

 compte-rendu	2 ^{ème} réunion « Data Rescue »	Rédacteurs Jean Le Fur & Peggy Bretaudeau	Date création : 13.01.05	Référence: 21.003CR/1
			Dernière modif. : 12/04/2005 10:54	Pages 5

Diffusion : participants à la réunion
 Core Group Décisionnel
 Documents du projet – web

Résumé : une réunion restreinte des personnels de l'US Osiris a été réalisée pour faire le point et programmer la suite de l'opération Data Rescue. Une liste d'actions à réaliser a été identifiée et discutée.

Un certain nombre de recommandations ont été établies (encadrées dans le compte-rendu).

Lieu de réunion : Sète, CRHMT, Salle Méditerranée

Date : Mercredi 13 janvier 2005, de 9H à 11H00

Organisateur : Pierre CHAVANCE

Participants : Peggy BRETAUDEAU / Pierre CHAVANCE / Jean LE FUR / Alain DAMIANO

Objectifs : bilan et poursuite de l'opération « Data Rescue »

Ordre du jour :

- 1/ bilan de la première récolte des données
- 2/ identification des actions à mener

Compte-rendu

1/ bilan de la première récolte de données

A partir du formulaire diffusé, un certain nombre de réponses ont été obtenues. Les résultats ont été compilés par A.Damiano et P.Bretaudeau dans un fichier Excel (doc n° 21.004)

La situation est globalement la suivante :

- GAO (F.Marsac), informations fournies, données non fournies
- P.Fréon et A.Bertrand (données Upwelling) : les données seraient confidentielles et non intégrables et de plus elles auraient déjà leur propre système de sauvegarde.
Jean Le Fur propose que même si des données ne peuvent être physiquement incluses dans l'entrepôt, il faut qu'elles soient répertoriées pour savoir qu'elles existent.
- F.Ménard a fourni un CD de données sur la pêche sardinelles en Côte d'Ivoire et des données d'environnement

- P.Bach a fourni des informations sur les programmes EcoTap sur la pêche expérimentale à la palangre,
- A. Damiano a intégré dans la compilation Excel les différents jeux de données disponibles dans StatBase

Jean Le Fur demande à ce que, en règle générale, les données soient physiquement récoltées dès que possible. Les premières données disponibles serviront d' « échantillon » pour montrer au plus vite la faisabilité de l'opération « data rescue » aux thématiciens susceptibles d'y contribuer.

Peggy Bretaudeau dit qu'actuellement on dispose concrètement du CD de F.Ménard, de celui de J.J. Levenez, des données de M.Laurans et de cassettes non lues de A.Fonteneau.

Il faut au plus vite récupérer sous un format « moderne » les données des cassettes de A.Fonteneau.

Pierre Chavance présente le résultat d'une réunion tenue¹ sur les données disponibles au Sénégal. Les résultats sont encore préliminaires mais un descriptif assez complet de la qualité des données disponibles, des thématiques et des lacunes a été réalisé.

Une discussion est née sur les niveaux différents de qualité de la documentation associée aux jeux de données. L'exemple sénégalais constituerait une extrémité de l'échelle, l'autre extrémité comprenant des séries de chiffres correspondant à des données inconnues (sans date, lieu ou autre, donc inutilisables en l'état).

Dans la description des jeux de données qui sera effectuée, il faudrait élaborer une échelle de classification de la qualité de la documentation associée aux données.

Peggy Bretaudeau indique que la cellule n'a pas compétence pour évaluer cette qualité

Pierre Chavance et **Jean Le Fur** répondent que c'est de la responsabilité du thématicien proposant le jeu de données de préciser (de façon circonstanciée) quel est le niveau de confiance que l'on peut accorder

Jean Le Fur propose que l'on constitue une grille de critères qui permettront de calculer un indicateur du niveau de qualité des données

Il y a, de même, différents niveaux de restitution. Il est décidé de conserver tous les jeux de données dans le data rescue, même s'ils sont très incomplets en terme de documentation (J. Le Fur donne l'exemple d'un jeu de données non documentées qu'un stagiaire intéressé a fini par re-documenter après une enquête auprès de différents experts impliqués).

Par contre, on définira un niveau minimum de qualité pour les jeux de données qui seront accessibles aux requêtes dans l'Ecoscope.

Dans le cas où deux séries de données portant sur le même objet (au même moment, etc.) seront disponibles, il faudra identifier quelle est la série de référence. Ce point sera éventuellement à préciser lorsque le cas se présentera concrètement.

¹ Participants : M.Laurans, V.Fonteneau, A.Damiano, P.Bretaudeau, P.Chavance, A.Caverivière

Pierre Chavance demande à ce qu'un schéma du processus du traitement des jeux de données soit fait décrivant les différentes opérations entre le moment où un jeu de données est soumis et celui où il est rendu disponible et qualifié au sein du Centre d'information. Ce schéma permettra de rendre transparente cette procédure de traitement des jeux de données auprès des thématiciens et des partenaires.

Jean Le Fur estime qu'il y aura deux façons d'explorer les jeux de données. La première se fera de façon directe par l'intermédiaire de requêtes réalisées sur les méta-données disponibles, ceci pour les personnes souhaitant travailler directement sur les données. L'autre méthode se fera par l'intermédiaire de mots-clés dans le centre d'information de façon à rendre disponibles les données au même titre que les informations, connaissances, outils pour une requête donnée et les mettre en relation avec ces autres types de savoirs.

Décisions - conclusions

Une liste des actions à entreprendre est dressée (classées par ordre approximatif de priorité beaucoup d'actions étant à réaliser en parallèle) :

1. Entamer une recherche sur les différentes normes de méta-information susceptibles d'être utilisées en particulier sur la norme ISO 19115 concernant les méta-données (voir annexe),
2. Récupération physique des données et de l'information associée (un code de référencement des jeux de données sera mis en place)
3. Sur la base de la norme, identification de critères de qualification supplémentaires spécifiques au domaine halieutique et à la problématique data rescue
4. Mise en place de la structure des données (schéma relationnel des tables de description des jeux de données)
5. Schéma fonctionnel (processus de traitement des jeux de données). Ce schéma sera développé de façon concomitante avec celui du Centre d'Information dont il partage une section
6. Identification et initialisation d'indicateurs de suivi de l'opération²
7. lister les critères qui ont servi à structurer StatBase et présenter le modèle de données (Pierre et Alain)

Rédaction : Jean Le Fur & Peggy Bretaudeau / corrections P.Chavance et A.Damiano / saisie : JLF

² par exemple : taille en octets des données stockées, des données référencées, nombre de tables, jeux de données, etc.)

Annexe : note sur la norme ISO 19115

Document tiré de : http://eden.ign.fr/std/iso_19115/index.html

La norme ISO 19115:2003 est la norme ISO de métadonnées pour l'information géographique

Une référence

La norme ISO 19115:2003 a le statut de norme internationale depuis 2003. Cette norme abstraite de contenu définit en les organisant par classes toutes les informations que l'on peut mettre à disposition pour décrire la donnée.

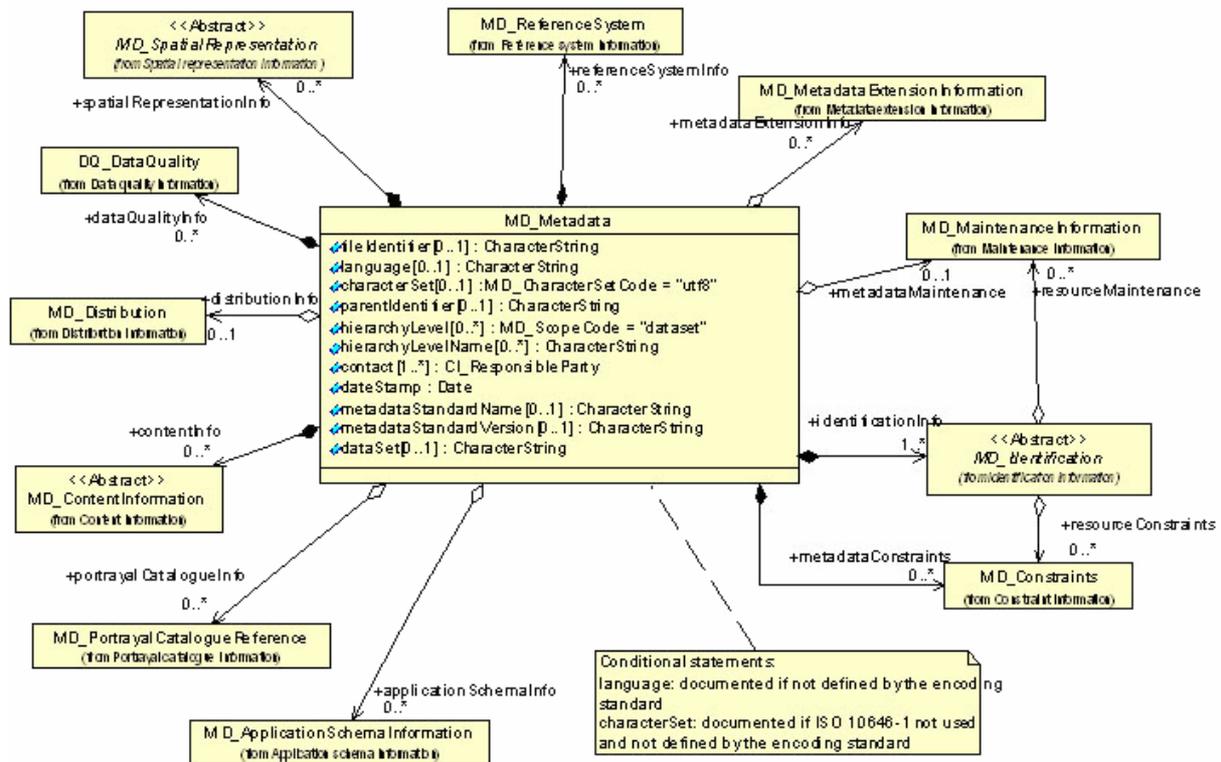
Cette norme s'est affirmée comme une référence pour l'information géographique dans le domaine des métadonnées. Outre le fait qu'elle est enfin disponible après une longue gestation, ses atouts principaux pour la communauté résident dans son caractère modulaire et extensible qui la rend aisément adaptable.

Contenu

Au sens de la norme ISO 19115, les éléments de métadonnées sont les suivants :

- Informations **d'identification**:
 - intitulé, description, dates de référence, version, résumé, intervenants, ...
 - L'étendue des données
 - Des aperçus sur les données
 - Informations sur les emplois possibles
 - Contraintes légales et de sécurité, ...
- La **description du contenu**
- Le **système de coordonnées**
- Les informations de **géolocalisation** et **d'organisation des données**
- Des **informations de qualité** (ou de qualification)
- Des **mesures de qualité**
 - précision géométrique, temporelle et sémantique,
 - exhaustivité,
 - cohérence logique
- Des informations de **généalogie** :
 - Description des sources
 - Description des processus appliqués aux sources
- Les modalités d'**affichage** (légendes)
- Les modalités de **diffusion**
- Les modalités de **maintenance**

Schéma de données



Mise en oeuvre

La norme ISO 19115 est une norme abstraite qui ne précise pas les modalités de sa mise en oeuvre, d'où un risque de non-interopérabilité. Pour cela, l'équipe Projet [19139](#) de l'ISO travaille à la définition d'une implémentation en XML de la norme 19115.

De plus, la norme ISO 19115 ne couvre pas les données image. Celles-ci seront prises en compte dans le cadre de la [19115-part 2](#). Par ailleurs elle ne prend pas non plus en compte les données de renseignement ni de simulation.