

Open source CeCILL

MDweb

Outil de catalogage et de localisation de l'information environnementale

Notions générales à destination du petit catalogueur

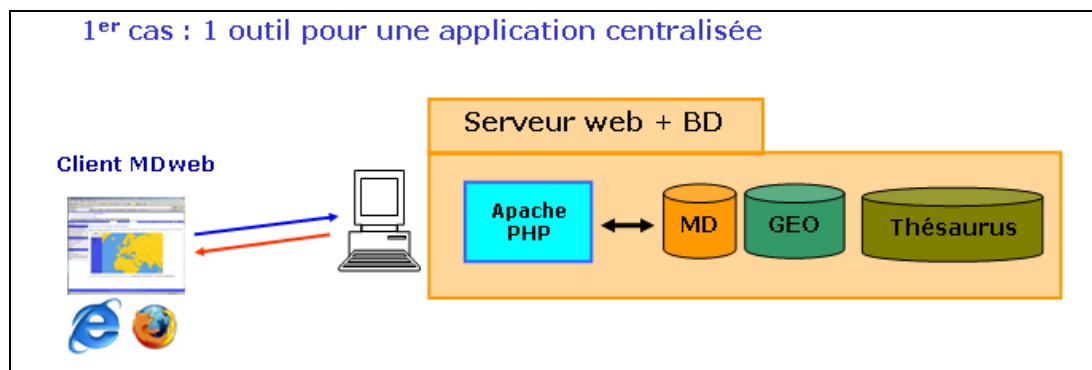
Version	Commentaires
1	Auteurs : Jean Christophe Desconnets (jcd@teledetection.fr) Date de création : 19/06/2007 22:31:00

Qu'est ce que MDweb ?	2
Notion de rôles et fonctionnalités associées	4
Notion de gabarit ou metadata profile.....	5
Notion de ressource	6
Notion de fiches de métadonnées ou références.....	9
Notion de niveau de saisie d'une fiche de métadonnées.....	10

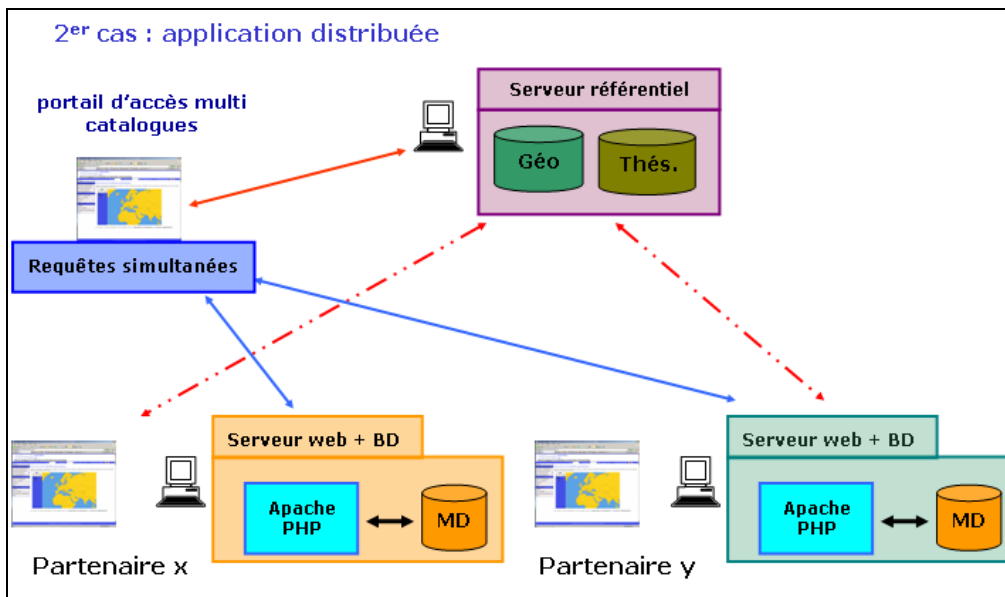
Qu'est ce que MDweb ?

MDweb est un outil générique, multi-langues, multi-normes, destiné au catalogage et à la localisation des données et documents (spatiales et non spatiales). Cette application est conçue pour constituer, gérer, administrer et consulter des catalogues de données via le web et d'accéder si besoin aux données et documents référencés. MDweb s'appuie sur les standards internationaux de métadonnées (ISO 19115 et 19139) et de communication (OGC – Catalogue service Specification – Z39.50) qui permettent d'assurer des interrogations entre catalogues. Les spécifications du web service CSW-2 seront implémentées dans la future version. De fait, MDweb répond aujourd'hui aux besoins de fournitures de métadonnées dans le contexte des infrastructures de données spatiales telles que spécifiées au niveau français (initiatives géocatalogue et géoportail) et européen (Directives INSPIRE).

La généricité et l'originalité de l'outil repose, d'une part, sur sa base de données qui permet la prise en compte de l'évolution des normes en offrant la possibilité d'insérer de nouvelles normes, de modifier ou d'étendre celles existantes sans pour autant remettre en cause le schéma et les traitements. Il en est de même en ce qui concerne les éléments de saisie des données.



Architecture 3 tier centralisée



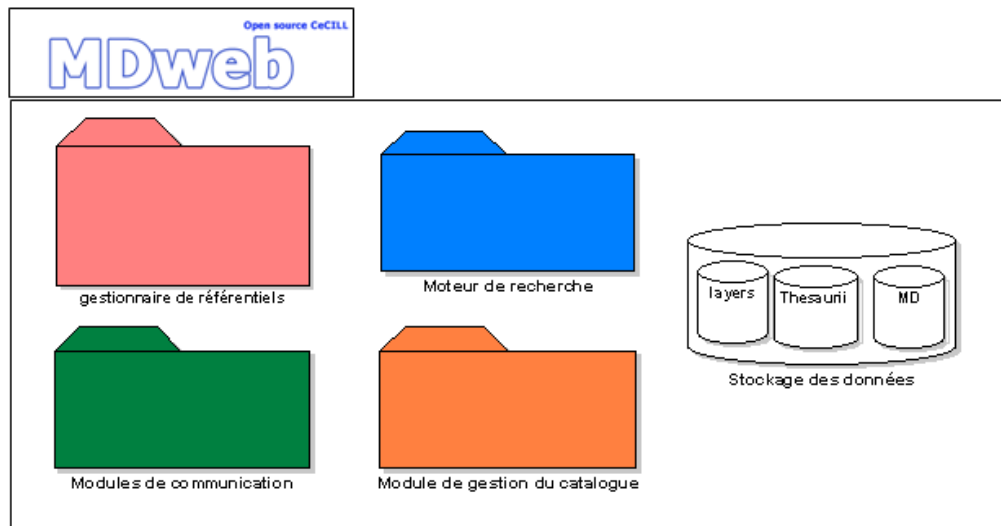
Architecture 3 tier avec référentiels distribués

MDweb permet, à travers la notion de profil de métadonnées, de choisir et spécifier les propriétés des éléments de métadonnées qui seront utilisés pour la documentation d'une référence. Enfin, afin d'améliorer les performances du moteur de recherche et d'offrir à l'utilisateur des possibilités de reformulations des interrogations, une base de données de thésaurus (support à la saisie des mots clés) et une base de données spatiales (support à la recherche par sélection spatiale) complète l'outil et sont personnalisable en fonction du contexte d'utilisation de l'outil. Les liens entre termes du thésaurus sont notamment utilisés pour proposer l'expansion ou l'affinement d'une requête selon le niveau de réponse donné par le moteur de recherche.

MDweb est utilisé de manière opérationnelle dans des contextes institutionnels et thématiques variés, en France et à l'international. Nous en citons quelques uns : observatoire de la Désertification (ROSELT), gestion Intégrée des Zones Côtières (PRCM en Afrique de l'Ouest, SYSCOLAG en Languedoc-Roussillon), inventaire national sur l'environnement (SIA au Cap Vert), glaciologie-paléoclimatologie (Great-ICE IRD), hydrosystèmes (GIP Loire), risques naturels et développement territorial et dans d'un grand ensemble de services et de collectivités territoriales.

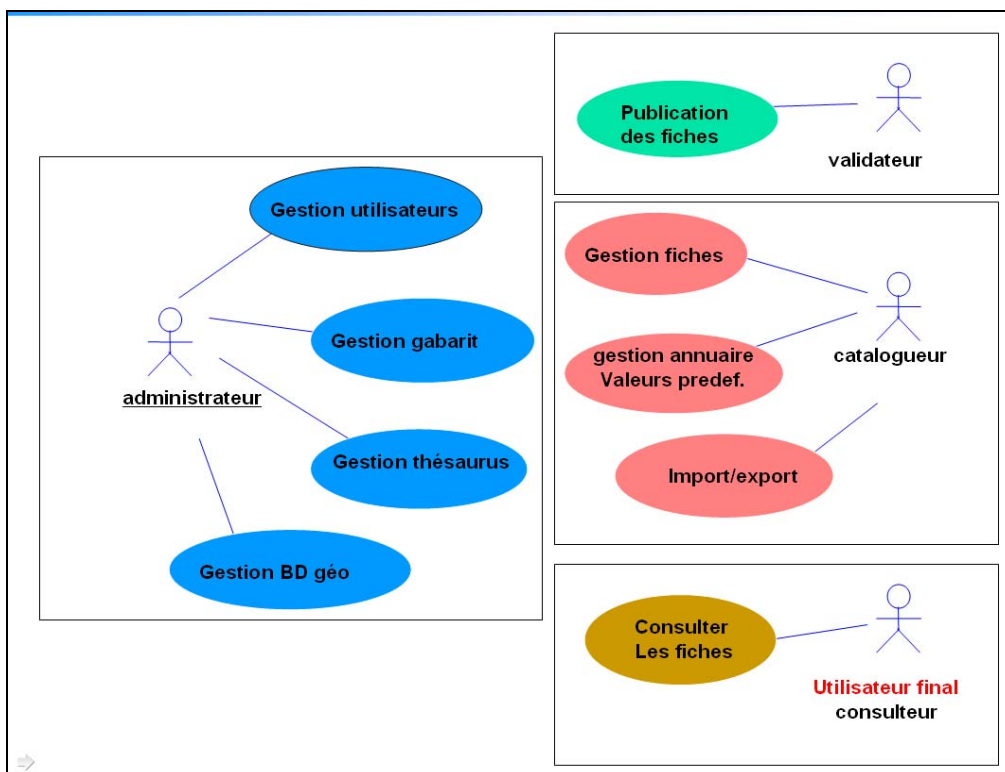
Dans sa version actuelle, MDweb 1.5 propose trois modules standard qui couvrent l'ensemble du cahier des charges fonctionnels :

- un module « de gestion » qui assure via une authentification par mot de passe la *saisie et la mise à jour de métadonnées* et des fichiers de données et images attachées aux métadonnées, *l'importation et l'exportation* de métadonnées au format XML, des jeux de valeurs prédéfinies pour automatiser la saisie de certains éléments (coordonnées, éléments techniques des normes)
- un moteur « de recherche » des métadonnées qui offre un mode de recherche par *sélection spatiale* (client cartographique WMS, WMC) par *type de données*, ou un mode de recherche avancée ou *recherche multi-critères*.
- un module « d'administration » ou **gestionnaire de référentiels** qui assure le paramétrage de l'outil, la gestion de toutes les métadonnées des catalogues existants, la personnalisation des interfaces, la gestion des gabarits, la gestion de la base de données spatiales (support au client WMS, WMC) et des thésaurus, la configuration du client cartographique (choix des couches et des styles, éditeur SLD)



Notion de rôles et fonctionnalités associées

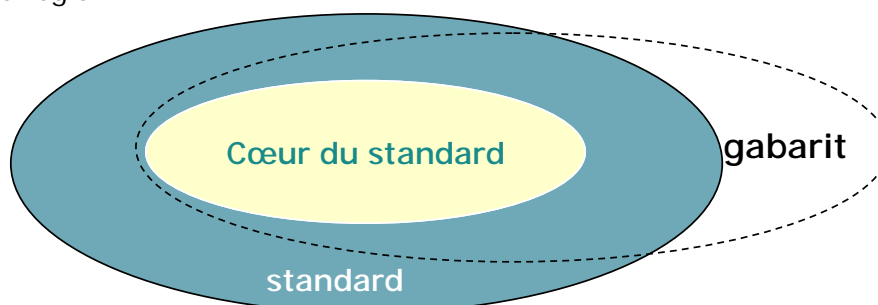
Par nature, MDweb est une application multi-utilisateur qui demande pour l'utilisation de ses différentes fonctionnalités la définition de rôles bien distincts. Chaque utilisateur se voit attribuer un compte pour lequel un rôle est donné par l'administrateur. Cela lui donne l'accès aux modules correspondants aux tâches qui lui sont assignées dans l'outil. quatre rôles sont distingués :



- l'**administrateur** est le super-utilisateur qui a accès à l'ensemble des modules pour assurer la gestion des catalogues, des utilisateurs et la configuration de l'outil.
- le **validateur** est un expert qui assurera la validation du contenu des références. Il sera donc garant de la qualité des références et des données attachées aux catalogues. Pour cela, ce type d'utilisateur a accès à l'ensemble des références d'un catalogue ou non par lui-même.
- le **catalogueur** est celui qui décrit les données. Cela doit être la personne la plus à même de décrire le contenu et les caractéristiques des données qu'il a produites. Ce type d'utilisateur a accès aux modules lui permettant de saisir, mettre à jour et gérer ses propres références.
- l'**utilisateur final** a accès au module de recherche - consultation du catalogue. Deux cas de figures sont possibles. Si il est identifié comme utilisateur privilégié, il pourra à travers une authentification, avoir accès au module de consultation privée. Dans le cas contraire, l'utilisateur a un accès au module de recherche - consultation public (sans authentification)

Notion de gabarit ou metadata profile

Un gabarit, ou encore adaptation, est un document ou un schéma (au sens structure de données) qui précise les options de mise en place d'une norme afin de répondre à un besoin particulier. Par essence, un gabarit ne contredit pas la norme à laquelle il se réfère et n'introduit pas, en principe, de nouveaux concepts. Il décrit plutôt la norme ou une partie de celle-ci afin qu'elle puisse être mise en place et utilisée dans un contexte particulier. Par contre, des éléments n'existant pas dans la norme (éléments étendus) peuvent y être intégrés. Ils viennent compléter celle-ci en y apportant des éléments de description utiles dans le contexte précis pour lequel est utilisé le gabarit. Un gabarit d'une norme permet, en outre, d'adapter culturellement et linguistiquement une norme internationale aux particularités d'un pays ou d'une région.



Une communauté peut ainsi définir des gabarits pour des types particuliers de jeux de données. Par exemple, un gabarit pour des jeux de données matricielles ou « raster » ne conservera que les éléments de métadonnées propres à ce type de données. Un gabarit peut aussi gérer certaines particularités ou règles qu'une organisation aimerait appliquer aux

éléments de métadonnées. Un gabarit, par exemple, pourra identifier les éléments obligatoires et optionnels à renseigner.

Dans MDweb, 9 gabarits sont proposés en standard. Ils correspondent à 9 types de données :

• **De type collection de données :**

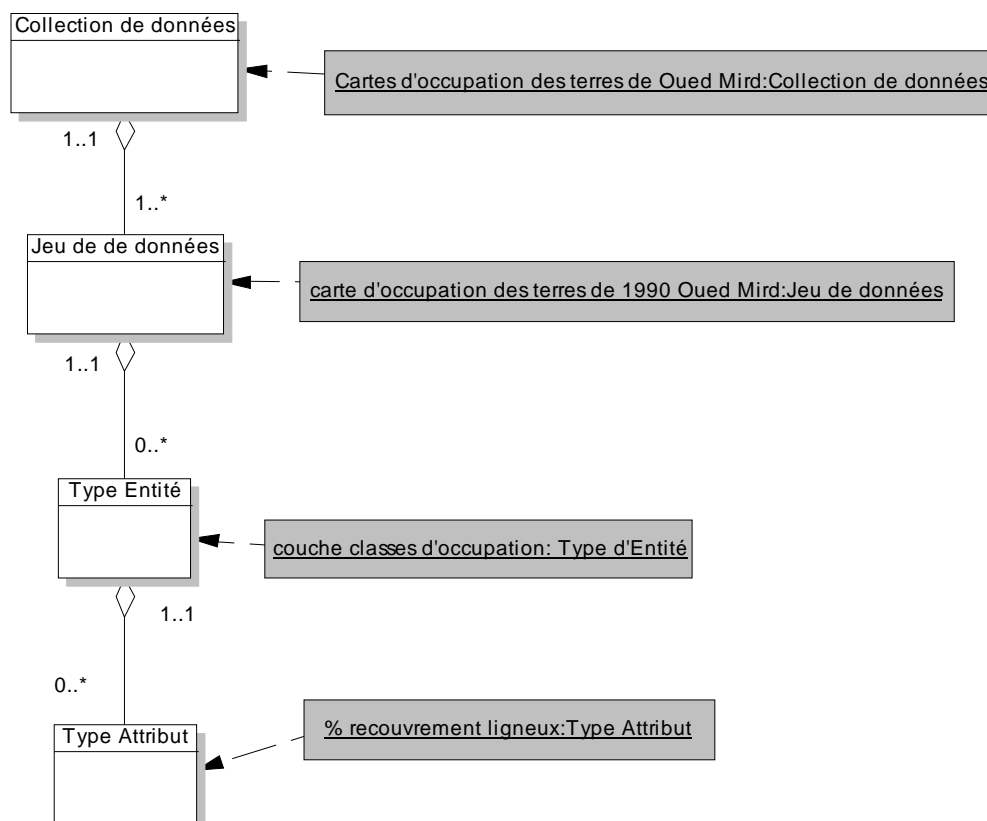
- BD géographique ou géodatabase
- Base de données temporelles
- Carte numérique

De type jeu de données

- Carte papier
- Données vecteur
- Image satellitale – photos aériennes
- Documents texte
- Tableaux de données
- Référence bibliographique

Notion de ressource

Les normes de métadonnées, et le standard international en particulier, s'appliquent en principe à des données numériques mais peuvent aussi s'appliquer à des documents analogiques tels que les cartes, plans, photographies aériennes, etc. Dans ce dernier cas, la documentation de la donnée et son catalogage font toujours référence au document lui-même. De plus, les jeux de données de ce type comprennent en général un ensemble bien identifiable de documents. En revanche, pour les données numériques, la définition de ce que constitue une donnée, ou un jeu de données, est plus difficile et reflète souvent le contexte institutionnel et technologique de l'organisation de laquelle est issue la donnée. De façon générale, on se rend compte que les données numériques peuvent être décomposées en une hiérarchie présentée dans le schéma suivant (figure 4), allant des attributs à la collection de données, en passant par les types d'entités et les jeux de données. Pour définir plus simplement cette vue des données, le terme général de ressource peut être utilisé. Il recouvre toutes les notions associées à la hiérarchie des données présentée en figure.



Pour illustrer cette notion, nous avons pris comme exemple, en figure ci-dessus, les cartes d’occupation des terres dressées sur un territoire, et en particulier celles du territoire de Oued Mird (Maroc). Cette ressource de type «carte numérique» peut être décomposée selon la hiérarchie des ressources proposée. Cette décomposition hiérarchique est proposée sous l’angle du formalisme UML. Si on la considère au plus haut niveau, on peut parler d’une **collection de données**. Il s’agit en effet, d’un ensemble de cartes portant sur le même thème et produite à différentes périodes d’observations, celle de l’occupation des terres des années 90, celle qui concernera l’occupation des terres pour la décennie 2000, etc. Au niveau **jeu de données**, on considérera, un élément de cette collection, par exemple, la carte d’occupation des terres des années 90. Dans ce cas, le niveau inférieur ou **type d’entité** correspond à l’ensemble des couches thématiques qui composent la carte d’occupation des terres des années 1990. Dans notre exemple, nous avons pris la couche « polygone » des classes d’occupation des terres. D’autres couches telles que les villages peuvent faire partie de la carte. Enfin, le niveau le plus élémentaire, ou **type d’attribut**, est l’ensemble de propriétés de la couche « polygone », un exemple de type d’attribut est donné avec l’attribut « % de recouvrement de ligneux ».

Les niveaux traités dans MDweb se limitent à :

- la collection de données
- au jeu de données

Définitions

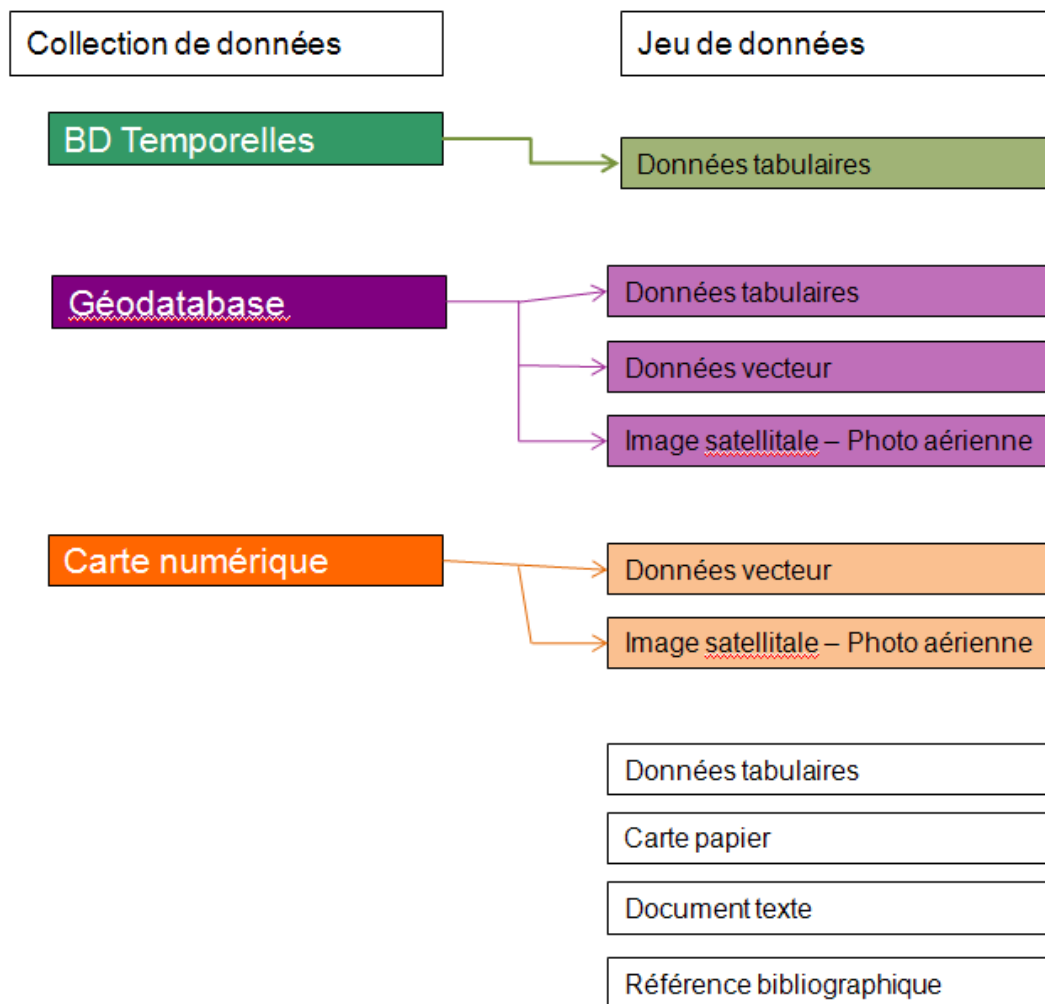
Collection de données : ensemble de jeux de données distincts reliés les uns aux autres par des caractéristiques communes telles que leur mode d’acquisition ou de

traitement (images satellites), leur extension spatiale, la nature de leur contenu, par exemple, une série de données est synonyme de collection de données. Cette dénomination est utilisée dans MDweb pour les données de type « **Carte numérique** », « **géodatabase** », « **base de données relationnelles** » .

Jeu de données : ensemble de données apparentées, identifiable sans équivoque reliées les uns aux autres par des caractéristiques communes telles que leur mode d'acquisition ou de traitement, leur extension spatiale, etc. Un jeu de données peut être considéré comme un petit ensemble de données ou une partie de cet ensemble. Cette dénomination est utilisée dans MDweb pour les données de type « **Carte papier** », « **Données vecteur** », « **Image satellitale – Photo aérienne** », « **Document texte** », « **Tableaux de données** », « **référence bibliographique** ».

Hierarchie entre les collections et les jeux de données

MDweb établie une hiérarchie entre les types de données notamment par la définition de gabarit parent et enfant (voir notion de gabarit ou metadata profile) . Dans la version standard, cette hiérarchie est la suivante :



Notion de fiches de métadonnées ou références

Dans ce document est utilisé de la même manière la notion de fiche de métadonnées et de référence. Ces deux notions portent sur le même objet. Une fiche de métadonnées ou référence se définit comme un **ensemble d'éléments de métadonnées renseignés par un utilisateur afin de décrire une collection de données ou un jeu de données, plus généralement une ressource.**

La **notion de fiche de métadonnées** se rapporte à la structure et de la nature des éléments qui la compose, les éléments de métadonnées issus de la norme ISO 19115.

La **notion de référence** se rapporte davantage à une vision des métadonnées comme un élément du catalogue de données que gère MDweb.

Notion de niveau de saisie d'une fiche de métadonnées

Le niveau de saisie fait référence à la quantité d'éléments et à leur nature pour décrire une ressource. Il correspond aux différents niveaux d'utilisation des métadonnées. En effet, selon les objectifs d'utilisation des métadonnées, les informations requises pour décrire les ressources ne sont pas les mêmes. Par exemple, des objectifs de recherche des ressources nécessitent des informations moins détaillées et moins complètes que des objectifs de documentation en vue de la diffusion et le transfert des ressources. Ainsi, pour le catalogage des ressources, qui sert de support à leur recherche, des métadonnées simplifiées peuvent s'avérer suffisantes.

Ces différents contextes ou niveau d'utilisation des métadonnées peuvent conduire à définir plusieurs niveaux de détail des métadonnées. Le standard international définit deux niveaux de détail ou niveau de conformité. Le premier niveau de conformité ou niveau « élémentaire » correspond à des objectifs de catalogage des ressources. Pour cela, il propose un ensemble d'éléments obligatoires ou « metadata core profile » qui regroupe les éléments nécessaires à identifier la ressource et à connaître sommairement son contenu. Il peut être uniquement utilisé à des fins de catalogage et de support aux services de métadonnées permettant la recherche de ressources. Un deuxième niveau conformité ou niveau « complet » fournit les éléments de métadonnées nécessaires pour documenter une ressource de façon complète. Ce niveau de conformité définit les éléments de métadonnées requis pour identifier, évaluer, extraire utiliser et gérer les ressources de type géographique.

Nous nous sommes inspirés de la proposition du standard international pour identifier trois niveaux de détail dans le gabarit de métadonnées ROSELT :

- ✓ **un niveau élémentaire**
- ✓ **un niveau étendu,**
- ✓ **et un niveau complet.**

Le **niveau élémentaire** est construit à partir des éléments de métadonnées minimaux du standard.

Le **niveau étendu** est construit sur la base du précédent en incluant des éléments de métadonnées permettant d'assurer l'échange et le transfert de la ressource, et de décrire plus précisément l'origine de la ressource (données sources et traitements mis en œuvre). Cette dernière description s'avère indispensable pour la réutilisation d'une ressource à des fins scientifiques. Pour les types de données proposés dans la version standard, les niveaux étendu et complet sont confondus.

Contacts



IRD / US ESPACE (US 140)
500, rue Jean François Breton, 34093 Montpellier Cedex 05
TEL : +33 (0)4 67 54 87 02
J.C Desconnets jcd@teledetection.fr

Site du projet MDweb : www.mdweb-project.org
Démonstrateur en ligne : demo16.mdweb-project.org