



Catalogage et services web associés

Notions, outils et initiatives

Jean Christophe Desconnets (IRD-ESPACE)



Plan

- Les enjeux
- Notion relatives au catalogage
- Service de découverte
- Outils de catalogage
- Présentation MDweb

Les enjeux



Le Contexte

Les Systèmes d'information dans le domaine
environnemental (science de la terre, du vivant, société)

- **Multi sources : Variées, hétérogènes, en constante augmentation**
- **Multi - Acteurs : (société civile, collectivités, scientifiques, Etat)**

Les enjeux

Au niveau de la conception des SI

- Proposer une vision globale et homogène des SI
 - des données et des traitements (ressources)
 - Des concepts du domaine (connaissance)

Au niveau de la production des données

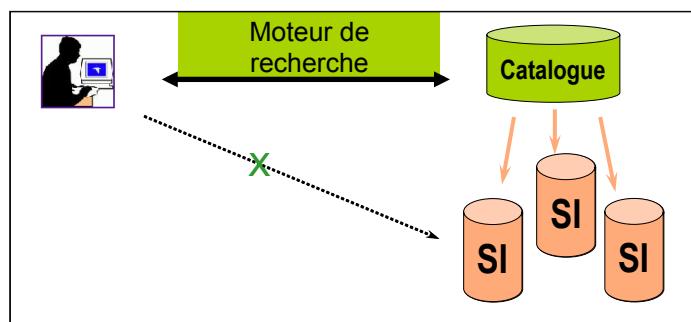
- Améliorer la gestion du capital d'information
- Assurer l'autonomie des producteurs et diffusion des données

Au niveau réglementaire

- Directives européennes INSPIRE sur la diffusion et l'accès à l'information environnementale dans les pays de l'Union
- Accès de l'information vers le citoyen (convention Aarhus)

Approche adoptée

Introduction d'un niveau de description des SI pour localiser et accéder aux ressources informationnelles



Notions relatives au catalogage



Notions de catalogue de données

- Catalogue informatisé de données géoréférencées
- Objectifs :**
- Trouver une carte, une image** (fonction de localisation)
- Mais aussi
- Donner des informations pour juger de la qualité et de l'adéquation avec son besoin (**fonction de description**)
 - Visualiser et découvrir le jeu de données (**fonction de représentation**)
 - Le télécharger (**fonction d'accès**)

Notions de catalogue de données

- Catalogue informatisé de données géoréférencées
Le Contenu des catalogues doivent permettre de répondre aux questions :
 - Quels sont les données disponibles sur un territoire x ?
 - Comment puis-je les obtenir ?
 - Combien coûtent-ils ?
 - Que contiennent-ils ?
 - Sous quels formats sont-ils disponibles ?
 - Quelle est la fréquence de mise à jour ?

Nécessité de structurer le contenu des catalogues : les métadonnées

Les métadonnées : un support au catalogage

Définition :

Une métadonnée est une donnée qui renseigne sur la nature de certaines autres données et qui permet ainsi leur utilisation pertinente.

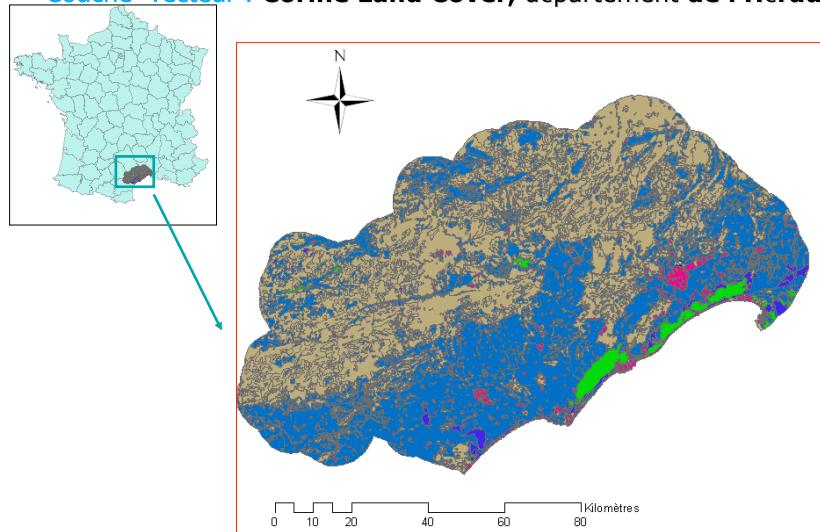
Spécificité de l'information géoréférencée

En plus des descripteurs classiques (identification, qualité, contenu, format)

- Informations sur la dimension spatiale :
 - emprise du jeu de données (toponymie, rectangle englobant)
 - système de référence
 - mode de représentation

Les métadonnées : un support au catalogage

Couche vecteur : Corine Land Cover, département de l'Hérault .

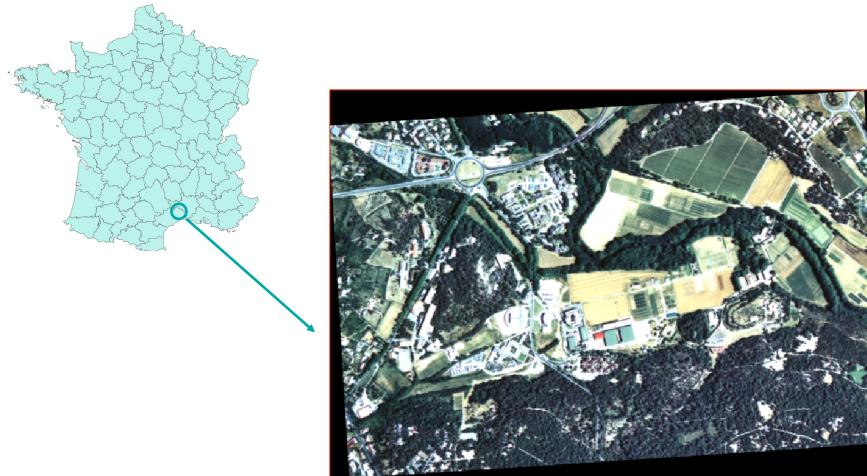


Les métadonnées : un support au catalogage

Identification du jeu de données	
Résumé	Corine Land Cover est un inventaire homogène de l'occupation du sol au niveau européen caractérisé par des choix techniques spécifiques : une méthodologie de photo-interprétation d'images satellites assistée par ordinateur ; une échelle de travail au 1:100 000 ; un seuil minimal des unités cartographiées de 25 hectares ; Une nomenclature en 3 niveaux et 44 postes.
Titre	Corine LandCover – Extrait pour le département de l'Hérault
Thèmes généraux	Carte de référence de la surface terrestre Environnement
Mots clés	Occupation des terres Ecosystèmes paysage
Date de création	1996
Type de représentation	vecteur
Extension géographique du jeu de données (degrés décimaux)	
Toponyme	Hérault
Format	fichier de forme shape
Version du format	8.3
Représentation spatiale du jeu de données	
Topologie	Géométrie seulement
Type de géométrie	polygone
Nombre d'objets	4755
Système de référence spatial du jeu de données	
Système de projection	NTF Lambert Carto II

Les métadonnées : un support au catalogage

Image : simulation SPOT 5, Maison de la télédétection.



Les métadonnées : un support au catalogage

Identification du jeu de données	
Résumé	Image de simulation du capteur multispectral SPOT 5 sur le site de la Maison de la Télédétection
Titre	Simulation SPOT 5 - vue de la maison de la télédétection
Thèmes généraux	localisation
Mots clés	Centre de recherche Centre de formation Télédétection
Date de création	2005
Type de représentation	raster
Taille du pixel (mètres)	2.5
Extension géographique du jeu de données (degrés décimaux)	
Toponyme	Montpellier
Format	Raster Dataset - IMG
Version du format	Inconnu
Représentation spatiale du jeu de données	
Nombre de dimension (axe)	2
Nom de l'axe	Axe X
Nombre de ligne	646
Taille du pixel (mètres)	2.5
Nom de l'axe	Axe Y
Nombre de ligne	449
Taille du pixel (mètres)	2.5
Géométrie de la cellule	Pixel
Système de référence spatial du jeu de données (3 ^{eme} rubrique)	
Système de projection	NTF Lambert Carto II

Les métadonnées : un support au catalogage

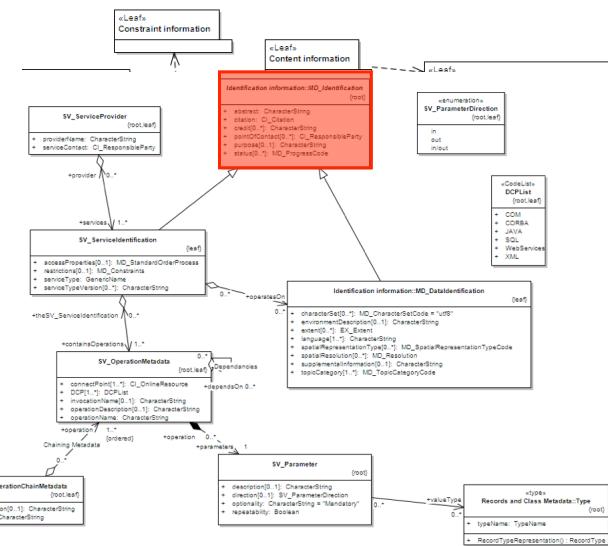
Intérêt des métadonnées

- Rôle du producteur
 - Gérer, faire connaître et valoriser en interne et en externe le patrimoine d'information (et l'expertise associée)
 - ... tout en maîtrisant l'accès à ces informations
- Rôle de l'utilisateur
 - Connaître l'existence de données, évaluer leur qualité (/ à un usage donné)

Standards de métadonnées

ISO 19115:2003 Geographic Information - Metadata

ISO 19119:2005 -
Geographic Information - Services



Standards de métadonnées

ISO 19115:2003 Geographic Information - Metadata

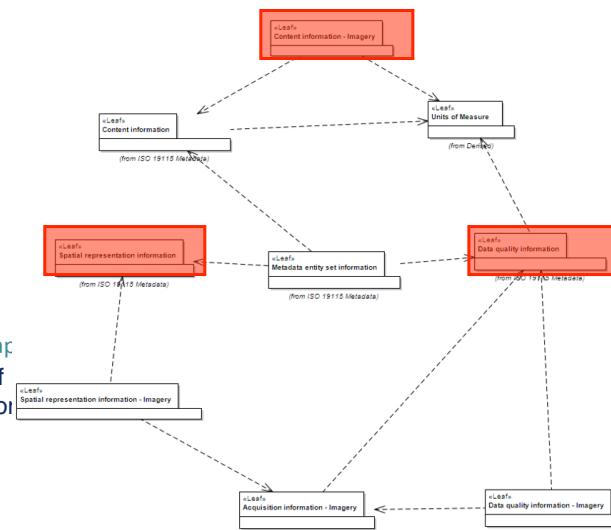
	Name / Role Name	Short Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
344.	westBoundLongitude	westBL	western-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east)	M	1		Angle (B.4.3) -180.0 <= West Bounding Longitude Value <= 180.0
345.	eastBoundLongitude	eastBL	eastern-most coordinate of the limit of the dataset extent, expressed in longitude in decimal degrees (positive east)	M	1	Class	Angle (B.4.3) -180.0 <= East Bounding Longitude Value <= 180.0
346.	southBoundLatitude	southBL	southern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north)	M	1	Class	Angle (B.4.3) -90.0 <= South Bounding Latitude Value < 90.0; South Bounding Latitude Value <= North bounding Latitude Value
347.	northBoundLatitude	northBL	northern-most coordinate of the limit of the dataset extent, expressed in latitude in decimal degrees (positive north)	M	1	Class	Angle (B.4.3) -90.0 <= North Bounding Latitude Value <= 90.0; North Bounding Latitude Value >= South Bounding Latitude Value
348.	EX_GeographicDescription	GeoDesc	description of the geographic area using identifiers	Use obligation/condition from referencing object	Use maximum occurrence from referencing object	Specified Class (EX_GeographicExtent)	Line 349 and 340
349.	geographicIdentifier	geoid	Identifier used to represent a geographic area	M	1	Class	MD_Identifier (B.2.7.3)

extrait ISO 19115:2003

Standards de métadonnées

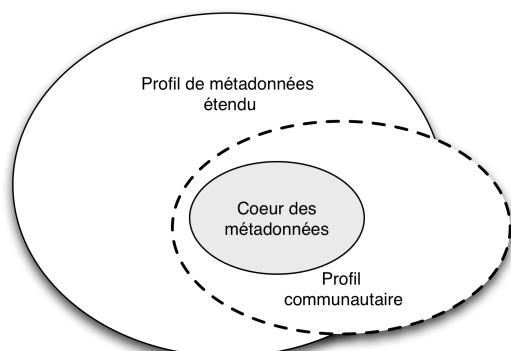
ISO 19115-2:2005 - Geographic Information - Imagery

ISO/TS 19139:2007 - Geograph XML encoding of specification for



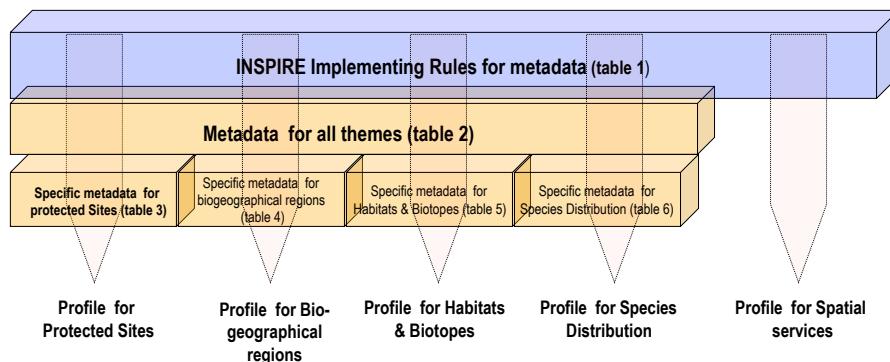
Profil de métadonnées

Définition : Un profil de métadonnées peut être considéré comme une spécialisation du standard. Une communauté d'usagers peut donc sélectionner des éléments obligatoires dans un profil de métadonnées et ajouter des éléments additionnels non standard.



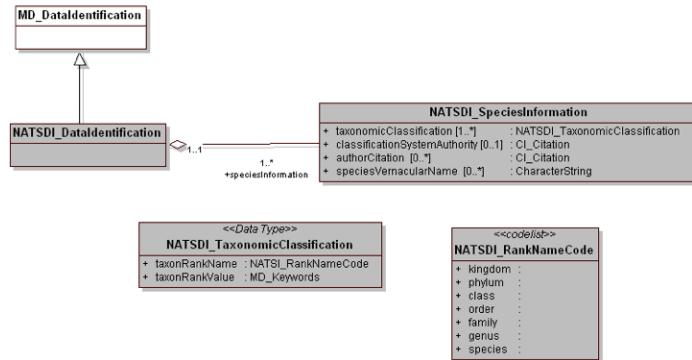
Profil de métadonnées : exemple NatureSDI plus

5 profils de métadonnées définis correspondants aux thèmes des annexes INSPIRE



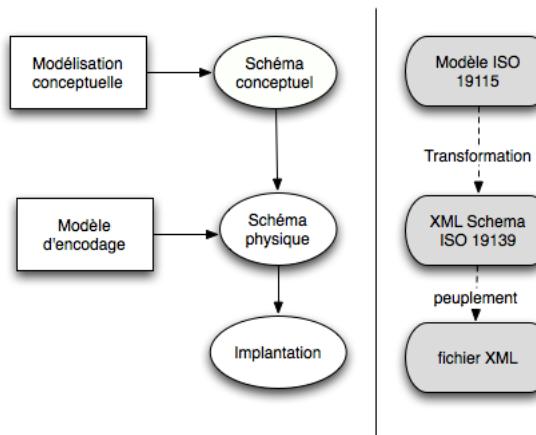
Profil de métadonnées : exemple d'extension ISO

- Pour le thème Distribution des espèces :
 - Identifier dans les métadonnées les espèces ou classes d'espèces décrites dans le jeu de données
 - Construire l'index du catalogue sur le nom des espèces et leur rang taxonomiques



Extension de la classe DataIdentification ISO 19115.

Du modèle conceptuel au format d'échange



Exemple de métadonnées standardisées

Métadonnées au format XML ISO 19119 (ISO 19115)

```

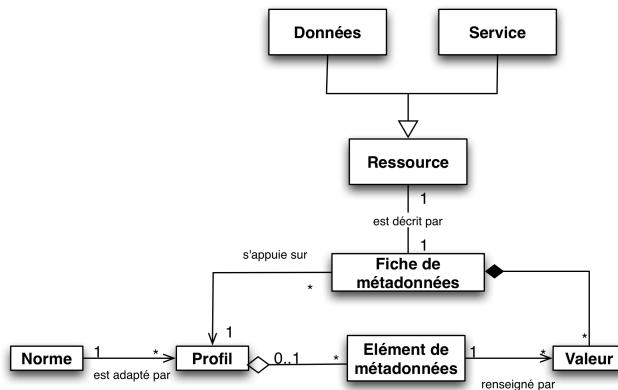
<fra:FRA_DataIdentification gco:isoType='MD_DataIdentification'>
  <gmd:abstract><gco:CharacterString>Image de simulation du capteur multispectral SPOT 5 sur le site de la Maison de la Télédétection</gco:CharacterString>
  <gmd:credit><gco:CharacterString></gco:CharacterString></gmd:credit>
  <gmd:status><gco:CharacterString></gco:CharacterString></gmd:status>
  <gmd:fileIdentifier><gco:CharacterString>simul_spot5_98.img.1</gco:CharacterString></gmd:fileIdentifier>
  <gmd:parentIdentifier><gco:CharacterString></gco:CharacterString></gmd:parentIdentifier>
  <gmd:language><gco:CharacterString>fra</gco:CharacterString></gmd:language>
  <gmd:characterSet><gmd:MD_CharacterSetCode codeList="MD_CharacterSetCode" codeListValue="8859part1" /></gmd:characterSet>
  <gmd:metadataStandardName><gco:CharacterString>ISO 19115 FRA 1.0</gco:CharacterString></gmd:metadataStandardName>
  <gmd:metadataStandardVersion><gco:CharacterString>Image satellitaire - Photos aériennes</gco:CharacterString></gmd:metadataStandardVersion>
</gmd:MD_DataIdentification>

<gmd:MD_GridSpatialRepresentation>
  <gmd:numberOfDimensions><gco:Integer>002</gco:Integer></gmd:numberOfDimensions>
  <gmd:axisDimensionsProperties>
    <gmd:MD_Dimension>
      <gmd:dimensionName><gmd:MD_DimensionNameTypeCode codeList="MD_DimensionNameTypeCode" codeListValue="column" /></gmd:dimensionName>
      <gmd:dimensionSize><gco:Integer>646</gco:Integer></gmd:dimensionSize>
      <gmd:resolution><gco:Measure>2.5</gco:Measure></gmd:resolution>
    </gmd:MD_Dimension>
    <gmd:MD_Dimension>
      <gmd:dimensionName><gmd:MD_DimensionNameTypeCode codeList="MD_DimensionNameTypeCode" codeListValue="row" /></gmd:dimensionName>
      <gmd:dimensionSize><gco:Integer>449</gco:Integer></gmd:dimensionSize>
      <gmd:resolution><gco:Measure>2.5</gco:Measure></gmd:resolution>
    </gmd:MD_Dimension>
  </gmd:axisDimensionsProperties>
  <gmd:cellGeometry><gmd:MD_CellGeometryCode codeList="MD_CellGeometryCode" codeListValue="area" /></gmd:cellGeometry>
  <gmd:transformationParameterAvailability><gco:Boolean>false</gco:Boolean></gmd:transformationParameterAvailability>
</gmd:MD_GridSpatialRepresentation>
<gmd:spatialRepresentationInfo>
<fra:FRA_DirectReferenceSystem gco:isoType='MD_ReferenceSystem'>
  <gmd:referenceSystemIdentifier>
    <gmd:RS_Identifier>
      <gmd:code><gmx:Anchor xlink:href='27582'>NTF Lambert II Carto</gmx:Anchor></gmd:code>
    </gmd:RS_Identifier>
  </gmd:referenceSystemIdentifier>
</fra:FRA_DirectReferenceSystem>
</gmd:spatialRepresentationInfo>

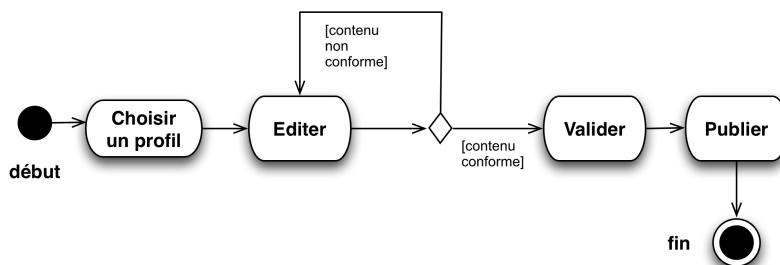
```

Notions de fiches de métadonnées

- Définition :** Une fiche de métadonnées s'appuie sur un profil et correspond à l'ensemble des valeurs qui renseignent les éléments de métadonnées de ce profil



Cycle de vie d'une fiche de métadonnées



Cycle de vie d'une fiche de métadonnées
(diagramme d'activité UML)

En résumé : Pourquoi cataloguer ?

- Pour le producteur

- Décrire de manière rigoureuse leur contenu ainsi que les conditions d'accès/utilisation
- Faire connaître ses données en interne et à l'extérieur
- Répondre aux exigences réglementaires

Permet une valorisation de ses données...

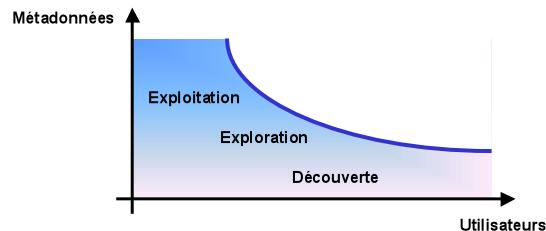
Mais très chronophage

- Nécessite une saisie manuelle car peu intégré au processus de production des données
- Les métadonnées doivent être mises à jour de manière concomitante aux données

En résumé : Pourquoi cataloguer ?

- Pour le consommateur

- Localiser la ressource
- Identifier et évaluer la ressource selon ses besoins > **Métadonnées de découverte**
- Explorer les caractéristiques de la ressource > **Métadonnées d'exploration**
- Utiliser, exploiter la ressource > **Métadonnées d'exploitation**



En résumé : Qui ? : les acteurs du catalogage

- Les producteurs

- Théoriquement au moment de la création de la donnée

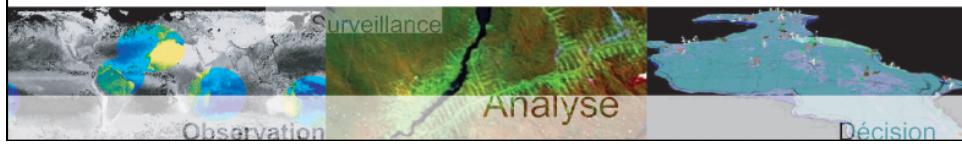
- Les « prosumers »

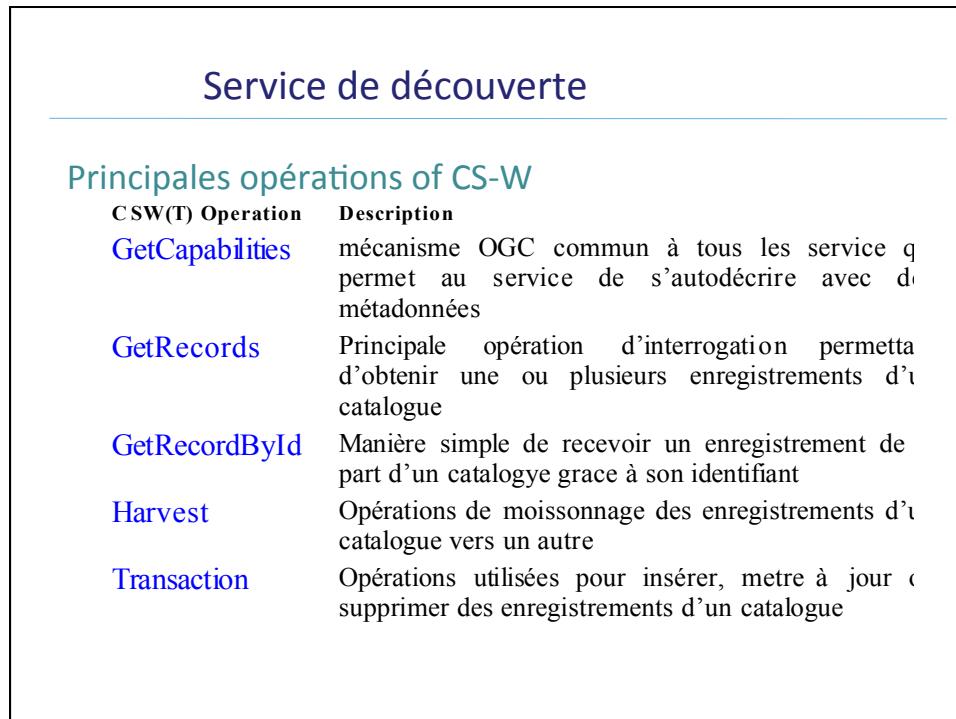
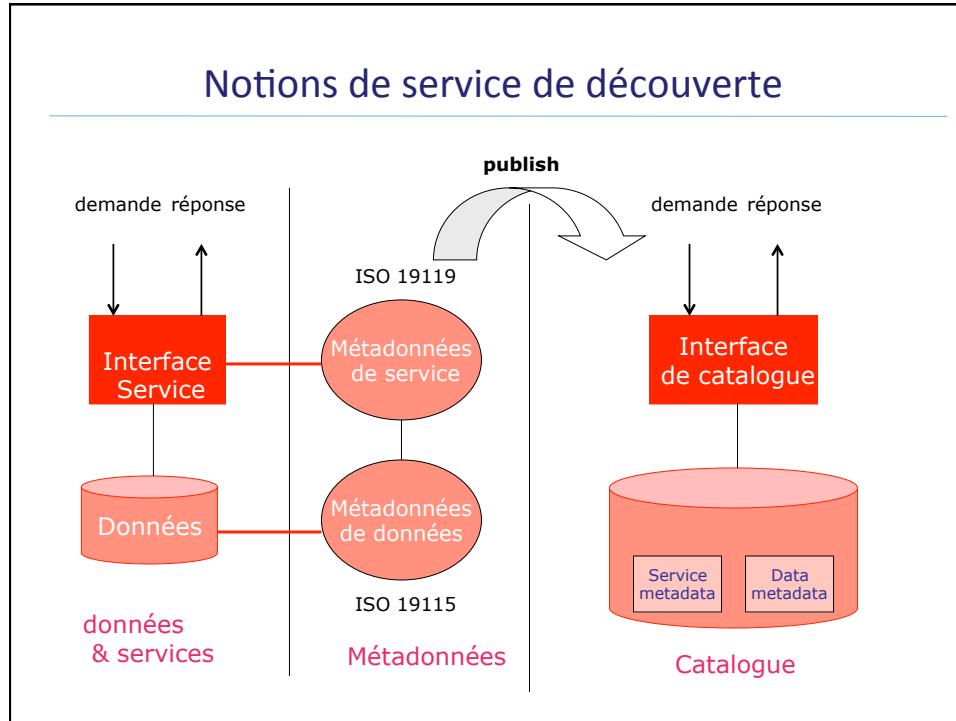
- Après avoir traiter (« transformer ») une donnée acquise

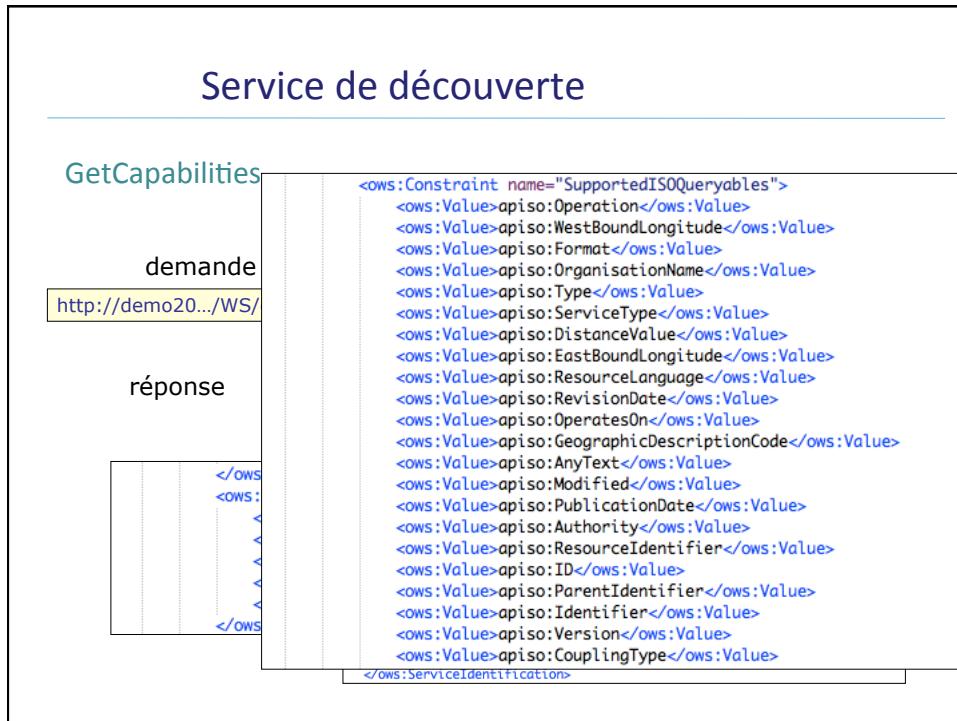
En résumé :Les enjeux et les écueils du catalogage

- Enjeux : Fournir une métadonnée de qualité
 - Efficace
 - Fraîche
 - Complète
 - Conforme
- Écueils : une processus de production délicat
 - Peu ou pas intégré au processus de production des données
 - Outils de catalogage faiblement couplés
 - Nécessite une saisie manuelle
 - Les métadonnées doivent être mises à jour de manière concomitante aux données

Service de découverte







Interrogation des métadonnées

- ◆ **Demande de l'utilisateur**
« Je cherche de le territoire de l'Appellation d'Origine Contrôlée Pélardon »
 - ◆ **Expression au sein du client de recherche**
 - ◆ champ **Quoi ?** : « Pélardon »

Interrogation des métadonnées

Requête GetRecords

Avec une expression CQL de sélection sur le titre

```
http://demo.geomatys.com/mdweb/WS/csw?
Request=GetRecords&SERVICE=CSW&Version=2.0.2&TYPENAMES=g
md:MD_Metadata&RESULTTYPE=results&ELEMENTSETNAME=full&OU
TPUTSCHEMA=http://www.isotc211.org/2005/
gmd&CONSTRAINTLANGUAGE=CQL_TEXT&CONSTRAINT_LANGUAGE_
VERSION=1.1.0&CONSTRAINT=apiso:Title==%27P%C3%A9lardon
%27
```

Exemple de réponse du service

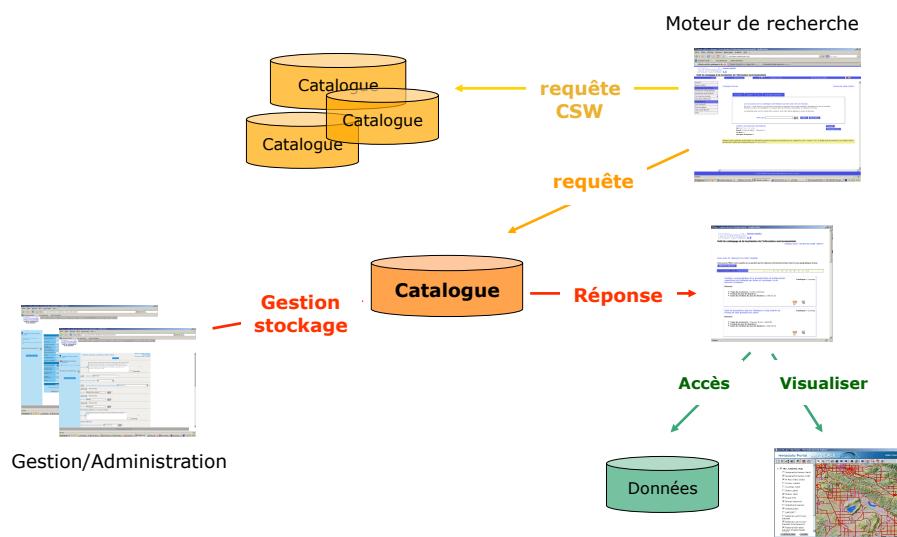
```
<csw:GetRecordsResponse version="2.0.2">
<csw:SearchStatus timestamp="2009-11-16T20:45:05.130+01:00"/>
- <csw:SearchResults nextRecord="0" numberOfRecordsReturned="3" numberOfRecordsMatched="3" elementSet="summary">
- <csw:SummaryRecord>
- <dc:identifier>
    mdweb_2_catalog_CSW Data Catalog_profile_inspire_core_dataset_vector_12
</dc:identifier>
- <dc:title>
    Lista de lugares de Importancia Comunitaria (LIC). Nacional
</dc:title>
<dc:type/>
<dc:subject>MEDIO AMBIENTE</dc:subject>
<dc:subject>áreas naturales</dc:subject>
<dc:subject>espacios protegidos</dc:subject>
<dc:subject>espacios protegidos</dc:subject>
<dc:format/>
<dct:modified>2009-10-02</dct:modified>
- <dct:abstract>
    La Directiva de Hábitat 92/43/CEE tiene por objetivo principal el mantenimiento de la biodiversidad. Esta norma comunitaria obliga a entregar una Lista Nacional de lugares (pLIC), la cual, en sucesivas fases, se transformará en Lista de Lugares de Importancia Comunitaria de Conservación (ZEC). Tales ZEC, junto con las Zonas de Especial Protección para las Aves (ZEPA), de la Directiva 79/409/CEE, o la Lista Nacional de lugares está estructurada en cuatro regiones biogeográficas (alpina, atlántica, mediterránea y macaronésica) y la propuesta territorial a la Dirección General para la Biodiversidad (DGB) del Ministerio de Medio Ambiente, quien actúa como coordinador de la transmisión oficial a la Comisión Europea. La lista macaronésica fue aprobada por Decisión de la Comisión de 28 de Diciembre de 2003 de la Comisión Europea con fecha de 22 de Diciembre de 2003. La lista atlántica fue aprobada por Decisión de la Comisión Europea con fecha de 19 de Julio de 2006. Una vez culminada la declaración de protección, cada Estado deberá designar las ZEC (fase de aplicación de medidas de gestión y conservación para hábitat y especie que se encuentran en formato ARCE y Dxf)
</dct:abstract>
```

Les outils de catalogage

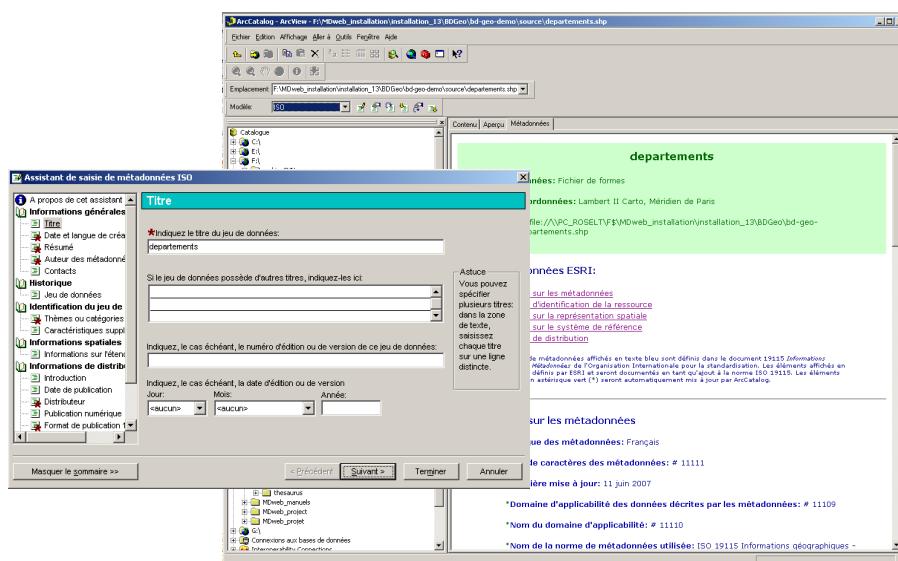
Quelques exemples



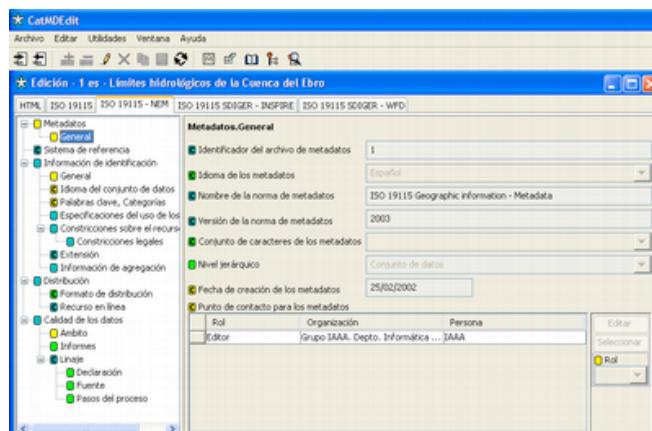
Fonctions essentielles d'un outil de catalogage



Quelques exemples : ArcCatalog (Esri)



Quelques exemples : CatMDedit (IGN).



Quelques exemples : ESRI GEOportal

The screenshot shows the ESRI GEOportal homepage. It features a search bar at the top right, followed by links for HOME, BROWSE, GALLERY, and SEARCH. Below the search bar is a message: "Records shown from: This Site Click here to select different site or configure search." Under "Additional Options", there are radio buttons for "Anywhere", "Intersecting", and "Fully within". A "WHERE" section contains a map of the world with several points marked. A note below says: "This Geoportal was built using the Geoportal Si...". On the right side, there is a "GEOVIEWER" section with a "Catalog" tree view. The tree includes categories like Applications, Clearingshouses, Documents, Downloadable Data, Geographic Activities, Geographic Services, Live Map Services, Map Files, Offline Data, Static Map Images, ISO Topic Category, Administrative and Political Boundaries, Agriculture and Farming, Atmospheric and Climatic, Biology and Ecology, Business and Economic, Cadastral, Cultural, Society and Demographic, Elevation and Derived Products, Environment and Conservation, and Facilities and Structures. A specific entry for "Web Map Service nasa" is expanded, showing its details and a preview link.

Quelques exemples : GeoNetwork (FAO)

The screenshot shows the FAO GeoNetwork interface. At the top, it displays a search result for "Major dominant soil units of Africa (1993)". The results page includes a map of Africa with color-coded soil units, and links for "Metadata", "Download", "Create Map", and "View in Google Earth". Below this, another search result for "Major climates of Africa" is shown, also with a map and similar download links. To the right, there is a "CREATE" tool for map compositions, which allows users to "Build your own map composition" by selecting base layers (e.g., Major Cities, Rivers of the World, Inland Water (perennial), Coastal Lines) and thematic layers (e.g., core). The bottom of the screen shows a taskbar with multiple open windows, including Microsoft Office applications and a terminal window.

Quelques exemples : Géosource

The screenshot shows the Géosource software interface running in Mozilla Firefox. The title bar reads "Géosource - Outil de gestion des métadonnées cartographiques - Mozilla Firefox". The main window displays a search interface for geospatial metadata records. It includes fields for simple and advanced search, and a sidebar with administrative links like "Accueil", "Contact", "Liens", "A propos", and "Aide". A "LOG-IN" section is visible on the right. The bottom of the window shows a footer with links to "Aide et Support", "Derniers ajouts", "RSS", and "GeoRIS". The status bar at the bottom of the browser window shows various tabs and icons.

MDweb

Pour en savoir plus sur les outils ...

Panorama

Blog Neogeo : <http://www.neogeo-online.net/blog/archives/517/>

FGDC : <http://www.fgdc.gov/metadata/iso-metadata-editor-review>

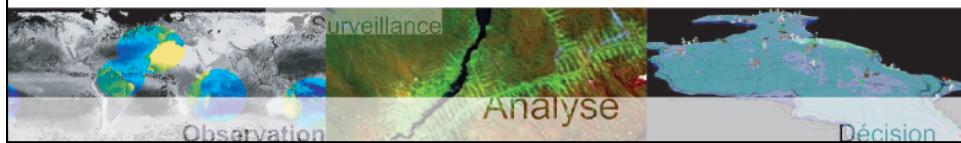
Comparatifs

- IRD DSI : http://www.ird.fr/informatique-scientifique/outils/outils_is/catalogages/
- Forum SIG : www.forumsig.org/showthread.php?t=19745
- Géorezo : http://georezo.net/wiki/main/logiciels/comparatif_metadatasoft

Sites

- MDweb : <http://www.mdweb-project.org/>
- GeoNetWork : <http://geonetwork-opensource.org/>
- Géosource : <http://www.geosource.fr/>
- CatMDedit : <http://catmdedit.sourceforge.net/>
- ESRI Geoportal : <http://www.esri.com/software/arcgis/geoportal/>
- Expire : <http://sourceforge.net/projects/expire/>
- eXcat : <http://www.neonet.nl/gdsc/en/tools/excat>
- M3Cat : <http://www.intelec.ca/html/fr/technologies/m3cat.html>
- EUOSME : <http://www.inspire-geoportal.eu/EUOSME/>

MDweb, outil libre de catalogage et de localisation

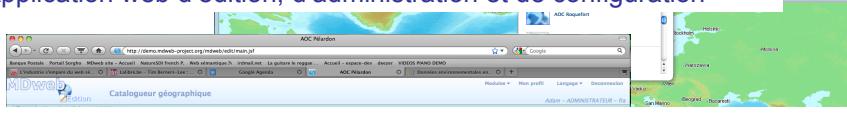


Un outil libre de catalogage et de localisation

- Un service de recherche des données géographiques

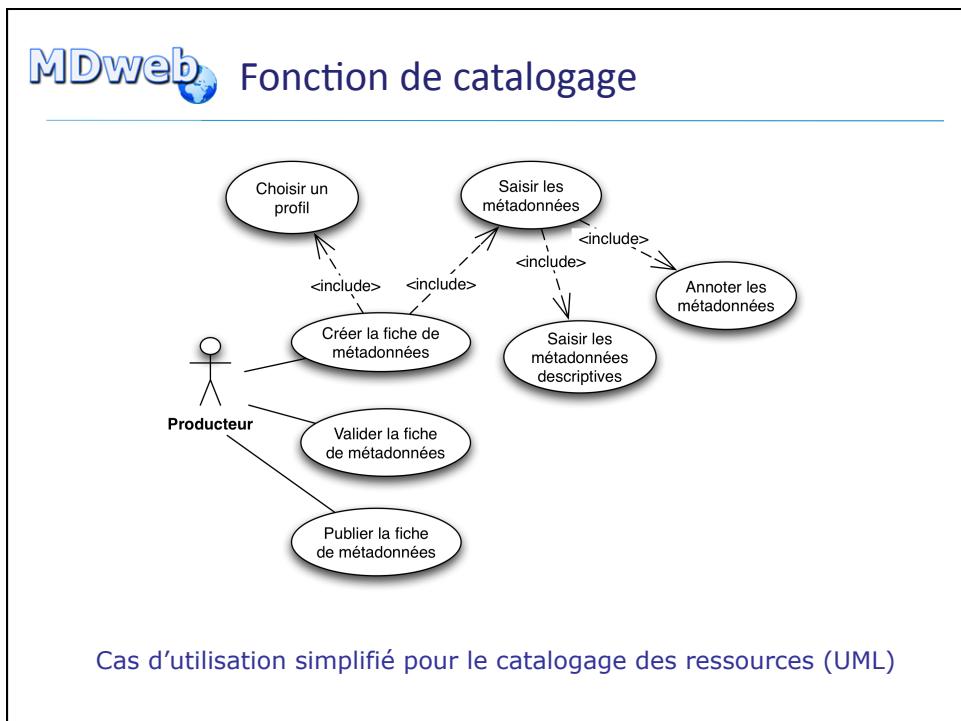
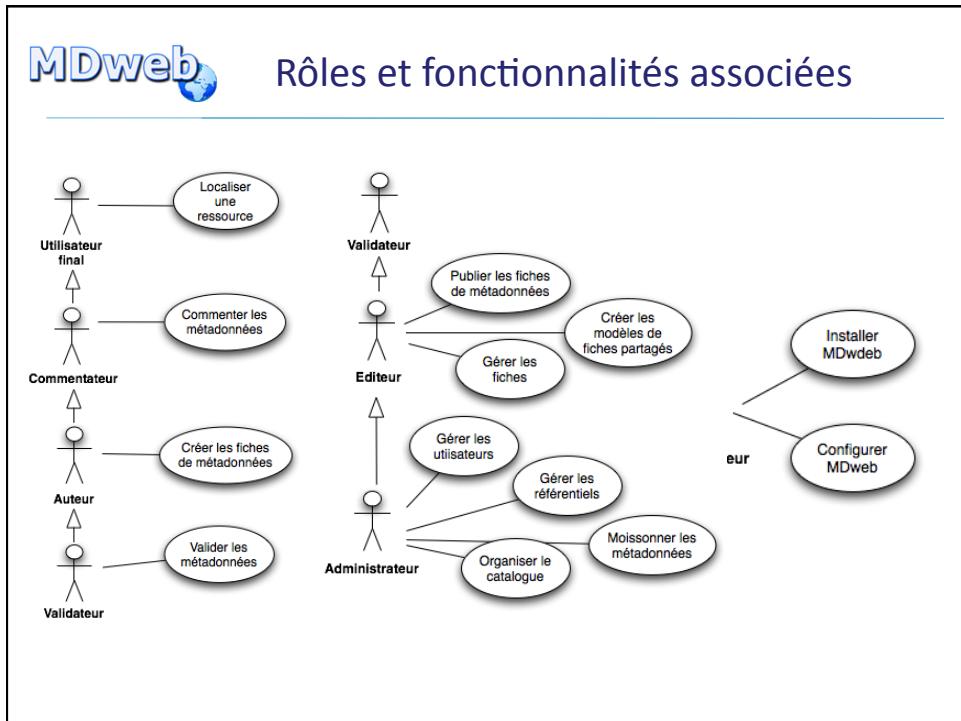


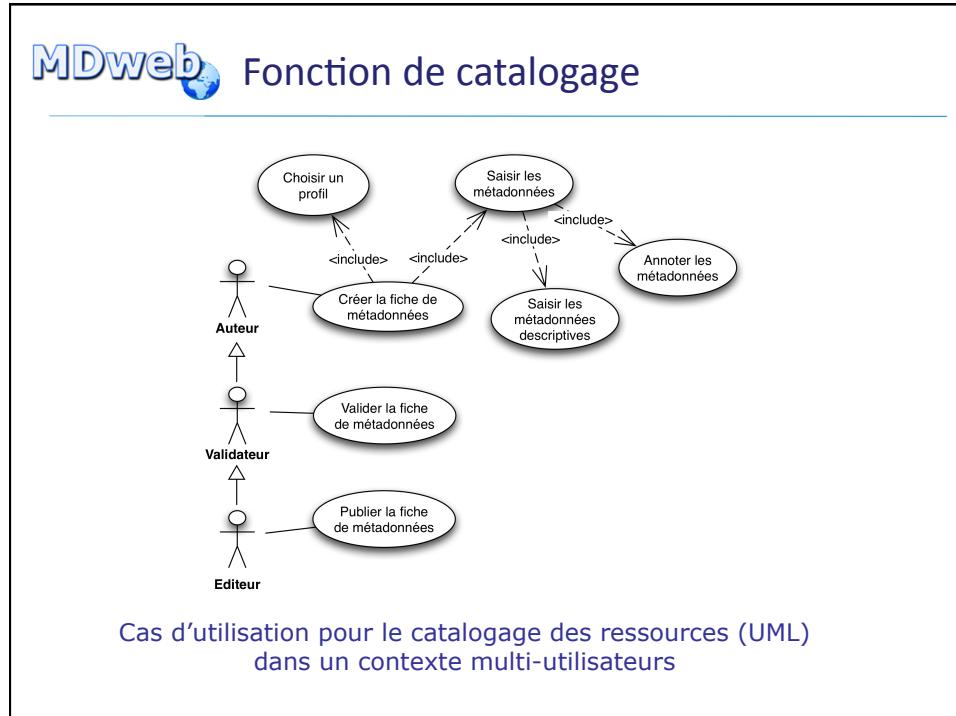
- Application web d'édition, d'administration et de configuration



- **Approche normalisée** par l'implémentation des standards ouverts (W3C, ISO, OGC) visant l'intéropérabilité IDE/SDI (EU,GEO)
Dans un contexte multi-langue



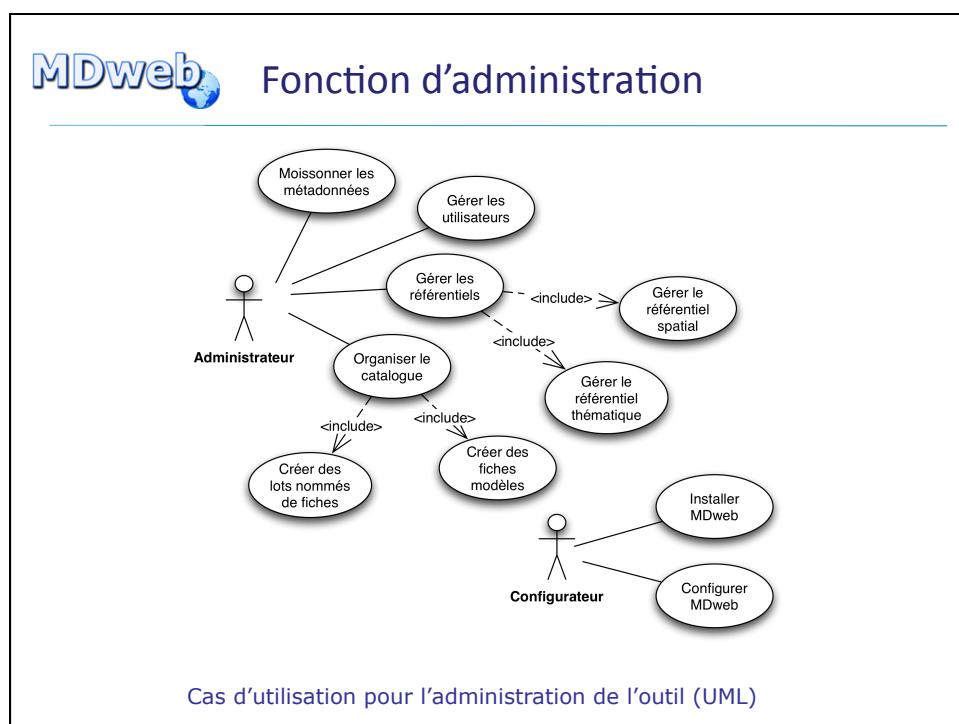
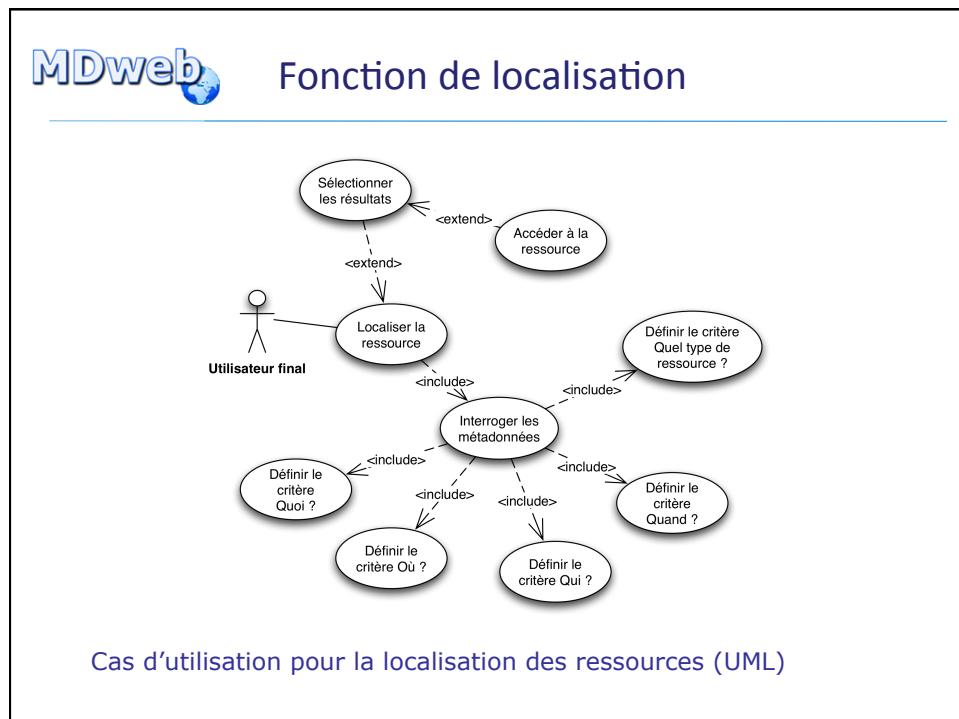


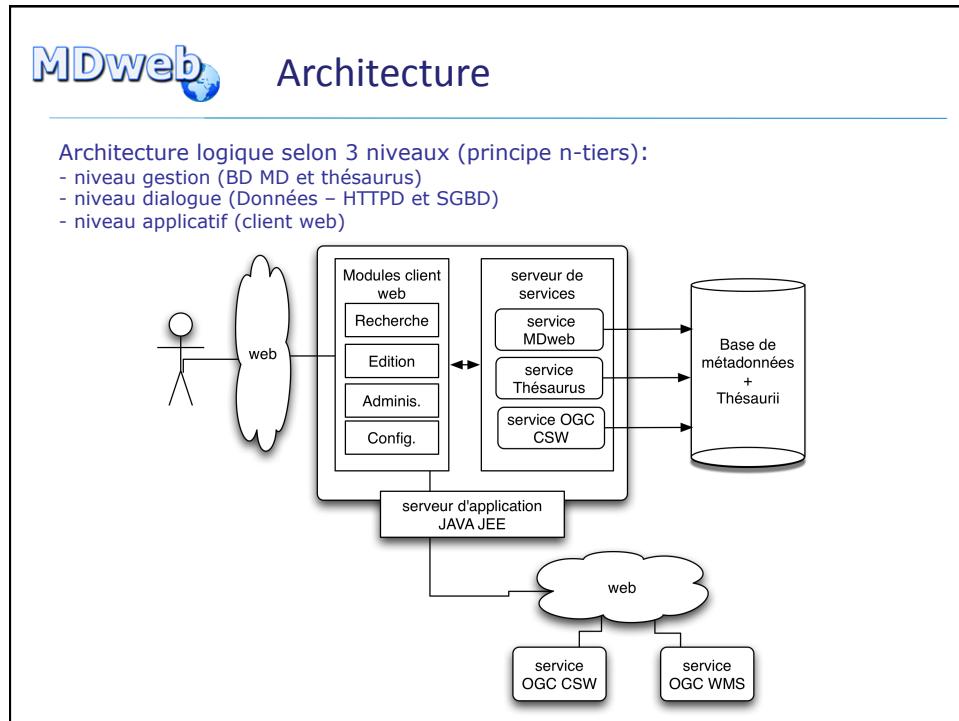


MDweb Extension à la fonction de catalogage

- Faciliter la saisie des métadonnées

Composant d'auto-complétion utilisé lors de la phase d'édition pour sélectionner les mots clés issus des théâtres embarqués assistez l'utilisateur dans la saisie de l'extension spatiale de la ressource







Un outil libre de catalogage et de localisation

Site web : <http://www.mdweb-project.org/>

Documentation :

- Guide d'installation :
<http://www.mdweb-project.org/files/documentation/2010.10/QuickStart-MDweb-2010.10-fr.pdf>
- Guide de l'utilisateur :
<http://www.mdweb-project.org/files/documentation/2010.10/MDweb-UserManual-2010.10-fr.pdf>
- TP d'installation :
http://www.mdweb-project.org/files/longterm/BDSpatiales/TP_Catalogage-installation.pdf
- TP – gestion des métadonnées :
http://www.mdweb-project.org/files/longterm/BDSpatiales/TP_Catalogage-edition-recherche.pdf

Site de démonstration : <http://demo.mdweb-project.org/mdweb/>