

 enregistrement	Bilan de l'opération Data Rescue pour le prototype Ecoscope	rédaction Bretaudeau Peggy	Date création : 27/02/2007	Référence: 21.013E/1
			Dernière modif. : 29/05/2007	Pages 4

Diffusion : Documents du projet, J. Le Fur, site web

Résumé

Ce document constitue un plan détaillé de la restitution de l'opération conservatoire de données pour le bilan du prototype Ecoscope. Il a été conçu à l'issue d'une concertation entre J.Le Fur et P.Bretaudeau vis-à-vis de cette action

Sommaire

1. Objectif et attendus	1
2. mises en œuvre d'un circuit des données	2
<i>a) mise en place d'un inventaire des données existantes (niv.0).....</i>	<i>2</i>
<i>b) circuit des données.....</i>	<i>2</i>
i) proposition initiale	2
ii) proposition pour le projet Ecoscope.....	2
<i>c) formulaires d'enregistrements des jeux de données.....</i>	<i>2</i>
<i>d) description et utilisation de la base.....</i>	<i>3</i>
<i>e) schéma entités-relations du catalogue (base Access)</i>	<i>3</i>
3. problèmes rencontrés (hétérogénéité contraignant le remplissage des champs).....	3
4. solution proposée (notion de 'nom d'usage').....	3

1. Objectif et attendus

Le « Data Rescue » (« conservatoire de données ») a pour but de rassembler les données du domaine halieutique de façon à les conserver dans le temps et à les rendre accessibles à la communauté scientifique.

L'objectif est donc de recenser et de capitaliser le plus grand nombre de jeux de données sur les écosystèmes marins exploités.

Un autre attendu du projet est de pouvoir documenter les protocoles d'acquisition des données et de définir les différents attributs et paramètres d'un jeu de données, afin qu'un scientifique puisse y accéder de façon autonome.

2. mises en œuvre d'un circuit des données

a) mise en place d'un inventaire des données existantes (niv.0)

Au cours d'une première réunion une ébauche d'inventaire a été obtenue (cf document de référence n° 21.002CR/1) et a servi de base à la constitution d'une liste provisoire de jeux de données à sauvegarder (avec différents degrés de priorité) présentée dans le fichier de référence n° 21.004E/1. Cette liste recense les jeux de données selon différents critères tels que l'écosystème, les partenaires, l'origine des données, la thématique, la période concernée, les droits d'utilisation etc. Une centaine de jeux de données ont été identifiés.

b) circuit des données

i) proposition initiale

Il a été nécessaire de définir les principes généraux de gestion des jeux de données. Ainsi plusieurs niveaux d'intégration ont été élaborés au cours de groupes de travail et ont donné lieu au document de référence n° 21.012E/1 contenant les points suivants :

- Description des niveaux d'intégration
- Une procédure normalisée qui décrit les critères de validation et les différentes étapes d'intégration.
- Description des champs indispensables à la définition d'un jeu de données

ii) proposition pour le projet Ecoscope

Dans la phase de prototypage, le projet Data Rescue est une composante indépendante de l'Ecoscope. Pour la seconde phase du projet Ecoscope, Le Data Rescue pourrait être intégré au circuit des savoirs. Les données du Data Rescue seraient alors traitées comme des savoirs tout en gardant la partie des traitements spécifiques aux données. Dans le circuit de l'Information, cela correspond à donner une particularité aux données lors des étapes de validation et de mises en forme des savoirs (cf document de référence n° 11.008P/1 – Synthétique du circuit de l'information).

c) formulaires d'enregistrements des jeux de données

Pour attester le passage des jeux de données d'un niveau à un autre, deux formulaires ont été constitués.

Le premier (cf document de référence n° 21.008F/1) est un accusé de réception d'un jeu de données (passage au niveau I) et le second (cf document de référence n° 21.009F/1) est une attestation de conformité des données (passage au niveau II).

d) description et utilisation de la base

La base de données contient des informations sur les jeux de données sauvés dans le cadre du module 2.1_entrepôt de données (Data Rescue). Le but de cette base est d'effectuer le suivi des opérations réalisées sur les différents jeux de données, de la réception à leur éventuelle mise à disposition.

La base contient les jeux de données à partir du niveau I (cf document de référence n° 21.010P pour l'explication des différents niveaux).

Le manuel d'utilisation est contenu dans le document de référence n° 21.012E

e) schéma entités-relations du catalogue (base Access)

Cette base de données est une base Microsoft Access.

Le schéma de base de données est décrit en Annexe 1 du document de référence n° 21.012E/1.

3. problèmes rencontrés (hétérogénéité contraignant le remplissage des champs)

La difficulté principale du projet est la définition des contours d'un jeu de données. Les fournisseurs de données donnent un nombre importants de fichiers concernant parfois de nombreux domaines (données de captures, données économiques, données environnementales, etc.) et pays. Il est donc difficile de cerner si il faut faire un découpage ou non. La notion de pays peut permettre de séparer les données. Quant aux différents domaines abordés, ils ne sont pas complètement indépendants les uns des autres donc le découpage s'avère plus délicats. Le travail avec les thématiciens et fournisseurs de données est à ce point crucial.

La séparation ou découpage des jeux de données doit se faire entre le passage du niveau I au niveau II mais peut difficilement avant.

4. solution proposée (notion de 'nom d'usage')

Une des solutions au problème posé au point 4, est la notion de nom d'usage.

Les jeux de données, lorsqu'ils sont transmis à l'équipe du « Data Rescue » est défini par un nom global qui correspond parfois à la personne qui a récolté les données répertoriés ou à un nom donné par les utilisateurs des données. Ce nom n'était pas conservé en tant que référence mais dans le descriptif général du jeu de données.

La notion de nom d'usage permet de garder cette information et de s'en servir comme référence. Ainsi, au niveau I de conservation des données, il n'est plus nécessaire de définir des contours stricts aux jeux de données. Cela ne veut pas dire qu'il ne faudra pas le faire. Cependant, cela permet de le faire lors du passage au niveau II voire au niveau III sans remettre en cause la conservation des jeux de données de niveau I.

Rédaction : Peggy bretaudeau / saisie : PB / corrections : JLF