

---

# Des indicateurs et des métadonnées pour les décrire : intégration au sein d'atlas géomatique en ligne

Ely Beibou<sup>1,3,4</sup>, Jérôme Guitton<sup>2</sup>, Julien Barde<sup>3</sup>, Thérèse Libourel<sup>4</sup>

1. Institut Mauritanien de recherches Océanographiques et des Pêches (IMROP)  
BP 22, Nouadhibou, Mauritanie, beibou\_es@yahoo.fr;
2. UMR ESE, Ecologie et Santé des Ecosystèmes  
Agrocampus ouest, Rennes, France, jerome.guitton@agrocampus-ouest.fr
3. UMR 248 MARBEC (IRD)  
34203 Avenue Jean Monnet, CRH, Sète, France, julien.barde@ird.fr
4. UMR 228 ESPACE-DEV (IRD - UM)  
500 rue Jean-François Breton, Montpellier, therese.libourel@univ-montpe2

---

*RESUME.* S'appuyant sur des indicateurs et tableaux de bords de suivi des phénomènes observés, les outils d'aide à la décision proposent des documents prenant différentes formes (synthèses, sites web, fiches thématiques), faciles à utiliser par les acteurs du domaine. Pour permettre à ces derniers d'interpréter correctement l'information synthétisée, la description des jeux de données et traitements sous-jacents constitue un apport indéniable.

Nous proposons une approche innovante, qui, dans le contexte de la construction de tableaux de bord conçus autour d'entrepôts de données, tels que les atlas en ligne, permet la mise en place d'un système intégrant la génération, la gestion et la publication de métadonnées. Nous illustrons ce propos par l'expérience d'un atlas en ligne pour la gestion durable des pêches et de l'environnement en Mauritanie.

*ABSTRACT.* Based on indicators and monitoring dashboards for the observed phenomena, the decision support tools provide documents taking different forms (summaries, websites, fact sheets), easy to use by players in the field. To allow users to correctly interpret the synthesized information, description of the underlying data sets and processing is an undeniable contribution.

We propose an innovative approach, which, in the context of the construction of dashboards designed around data warehouses, such as online atlas allows the establishment of a system integrating the generation, management and publication of metadata. We illustrate this point by the experience of an online atlas for the sustainable management of fisheries and the environment in Mauritania.

*MOTS-CLES :* Indicateurs, Métadonnées, Catalogage, Systèmes d'information, Atlas.

*KEYWORDS :* Indicators, Metadata, Cataloging, Information Systems, Atlas.

---

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

## 1. Introduction

L'accès transparent à l'information constitue, à la fois, un défi persistant pour la communauté des informaticiens, et une condition pour que, globalement, la science soit pleinement reproductible (Coles et al., 2013), ce qui constitue un challenge pour la communauté scientifique en général. Si, à l'origine, ce challenge se justifiait par la capacité limitée des moyens de communication (réseaux informatiques et web), aujourd'hui, il s'explique par l'accroissement rapide du nombre de réseaux et d'outils de communication en plus de l'explosion du volume de données à échanger (concepts d'Open Data : <http://opendefinition.org/> et Big Data). En effet, (Berners-Lee, 1998) a prévenu : « Le web de documents lisibles par l'homme sera fusionné avec une toile de données compréhensibles par les machines. Le potentiel du mélange des hommes et des machines qui travaillent ensemble et communiquent via le web pourrait être immense ». De ce fait, les utilisateurs finaux se trouvent actuellement confrontés à des problèmes de localisation et d'accès à l'information pertinente dans un environnement caractérisé par l'hétérogénéité, la dispersion et la volatilité de ressources souvent mal référencées ou contextualisées (concept de Linked Data : <http://www.w3.org/DesignIssues/LinkedData.html>). En plus, les chercheurs ne sont pas experts dans la tâche complexe de la création de métadonnées et encore moins quand il s'agit d'en créer pour les autres (Borgman, 2008). Tous ces constats représentent autant de verrous pour lesquels, les applications de gestion de métadonnées constituent une réponse pertinente.

Le terme de métadonnées signifie « données sur les données ou données qui renseignent sur certaines données et qui permettent ainsi leur utilisation pertinente » (Desconnets et al., 2001). Une méta-information permet de donner du sens au contenu des ressources, en les décrivant sous différents aspects, en répondant notamment aux questions : Qui ? Quoi ? Où ? Quand ? Comment ? Et ce afin de permettre aux utilisateurs, de les localiser, interroger et exploiter. On peut citer quelques exemples de métadonnées : table d'allocation de fichiers ; le schéma d'une base de données, le dictionnaire de données, les index ; toute description associée à une ressource permettant d'identifier son auteur, sa date de création ; la spécification, la signature d'un composant logiciel ; des annotations relatives au contenu de la ressource ; des commentaires exprimant un point de vue sur la ressource.

Pour implémenter les métadonnées, deux approches coexistent. La première, qu'on peut qualifier de méthode *ad hoc*, consiste à ce que le développeur définisse avec le producteur des ressources des éléments de métadonnées propres à leurs besoins, ceux qu'il juge importants pour mieux décrire les ressources qu'ils gèrent ou plus simplement ceux pour lesquels ils disposent de l'information, et qui seront utilisés pour générer les métadonnées. Cette méthode non standardisée est généralement mise en œuvre pour répondre à des besoins spécifiques : nous pouvons citer par exemple le cas d'entrepôts de données dédiés à la pêche, comme

l'application StatBase (Thibaut et al., 2002) développé par l'IRD dans les années 2000 et le prototype ISTAM<sup>1</sup> (Guitton and Gascuel, 2002) qui lui a succédé.

Dans ces contextes particuliers, les métadonnées restent accessibles uniquement au sein de l'application mère, ce qui pose le problème de leur exploitation par des agents logiciels externes et celui de leur interopérabilité avec d'autres applications.

La seconde approche consiste à utiliser des solutions de catalogage génériques et standardisées. En effet, vu le caractère transversal des métadonnées et leur importance pour faciliter l'accès à l'information, des groupes de travail ont proposé des normes (W3C & RDF, Dublin Core / ISO 15836:2009, FGDC, ISO TC 2011, etc.). Celles-ci définissent un cadre structuré et standardisé pour décrire les ressources (données, informations, connaissances, traitements). Dans ce cas, deux solutions se présentent :

1. les solutions qui se restreignent au catalogage, comme GeoNetwork (<http://geonetwork-opensource.org>) et MdWeb (<http://www.mdweb-project.org>), pour publier les métadonnées et en assurer l'accès indépendamment de leur contexte de localisation et de gestion. Cette approche, si le catalogue est mal intégré aux applications, peut présenter divers inconvénients : double saisie (saisie dans l'application mère et saisie dans les plates formes externes) et défaut de mise à jour (si les métadonnées sont indépendantes des données sources et gérées dans des contextes séparés et différents).
2. les solutions qui permettent d'intégrer les métadonnées au sein d'une démarche plus globale de gestion et diffusion de l'information spatiale, comme Geotoolkit ([www.geotoolkit.org](http://www.geotoolkit.org)), Geotools ([www.geotools.org](http://www.geotools.org)) ou d'autres API. Ces solutions plus génériques permettent de décrire en RDF tous types de ressources et de les rendre accessibles avec des SPARQL endpoint (par exemple l'API Java JENA et le serveur Joseki) (Prud'hommeaux and Seaborne, 2008).

Nous proposons de développer une solution normalisée, respectant les standards de génération de métadonnées directement au sein de l'application mère et du coup assurant l'interopérabilité avec les agents logiciels qui voudront venir moissonner les métadonnées, les acquérir ou les partager. Après cette introduction, nous étayons, en section 2, la problématique de recherche. En section 3, nous présentons un aperçu historique des travaux portant sur le catalogage et un panorama des outils de catalogage existants. Notre proposition est présentée en section 4 et les conclusions et perspectives sont données en section 5.

## 2. Problématique

Le service de catalogage, dans les approches non modulaires de conception des systèmes d'information est perçu au mieux<sup>2</sup> comme une fonction externe. Ce qui fait

---

<sup>1</sup> Improve Scientific and Technical Advice for Management

<sup>2</sup> très souvent ce service est omis et rajouté au système a posteriori.

que les concepteurs n'accordent pas suffisamment de temps à son développement et s'adossent généralement à des solutions *ad hoc* ou génériques mais sans lien direct entre les métadonnées et les ressources qu'elles décrivent. Cette approche a le mérite de simplifier la vie des développeurs en leur permettant de passer outre le développement d'un nouvel outil, qui peut être remplacé par un équivalent existant et testé. Elle permet également de gagner du temps et économiser les ressources (matérielles et humaines).

Cependant, cette approche présente certains inconvénients sur les plans technique et fonctionnel. Sur le plan technique, les choix opérés par un développeur peuvent souvent s'avérer différents de ceux utilisés dans le cadre des solutions génériques qui sont souvent plus lourdes à prendre en main. Ce qui peut compromettre la compatibilité et l'interopérabilité entre celles-ci et l'application mère. Sur le plan fonctionnel, les solutions génériques essaient généralement de couvrir les fonctionnalités jugées essentielles et du coup, on ne trouve pas forcément certaines fonctions « métier » qu'on aimerait utiliser, dans des cas particuliers d'implémentation liés à des thématiques particulières. Créer des métadonnées est bien pris en compte mais automatiser une partie de la saisie des contenus de celles-ci reste en général complexe. Il est toutefois possible de profiler les standards (schémas XML) et d'adapter les codes sources des applications qui les implémentent. On peut ainsi hériter d'un noyau d'éléments de métadonnées et de fonctionnalités communes sur la base duquel on peut s'adapter au contexte métier en ajoutant des éléments plus spécifiques.

Développant une application dénommée « atlas en ligne », dédiée à la mise en œuvre d'indicateurs et de tableaux de bord relatifs à des ressources halieutiques, nous sommes confrontés à la gestion de métadonnées directement au sein de cet observatoire « virtuel » qui repose sur un entrepôt de données issues de sources réparties et hétérogènes. Les extractions de données contenues dans cet entrepôt servent de paramètres d'entrée aux traitements qui produisent les indicateurs du tableau de bord. Le verrou essentiel est donc de produire de manière la plus automatisée possible les métadonnées pour faciliter la découverte et l'accès des ressources par divers utilisateurs et diverses applications.

### **3. Etat de l'art**

#### ***3.1 Aperçu historique sur le catalogage***

Les premiers catalogues de métadonnées informatisés remontent aux années 70 où, ils trouvent leur origine dans les tables d'allocation de fichiers maintenues par les systèmes d'exploitation et qui, sauvegardant le nom logique du fichier et son adresse physique sur disque (métadonnées), permettaient de localiser les fichiers et d'y accéder ultérieurement.

Pour le Web, on peut distinguer deux grands types de catalogues qui sont les annuaires et les catalogues basés sur les normes descriptives.

### 3.1.1 Annuaires

Dans le contexte du Web, les premiers outils et travaux utilisant les métadonnées, ont été les annuaires (Yahoo) et les moteurs de recherche (AltaVista, Google, Inktomi, Northern Light). Initialement, les premiers annuaires accessibles sur Internet demandaient au gestionnaire de l'annuaire de définir les métadonnées qui restaient à l'état de mots-clés, voire de description textuelle plus ou moins structurée et conforme à des modèles propriétaires. Remarquons que l'ère des services Web a aussi introduit de nouveaux types d'annuaires (UDDI Universal Description Discovery and Integration), mais celui-ci aussi demande une opération de publication du service concerné, c'est-à-dire une description de ce que l'on peut considérer comme des métadonnées conformes au modèle de données propre et défini par l'UDDI (Cf. Figure 1).

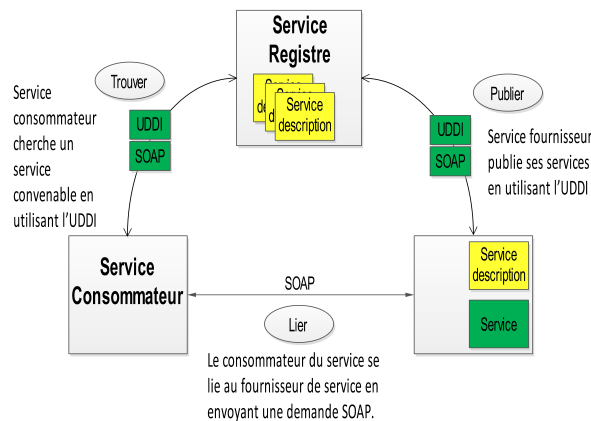


Figure 1 : Modèle général pour une architecture web service, exemple de combinaison (SOAP+WSDL+UDDI), utilisés dans le cadre de l'atlas géomatique (Beibou et al., 2014)

### 3.1.2 Catalogues basés sur des normes descriptives

Une deuxième étape est franchie dans le contexte du catalogage et s'appuie sur l'adoption des normes descriptives standardisées définies par divers organismes ou groupes de travail qui ont analysé et répondu aux besoins métiers de différentes communautés d'utilisateurs : Dublin Core (Wittenburg and Broeder, 2002) pour tout type de ressources (<http://dublincore.org>), pour l'information spatiale : CEN Comité Européen de Normalisation TC 287 (<http://centc287.eu/>), FGDC Federal Geographic

Data Committee (<https://www.fgdc.gov/>), ISO International Organisation for Standardisation TC 211 ([www.isotc211.org/](http://www.isotc211.org/)), ANZLIC ([www.anzlic.gov.au](http://www.anzlic.gov.au)) the Spatial Information Council Home, l'OpenGIS Consortium : OpenGIS, ([www.opengeospatial.org](http://www.opengeospatial.org)), EML / Darwin Core pour la biodiversité ...

Indépendamment de leurs différences, l'objectif de ces normes est de structurer et de donner un sens aux éléments de métadonnées ainsi qu'à leurs valeurs pour permettre ainsi une interopérabilité syntaxique et sémantique. Il s'agit d'uniformiser la manière d'effectuer la description/indexation des ressources et par voie de conséquence d'améliorer les échanges et le partage a posteriori en dehors du contexte de production. De manière générale, les normes proposent un guide de structuration des métadonnées nécessaires à la description d'une ressource. Les métadonnées sont présentées sous forme d'éléments (sections ou rubriques) lesquels peuvent, selon leur sémantique, être regroupés en catégories. Par exemple, la norme Dublin Core, dans sa version la plus légère, propose 15 éléments de description relatifs au contenu d'une ressource classifiés en trois catégories concernant :

- Le contenu de la ressource : titre, sujet ou codes de classement, description, source, langue, lien vers une autre ressource, couverture spatiale et temporelle.
- La propriété intellectuelle : créateur, éditeur, collaborateur, droits d'utilisation.
- La matérialisation de la ressource : cycle de vie, type, format, identificateur.

Ce noyau d'éléments génériques peut être appliqué à n'importe quelle ressource et se retrouve donc dans les normes métiers qui le complètent par des éléments plus spécifiques au domaine d'application. Par exemple, la norme ISO19115/19139 dédiée aux métadonnées pour l'information géographique, tout comme le standard FGDC, complète la proposition du Dublin Core pour prendre en compte les spécificités des ressources concernées dans des packages de métadonnées plus spécifiques : système de référence spatial, étendue spatiale et temporelle de la ressource, modalités de distribution (résolution, mode raster ou vecteur...).

La norme ISO 19115 incite explicitement à produire des métadonnées de qualité (mise à jour et traçabilité / généalogie / historique) et révèle plusieurs problèmes d'adaptation au contexte d'implémentation qui sont :

- la connexion avec la sémantique du domaine concerné, comment aider le producteur de métadonnée à décrire la thématique de la ressource. Pour éviter les difficultés d'interprétation des valeurs saisies pour renseigner ces divers éléments, les guides associés aux normes recommandent l'utilisation de formats prédéfinis. Ces formats relèvent de notations, contraintes ou de structures plus ou moins élaborées comme les Code Listes ou autres vocabulaires contrôlés (pour les mots clés en particulier). En effet, une recherche de ressource sera d'autant plus pertinente et complète pour l'utilisateur s'il s'appuie sur des vocabulaires contrôlés (des glossaires voire des thésaurus ou des ontologies).
- l'extensibilité. De nombreux travaux actuels tendent à définir, au-delà du guide préconisé, des ensembles personnalisés d'éléments de métadonnées (il s'agit de définir des sous-ensembles judicieux d'éléments regroupés en ce que l'on

dénommée « profils »). Cette démarche est adoptée au niveau européen dans le cadre de la directive INSPIRE Infrastructure for Spatial Information in Europe pour profiler la norme ISO 19115 (<http://www.ec-gis.org/inspire/>) et l'adapter aux besoins des états membres et de l'UE.

Au-delà du choix des normes, le producteur de métadonnées doit décider de la forme qu'il souhaite donner à son catalogue. Les métadonnées peuvent être directement décrites et intégrées dans la ressource (par exemple dans un fichier de séquence génomique ou dans un fichier netCDF, où métadonnée et donnée sont regroupées), ou elles peuvent être dissociées des ressources et simplement les référencer depuis un système extérieur. La frontière reste souvent complexe à définir. Dans le cas où métadonnées et ressources sont disjointes la localisation de la ressource devient plus rapide (indexation) et assure, de plus, un niveau de protection exigé par les communautés scientifiques. Mais on court alors le risque, d'une part, de dissocier les cycles de vie de la ressource et de ses métadonnées (problème de maintenance des mises à jour) et, d'autre part, de complexifier l'accès direct à la ressource elle-même. Si les métadonnées et les ressources décrites sont entreposées dans un système de gestion de base de données, le service de catalogage bénéficie alors de la puissance d'interrogation du SGBD sous-jacent. Cette solution est retenue par de nombreux services de catalogage constituant l'ossature essentielle de portails dédiés.

### 3.2 Aperçu des outils de catalogage existants

Nous allons analyser les outils de catalogage existant en considérant les normes utilisées, l'architecture (web ou *stand-alone*), la couverture de la sémantique, l'utilisation d'interface cartographique, les licences associées, l'architecture de stockage, l'interopérabilité. Dans notre analyse, nous serons guidées par la nature spatiale des données à traiter, tout en tenant compte du besoin de normalisation (standards ISO TC 211 produits par l'OGC, notamment le CSW : OGC Catalog services for the web). Les principaux standards pour les Web Services de l'OGC sont répertoriés et détaillés dans (Pornon et al., 2008) et sur le site de l'OGC (<http://www.opengeospatial.org/standards>). Nous analyserons également l'utilisation des services Web pour la communication entre les systèmes partenaires à travers un accès internet normalisé, dans un environnement distribué et hétérogène.

Dans le monde *open source*, le service de catalogage est aujourd'hui souvent basé sur différentes solutions comme GeoNetwork (<http://geonetwork-opensource.org/>) et MdWeb (<http://www.mdweb-project.org/>). Ces outils sont performants et largement utilisés dans les domaines de l'agriculture, au sens large du terme, et de l'environnement. Ils sont des catalogues de métadonnées liées à des ressources à composantes géographiques qui permettent la recherche et l'accès aux catalogues géospatiaux locaux ou distribués. Ils permettent de télécharger non seulement les données spatiales mais tout type de données et disposent d'interfaces web interactives pour dialoguer avec les services CSW de serveurs distribués.

Après l'étude de ces outils, et au regard de précédentes expériences (prototype ISTAM) nous avons vérifié que les fonctions d'interrogation des métadonnées (la partie cliente) souhaitée pour notre application sont prises en charge par ces outils et permettent de proposer des *front-end* faciles à mettre en place et accessibles sur le Web.

#### 4. Notre proposition

##### 4.1. Conception

En réponse à la problématique présentée dans la section 2, nous proposons une démarche modulaire intégrant la question des métadonnées et considérant le catalogage comme composant ou service accessible au sein de l'application mère, mais moissonnable à partir d'agents externes.

Notre proposition consiste donc, à concevoir un système d'information pour assurer un accès transparent aux données, issues de sources diverses, hétérogènes et distribuées, aux traitements et aux indicateurs produits à partir des ces données et qui alimentent un tableau de bord pour la gestion durable des pêches et la préservation de l'environnement. Ce système est basé sur différents composants incluant des services pour i) la recherche dans un catalogue de métadonnées, ii) l'accès aux données iii) la construction des indicateurs du tableaux de bord, et iv) l'interaction entre acteurs et utilisateurs dudit tableau de bord. Nous présentons dans ce qui suit le schéma conceptuel d'un tel système.

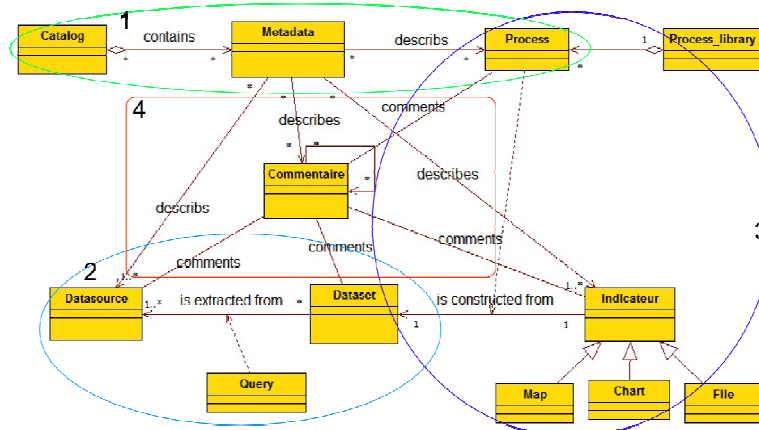


Figure 2 : Modèle conceptuel du Système SICP (Beibou et al., 2014). Dont le module de catalogage 1 sert à la recherche et la localisation de l'information pertinente ; le module 2 est pour l'accès et l'extraction des jeux de données à partir



*des sources ; le module 3 permet de construire les indicateurs à partir de ces jeux de données et, le module 4 permet aux acteurs d'interagir avec un système de commentaires portant sur les différents composants.*

Selon ce modèle, plusieurs modules rentrent en jeu (Catalogage, Extraction, Traitement & Indicateur, Commentaire). Le module qui nous intéresse est celui du catalogage (ici, 1). Celui-ci, permet d'effectuer la recherche et la localisation de l'information pertinente en prenant en compte les différents niveaux de granularité (indicateur ou fiche thématique composée de plusieurs indicateurs et de leur analyse). En effet, l'atlas produit des fiches thématiques composées de différentes ressources : des indicateurs (résultats de traitements appliqués à des jeux de données) et des commentaires d'experts (Beibou et al., 2014). Du coup, la structure de l'information à décrire par des métadonnées est composée de plusieurs niveaux de description portant sur :

- les sources de données de base (producteur, mise à jour, description du mode de collecte ...) ;
- les fiches thématiques (auteur(s), titre, objectif de la fiche ...) ;
- les indicateurs (source de données, traitement, objectif, guide de lecture, date de mise à jour...) ;
- le traitement appliqué à la donnée.

#### *4.1.1 Articulation des métadonnées avec les entrepôts de données*

Les entrepôts de données sont devenus incontournables pour intégrer des sources d'informations hétérogènes et permettre le traitement analytique en ligne (Inmon, 1996 ; Codd, 1970 ; Silhadi-Hacid and Tarafi, n.d.).

Etant donné qu'un entrepôt de données reflète le modèle de l'organisation, un élément essentiel d'une architecture de stockage est la gestion des métadonnées (Surajit and Umeshwar, 1997). Mais, comme schématisé dans la figure 3, la conception traditionnelle de l'architecture générale des entrepôts, traite les métadonnées comme étant juste un référentiel, non intégré à l'entrepôt. Ce qui fait que leur gestion, reste reléguée au second niveau d'importance.

Par contre, suivant notre conception (figure 2), nous proposons d'en faire un module indépendant mais intégré au système et, du coup, notre entrepôt prend en charge la production, la génération, la gestion et la publication au sein d'un serveur dédié, des métadonnées associées aux sources de données, traitements, indicateurs et commentaires associés à ces différents éléments.

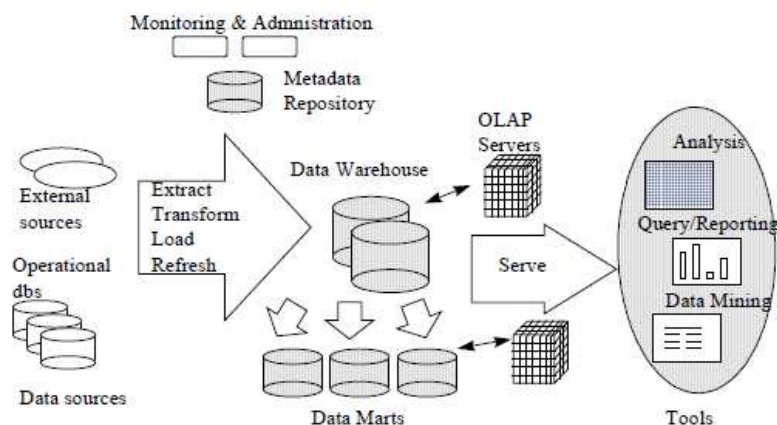


Figure 3 : Architecture générale d'entrepôt de données d'après (Surajit and Umeshwar, 1997)

#### 4.2. Opérationnalisation

D'une part, l'atlas en ligne génère automatiquement un certain nombre de métadonnées sur les indicateurs au fur et à mesure que ces derniers sont créés et intégrés (Beibou et al., 2014). D'autre part, nous disposons d'outils (clients) de catalogage comme GeoNetwork, qui permettent de parcourir les métadonnées générées conformément à des éléments de métadonnées normalisées (fiche). Notre méthode consiste à établir la liaison entre les deux (l'atlas et les outils génériques). Pour ce faire, nous avons commencé par mettre en correspondance les métadonnées nativement produites au sein de l'atlas avec celles qui sont recherchées par ces outils de catalogage et ensuite, nous avons déterminé les éléments de métadonnées manquants ou à renseigner pour se conformer à une fiche normalisée. C'est-à-dire faire en sorte que l'atlas produise des fiches de métadonnées qui soient moissonnables directement par ces outils. Trois catégories d'objets ont été considérées pour illustrer notre propos. Il s'agit de : i) « *sources de données* », ii) « *indicateurs* », produits à partir des jeux de données et des traitements qui leurs sont appliqués et enfin iii) « *fiches* » où, sont présentés les indicateurs sous différentes formes (graphiques, cartes, tableaux, etc.).

Pour l'opérationnalisation du moissonnage, le service web qui nous intéresse est le CS-W, parce qu'il permet la publication de catalogues de métadonnées (relatives à des données, des traitements ou des services) exploitables par des clients extérieurs au système pour rechercher parmi les entrées de catalogues (Pornon et al., 2008).

Pour l'implémentation de ce service au sein de l'atlas, notre choix s'est arrêté sur PyCSW ([www.pycsw.org](http://www.pycsw.org)), basé sur Python et qui correspond à une bibliothèque sur laquelle on peut s'appuyer pour exécuter des opérations de recherche, de production, de visualisation et d'interrogation de métadonnées (Kralidis, 2014). Ce choix est motivé par les raisons suivantes :

- *Interopérabilité et normalisation.* PyCSW est une implantation du serveur OGC CSW en Python, qui met en œuvre la clause 10 (liaison protocole HTTP, Catalogue Services pour le Web, CSW) de la spécification d'implémentation OpenGIS catalogue de services, la version 2.0.2.

- *Possibilité d'intégration du service à l'atlas en ligne.* PyCSW permet de publier des métadonnées soit à partir de son propre modèle intégré de métadonnées, ou à travers la configuration d'une liaison à un modèle existant de métadonnées.

- *Indépendance par rapport aux plateformes systèmes.* PyCSW est Open Source, publié sous licence MIT, et fonctionne sur toutes les grandes plates-formes (Windows, Linux, Mac OS X).

- *Compatibilité des environnements de développement.* En effet, PyCSW fonctionne grâce à une simple table Postgres et l'Atlas est lui-même développé sur ce SGBD. Cette table peut alors être intégrée directement à l'atlas et alimentée par ses métadonnées.

#### 4.3. Résultats obtenus

Dans l'atlas en ligne, les métadonnées sont présentes à différents endroits (Cf. Figure 4) :

- Au sein des fiches XML qui permettent la génération des fiches de synthèses et qui contiennent un ensemble de sous éléments (indicateurs, expertises)
- Au sein de la table de description des entrées thématiques où, sont stockées des informations transversales, valables pour toutes les fiches de la même thématique.
- Au sein de la table de description des bases de données ressources utilisées pour créer les indicateurs.

La proposition que nous avons mise en place consiste à créer pour les 3 types (Source, Fiche et Indicateurs) des métadonnées (correspondant à un profil ISO 19115), générées à partir des 3 types d'information décrits ci-dessus (grâce au composant 1 Catalog (moteur de catalogage) de la figure 2 qui les extrait des sources et les sauvegarde dans la table RECORD utilisée par PyCSW (Cf. Figure 4 et 5). Par la suite PyCSW permet à n'importe quel client (et ici à Geonetwork) de publier ces métadonnées.

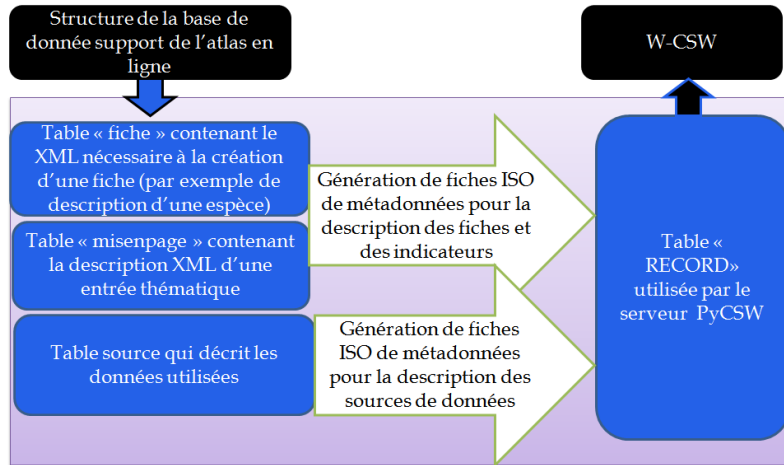


Figure 4 : Schéma d'intégration du composant Catalog au sein de l'atlas en ligne

Les résultats obtenus sont donnés dans la figure 5 qui suit.

La figure présente deux captures d'écran. À gauche, l'interface d'accès aux fiches (composant 3 de l'atlas) pour l'espèce *Octopus vulgaris* (Poulpe). Elle affiche un titre, une image de l'animal, et deux graphiques à barres empilées montrant la caractérisation temporelle des captures de l'espèce. À droite, les éléments de métadonnées moissonnés par Geonetwork, qui incluent : 'Meta donnée fiche' (contenant des informations de base), 'Meta donnée indicateur' (contenant des informations géographiques et techniques), 'Description indicateur' (contenant une description détaillée et un graphique), et 'Contact indicateurs' (contenant des coordonnées de contact).

Figure 5 : Capture d'écran d'une fiche thématique sur l'espèce Poulpe (à gauche) et les éléments de métadonnée afférents moissonnés par Geonetwork (à droite)

## 5. Conclusions et perspectives

L'atlas en ligne tel que nous le proposons apporte, nous le pensons, une réponse à la problématique soulevée qui consiste à mieux documenter les indicateurs des observatoires pour en faciliter l'interprétation par les acteurs des domaines, de l'environnement et de l'halieutique dans notre cas. Le fait que les métadonnées soient globalement générées par le composant Catalog de l'atlas en coopération avec le client catalogue Geonetwork permet de gérer les problèmes récurrents de mise à jour des métadonnées. En effet, si le producteur d'une source alimentant l'entrepôt, modifie les métadonnées de celle-ci, l'atlas donnera pour les indicateurs et les fiches, des métadonnées actualisées. Le service web CSW quant à lui permet le moissonnage par d'autres applications (autres moteurs de catalogage, clients ou serveurs CSW).

Restent encore cependant quelques problèmes : la mise à jour des métadonnées se fait en continu ce qui ne permet pas de vérifier l'historique des mises à jour, pour cela il faudrait avoir recours à des liens de filiation (Généalogie ou Lineage dans 19115). D'autre part, l'approche privilégie le calcul au stockage ce qui évidemment entraîne une surcharge du fonctionnement de l'atlas (notamment lorsqu'il est accédé par de multi utilisateurs) mais nous sommes dans l'éternel dilemme temps/espace. Il faut aussi considérer les problèmes posés par le calcul d'indicateurs complexes qui nécessitent idéalement d'exécuter directement les codes dans les langages de programmation utilisés par les chercheurs (R, Python, Java..). Cela pousse à l'intégration de composants dédiés aux traitements de données, idéalement en Web Service tel que le standard WPS (Web Processing Service) proposé par l'OGC. Ce standard implémenté par différentes applications Open Source comporte un modèle utile à la description des traitements et à la gestion des codes sources. Ainsi l'approche que nous proposons pour notre système gagnerait à implémenter ce standard pour fournir des métadonnées plus riches et des informations plus transparentes et totalement reproductibles.

## Bibliographie

- Beibou, E., Guitton, J., Libourel, T., 2014. Atlas géomatique collaboratif pour l'environnement et la gestion durable des ressources halieutiques, en Afrique de l'ouest, cas de la Mauritanie, in: Congrès INFORSID 2014 Lyon, 20 - 23 Mai 2014. Lyon.
- Berners-Lee, T., 1998. The World Wide Web: A very short personal history, May 1998 [WWW Document]. URL <http://www.w3.org/People/Berners-Lee/ShortHistory.html>
- Borgman, C.L., 2008. Data, disciplines, and scholarly publishing. *Learned Publishing* 29–38.
- Codd, E.F., 1970. relational model of data for large shared data banks. *Communications of the ACM* 13, 377–387.
- Coles, S.J., Frey, J.G., Bird, C.L., Whitby, R.J., Day, A.E., 2013. First steps towards semantic descriptions of electronic laboratory notebook records. *J. Cheminformatics* 5, 52.

Copyright © by the paper's authors. Copying permitted for private and academic purposes. Proceedings of the Spatial Analysis and GEOmatics conference, SAGEO 2015.

- Desconnets, J.-C., Rouge, T.L., Maurel, P., Miralles, A., Passouant, M., 2001. Proposition de structuration des métadonnées en géosciences : Spécificité de la communauté scientifique. Presented at the Journées Cassini' 2001 : Géomatique et espace rural, pp. 69–82.
- Guitton, J., Gascuel, D., 2002. TrawlBase-Siap : un outil de gestion des données de campagnes de chalutage scientifique, in: Actes Du Symposium International, Dakar (Sénégal), 24-28 Juin 2002. Dakar.
- Inmon, W.H., 1996. The data warehouse and data mining, vol. 39, no 11. ed. Communications of the ACM.
- Kralidis, T., 2014. pycsw Documentation. [readthedocs.org](http://readthedocs.org).
- Pornon, H., Yalamas, P., Pelegris, E., 2008. Services Web géographiques: état de l'art et perspectives. *Géomatique Expert* - N° 65.
- Prud'hommeaux, E., Seaborne, A., 2008. SPARQL Query Language for RDF [WWW Document]. W3C. URL <http://www.w3.org/TR/rdf-sparql-query/>
- Silhadi-Hacid, S., Tarafi, M., n.d. Du DataWarehouse au WebHouse : le croisement des réseaux et des bases de données [WWW Document]. URL <http://web.univ-pau.fr/~cpham/M2SIR/BIBLIO/DOC01-02/Datawarehouse.doc>
- Surajit, C., Umeshwar, D., 1997. An Overview of Data Warehousing and OLAP Technology. *ACM* 26, 65–74.
- Thibaut, L., Chavance, P., Damiano, A., 2002. StatBase, une approche générique pour la gestion de statistiques de pêche d'origines multiples, in: Actes du symposium international Dakar — Sénégal — 24-28 juin 2002, Collection des Rapports de recherche halieutique ACP-UE, numéro 15, Vol.1. Dakar.
- Wittenburg, P., Broeder, D., 2002. Metadata overview and the semantic web. Presented at the Proceedings of the International Workshop on Resources and Tools in Field Linguistics, Las Palmas.