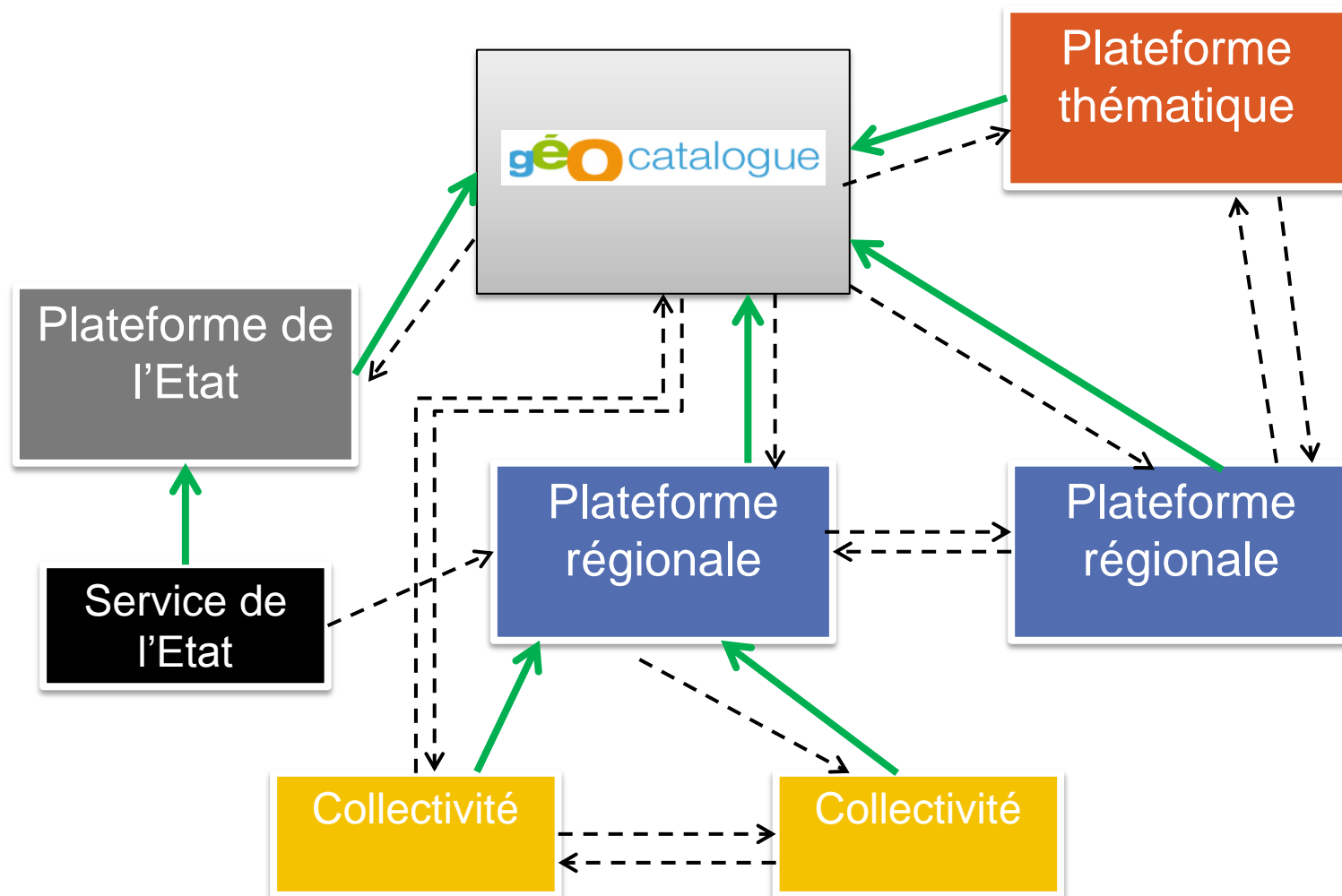
> Ce qui est prévu :

- faire évoluer GéoSource pour automatiser la création des fiches

La recherche dans le Géocatalogue

- > Une pondération ("pertinence") est appliquée au niveau des champs de métadonnées (titre, mots-clés, etc.)
- > Constat : les données « de référence » sont noyées dans toutes les autres
- > Ce qui a été réalisé :
 - intégration d'un mécanisme de suggestion
 - intégration d'un mécanisme de synonymie à partir des vocabulaires contrôlés GEMET, GéoVoc (QUOI ?) et entités administratives (OÙ ?)
- > Ce qui est prévu :
 - migration du moteur de recherche vers la technologie SolR (début 2016)
 - suppression semi-automatique des doublons (*cf. diapo dédiée*)
 - harmonisation des noms de producteurs et intégration dans les "facettes"
 - amélioration de la qualité par des actions d'accompagnement et de traitements automatiques (ex. CSW+, gestion des mots clés par thésaurus, etc.)

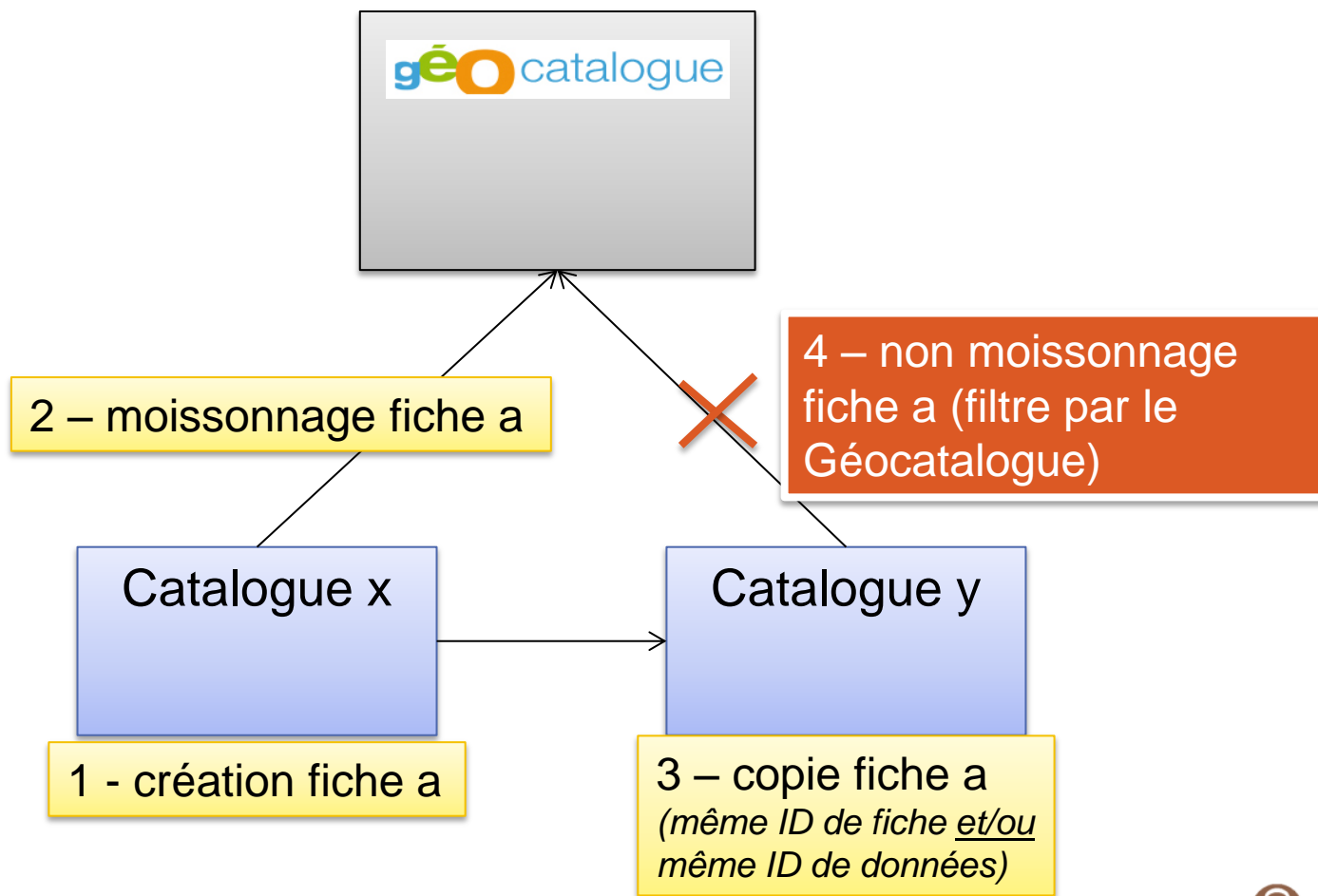
L'organisation réelle des catalogues du local au national



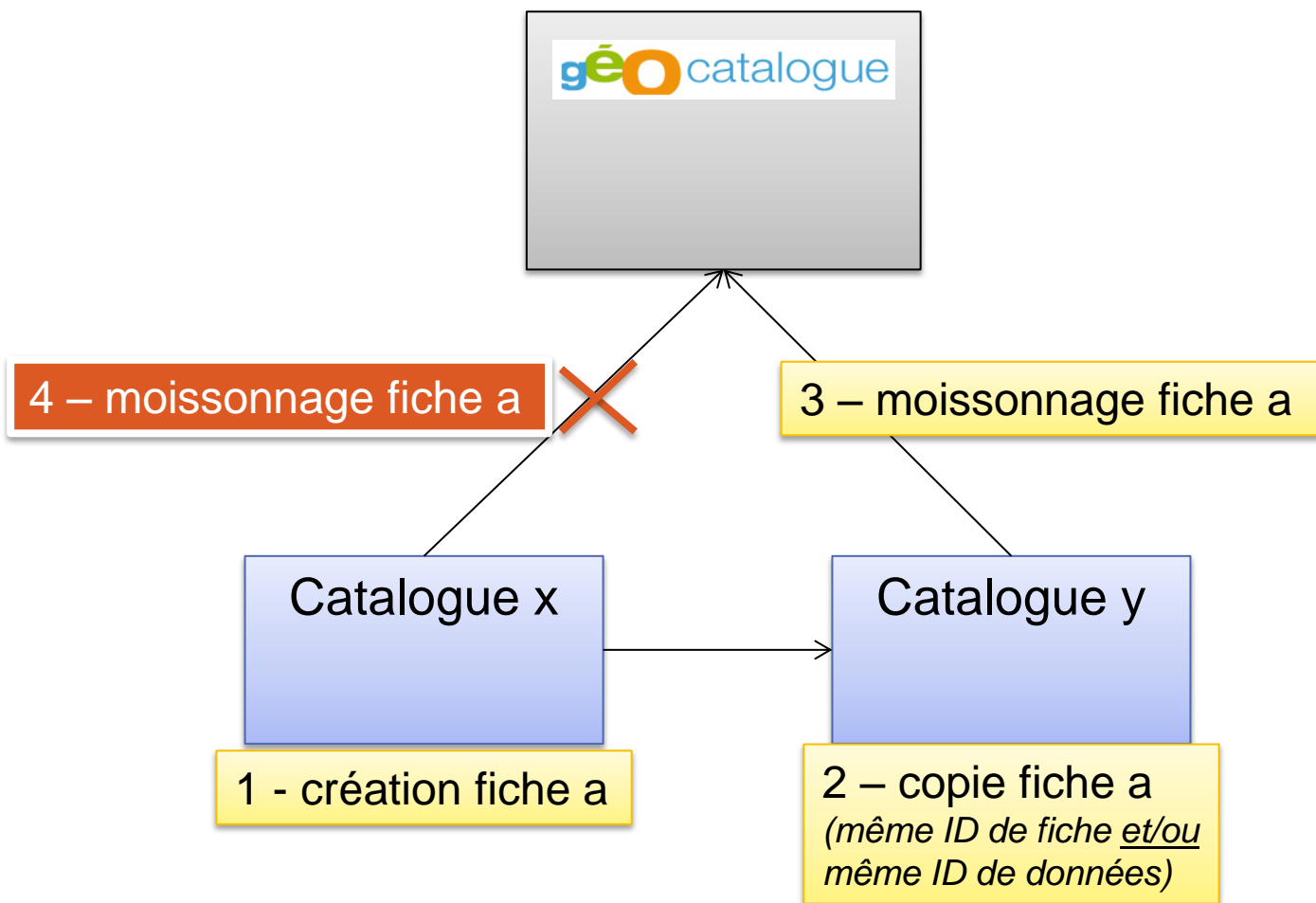
La gestion des doublons dans le Géocatalogue

- > On parle de doublon lorsqu'une donnée est référencée au moins deux fois dans un catalogue
- > Plus le catalogue est fédérateur, plus le risque d'y trouver des doublons est grand
- > Ainsi, dans le Géocatalogue certaines données peuvent être référencées plus d'une dizaine de fois (ex. SCAN IGN)
- > En pratique, les recommandations nationales sont peu suivies, ce qui pose des problèmes dans la recherche
- > Techniquement, les doublons sont gérés via des contraintes d'unicité :
 - sur l'identifiant des fiches (fileIdentifier)
 - Sur l'identifiant de la donnée (MD_Identifier) => identificateur de ressource unique dans INSPIRE

Cas 1 : la fiche est publiée par le producteur de la donnée

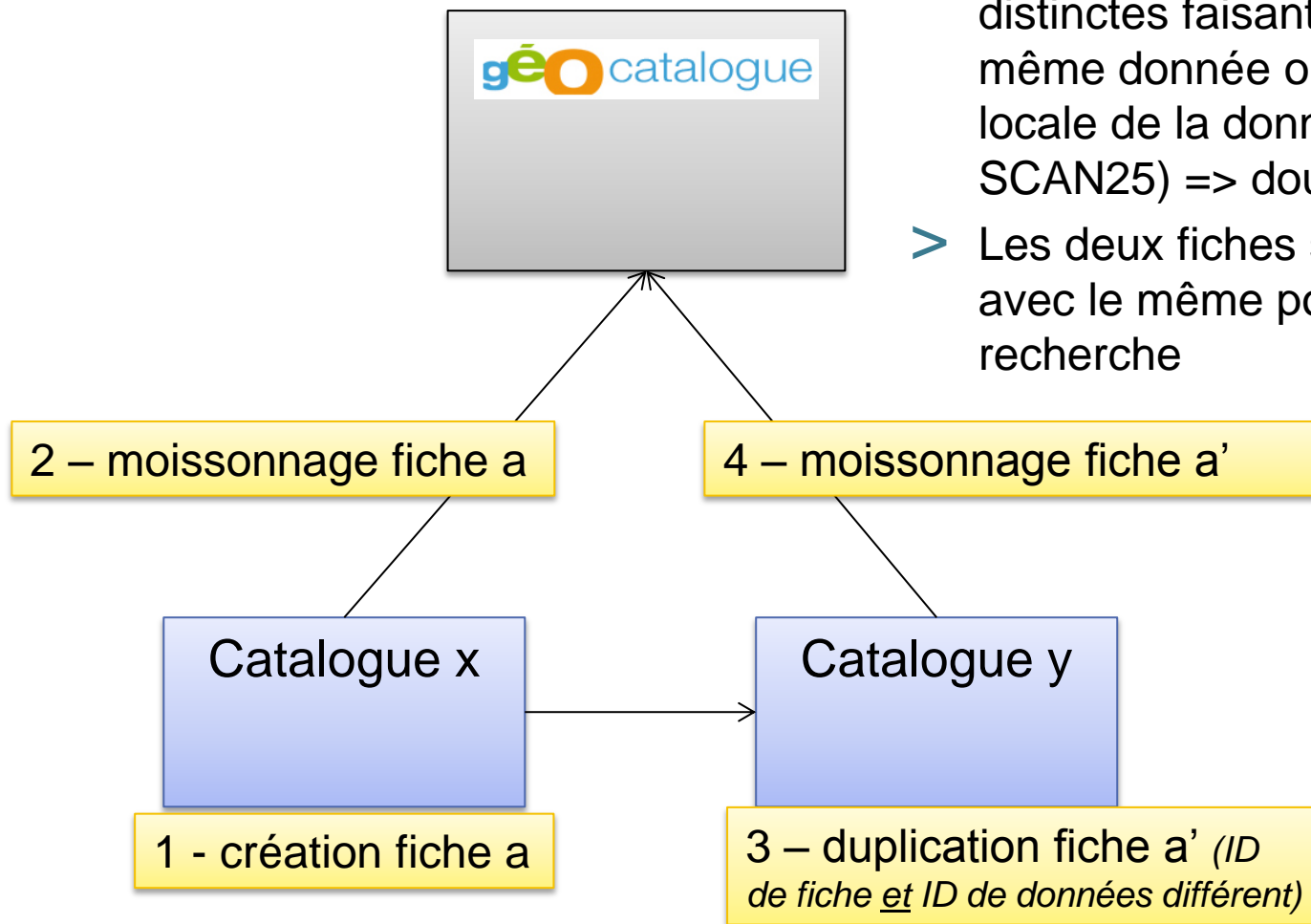


Cas 2 : la fiche n'est pas publiée par le producteur de la donnée



Cas 3 : la fiche est dupliquée

- > Présence de deux fiches distinctes faisant référence à une même donnée ou déclinaison locale de la donnée (ex : SCAN25) => doublons « métier »
- > Les deux fiches sont indexées avec le même poids dans la recherche



La gestion des doublons dans le Géocatalogue

> Ce qui a été réalisé :

- travail manuel de « dé-doublonnage » dans le rapport INSPIRE 2015
- retour aux diffuseurs uniquement sur les données nationales de référence (IGN, DGFIP, SANDRE)
- explication des bonnes pratiques aux adhérents
- possibilité de filtrer ses métadonnées dans GéoSource (« CSW virtuel »)

> Ce qui est envisagé :

- tout moissonner et faire des regroupements par fiche/donnée et afficher les différents catalogues diffuseurs dans le résultat de recherche
- identifier les fiches de référence (celles remontées par le producteur) manuellement ou à l'aide de mécanismes automatiques
- avertir les producteurs lorsqu'une de leurs données est ajoutée dans le Géocatalogue

L'harmonisation et la catégorisation des noms de producteurs

> Objectifs :

- améliorer la pertinence des résultats de recherche
- réalisation de synthèses et d'indicateurs (par ex. nombre de données téléchargeables pour chaque DREAL)

> Ce qui a été réalisé :

- fichier de correspondance entre les noms des producteurs déclarés et harmonisés, actualisé plusieurs fois par an

(cf. <http://www.geocatalogue.fr/#!HelpCatalogue>)

> Ce qui est prévu :

- fin 2015 : service CSW à valeur ajoutée fournissant les métadonnées avec les noms harmonisés des producteurs (ciblé sur certaines catégories de producteurs : services de l'Etat, etc.)
- janvier 2016 : prise en compte des coproducteurs

Le service à valeur ajoutée (CSW+)

> Objectifs :

- Corriger à la volée les métadonnées moissonnées et de les rendre disponibles dans un service CSW
 - Correction de erreurs de saisie (par ex. les thèmes INSPIRE)
 - Réduction des problèmes d'hétérogénéités des informations (par ex. les noms des producteurs)
- Accroître la pertinence de la recherche et de faciliter le calcul des indicateurs dans le tableau de bord
- Les informations renseignées dans les fiches d'origine restent accessibles

> Publication prévue le 18 novembre :

- Application sur l'harmonisation des producteurs
- Traitement appliqué sur une sélection de producteurs (par défaut services de l'Etat, établissements publics)
- Utilisation pour le rapportage 2016

> Ce qui est envisagé :

- Retour par email aux producteurs lorsqu'un traitement est effectué
- Application sur la correction des thèmes INSPIRE, ...

L'accessibilité des données

> Contexte :

- L'accessibilité des données est l'un des principaux objectifs visés dans l'IDG nationale, en particulier le téléchargement des données
- Elle est exprimée au travers de l'indicateur NSi2 dans le cadre du rapportage INSPIRE mais surtout de façon concrète dans les applications
- Les taux d'accessibilité sont de 43% de données téléchargeables et de 66% visualisables
- Constats :
 - Les métadonnées de services, dont la fourniture est rendue obligatoire par le règlement métadonnées ne sont pas toujours disponibles
 - Peu de données sont réellement visualisables dans le Géocatalogue (règles INSPIRE non suffisantes)
 - Les règles de saisie des métadonnées sont complexes et peu claires

> Ce qui a été réalisé :

- Tutoriel expliquant comment bien renseigner ses métadonnées (disponible sur <http://www.geocatalogue.fr/#!/HelpCatalogue>)

> Ce qui est envisagé :

- Réaliser une analyse approfondie de chaque service de visualisation
- Contrôler et simplifier les métadonnées dans un service à valeur ajoutée

Gestion des Identificateurs de ressource uniques (IRU)

- > Les IRU permettent :
 - 1/d'identifier de manière unique une ressource
 - 2/de faire le lien entre les métadonnées de données et de services, et entre les capacités des services et les ressources correspondantes
- > Les IRU doivent être retournés (on dit « résoudre ») la fiche de métadonnées de la ressource
- > Scénario retenu : chacun choisit sa propre politique d'identifiant de ressources tout en respectant les bonnes pratiques définies au niveau national.
 - *Avantages* : Les organismes ayant déjà fait le choix d'un système de gestion d'identifiants peuvent le conserver moyennant le respect de bonnes pratiques
 - *Inconvénient* : Aucune solution clé en main n'est offerte aux organismes n'ayant pas fait le choix de politique d'identifiants
- > Il a été décidé de proposer dans le Géocatalogue un service permettant de fournir un IRU résolvable à ceux qui le souhaitent
- > Cf. guide de recommandations du CNIG sur les IRU

Registre national d'extensions de listes de codes

- > L'infrastructure INSPIRE nécessite la définition de certains éléments centralisés dans des registres (cf. <http://inspire.ec.europa.eu/registry/>)
- > L'article 6 du règlement (UE) N° 1089/2010 modifié par le règlement (UE) N° 1253/2013 offre la possibilité aux fournisseurs de données de définir pour certains attributs des extensions aux listes de codes définies par le règlement. L'article définit 4 types de listes de codes :
 - Non extensible
 - Extensible avec des valeurs spécifiques
 - Extensible avec des valeurs à tout niveau
 - Liste de codes vide
- > L'article 6.3 précise : « Lorsque, pour un attribut dont le type est une liste de codes visée au paragraphe 1, point b), c) ou d), un fournisseur de données communique une valeur qui n'est pas spécifiée dans le présent règlement, cette valeur et sa définition sont mises à disposition dans un registre. »
- > **Il a été décidé de proposer dans le Géocatalogue un registre de listes de codes national qui étende le registre INSPIRE**

Les indicateurs / tableau de bord

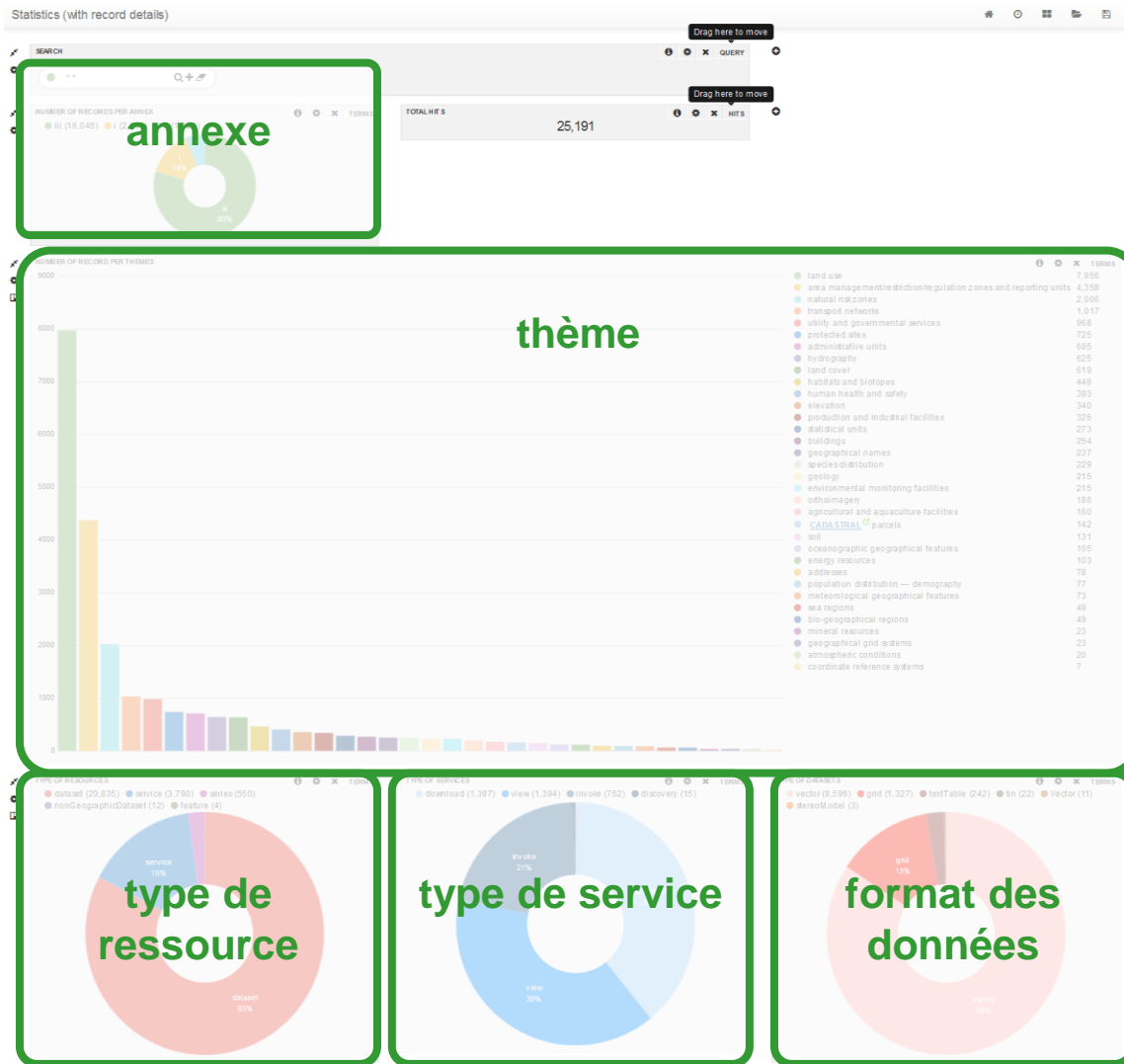
> Contexte :

- Réalisation de synthèses et d'indicateurs pour produire des rapports à destination, des producteurs, des diffuseurs, de la Commission européenne
- Permet de valoriser le travail de référencement des données
- Permet de fournir les indicateurs INSPIRE au format XML à la CE

> Ce qui est prévu (fin 2015) :

- Mise en place d'un tableau de bord avec des indicateurs au-delà d'INSPIRE
- A terme, doit remplacer l'outil de supervision <http://geocat.docinspire.eu>
- Basé sur le prototype réalisé dans le cadre du Groupe de maintenance INSPIRE sur le reporting & monitoring (MIWP-16)
- Ouverture au public début 2016

Un exemple de tableau de bord dans le Géocatalogue

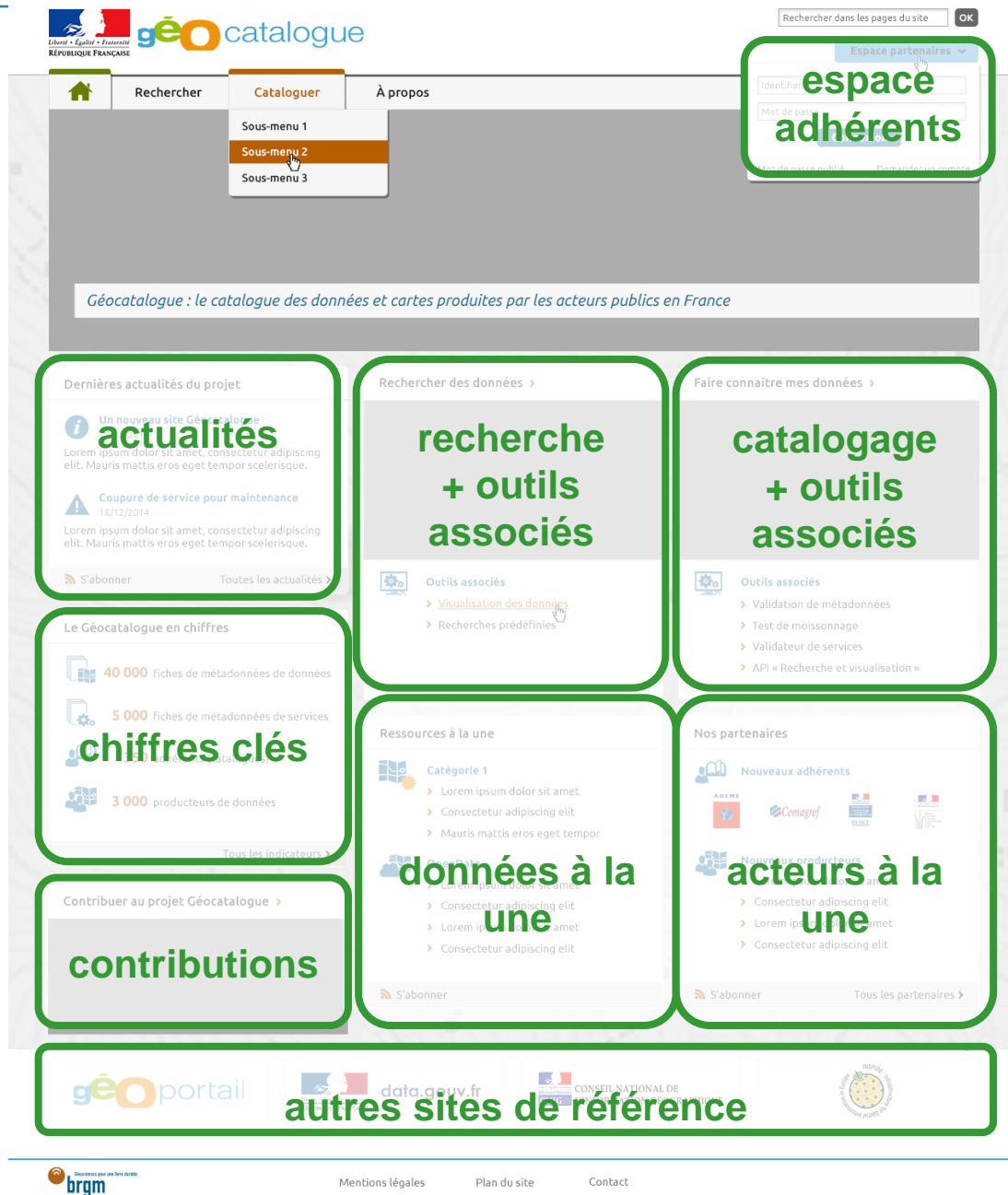


La refonte du site éditorial

page d'accueil

(maquette - projet)

- > 2 grandes entrées
 - rechercher des données + requêtes prédéfinies
 - publier des métadonnées + outils associés (ex. validation métadonnées)
- > valorisation des producteurs et données
 - "à la une"
 - chiffres et indicateurs (dont rapportage)
- > actualités
 - nouvelles fonctionnalités
- > retours utilisateurs
 - sondages, enquêtes
 - réutilisations



La refonte du site éditorial

page interne
exemple : aide aux utilisateurs
(maquette - projet)

Rechercher dans les pages du site OK

Espace partenaires ▼

Rechercher Cataloguer À propos

Géocatologue : le catalogue des données et cartes produites par les acteurs publics en France

Accueil > À propos > Foire aux questions

À propos

- > Sous-menu 1
- > Formulaire de demande de support
- > Foire aux questions
 - > Sujet 1
 - > Sujet 2

Foire aux questions - Sujet 1

foire aux questions

A qui s'adresse le Géocatologue ?

Le Géocatologue s'adresse en priorité aux professionnels de la géomatique (établissements publics, collectivités, bureau d'études, etc.) ayant besoin d'accéder à des données géolocalisées sur des thématiques particulières.

Comment est alimenté le Géocatologue ?

Comment accéder aux données référencées dans le Géocatologue ?

Les données sont-elles hébergées par le Géocatologue ?

Quel lien avec la directive INSPIRE ?

tutoriels

support

géoportail

data.gov.fr

CONSEIL NATIONAL DE L'INFORMATION GEOGRAPHIQUE

brgm

Mentions légales Plan du site Contact

La refonte du site éditorial

> Objectifs

- Valoriser les nouveautés et les actualités du projet, les contenus (communication, fonctionnalités, rapportage), les partenaires (producteurs et diffuseurs) et les réutilisations
- Simplifier l'accès aux contenus via des requêtes prédéfinies (accès par thématiques, par territoire, etc.)
- Rendre plus facilement accessibles les supports d'accompagnement (tutoriels, FAQ, etc.)
- Permettre de recueillir les retours des utilisateurs par exemple au travers de sondages

> Échéance : début 2016 (en cours de spécifications)

Conclusions

- > Principaux défis du Géocatalogue pour 2016 :
 - poursuivre les actions en lien avec la mise en œuvre de la directive INSPIRE
 - collecter les informations de manière plus exhaustive tout en améliorant la pertinence de la recherche
 - améliorer la qualité globale des métadonnées

- > Avec une implication plus importante des contributeurs pour apporter des réponses concertées et coordonnées aux difficultés
 - le Géocatalogue doit être utile au plus grand nombre
 - Volonté de valoriser les acteurs et les contenus.
 - Améliorer les contenus et mettre en oeuvre des processus de validation des contenus (MDD).
 - le Géocatalogue ne peut pas tout faire et besoin d'une amélioration des catalogues locaux
 - mise en place d'un groupe contributeurs

