

Efforts de Médias-France dans le domaine des métadonnées

Atelier CNES sur les
métadonnées
24 et 25 novembre 2004



Contexte

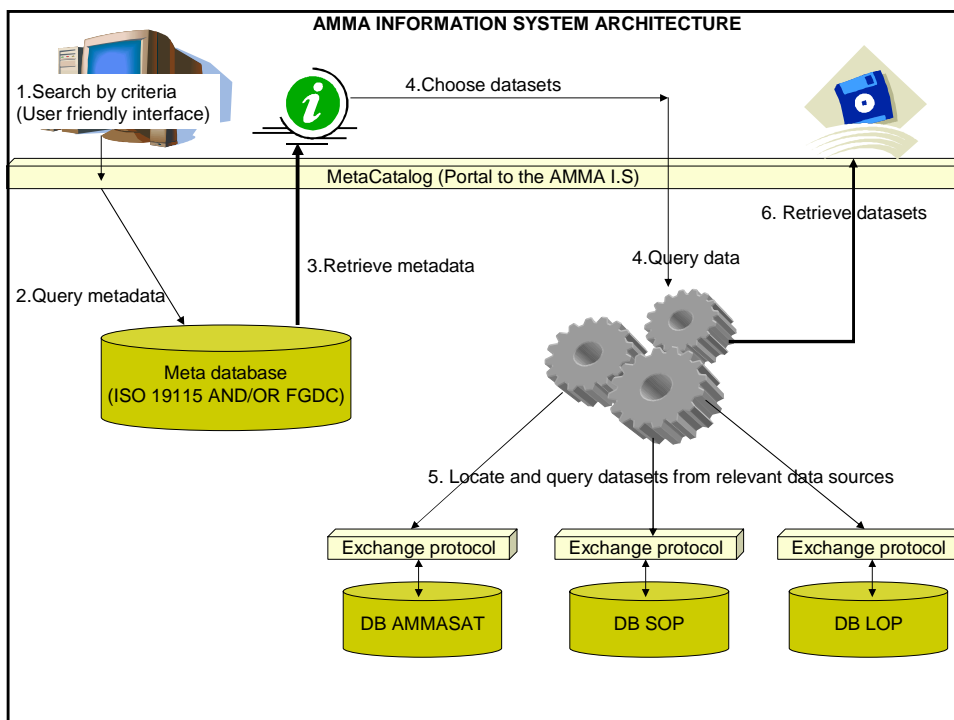
- Métier clé de Médias-France
 - Structuration et gestion des données scientifiques
 - Pérennisation des données
 - Services de découverte des données
 - Services d'accès aux données
- Divers domaines thématiques couverts
- Divers structures de données gérées



Cas d'utilisation type: AMMA



- Campagne multidisciplinaire d'analyse de la mousson africaine
- Couvre les aspects océan, atmosphère, précipitations, hydrologie, végétation, sols
- Mesures satellitaires, radiosondages, campagnes de terrain, sorties de modèles
- Equipes multidisciplinaires, internationales
- Campagne sur 10 ans



Besoins en métadonnées



- Besoin de décrire les données
- Besoin de découvrir les données
- Besoin d'accéder à la donnée
- Besoin de pouvoir exploiter la donnée
- Besoin de valoriser la donnée
- Besoin de pérenniser la donnée (Son contexte d'utilisation)

Approche préliminaire



- En 2001, début des investigations sur les standards existants
- Etude des besoins des laboratoires scientifiques en termes de services de données dans le cadre d'une action R&T CNES
- Développement d'un premier catalogue des métadonnées basé sur un profil simple du standard du FGDC pour l'OSS

IGN | Revenir sur l'éditeur de métadonnées pour le projet 60-050

Métadonnées | Les champs avec * sont obligatoires

Jeu de données

- Thèmes
- Données

Informations annexes

- Productions
- Centres
- Adresses

Ajout/Modification d'un jeu de données

Titre *	Base cartographique CACT
Producteur *	CACT
Contact	SAPAR-STOUR
Date de production *	20/05/1995
URI Web	A
Résumé *	Ensemble des informations sur l'état de la cartographie et sciences annexes dans 27 pays d'Afrique (Pays ACP)
Objets 1/5 *	Principaux des travaux, structures et potentialités.
Date de début des mesures *	01/01/1990
Date de fin des mesures *	01/01/1995
Etat d'avancement *	Complet
Longitude Ouest *	20.147
Longitude Est *	23.995

Accueil | **Accueil**

Critère de recherche des métadonnées/données

Sélectionner la zone et les critères de recherche

Sélection par requêtes spatiales

Sélection par attributs

Attributs de jeu de données

Données / Thèmes / Sous-Thèmes

- Titre
- Auteurs
- Appartenances
- Atmosphère
- Biogènes
- Crocoïtes

Producteur: [Titre]

Date de production: de [] à []

Attributs des mesures

Type de données: [Titre]

Echelles de: de [] à []

Résolutions: de [] à []

(1) Entrer une date sous la forme jj/mm/aaaa
(2) entrer un nombre

Sélection par expertise géographique

Coordonnées géographiques de la zone sélectionnée

Long min: [19] | Lat min: [36] | Long max: [51] | Lat max: [41]

(3) En degré décimale. Pour une longitude = 0° ou une latitude = 0°, les nombres à entrer sont négatifs

Metadata
Identification_Information
Citation
Citation_Information
Originator: OSS
Publication_Date: 2002-02-12
Title: Archive NDVI SPOT4/VEGETATION
Online_Linkage:
Description
Abstract: Exemple archive NDVI SPOT4/VGT
Purpose: Illustration
Time_Period_of_Content
Time_Period_Information
Range_of_Dates/Times
Beginning_Date: 2001-02-01
Ending_Date: 2001-07-31
Currentness_Reference: Publication date
Status
Progress: En cours
Maintenance_and_Update_Frequency: IRREGULAR
Spatial_Domain
Bounding_Coordinates
West_Bounding_Coordinate: -18.000
East_Bounding_Coordinate: 25.000
North_Bounding_Coordinate: 36.000
South_Bounding_Coordinate: 4.000
Keywords
Theme
Theme_Keyword_Thesaurus: GCMD
Theme_Keyword: Biosphere->Vegetation->Indice de vegetation
Place
Place_Keyword_Thesaurus: GCMD

Leçons de la première approche



- Pendant longtemps, l'ISO 19115 est resté sous forme de draft
- La politique de l'ISO n'encourage pas les développements « libres » et donc une mutualisation des efforts au niveau international
- Le standard Content Standards for Digital Geospatial Metadata du FGDC est largement utilisé et un ensemble de développements libres sont disponibles

Conclusion: Utiliser le standard du FGDC en attendant la maturité du standard de l'ISO

Questions



- Jusqu'où doit on aller dans la granularité pour décrire les jeux de données ?
- Pour quelle finalité ?
- Qui doit créer les métadonnées ?
- Quels outils pour la création des métadonnées ?
- Outils simplifiés mais restreints ou exhaustifs et complexes ?
- Des profils adaptés à chaque thématique ou génériques ?
- Quelle forme de stockage: XML natif ou mapping vers le relationnel ?

Contraintes



- Capacité d'accepter tout profil dérivé du standard
- Capacité de gérer les instances de métadonnées et leur intégrité (Création, MAJ, Suppression)
- Capacité de pérennisation des instances de métadonnées
- Capacité de recherche rapide et de restitution
- Les outils de création ne doivent pas bloquer le créateur
- Le catalogue de métadonnées doit être interopérable avec les catalogues existants

Choix techniques



- Le standard du FGDC comme canevas dans un premier temps
- Le profil « Schéma XML » modifiable selon le besoin
- Les formulaires d'édition configurables selon le profil
- Le stockage des métadonnées dans une base de données XML « Open source »
- Un service de catalogue basé sur Z39.50

Les outils



- Le W3C XML Schema pour définir les profils
- XSLT pour générer l'ensemble des formulaires de saisie
- Castor pour générer les classes java à partir du schema XML
- Xindice comme base de données XML avec xpath et xupadte
- Zebra comme serveur Z39.50
- ZAP comme module client Z39.50

Résultats obtenus



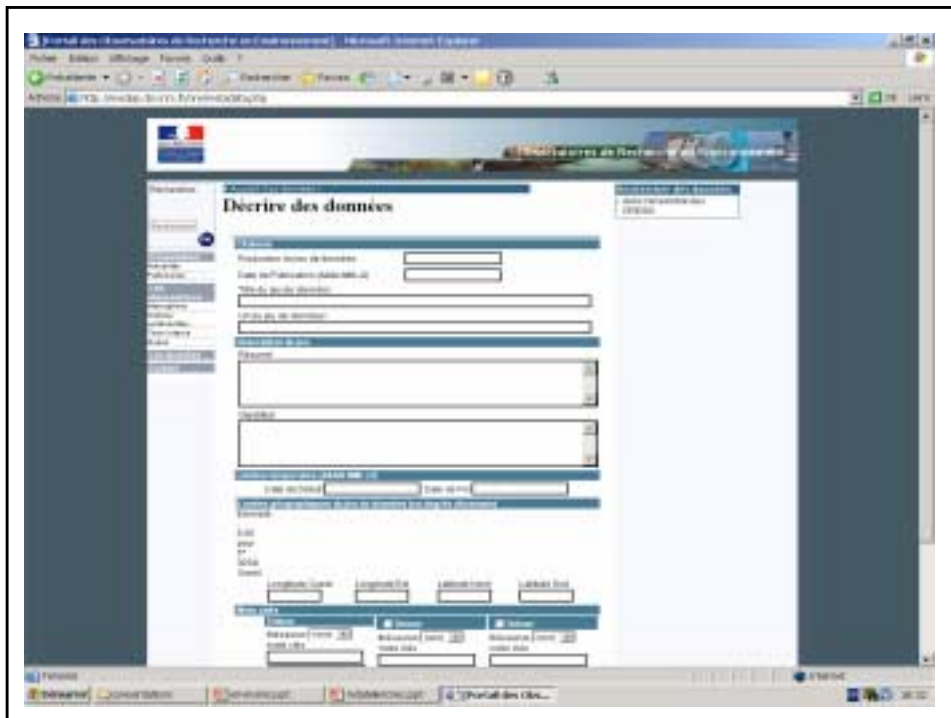
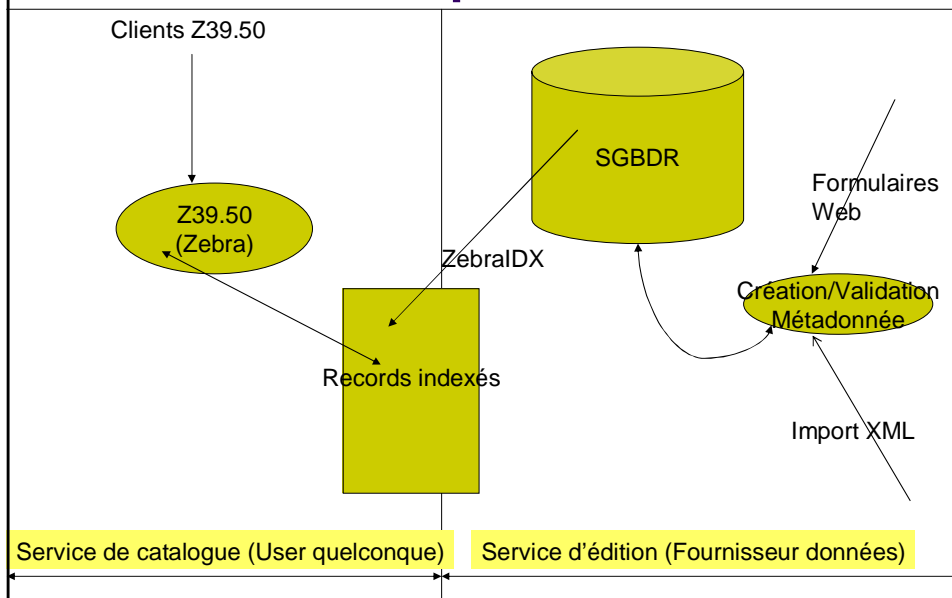
- Xindice nous a posé des problèmes à cause de sa jeunesse: Bugs et changements de version
- Xindice a été abandonné au profit d'une solution mixte: Fichiers XML + BD relationnelle
- Cette solution est en partie développée et les résultats sont satisfaisants
- Pour la génération automatique de formulaires, la solution est en état de prototype non encore opérationnel

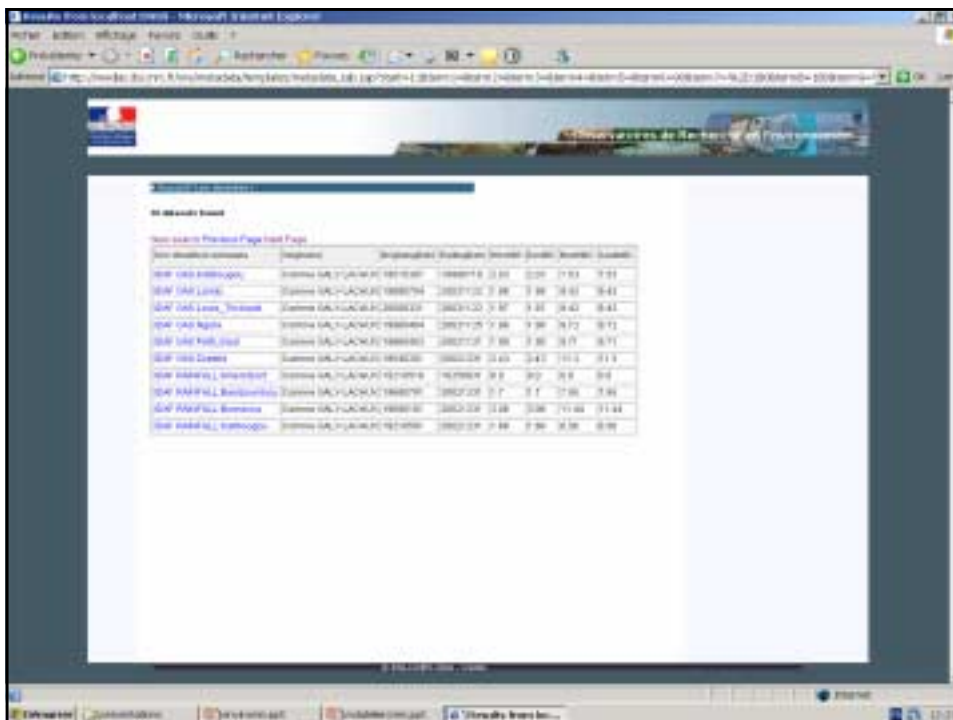
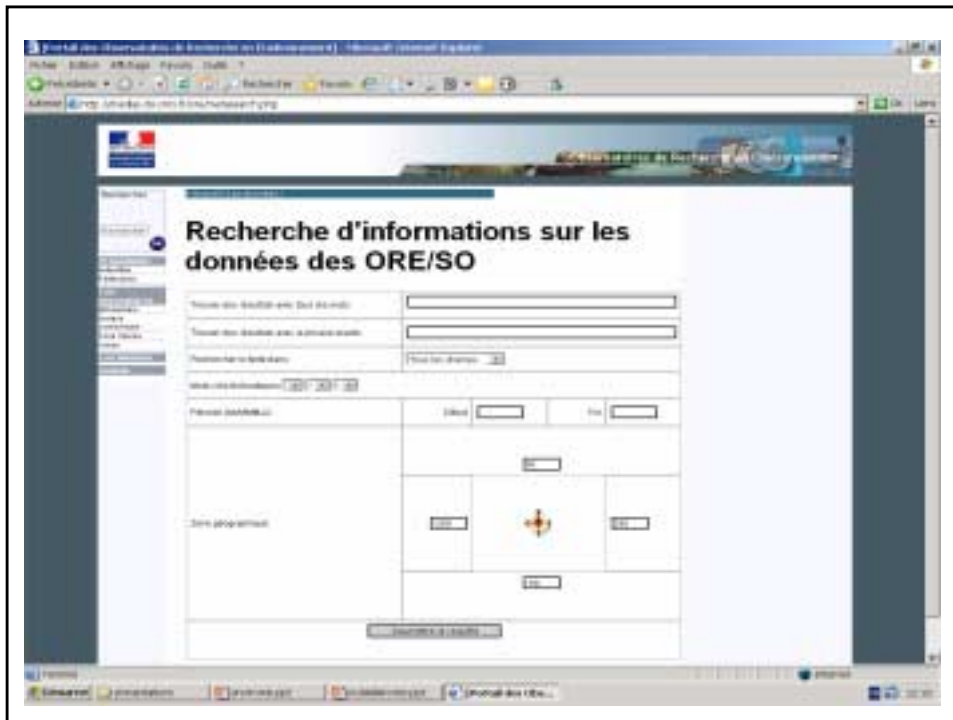
Solution actuelle



- Formulaires légers basés sur FormSix: Outil de génération automatique à partir d'une DTD
 - Structure moins typée
 - Code source non disponible
- Le stockage des instances XML dans une BD relationnelle sous forme d'un champ texte
 - Perte des facilités SQL du SGBD
 - Découplage entre les « records » en base et les « records » indexés

Schéma technique





Efforts futurs



- Mettre en place une architecture de gestion des « records » de métadonnées cohérente
- Automatisation de la création de formulaires à partir de schémas XML
- Gestion multi-normes des métadonnées
- Faire le lien entre le service de catalogue et les services d'accès aux données