

# **Manuel technique pour l'implémentation de la metabase de données sur le développement durable**

Version 1.01

**Services fédéraux des affaires Scientifiques, Techniques et Culturelles (SSTC).**

**Ir. Bruno Kestemont\* et Ir. Dirk Le Roy\*\***  
sous la direction de:  
**Prof. Walter Hecq\* et Dr.Ir. Paul Vanhaecke\*\***

**\*CEESE-ULB**, 44 avenue Jeanne, CP 124, B-1050 Bruxelles

**\*\*ECOLAS N.V.**, Lange Nieuwstraat 43, B-2000 Antwerpen

# TABLE DES MATIERES

<b>TABLE DES MATIERES</b>	<b>2</b>
<b>MANUEL TECHNIQUE POUR L'IMPLÉMENTATION DE LA METABASE DE DONNÉES SUR LE DÉVELOPPEMENT DURABLE.</b>	<b>4</b>
<b>Introduction</b>	<b>4</b>
<b>Activités</b>	<b>4</b>
<b>Structure de la base de données.</b>	<b>7</b>
<b>Mode de représentation de la fiche-type.</b>	<b>7</b>
<b>Fiche type</b>	<b>8</b>
données techniques:	8
type de source:	10
titre:	10
Objectifs:	11
Contenu:	11
Couverture géographique:	12
Période couverte et fréquence des données (si applicable):	14
Qualité:	14
Référence éventuelle du produit, ou copyright:	14
Versions	15
Identificateur unique de référence pour la source	16
Accès direct à la source ou à sa description (sites miroirs)	17
Mode de construction, sources ou références utilisées	17
Politique de distribution:	19
Déontologie:	20
Volume approximatif de la source:	20
Organisation ou personne responsable pour la gestion de la source	20
Source des metadonnées (qui a rempli ce questionnaire?)	20
Remarques et autres:	21
<b>Exemple de fiche simplifiée pour les organisations parentes.</b>	<b>22</b>
<b>Exemple de fiche pour décrire la source des metadonnées</b>	<b>23</b>
<b>Codification de la meta-banque de données.</b>	<b>24</b>
<b>Remarques techniques sur chaque champs (thesaurus utilisés, ...)</b>	<b>29</b>
Considérations générales.	29
données techniques:	32
type de source:	33
titre:	35
Objectifs:	36
Contenu:	36
Couverture géographique:	38
Période couverte et fréquence des données (si applicable):	40
Qualité:	41
Référence éventuelle du produit, ou copyright:	41
Versions	41
Identificateur unique de référence pour la source	42
Accès direct à la source ou à sa description (sites miroirs)	44
Mode de construction, sources ou références utilisées	44
Politique de distribution:	48
Déontologie:	49

Organisation ou personne responsable pour la gestion de la source	49
Source des metadonnées (qui a rempli ce questionnaire?)	50
Remarques et autres:	50
<b>Explication d'une solution informatique opérationnelle (logiciel et tâches techniques).</b>	<b>51</b>
Formats et jeux de caractères	51
Logiciels	51
Matériel informatique	53
Accès au serveur WWW	54
Changements futurs de configuration	54
<b>Standards et programmes pouvant être utilisés dans certains domaines spécifiques</b>	<b>55</b>
Global change, Earth Observation	55
Earth Observation, outil futur	56
Environnement, outils futurs	56
Suivi des initiatives analogues	56
<b>Limitations pour le passage d'un standard à un autre et problèmes d'exportation des données vers d'autres standards.</b>	<b>56</b>
<b>Où trouver le thesaurus GEMET, et les autres standards?</b>	<b>57</b>
<b>Comment structurer les documents-maîtres?</b>	<b>58</b>
<b>Création d'indexes</b>	<b>59</b>
<b>Création de meta-tags</b>	<b>59</b>

# **Manuel technique pour l'implémentation de la metabase de données sur le développement durable.**

## **Introduction**

Ce manuel s'adresse au gestionnaire de la metabanque de données sur le développement durable, basé au SSTC, et oriente ses premiers mois de travail. La fiche-type et les annexes techniques peuvent également être utilisées par tout organisme désirant utiliser le même standard pour la constitution ou la traduction de son catalogue interne.

## **Activités**

1. S'équiper du matériel nécessaire (PC haut de gamme en matière de traitement de texte e de base de données).
2. En accord avec les principaux chefs de programme, standardiser la récolte de metadonnées
  - si de nouveaux questionnaires ne sont plus envisagés, demander une copie digitale des documents contenant les metadonnées (propositions de projets par exemple). Si nécessaire, procéder à une digitalisation par scanning et OCR (reconnaissance de caractère).
  - si de nouveaux questionnaires sont envisagés (par exemple si chaque chef de programme veut indexer ses projets en vue d'une publication spécifique):
    - récolter la liste des mots-clés (descripteurs) spécifiques à chaque programme auprès des chefs de programmes (ou leur proposer des descripteurs en provenance du GEMET (Thésaurus multilingue de l'environnement) - . Faire au besoin une table de conversion vers le GEMET et utiliser si possible ce dernier dans les formulaires.
    - réaliser un avant-projet de questionnaire et le discuter avec les chefs de programme.
    - fournir à chaque chef de programme formulaire et descripteurs spécifiques (sur papier et disquette).
3. Sonder les metadonnées existantes (IWETO, CORDIS, FEDRA, autres programmes SSTC dont les metadonnées sont déjà encodées, WWW-Internet).
4. Importer les metadonnées existantes en construisant des scripts ou macros (sous WordPerfect par exemple) capables de remplacer les noms de champs et la structure d'origine par les noms de champs et la structure standard (voir tableau 1).
5. Editer manuellement les fiches:
  - Ajouter aux mots-clés d'origine des mots-clés standards. Cette opération est semi-automatique (remplacements) mais aussi manuelle.
  - Ajouter les champs Source des Metadonnées (IWETO, CORDIS etc.) et Date d'enregistrement (d'origine).

-Corriger les erreurs et supprimer les doublons (par recherche de chaînes de caractères par exemple).

rem: en réalisant ces opérations, il faut réfléchir à une manière de les automatiser. Non seulement l'automatisation permettra une mise à jour plus rapide, mais dans le futur, un outil de recherche sur plusieurs sites 'au vol', style SavvySearch sur <http://guaraldi.cs.colostate.edu> sera implémenté.

5a. Parallèlement et en guise d'alternative possible à la gestion de données sous format textuel (html) pour le document maître, créer un fichier Access sur base d'un seul tableau pour la fiche-type et une série de tableaux liés pour gérer la multiplication possible des champs ainsi que les liens inter-fiche (un tableau pour gérer les liens et leur qualité). Si cette voie est choisie, les champs textuels de longueur variable ainsi que les champs peu courants (importés) seront renvoyés à des champs 'memo'. Il faut garder en tête les fonctionnalités finales du système qui doit rester ouvert, extensible et rempli de liens hypertextes de fiches à fiches. Pour la gestion des langues, les tableaux liés prévoieront un champs indiquant la langue de chaque réponse. La création d'un formulaire délimité sous Word ('fusions et publipostages') peut être une alternative avantageuse sinon préférable.

6. Récolter les questionnaires remplis. Chaque chef de programme est supposé digitaliser lui-même ses metadonnées.

7. Ajouter à la metabase de données existante.

8. Par une succession de 'remplacer' dans le(s) documents maîtres, remplacer les objets (sources, noms, ...) par leur équivalent accompagné d'un lien hypertexte. D'une part, la plupart des liens hypertextes seront définis par une référence au sein même de la meta-banque de données (fiches au format standard), d'autre part, chaque fiche pointerà vers un ou plusieurs liens externes (champs URI). Par exemple, si une fiche générale décrit la KUL, chaque fois que la KUL apparaîtra dans une des autres fiches

9. Développer des modes de présentation alternatives des metadonnées en langage html (navigation par indexes divers) et des extraits pour publication sur papier (par exemple projets d'un même programme SSTC) ou pour téléchargement dans quelques formats courants (exemple: .pdf). Il est important de présenter dès le début les données déjà récoltées, de manière à susciter l'intérêt des fournisseurs de données. Les données peuvent être présentées par paquet d'informations relatives à chaque programme.

10. Développer un questionnaire en ligne (WWW) pour metadonnées candidates (URLs par exemple). Le but de ce questionnaire est d'identifier des sources importantes potentielles non prévues par d'autres biais. Pour faciliter ce travail, des questionnaires existants seront utilisés et éventuellement transformés. Par exemple, un créateur d'entrées DIF (très proche de ce que nous proposons) est accessible à l'URL <http://gcmd.gsfc.nasa.gov/difweb.html>.

Un logiciel d'entrée DIF pour Windows peut être obtenu à l'URL [www.smithsys.co.uk/dif](http://www.smithsys.co.uk/dif).

On peut proposer à l'IDN une traduction de l'interface de ces outils vers le français et le néerlandais. En outre, on devrait leur proposer une méthode pour l'enregistrement de metadonnées textuelles en français/néerlandais. Par exemple l'ajout d'un champs

DIF\_language (le défaut étant l'anglais) ou la possibilité d'ajouter des champs en fonction de leur langue, en utilisant le code ISO pour la langue (exemple Title\_fr: pour le champs Title:).

11. Contacter les institutions suivantes pour la description active de sources de données importantes non encore répertoriées par ailleurs. Il faut négocier avec chacun de ces instituts la manière dont peut se faire les inventaires: soit les instituts réalisent les inventaires eux-mêmes, soit (solution la plus probable), il faut se procurer leurs catalogues et/ou décrire les sources de données soi-même et/ou envoyer des questionnaires (éventuellement pré-remplis grâce aux métadonnées déjà collectées par ailleurs) aux responsables de service.

- Tous les instituts fédéraux (INS, IRScNB, IRM, ING, Bibliothèque Royale)
- administrations fédérales compétentes (Bureau du Plan, Environnement, ...)
- Point Focal Belge de l'Agence Européenne de l'Environnement (qui devrait logiquement collaborer à cet inventaire)
- CNDD
- Instituts et administrations régionales (DGRNE, IBGE, VMM etc.)
- Conseils régionaux du développement durable

De fil en aiguille, d'autres institutions pourraient être sondées pour combler des manques importants d'informations:

- certains centres universitaires (identification des thèses de doctorat par exemple);
- bureaux d'étude;
- ONGs;
- autres organismes privés (entreprises, associations professionnelles).

Une liste d'institutions fournissant des données a été constituée lors de l'étude de faisabilité.

12. Le travail peut être segmenté de la manière suivante:

- récolte tous azimuts des métadonnées facilement importables, ainsi que des métadonnées générales;
- extraits thématiques suivant les grands axes du programme (changement climatique, transport, etc.) et approfondissement de chacun de ces thèmes les uns après les autres ou en parallèle; chaque fois qu'un thème est suffisamment documenté, une publicité spécifique est opérée, comme si il s'agissait d'une meta-base de données indépendante ('Meta-banque de données belges sur le changement climatique' par exemple). Dans certains cas, il est même possible de décentraliser la gestion des domaines spécifiques (exemple: mobilité durable). Le principe de la Virtual Library peut être adopté, des volontaires pouvant se charger de l'identification initiale des sources spécifiques. Pour les sources régionales, cette manière de procéder est sans doute la plus acceptable.

13. Progressivement, la fonction doit évoluer vers un carrefour d'informations détenues et mises à jour par des tiers (volontaires ou institutions compétentes, comme CELINE ou les Régions). L'étape de récolte initiale d'informations a pour but essentiel de banaliser (et démontrer) l'usage d'un standard d'échange commun, et de permettre au gestionnaire de la meta-banque de données de se doter de l'expérience et des outils nécessaires à la constitution

d'un carrefour réel d'informations, où les informations ne sont pas stockées mais construites 'au vol' à partir de sources extérieures grâce à l'usage de 'traducteur' appropriés.

14. Développement des outils de recherche sur des sources de métadonnées distribuées (recherche inter-catalogues, recherche de pages WWW pré-indexées, ...).

## **Structure de la base de données.**

On se reportera à l'étude de faisabilité (Kestemont et al, 1996), chapitre 6 pour les principes généraux de respect des standards existants. La structure présentée ici sera utilisée au démarrage de l'activité.

## **Mode de représentation de la fiche-type.**

La fiche type (standard d'échange) est présentée sous forme d'un formulaire utilisable tel quel. Cependant, dans la pratique, ce formulaire sera adapté en fonction:

- du type de source (supprimer les questions non pertinentes)
- du thème traité (ne fournir que les mots-clés pertinents).

Le format suivant est utilisé:

### Titre de section souligné:

n'a pour but que de présenter les champs par groupes logiques afin d'accroître la lisibilité du questionnaire.

### *Remarques en italique:*

aide à l'utilisateur du questionnaire.

**QUESTION:** questions posées (correspondant le plus souvent à un champs dans la base de données). Par défaut, toute question peut avoir plusieurs réponses, de longueur libre mais de préférence courte.

**\*QUESTION MARQUEE D'UNE ETOILE:** une seule réponse admise (le champ ne peut pas être répété)

**QUESTION EN CARACTERE GRAS:** réponse minimale pour constituer une fiche séparée.

case à cocher suivie de « : »: on peut cocher et proposer librement un sous-descripteur

case à cocher simple : cocher seulement (les notes ne seront pas encodées)

Les questions sont numérotées temporairement à l'usage du gestionnaire de base de donnée pour renvoi à la table des noms de champs présentée plus loin.

Les sous-séparateurs à utiliser en cas de plusieurs réponses à un même champ devraient être clairement identifiables: point-virgule pour séparer des mots-clés, auteurs ou noms, retour à la ligne et tiret '-' pour séparer des éléments de listes.

## **Fiche type**

Seules les questions posées sont représentées ici, les codes sous-jacents faisant l'objet du chapitre suivant.

*Remarque générale: en tout lieu du formulaire, on peut spécifier un lien hypertexte vers un document existant sur internet ou vers un formulaire déjà rempli dans cet inventaire (exemple: dans 'source de données parente', on peut mettre l'URL d'une description de cette source).*

*Les questions en caractère gras doivent avoir une réponse impérativement. Pour les autres questions, ne répondre que si elles s'appliquent à la source décrite.*

*Les questions marquées \* ne peuvent avoir qu'une réponse (sans compter les traductions bien sûr). Toutes les autres questions peuvent avoir plusieurs réponses (par exemple AUTEUR est au singulier mais on peut lister plusieurs auteurs dans leur ordre logique de publication en les séparant par un point-virgule).*

*Les questions ou choix multiples se terminant par deux points ':' indiquent qu'on peut spécifier par un texte libre en regard du choix.*

*En cas de réponse dans le formulaire sous forme digitale, cocher vos choix en double-cliquant sur la case pour la remplacer par un 'x' (devant la ou les réponses choisies).*

*N'hésitez pas à fournir des informations supplémentaires en regard de chaque question, en annexe ou via une référence présente ou non sur Internet.*

*Si vous avez déjà décrit cette source dans d'autres questionnaires de même type, envoyés par des tiers (CORDIS, AEE, etc.), inutile de recommencer: renvoyez-nous simplement une copie de ce questionnaire, et ne répondez qu'aux questions qui vous semblent apporter un complément d'informations (par exemple, la traduction du nom officiel ou du résumé dans une autre langue).*

### **données techniques:**

**1.1.\*NUMERO D'ORDRE** (code auquel le répondant peut faire référence sur les documents éventuellement annexés: traductions, illustrations, copies d'autres formulaires. Il est recommandé d'utiliser une chaîne courte de caractères permettant d'identifier uniquement un objet décrit, comme une série d'acronymes en ordre hiérarchique, par exemple: ULB\_CEESE\_CDS ou SSTC\_FEDRA).

### **1.2.\*LANGUE DU FORMULAIRE** (langue utilisée par défaut dans les réponses)

anglais:      néerlandais:      français:      allemand:      autre:





type de source:

**2.1.TYPE DE SOURCE** (*cocher une ou plusieurs cases et spécifiez au besoin*)

*Pour les produits, les questions sur le contenu et les formats permettent une délimitation plus précise.*

institution:

académique, université, centre de recherche  
administration fédérale      communautaire      régionale      locale  
organisation internationale  
organisation non gouvernementale (ONG)  
consultants, bureaux d'étude privé  
autre privé (industriel, commercial, ...)  
responsables politiques

autre:

activité, programme, projet:

produit, base de donnée, publication:

station:

personne physique:

autre:

titre:

**\*3.1.NOM OU TITRE:** (*des traductions officielles sont ici recommandées. Un nom par langue*)

(si une personne:

3.1.1.TITRE: \*3.1.2.NOM DE FAMILLE:      3.1.3.PRENOM:      )

3.2..ACRONYME:

3.3.SOURCE DE DONNEES PARENTE: (*base de données, série, programme dont elle est éventuellement une partie*). *On peut remplir une fiche minimale pour cette référence (voir plus bas).*

3.4.REMARQUES EVENTUELLE SUR L'ORGANISATION INTERNE DE LA SOURCE DE DONNEES:

3.5.LIEN AVEC UNE AUTRE SOURCE DE DONNEES:

*Spécifier le type de lien et le nom des sources, ainsi qu'un élément d'identification, comme l'URL par exemple.*

Objectifs:

4.1.OBJECTIF:

4.2.DESCRPTION GENERALE (RESUME):

4.3. UTILISATION OU SERVICE RENDU:

Contenu:

5.1.DOMAINES GENERAUX TRAITES, MOTS-CLES:

*(Les mots-clés généraux d'un Thésaurus Multilingue courant sont présentés par défaut en fonction de l'inventaire spécifique, mais le répondant peut utiliser des mots-clés (descripteurs) provenant d'un autre système de classification, qu'il doit alors spécifier). Exemple minimum:*

environnement:

économie:

sociologie:

institutions:

autre:

5.2.PARAMETRES, UNITES ET METHODES:

*(brève description du produit, reprenant surtout les éléments de détail décrivant le mieux la quantité et la qualité des données ainsi que leur désagrégation. Si possible, en cas de classification, lister les classes. Joindre en annexe une illustration):*

*[exemple: Emissions de 4 polluants atmosphériques - NOx, SO2, N2O et CO2 - en tonnes par an par 44 secteurs économiques et 33 carburants dans 2 pays (Pays-Bas et Belgique) sur une période de 10 ans (1980-1990). Calculé en multipliant des facteurs spécifiques d'émission par les utilisation d'énergie finale. Suit la liste des secteurs et carburants: ...].*

**5.3.ILLUSTRATION EN ANNEXE** (*donner une référence*):

*(illustration ou brochure descriptive pour le produit, telle que carte, graphe, exemple de fiche, modèle de données, formules, photo, etc. sous forme digitale en standard d'échange GIF, TXT ou HTML ou maximum 2 pages A4. Pour une illustration fournie sur support digital, spécifier le chemin d'accès et le nom de fichier)*

**5.4.STANDARD UTILISE:**

Couverture géographique:

**6.1.COUVERTURE GEOGRAPHIQUE:** (*à laquelle se rapporte la source*)

*Des descripteurs standards sont proposés, des mots-clés libres peuvent être ajoutés.*

*Administrative:*

*spécifier (utiliser les codes ISO et NUTS si la place manque)*

Internationale:

Nationale:

Régionale:

Locale:

*Non administrative:*

Cours d'eau:

Plans d'eau:

Océans et mers:

Côtes:

Bassins de drainage:

Zones urbaines:

Zones industrielles:

Zones rurales:

Autres: (voir DIF location keywords, par discipline: astronomie, sciences de la terre, sciences planétaires, physique solaire, physique de l'espace)

Zone géographique couverte (*rectangle, volume, ou point central*)

*Les données qui suivent délimitent un volume, une surface ou un point sur la mappemonde, dans laquelle se situe l'aire de spécialisation de la source décrite. Pour un point par exemple, les maxima et minima sont confondus.*

6.2.LONGITUDE LIMITE OUEST *ddd mm ss E/W*:

6.3.LONGITUDE LIMITE EST *ddd mm ss E/W*:

6.4.LATITUDE LIMITE SUD *ddd mm ss N/S*:

6.5.LATITUDE LIMITE NORD *ddd mm ss N/S*:

6.6.ALTITUDE MIN (*mètres, positif au-dessus du niveau de la mer, négatif pour la profondeur*):

6.7.ALTITUDE MAX (*mètres, positif au-dessus du niveau de la mer, négatif pour la profondeur*):

6.8. EPAISSEUR MIN (*mètres*):

6.9. EPAISSEUR MAX (*mètres*):

6.10. ALTITUDE/STRATE (*mots clés*):

6.11.ZONE GEOGRAPHIQUE COUVERTE (FORMAT LIBRE):

*On peut également spécifier une zone plus complexe par une série de coordonnées d'objets (points, lignes, polygones, volumes) sous le format suivant(unités: degrés, mètres): (LAT11, LONG11, ALT11)(LAT2, LONG2, ALT2)...(LAT1n, LONG1n, ALT1n)...(LAT11, LONG11, ALT11). Il faut fermer la boucle en répétant la première valeur s'il s'agit d'un polygone*

6.12.RESOLUTION SPATIALE OU ECHELLE:

*Format libre [ex: '1:1000000', '25 ha/pixel', 'arrondissements', 'pays', 'villes de plus de 10000 habitants', 'unités de plus de 2 ha', 'résolution de 10 mètres' etc.]*

Période couverte et fréquence des données (si applicable):

*(format YYYYMMDD hhmmss, heures GMT, ou spécifier le fuseau horaire)*

7.1. TEMPS D'ACQUISITION *(pour une image par exemple):*

7.2.PERIODE DE:                      7.3.A:

*(Pour une série de données, deux temps sont donnés. En cas de périodes multiples, il faut répéter ces deux champs par paire)*

7.4. PERIODE *(format libre):*

7.5.FREQUENCE DES DONNEES

continue

régulière, tou(te)s les ..... secondes

(nombre) minutes

heures

jours

mois

ans

irrégulière

autre:

Qualité:

8. QUALITE:

Référence éventuelle du produit, ou copyright:

9.1.AUTEUR *(répéter le champs si plusieurs):*

9.2.ORIGINE *(organisation):*

9.3.DATE DE PUBLICATION:

9.4.SERIE:

9.5.EDITEUR:

9.6.PLACE DE PUBLICATION:

Versions

\*10.1.DATE DE CREATION:

10.2.FREQUENCE DE MISE A JOUR:

continue

régulière, tou(te)s les ..... secondes  
(nombre) minutes

heures

jours

mois

ans

irrégulière

autre:

\*10.3.DERNIERE MISE A JOUR:

10.4.PROCHAINE MISE A JOUR:

10.5.STATUT: en construction: fixée:

*(la source de données est-elle déjà ou encore enrichie ou est-elle fixée?)*

10.6. VERSION:

### 11.0.1.

#### Identificateur unique de référence pour la source

*Cette section est destinée à identifier la 'source-mère', référence unique dans des bases de données ou sur internet, où se trouve l'original de la source ou de sa description officielle. Plusieurs types de références uniques sont possibles sur différents supports. Ne donner qu'une seule référence par type, les autres copies pouvant être indiquées à la question suivante. Les autres identificateurs peuvent être des identificateurs uniques dans des metabases de données existantes.*

*Le cas échéant, indiquer des remarques (ex: login, password, configuration)*

#### Adresse

- \*11.1.1.RUE/N°/BOITE POSTALE:
- \*11.2.1.CODE POSTAL:
- \*11.3.1.VILLE:
- \*11.4.1.PAYS:
- \*11.5.1.LOCALISATION (*longitude-latitude*):
- \*11.6.1.CARTE D'ACCES (*joindre carte ou fichier*):

#### Téléphone

- \*12.1.1.TELEPHONE (voix):
- \*12.2.1.FAX:
- \*12.3.1.MODEM:

\*13.1.1.URL:

\*13.2.1.E-MAIL:

13.3.1.AUTRE INTERNET (*spécifier, une seule adresse par type supplémentaire*):

*autres*

\*14.1.1.ISBN:

\*14.2.1.ISSN:

15.1.1.AUTRE (*spécifier, une seule adresse par type supplémentaire*):



Accès direct à la source ou à sa description (sites miroirs)

*Indiquer ici les différents endroits où la source peut être consultée directement (plusieurs réponses possibles par type).*

*Le cas échéant, indiquer des remarques (ex: login, password, configuration)*

Adresse *(si plusieurs adresses, répéter l'ensemble du groupe)*

11.1.2.RUE/N°/BOITE POSTALE:

11.2.2.CODE POSTAL:

11.3.2.VILLE:

11.4.2.PAYS:

11.5.2.LOCALISATION:

11.6.2.CARTE D'ACCES *(joindre carte ou fichier):*

Téléphone

12.1.2.TELEPHONE (voix):

12.2.2.FAX:

12.3.2.MODEM:

13.1.2.URL:

13.2.2.E-MAIL:

13.3.2.AUTRE INTERNET *(spécifier):*

*autres*

14.1.2.ISBN:

14.2.2.ISSN:

15.1.2.AUTRE *(spécifier):*

Mode de construction, sources ou références utilisées

16.1.SOURCE

16.2.METHODE DE CONSTRUCTION:

16.3.FORME CONTENU *(cocher une ou plusieurs options et spécifier):*

données factuelles, tableaux, diagrammes

textes

graphes, dessins, images, cartes

films, vidéo, animations

échantillons, pièces de collection, maquettes, ...

autres:

#### 16.4.LANGUE DISPONIBLE (*plusieurs réponses possibles*)

Néerlandais

Français

Allemand

Anglais

autres ou remarques:

#### 16.5.JEU DE CARACTERES:

#### 16.6.SUPPORT (cocher et spécifier les formats disponibles)

digital:

disquette:

CD-ROM:

bande magnétique:

en Ligne (Internet, ...):

disque dur:

analogique:

papier (manuscrit, imprimé, photographie):

disque

transparent (négatif, microfilm, film, ...):

bande magnétique:

‘en Ligne’ (Télétexte, téléphone, TV, ...):

ondes (TV, radio, gsm...):

échantillon:

maquette, ...:

autre:

Politique de distribution:

17.1.DISTRIBUTION

seulement usage interne:

usage externe:

gratuit:

payant:

limité (préciser):

Déontologie:

18.1.AVEZ-VOUS DEFINI DES REGLES ?

- pour la récolte des données
- pour le traitement des données
- pour l'accès aux données
- pour la publication et la dissémination des données
- pour le paiement d'informations ou services
- autres

18.2.TEXTE EN ANNEXE (*si oui, donner une référence*)

Temps d'accès:

19.HORAIRE D'ACCES (*en temps local, ou préciser*):

Volume approximatif de la source:

20.VOLUME APPROXIMATIF (nb d'enregistrements, de pages, de Kbs, de valeurs, de membres du personnel d'une organisation, budget annuel, ...)

Organisation ou personne responsable pour la gestion de la source

21.RESPONSABLE:

(*donner un URI*)

*Référence à un questionnaire déjà complété:*

*Le cas échéant, remplir un questionnaire pour l'organisation en question. Voir exemple de questionnaire minimum pour les organisations ci-dessous.*

*D'autres liens peuvent être définis à divers endroits de la fiche.*

Source des metadonnées (qui a rempli ce questionnaire?)

*Prévoir un lien URL ou une fiche minimum séparée pour le documentaliste et/ou son organisation, ou pour la source de metadonnées déjà publiées (voir fiche minimale ci-dessous)*

*Personne, organisation, metabase de donnée, catalogue ou/et document ayant servi de source pour les informations fournies dans ce questionnaire ... Donner des éléments permettant d'identifier les fiches décrivant les sources mentionnées.*

22.2.DATE:

*Date de dernière mise à jour de cette fiche.*

22.3.REMARQUE SUR LA SOURCE DES METADONNEES:

22.4.AUTRES META INFORMATIONS:

Remarques et autres:

23.REMARQUES ET AUTRES INFORMATIONS (*vous pouvez recopier ici, ou joindre des extraits d'autres questionnaires auxquels vous avez répondu pour la même source*)

***(FIN de la fiche de base.)***

## **Exemple de fiche simplifiée pour les organisations parentes.**

(l'organisation parente peut être décrite une fois pour toute et être rappelée par son nom par la suite)

NOM

RUE/N°/BOITE POSTALE

CODE POSTAL

VILLE

CATEGORIE (*il s'agit ici d'un sous-groupe du type de source: institution*)

académique, université, centre de recherche

administration fédérale      communautaire      régionale      locale

organisation internationale

organisation non gouvernementale (ONG)

consultants, bureaux d'étude privé

autre privé (industriel, commercial, ...)

responsables politiques

autre:

personne de contact:

TITRE

NOM

PRENOM

TELEPHONE

FAX

E-MAIL

Organisation parente ou hôte éventuelle (si l'institution ci-dessus fait partie d'une entité plus large)

*Identificateur unique éventuel: URL:*

NOM:

## **Exemple de fiche pour décrire la source des metadonnées**

(le documentaliste ou la source de metadonnées peut être décrite une fois pour toute et être ensuite référencé automatiquement dans chaque fiche fournie).

Source de la metainformation (personne ou institution qui a rempli le formulaire) (dernière édition)?

TITRE

NOM

PRENOM

FONCTION

ORGANISATION

TELEPHONE

FAX

E-MAIL

DATE

Autres metainformations

*(libre à vous d'ajouter des champs descriptifs de votre source de données, si vous le jugez indispensable, et en faisant si possible référence à des standards existants).*

## **Codification de la meta-banque de données.**

Les noms ou codes donnés aux champs descriptifs n'ont aucun intérêt thématique et peuvent être donnés par le gestionnaire de la banque de donnée une fois qu'elle sera mise sur pied. Cependant, il est intéressant d'utiliser des noms ou codes repris dans des standards existants, notamment parce que cela peut faciliter les échanges ultérieurs. La règle est donc de faire référence à au moins un standard quand c'est possible.

La méthode à suivre peut s'inspirer de l'annexe 8 (exemple de meta-informations dans le header d'une page html).

### **Règle pour les noms de champs (META NAME)**

Nous proposons d'adopter la règle suivante:

Les standards de metadonnées émergeant sur le WWW (Dublin core, IAFA, Meta-Tags html) devraient être adoptés en priorité.

Si le champ n'est pas standardisé à ce niveau, les standards suivants seront adoptés dans l'ordre: CDS-EEA quand il existera, DIF (dernière version), FGDC (ou ISO quand il existera), GILS, US-MARC.

Provisoirement, un nom de champs libre peut être utilisé, par exemple en cas d'importation de données, mais ce n'est pas recommandé.

Le tableau 2 montre une proposition de noms de champs standards (note: le prototype utilise surtout des noms de champs non standard).

Les numéros correspondent aux questions du questionnaire "Fiche type" ci-dessus pour des raisons de lisibilité. Ces numéros ne font pas partie du standard. Moyennant l'utilisation d'un standard existant, tout type de champs peut être ajouté, de même que les champs non marqués d'un \* peuvent être répétés.



**Tableau 2: Noms de champs proposés pour les principales questions de la fiche type.**

**Tableau 2 (suite):**

### **Spécification de la langue du contenu du champs**

Les standards actuels supposent l'utilisation de l'anglais (ou d'un code) dans le contenu des champs. Pour représenter une autre langue pour les informations fournies dans un champ, il faut le spécifier en utilisant une des trois méthodes suivantes:

- ajouter un suffixe au nom de champs (l'absence de suffixe indiquant l'anglais par défaut), par exemple: "NAME fr:" si le nom est donné en français; c'est la méthode utilisée pour le prototype;

- ajouter un qualificatif (possible en html), par exemple:

META NAME="Keywords" LANGUAGE="French" CONTENT="liste en français"

- utiliser le champ "Langage du formulaire" comme langue par défaut pour toutes les metadonnées. Dans ce cas, il est impératif de construire une fiche séparée pour chaque traduction. Cette solution sera retenue s'il n'y a pas de possibilité d'accord international sur une méthode de gestion du multilinguisme.

### **Codification des mots-clés utilisés au sein des champs eux-même.**

La règle est la suivante:

- priorité aux codes utilisés dans le standard choisi pour le nom du champ; cependant, d'autres listes peuvent être utilisées à condition d'en donner la référence;

- pour le thésaurus-sujet général: Thésaurus Général Multilingue de l'Environnement quand il existera (Milieuthésaurus en attendant) devrait être utilisé en priorité; les autres thésaurus ou listes de codes qui peuvent également être utilisés sont INFOTERRA pour les matières environnementales, le Thésaurus de la Library of Congress (US-MARC) ou le Thésaurus EUROVOC pour les matières non couvertes par les précédents. Des systèmes de mots-clés plus spécifiques sont également autorisés à condition d'en donner la référence (numéro CAS pour les produits chimiques, listes d'espèces, codes NUTS etc.). Enfin, des mots-clés spécifiques à un sous groupe particulier du SSTC sont admis si jugés indispensables par des groupes d'utilisateurs. Une référence au SSTC devra dans ce cas les accompagner.

Tout mot-clé dont aucune source n'est spécifiée est considéré comme mot-clé libre.

Lors d'utilisation de bases de données sous format textuel (champ variable), il est possible de donner des mots-clés faisant appel à plusieurs systèmes de référence au besoin. La nomenclature suivante est par exemple utilisée sur Internet (mais le standard html pas encore fixé):

META NAME= "keywords" CONTENT="liste de mots clés"

pour des mots-clés libres (par défaut), ou

META NAME= "keywords" SCHEME="USMARC" CONTENT= liste de mots clés ou de codes"

pour donner la référence à un thesaurus existant (ici Library of Congress).

## **Remarques techniques sur chaque champs (thesaurus utilisés, ...)**

Cette section donne pour chaque champs des caractéristiques techniques, comme les formats et longueurs éventuelles des champs, ainsi que les thesaurus par défaut ou recommandés.

### Considérations générales.

Rappelons que chaque champs peut être répété autant de fois que nécessaire, à l'image du standard d'échange DIF et de la plupart des logiciels de gestion documentaires. En cas de valeurs multiples, le standard DIF recommande de répéter le nom du champs, et de ne lui faire succéder qu'une valeur. Par exemple:

Keyword: water

Keyword: quality

Cependant, pour les champs contenant en général des valeurs multiples, nous recommandons l'usage d'un champs unique au pluriel, suivi d'une série de valeurs séparées par un délimiteur déterminé (virgule pour les mots-clés). Par exemple:

Keywords: water, quality

Il est vivement recommandé de ne pas dépasser 31 caractères pour les noms et acronymes figurant dans des champs uniques (marqués d'un \*), ce pour faciliter une exportation éventuelle au format DIF.

Le format DIF se base sur des lignes de 80 caractères maximum. En cas d'exportation au format DIF, les champs dépassant 80 caractères seront coupés automatiquement tous les 80 caractères, le nom du champ étant répété en début de chaque ligne.

*En tout lieu du formulaire, on peut spécifier un lien hypertexte vers un document existant sur internet ou vers un formulaire déjà rempli dans cet inventaire (exemple: dans 'source de données parente', on peut mettre l'URL d'une description de cette source).*

Cette remarque est importante et devrait être appliquée autant que possible dans la mesure où la base de données est conçue comme une série de fiches reliées entre elles par des liens hypertextes.

*Les questions en caractère gras doivent avoir une réponse impérativement.*

Cette règle est plus un conseil qu'une obligation. Il est en pratique possible de décrire ou de référencer un objet sans y donner un nom ou un titre, mais cette méthode n'est pas élégante. En pratique, l'existence d'une fiche commence dès qu'elle contient quelque chose de lisible et une référence permettant de la connecter (adresse URL par exemple), le nom de l'objet décrit peut toujours être ajouté ou remplacé par la suite. En cas d'exportation vers d'autres formats d'échange (DIF par exemple), les fiches sans nom seront ignorées.

*Pour les autres questions, ne répondre que si elles s'appliquent à la source décrite.*

Certains champs n'ont aucune raison d'être pour certains types d'objets. Ils seront omis dans la fiche décrivant l'objet, et quand c'est possible, dans le questionnaire lui-même. Les inventaires sont en effet le plus souvent dirigés vers un type particulier d'objet (par exemple des bases de données). Dans ce cas, il est inutile d'afficher les champs spécifiques aux personnes (NOM, PRENOM, TITRE). Si on ne désire qu'une liste exhaustive (sites Internet par exemple) seuls quelques champs seront conservés (NOM/TITRE, URL, MOTS CLES et DESCRIPTION par exemple) pour atteindre plus rapidement l'exhaustivité recherchée. Si par contre la profondeur des informations est recherchée (description de bases de données par exemple), un maximum de questions seront posées. Chaque inventaire spécifique s'adaptera à ses objectifs propres et à la meilleure stratégie de recherche des informations.

*Les questions marquées \* ne peuvent avoir qu'une réponse (sans compter les traductions bien sûr).*

Dans certains cas, une réponse unique est impérative pour éviter toute ambiguïté dans la base de données. Cependant, des traductions sont permises dans les cas suivants:

-traduction dans une autre langue (par exemple, nom officiel donné dans plusieurs langues).

Dans ce cas, le champs portera la marque de la langue utilisée: soit le champs Metadata\_Language doit impérativement être rempli (s'il n'est pas renseigné, le système considère par défaut que la langue utilisée est l'anglais), et l'ensemble du formulaire est rempli dans la même langue, sauf indication contraire (les champs traduits en anglais doivent dans ce cas porter la marque 'en', par exemple "TITLE en:"). Donc, dans tous les cas où un champs est rempli dans une autre langue que la langue par défaut du formulaire (généralement l'anglais), il faut créer un nom de champ spécifique à la langue utilisée. Différentes traductions du nom de la source peuvent donc coexister dans un même formulaire, par exemple, si la langue du formulaire n'est pas spécifiée et est donc supposée être l'anglais:

Acronym: OSTC

Acronym nl: DWTC

Acronym fr: SSTC

Exemple si la langue du formulaire est spécifiée comme le français:

Metadata\_Language: fr

Acronym: SSTC

Acronym nl: DWTC

Acronym en: OSTC

Une autre méthode serait d'utiliser des attributs pour caractériser le contenu d'un champ:

Acronym (Language = en): OSTC

ou <Meta Name="acronym" Language="French" content="SSTC">

Il faut suivre l'apparition de standards à ce sujet sur Internet.

Dans le futur, on peut imaginer d'autres niveaux de traduction, comme des traductions pour des publics-cibles différents (voir Kestemont et al, IATAFI 1996). Ce cas de figure s'applique en particulier aux champs descriptifs comme les mots-clés ou le résumé. Une fois de plus, c'est l'usage d'attributs pour les noms de champs qui permettra d'envisager cette possibilité, par exemple:

Keywords (Scheme= GEMET): water, air

Keywords (Scheme= DIF): GLOBAL\_CHANGE>CLIMATE

Description (Users=expert; Language=French): Description précise.

Description (User=novice;Language=French): Description vulgarisée.

*Toutes les autres questions peuvent avoir plusieurs réponses (par exemple AUTEUR est au singulier mais on peut lister plusieurs auteurs dans leur ordre logique de publication).*

Cette une application de la caractéristique de répétabilité de chaque champ, particulièrement utile si les metadonnées proviennent de sources différentes (les attributs éventuels du champ peuvent dans ce cas varier).

*Les questions ou choix multiples se terminant par deux points ':' indiquent qu'on peut spécifier par un texte libre en regard du choix.*

Ce point est très important et est d'ailleurs utilisé dans le format DIF. Il permet de faire évoluer les thesaurus fournis par défaut en fonction du groupe d'utilisateur réel. Par rapport à

une liste de mots-clés libre, il offre l'avantage de permettre à l'utilisateur de situer ses mots-clés dans un contexte compréhensible par d'autres. Par exemple, le mot-clé libre "Brussel" entraîne la confusion entre plusieurs villes existant en Europe et sur le continent américain, mais si ce mot-clé est renseigné comme une sous-entité d'un thesaurus renseignant le pays "Belgique", l'ambiguïté disparaît.

*N'hésitez pas à fournir des informations supplémentaires en regard de chaque question, en annexe ou via une référence présente ou non sur Internet.*

*Si vous avez déjà décrit cette source dans d'autres questionnaires de même type, envoyés par des tiers (CORDIS, AEE, etc.), inutile de recommencer: renvoyez-nous simplement une copie de ce formulaire, et ne répondez qu'aux questions qui vous semblent apporter un complément d'informations (par exemple, la traduction du nom officiel ou du résumé dans une autre langue).*

Le système proposé doit permettre un maximum de souplesse. Le gestionnaire se chargera du transfert dans le format d'échange commun. En effet, cette metabase de données n'a a priori pas de moyen de pression pour obliger les fournisseurs d'informations à se conformer à ses standards. En tant que carrefour de l'information, l'accent est mis sur la possibilité de réutiliser des metadonnées existantes, sans perdre la richesse de description dans d'autres systèmes.

#### données techniques:

1.1.\*NUMERO D'ORDRE (*code auquel le répondant peut faire référence sur les documents éventuellement annexés: traductions, illustrations, copies d'autres formulaires. Il est recommandé d'utiliser une chaîne courte de caractères permettant d'identifier uniquement un objet décrit, comme une série d'acronymes en ordre hiérarchique, par exemple: ULB\_CEESE\_CDS ou SSTC\_FEDRA*).

Idéalement, ce numéro d'ordre ne devrait pas dépasser 31 caractères (DIF). Par défaut, on utilisera la méthodologie DIF.

#### **1.2.\*LANGUE DU FORMULAIRE** (*langue utilisée par défaut dans les réponses*)

Le code ISO-2 sera appliqué de préférence. A défaut, le nom de langue en anglais peut être choisi en fonction de l'évolution des standards les plus courants.

Pour rappel, le code ISO-2 s'écrit en minuscule pour le nom de langue, avec éventuellement un complément représentant le nom de pays pour un idiôme particulier. Par exemple "nl BE"



pour le flamand et ‘nl NL’ pour le néerlandais des Pays-Bas, nl pour l’ABN. Dans le cadre de la metabanque de données, il est recommandé de ne pas utiliser les particularités linguistiques nationales et seul le code général en deux caractères sera utilisé.

Quelques codes de langues et leur équivalent en anglais:

Référence: ISO 639, "Code for the representation of names of languages (1988-04-01).

<b>Code Name</b>	<b>Nom</b>	
ar	Arabic	arabe
de	German	allemand
en	English	anglais
es	Spanish	espagnol
fr	French	français
it	Italian	italien
nl	Dutch	néerlandais
la	latin	latin

type de source:

## **2.1.TYPE DE SOURCE**

Référence: Cette liste est le produit de types de sources utilisés dans de nombreux inventaires.

Sources: EEA-CDS, UNEP-HEM et CORDIS.

La liste peut être développée en utilisant des standards spécifiques (IAFA pour les documents Internet ou FGDC pour des bases de données géographiques). Plusieurs types de sources peuvent être utilisés (répétition du champ). Par exemple:

Object\_Type: institution

Object\_Type: academic

Voici quelques types d’objets pouvant être utilisés de manière standard (réf: EEA, 1994):

Catégories principales (utilisation vivement recommandée):

Institution

Person

Activity

## Product

Catégories secondaires (utilisation recommandée en plus de la catégorie principale, le code étant celui de l'AEE, il faut spécifier Sheme = AEE si on utilise ce code. On peut par contre utiliser les termes anglais proposés.)

### Sortes d'institutions:

PU Public (or public power dominant in the managing board)

PA Parapublic

PR Private (not specified)

AS Private association/foundation

NA Private non associate

### Types d'institutions:

IN Industrial, commercial (other than consultants)

RC Research centres, universities, academic institutions.

GO Governmental or public subnational institutions

NG Non governmental organisation (NGO)

CO Consultants

IG Inter-gouvernemental

PO Political responsables

ME Media, e.g. press

RL Religious

### Types de personnes:

Expert

Documentalist

Promotor

Researcher

etc.

### Types d'activités:

Program

Project

Monitoring network

Inventory

Statistical enquiry

### Types de produits:

Database  
Dataset  
Document  
Publication  
Standard  
Report  
Book  
Image  
Map  
Data  
Directory  
Bulletin/News  
etc.

Une codification plus complète ou l'appellation en anglais d'autres types de sources peut s'inspirer de ce qui est proposé pour le champ FORME, CONTENU. Une autre alternative est d'utiliser la codification ou les appellations de l'US-MARC qui est également très détaillée sur ce sujet.

titre:

**\*3.1.NOM OU TITRE:**

Maximum 80 caractères et de préférence moins de 31 caractères (DIF).

Dans le cas d'une personne, ce champ est construit à partir des champs suivants, en commençant de préférence par le nom de famille pour faciliter les tris. Exemple: Dupont, Jean.

(si une personne:

3.1.1.TITRE: \*3.1.2.NOM DE FAMILLE: \*3.1.3.PRENOM: )

On peut ajouter un champs Middel\_Name pour certains noms étrangers.

Le prénom est donné en entier ou à défaut sous forme d'initiale (exemple J.L. pour Jean-Luc).

S'il faut mettre plusieurs prénoms, les joindre dans le même champ.

**3.2..ACRONYME:**

Par défaut, les lettres sont de préférence en majuscule et sans points. Exemple: OCDE.

D'autres représentations sont admises s'il s'agit de représentations officielles. Exemples:

IIN/DPCSD ifen osf ATrBr On peut au besoin mettre plusieurs synonymes

On peut aussi donner la référence à un logo de petite taille (URL d'une image GIF).

3.3.SOURCE DE DONNEES PARENTE: *(base de données, série, programme dont elle est éventuellement une partie). On peut remplir une fiche minimale pour cette référence (voir plus bas).*

Appeler la source par son nom officiel (si possible suivant le format défini par NOM/TITRE) ou par un acronyme non ambigu, et/ou donner une référence précise (URL ou questionnaire joint).

3.4.REMARQUES EVENTUELLE SUR L'ORGANISATION INTERNE DE LA SOURCE DE DONNEES:

Texte et/ou référence (URL, ...) à un document extérieur explicatif (image, organigramme, ...).

3.5.LIEN AVEC UNE AUTRE SOURCE DE DONNEES:

*Spécifier le type de lien et le nom des sources, ainsi qu'un élément d'identification, comme l'URL par exemple.*

Utiliser de préférence la formulation suivante (respecter la casse et les espaces):

Link\_type: nom du lien (et URL)

Exemple:

Relation: Promotor: Prof.André Berger

Relation fr: Promoteur: Prof.André Berger

Cependant, une phrase plus complexe peut être utilisée, avec si possible une référence hypertexte vers les éléments liés (utiliser des tag html).

Objectifs:

Sauf en cas d'importation d'une base existante, rester bref!

4.1.OBJECTIF:

4.2.DESCRPTION GENERALE (RESUME):

4.3. UTILISATION OU SERVICE RENDU:

Contenu:

5.1.DOMAINES GENERAUX TRAITES, MOTS-CLES:

*(Les mots-clés généraux d'un Thésaurus Multilingue courant sont présentés par défaut en fonction de l'inventaire spécifique, mais le répondant peut utiliser des mots-clés (descripteurs) provenant d'un autre système de classification, qu'il doit alors spécifier). Exemple minimum:*

environnement:  
économie:  
sociologie:  
institutions:  
autre:

Utiliser le GEMET de l'Agence Européenne de l'Environnement, et dans un premier temps ses versions provisoires (voir document fourni en annexe).

On peut aussi utiliser le thesaurus "disciplines" du DIF, à charger sur le WWW.

Pour des formulaires spécifiques, le gestionnaire devra opérer une sélection des mots appropriés afin d'alléger les formulaires (le listing complet du GEMET prend 70 pages!).

Même en cas d'utilisation d'un vocabulaire spécifique au niveau de l'interface-utilisateur (par exemple CO2 au lieu de dioxyde de carbone), le gestionnaire s'assurera que le synonyme le plus proche en provenance du GEMET sera en pratique encodé.

## 5.2.PARAMETRES, UNITES ET METHODES:

*(brève description du produit, reprenant surtout les éléments de détail décrivant le mieux la quantité et la qualité des données ainsi que leur désagrégation. Si possible, en cas de classification, lister les classes. Joindre en annexe une illustration):*

*[exemple: Emissions de 4 polluants atmosphériques - NOx, SO2, N2O et CO2 - en tonnes par an par 44 secteurs économiques et 33 carburants dans 2 pays (Pays-Bas et Belgique) sur une période de 10 ans (1980-1990). Calculé en multipliant des facteurs spécifiques d'émission par les utilisation d'énergie finale. Suit la liste des secteurs et carburants: ...].*

Cette information peut également être renseignée en utilisant le thesaurus "Paramètres" du DIF. Dans ce cas, il vaut mieux utiliser un nouveau nom de champ (DIF):

Parameter

Toute liste spécifique peut être renseignée dans un champ Parameter (listes d'espèces, etc.) ou dans une liste séparée de mots-clés dans un champ supplémentaire Keywords (ou n champs Keyword).

### 5.3.ILLUSTRATION EN ANNEXE (*donner une référence*):

*(illustration ou brochure descriptive pour le produit, telle que carte, graphe, exemple de fiche, modèle de données, formules, photo, etc. sous forme digitale en standard d'échange GIF, TXT ou HTML ou maximum 2 pages A4. Pour une illustration fournie sur support digital, spécifier le chemin d'accès et le nom de fichier)*

Au besoin, le gestionnaire pourra être amené à scanner (digitaliser) des documents fournis sur papier, et les inclure sous format GIF. Cependant, il ne faut traiter en priorité que les images stockées sur d'autres serveurs.

### 5.4.STANDARD UTILISE:

Chaque standard peut éventuellement être décrit sur une nouvelle fiche, avec un type de source: produit, base de donnée, publication: publication: standard ou en pratique

Object\_Type: product

Object\_Type: document

Object\_Type: standard

ou seulement

Object\_Type: standard

### Couverture géographique:

#### 6.1.COUVERTURE GEOGRAPHIQUE: (*à laquelle se rapporte la source*)

*Des descripteurs standards sont proposés, des mots-clés libres peuvent être ajoutés.*

#### *Administrative:*

*spécifier (utiliser les codes ISO et NUTS si la place manque)*

Internationale:

Nationale:

Régionale:

Locale:

#### *Non administrative:*

Cours d'eau:

Plans d'eau:

Océans et mers:

Côtes:

Bassins de drainage:

Zones urbaines:

Zones industrielles:

Zones rurales:

Autres: (voir DIF location keywords, par discipline: astronomie, sciences de la terre, sciences planétaires, physique solaire, physique de l'espace)

Zone géographique couverte (*rectangle, volume, ou point central*)

*Les données qui suivent délimitent un volume, une surface ou un point sur la mappemonde, dans laquelle se situe l'aire de spécialisation de la source décrite. Pour un point par exemple, les maxima et minima sont confondus.*

6.2.LONGITUDE LIMITE OUEST *ddd mm ss E/W*:

6.3.LONGITUDE LIMITE EST *ddd mm ss E/W*:

6.4.LATITUDE LIMITE SUD *ddd mm ss N/S*:

6.5.LATITUDE LIMITE NORD *ddd mm ss N/S*:

6.6.ALTITUDE MIN (*mètres, positif au-dessus du niveau de la mer, négatif pour la profondeur*):

6.7.ALTITUDE MAX (*mètres, positif au-dessus du niveau de la mer, négatif pour la profondeur*):

6.8. EPAISSEUR MIN (*mètres*):

6.9. EPAISSEUR MAX (*mètres*):

6.10. ALTITUDE/STRATE (*mots clés*):

6.11.ZONE GEOGRAPHIQUE COUVERTE (FORMAT LIBRE):

*On peut également spécifier une zone plus complexe par une série de coordonnées d'objets (points, lignes, polygones, volumes) sous le format suivant(unités: degrés, mètres): (LAT11, LONG11, ALT11)(LAT2, LONG2, ALT2)...(LAT1n, LONG1n, ALT1n)...(LAT11, LONG11, ALT11). Il faut fermer la boucle en répétant la première valeur s'il s'agit d'un polygone*

6.12.RESOLUTION SPATIALE OU ECHELLE:

*Format libre [ex: '1:1000000', '25 ha/pixel', 'arrondissements', 'pays', 'villes de plus de 10000 habitants', 'unités de plus de 2 ha', 'résolution de 10 mètres' etc ]*

Période couverte et fréquence des données (si applicable):

(format YYYYMMDD hhmmss, heures GMT, ou spécifier le fuseau horaire)

7.1. TEMPS D'ACQUISITION (pour une image par exemple):

7.2.PERIODE DE:                    7.3.A:

(Pour une série de données, deux temps sont donnés. En cas de périodes multiples, il faut répéter ces deux champs par paire)

7.4. PERIODE (format libre):

7.5.FREQUENCE DES DONNEES

Cette question nécessite plusieurs champs, ou un format strictement défini. On peut alternativement utiliser US-MARC ou DIF, mais ces standards sont moins précis. Remarquons qu'il est possible de n'utiliser qu'un champ numérique en utilisant la convention suivante (ou toute autre équivalente):

Questionnaire		Code
		YYYY/MM/DD hh: mm: ss
continue		0
régulière, tou(te)s les .....	secondes	0,0000...1 à 58,99999... (pas 0)
(nombre)	minutes	1:00 à 59:00
	heures	1:00:00 à 23:00:00
	jours	1 00:00:00 à 31 00:00:00
	mois	1/00 00:00:00 à 12/00
00:00:00		
	ans	1/00/00 00:00:00 et plus
irrégulière		9999/12/31 23:59:59
autre:		9999/12/31 23:59:58

exemple: régulière tous les 3 jours, 20 heures 5 minutes: 3 20:05:00



Qualité:

8. QUALITE:

Format textuel libre (DIF). Des champs plus précis existent dans FGDC.

Référence éventuelle du produit, ou copyright:

9.1.AUTEUR (*répéter le champs si plusieurs*):

9.2.ORIGINE (*organisation*):

9.3.DATE DE PUBLICATION:

9.4.SERIE:

9.5.EDITEUR:

9.6.PLACE DE PUBLICATION:

Versions

\*10.1.DATE DE CREATION:

10.2.FREQUENCE DE MISE A JOUR:

(M<sup>^</sup>me format que fréquence des données)

\*10.3.DERNIERE MISE A JOUR:

10.4.PROCHAINE MISE A JOUR:

10.5.STATUT:            en construction:        fixée:

(*la source de données est-elle déjà ou encore enrichie ou est-elle fixée?*)

10.6. VERSION:

### 11.0.1.

#### Identificateur unique de référence pour la source

*Cette section est destinée à identifier la 'source-mère', référence unique dans des bases de données ou sur internet, où se trouve l'original de la source ou de sa description officielle. Plusieurs types de références uniques sont possibles sur différents supports. Ne donner qu'une seule référence par type, les autres copies pouvant être indiquées à la question suivante. Les autres identificateurs peuvent être des identificateurs uniques dans des metabases de données existantes.*

*Le cas échéant, indiquer des remarques (ex: login, password, configuration)*

#### Adresse

\*11.1.1.RUE/N°/BOITE POSTALE:

\*11.2.1.CODE POSTAL:

\*11.3.1.VILLE:

\*11.4.1.PAYS:

\*11.5.1.LOCALISATION (*longitude-latitude*):

\*11.6.1.CARTE D'ACCES (*joindre carte ou fichier*):

#### Téléphone

Entrer le code complet, y compris le code international.

Exemple: 32-2-650 35 88

\*12.1.1.TELEPHONE (voix):

\*12.2.1.FAX:

\*12.3.1.MODEM:

\*13.1.1.URL:

entrer l'URL complet, y compris le protocole.

Exemple

<http://www.ulb.ac.be> et non [www.ulb.ac.be](http://www.ulb.ac.be)

\*13.2.1.E-MAIL:

Exemple [info@ulb.ac.be](mailto:info@ulb.ac.be)

13.3.1.AUTRE INTERNET (*spécifier, une seule adresse par type supplémentaire*):  
*autres*

\*14.1.1.ISBN:

\*14.2.1.ISSN:

15.1.1.AUTRE (*spécifier, une seule adresse par type supplémentaire*):

Accès direct à la source ou à sa description (sites miroirs)

Indiquer ici les différents points de contacts où la source peut être consultée directement (plusieurs réponses possibles par type).

Même format que ci-dessus. La séquence peut être répétée autant de fois que nécessaire, éventuellement à la suite du champ correspondant ci-dessus (ce premier champ donnant l'accès à l'original).

Mode de construction, sources ou références utilisées

16.1.SOURCE

Utiliser si possible un URL.

16.2.METHODE DE CONSTRUCTION:

16.3.FORME, CONTENU (*cocher une ou plusieurs options et spécifier*):

Types généraux (recommandé comme minimum à renseigner de toute façon, même si des indications plus précises sont données dans une répétition du même champ). Il peut être opportun de s'aligner sur les standards WWW à ce sujet, ce qui n'est pas le cas de la liste proposée ici.

Questionnaire	Terme anglais (code)
données factuelles, tableaux, diagrammes	Facts
textes	Texts
graphes, dessins, images, cartes	Pictures
films, vidéo, animations	Animations
échantillons, pièces de collection, maquettes, ...	Samples
sons:	Sounds

Une représentation plus précise peut s'inspirer de la référence ci-dessous.

Référence: Alpine Convention/EEA: Inventory of alpine data sources (1994), extended and adapted considering CORINE codification and general use.  
EEA, 1994

**Code Name**

Type de produit ne spécifiant pas le support (contenu des systèmes):

FA Facts/tables  
 TXT texts  
 GE Geolocational information  
 MP Maps  
 BI Bibliography  
 SI Remote sensing/satellite images  
 PI Pictures  
 SO Sounds  
 SP Samples

MM Mapmodel  
 PH Photo  
 FI Film  
 VI Video  
 PA Painting

OT Others\*

Spécifiant le type de support:

(.P=printed; .D=digital; .A=analogic; .L=Actual\*; M=Manual\*; O=Others\*)

examples:

	EEA1994	Name
Alps		
PTX	TX.P	Printed text/books
	TX.M	Manuscript
PMM	MM.P	Printed map model
PMP	MP.P	Printed map
PFA	FA.P	Printed table/facts
PPI	PI.P	Printed picture
PSO	SO.P	Printed sounds (notes)
DTX	TX.D	Textfile
DMP	MP.D	Digital map
DFA	FA.P	Data file
DPI	PI.D	Digital picture
DSO	SO.D	Digital soundtrack
ASO	SO.A	Analogic soundtrack
	PA.L	Actual painting*
	PI.L	Actual picture*
	PI.O	Other kind of picture (e.g. hologram)*

FA.D? computerized datacollection  
 FA.P? manual datacollection  
 TX? document(s) \*  
 MP map(s) \*  
 SI remote sensing images \*  
 SP?object collection, samples

SN? object collection,specimens  
SM object collections, scale models (3-D maquette)

Support réel (combinant le type de support et le support réel demandé plus bas: P=Paper\*; L=Line; T=tape\*; D=Disk(ette)\*; R=CD-ROM\*; H=Hard; O=Other)

PP printed/paper  
DL on-line version  
DTdigit-(magnetic)tape  
DDdigit-disk(ette)  
DHdigit hard disk  
DRCD-ROM (read-only memory)  
DW optical disk (re-writable)  
DOother digital medium  
ODOther kind of disk

**Exemple de codes complets pour les produits, combinant le support et le contenu:**

<b>code</b>	<b>name</b>
TX.PP	Paper printed text
TX.MP	Paper manuscripted text
TX.MO	Manuscript on other medium (e.g. on a papyrus, sculpted on a monument, ...)
TX.DT	Digital text on tape
MP.DT	Digital Map on tape

16.4.LANGUE DISPONIBLE (*plusieurs réponses possibles*)

Dans la base de données, utiliser l'appellation anglaise ou de préférence le code ISO-2 (voir "Type de source").

16.5.JEU DE CARACTERES:

ASCII

ANSI

ISO Latin-1 (utilisé par html)

sont les plus courants, mais certainement pas les seuls!

16.6.SUPPORT (cocher et spécifier les formats disponibles)

Questionnaire	Base de donnée (anglais)
digital:	digital

disquette:	diskette
CD-ROM:	CD-ROM
bande magnétique:	digital tape
en Ligne (Internet, ...):	on-line
disque dur:	hard disk
analogique:	analogic
papier (manuscrit, imprimé, photographie):	paper
disque:	analogic disk
transparent (négatif, microfilm, film, ...):	film
bande magnétique:	analogic tape
‘en Ligne’ (Télétexte, téléphone, TV, ...):	analogic-line
ondes (TV, radio, gsm...)	hertz
échantillon:	
maquette, ...:	
autre:	

En pratique, deux champs au moins sont utilisés. L'un reprend une appellation standard (ci-dessus), l'autre précise le format en format textuel libre.

Pour les formats, on peut utiliser les listes fournies sur Internet en priorité, ou la liste suivante (EEA):

Les formats sont présentés en fonction des types de produits (voir plus haut)

<b>Type de produit</b>	<b>Système</b>	<b>Formats</b>
.PP:		A0, A1, A2, A3, A4
TX.D:	DOS	ASCII, TXT, WP5, DOC, ...
	MAC	ASCII, TXT, WP5, DOC, ...
	UNIX	
MP.D:	DOS	ARC-Export
	UNIX	ARC-Export, ARC-Ungenerate, Raster, BMP, ...
PL.D:	DOS	Raster, BMP, TIF, ...
.DT:		CCT, 8mm, 1/4"
.DD:		5", 3", ...

Politique de distribution:

17.1.DISTRIBUTION

Questionnaire	Code (non standard)
seulement usage interne:	internal
usage externe:	external
gratuit:	free
payant:	cost
limité (préciser):	limited

Il reste à chercher un code standard (voir DIF ou US-MARC)



Déontologie:

18.1.AVEZ-VOUS DEFINI DES REGLES ?

- pour la récolte des données
- pour le traitement des données
- pour l'accès aux données
- pour la publication et la dissémination des données
- pour le paiement d'informations ou services
- autres

Chercher un code standard ou proposer un code.

18.2.TEXTE EN ANNEXE *(si oui, donner une référence)*

Temps d'accès:

19.HORAIRE D'ACCES *(en temps local, ou préciser):*

Format: texte libre.

Volume approximatif de la source:

20.VOLUME APPROXIMATIF *(nb d'enregistrements, de pages, de Kbs, de valeurs, de membres du personnel d'une organisation, budget annuel, ...)*

On peut utiliser un format libre.

On peut aussi utiliser les champs spécifiques prévus à cet effet dans DIF notamment.

Organisation ou personne responsable pour la gestion de la source

21.RESPONSABLE:

*(donner un URI)*

*Référence à un questionnaire déjà complété:*

*Le cas échéant, remplir un questionnaire pour l'organisation en question. Voir exemple de questionnaire minimum pour les organisations ci-dessous.*

*D'autres liens peuvent être définis à divers endroits de la fiche.*

Source des metadonnées (qui a rempli ce questionnaire?)

*Prévoir un lien URL ou une fiche minimum séparée pour le documentaliste et/ou son organisation, ou pour la source de metadonnées déjà publiées (voir fiche minimale ci-dessous)*

22.1.DOCUMENTALISTE:

*Personne, organisation, metabase de donnée, catalogue ou/et document ayant servi de source pour les informations fournies dans ce questionnaire ... Donner des éléments permettant d'identifier les fiches décrivant les sources mentionnées.*

22.2.DATE:

*Date de dernière mise à jour de cette fiche.*

22.3.REMARQUE SUR LA SOURCE DES METADONNEES:

22.4.AUTRES META INFORMATIONS:

Remarques et autres:

23.REMARQUES ET AUTRES INFORMATIONS *(vous pouvez recopier ici, ou joindre des extraits d'autres questionnaires auxquels vous avez répondu pour la même source)*

Reprendre les champs importés tels quels, en précisant si possible leur origine par l'attribut Scheme= ...).

***(FIN de la fiche de base.)***

## **Explication d'une solution informatique opérationnelle (logiciel et tâches techniques).**

### Formats et jeux de caractères

Dans un premier temps, la base de données sera gérée sur un ou plusieurs fichiers textuels 'plats' sous format html. A ce propos, se référer par exemple au guide html (source: <http://www.w3.org/pub/WWW/TR/REC-html32.html>).

Pour éviter de devoir encoder en html, il est conseillé d'utiliser un logiciel d'édition courant (WordPerfect) et de sauver sous format html, voire d'utiliser un éditeur html comme FrontPage par exemple. Cependant, il sera nécessaire de travailler directement dans le fichier source (en format ASCII-texte) pour certaines opérations de mise en forme de la base de données.

Le jeu de caractère utilisé sera de préférence ANSI pour les documents venant d'UNIX ou de WordPerfect, (ce jeu de caractères est généré automatiquement lors de la sauvegarde en format html à partir de WordPerfect), ou ASCII (code-source utilisé sur Internet) pour les documents sources présentés d'origine sous ce format (les caractères accentués étant remplacés par leur code ISO Latin-1, générés automatiquement lors de la conversion par des logiciels d'édition html courants comme MS-Word Internet Assistant, et éventuellement ISO 8859-1 pour certains caractères spéciaux).

Caractères spéciaux pour le format html (source: [http://www.ucc.ie/doc/World-Wide\\_Web/htmlcard.html](http://www.ucc.ie/doc/World-Wide_Web/htmlcard.html)) sur base du jeu de caractères ASCII.

### Logiciels

Les logiciels de bureau utilisés par le SSTC sont suffisants. En fonction des choix logiciels de la maison, d'autres logiciels peuvent être utilisés dans le futur.

L'utilisation d'un serveur Windows-NT 4 est particulièrement recommandé pour la gestion des bases de données et indexes sans recours aux scripts CGI, ainsi que pour des applications

graphiques et bureautiques intégrées, largement plus conviviales et plus puissantes que sous UNIX (pour la cartographie interactive: voir par exemple <http://maps.esri.com/>).

Dans le cours terme (avant le passage à Windows-NT et MS-Office) WordPerfect sera donc utilisé sur PC pour la plupart des opérations de transfert de données, de maintenance et la création des indexes (fichiers séparés) et pages html. Les Macro de WordPerfect seront nécessaires pour la conversion de données provenant de sources diverses (les données importées via Internet seront souvent sous format textuel ASCII). En fonction de l'équipement du SSTC, le logiciel de base pourra être MS-Word, logiciel utilisé pour le prototype et permettant des fonctions Macro suffisantes pour le travail à réaliser.

En fonction des budgets disponibles, on peut envisager l'achat d'un logiciel documentaire, comme Texto-GED (Eurodoc, rue du Conseil, Bruxelles, env. 60000FB, également disponible pour serveur), particulièrement adapté à la gestion de bases de données textuelles à champs variables, ainsi qu'à la gestion de thesaurus. D'autres logiciels concurrents peuvent se trouver en consultant les responsables de bibliothèques informatisées.

Pour en revenir aux traitements de texte classiques, voici quelques exemples de macros.

Exemple de macro: importation de données reçues sous format ASCII et en provenance de la metabase de données d'IWETO. La macro opère des remplacements de noms de champs par leur équivalent dans le standard d'échange du projet. Elle allège également le fichier en supprimant par exemple des marques de paragraphe superflues.

Autre exemple de macro: par une succession de remplacements de chaînes de caractères, des délimiteurs de champs sont incorporés afin de pouvoir effectuer des tris suivant différents critères et isoler des groupes d'enregistrements par type de source, par langue, par mot-clé etc. Il est parfois utile de transférer les bases de données vers un gestionnaire de bases de données (FoxPro) ou un tableur pour les éditer (tris, intégrité des données textuelles etc.).

Autre exemple de macro: ajout de liens hypertextes dans des sources qui n'en disposaient pas. Tout lien hypertexte connu (interne ou externe) peut être ajouté à chaque mot concerné par un simple 'remplacer' dans le document source principal. Par exemple (après avoir harmonisé l'appellation de la KUL par une série de 'remplacer'): remplacer VUB par <A HREF=<http://www.vub.ac.be>>VUB</A>

Si le gestionnaire le préfère, il peut également utiliser un tableur (Lotus) ou le logiciel FoxPro pour gérer le document maître ou ses indexes.

Ce qui compte, c'est qu'un fichier maître (éventuellement coupé en morceaux, à l'extrême une fiche par fichier) serve à effectuer toutes les opérations de maintenance et d'édition de la métabase de donnée. C'est à partir de ce (jeu de) documents maîtres que tous les indexes, extraits ou fiches individuelles seront reconstruits à intervalle régulier. Vu la taille limitée d'une meta-banque de données, il ne faut pas craindre la duplication d'informations dans des fichiers séparés (par exemple un fichier html par type de source, par langue, etc. comme dans le prototype et en fonction des options les plus demandées par les utilisateurs).

Remarquons que certains champs (champs non répétés) peuvent avantageusement être gérés sous FoxPro et réimportés par la suite dans les fiches de base.

### Matériel informatique

Un PC de bureau 'haut de gamme' avec une grande puissance de calcul est indispensable en raison des opérations (macros) à réaliser sur de grands fichiers. Actuellement, un ordinateur à base de Pentium 200 Mhz (ou Pentium Pro), au moins 16 ou 32 Mb RAM, équipé de Windows-NT Workstation (plus stable) ou Windows 95 est un minimum. Windows 3.1 est fortement déconseillé en raison de la rapide perte de ressource interne de ce système lors de travail assidu sur de grands fichiers. Les processeurs à base d'architecture RISC sont actuellement plus puissants à vitesse d'horloge égale (PowerPC 604e pour Mac-OS ou IBM-OS2 en 32 bits) et plus appropriés pour l'édition multimédia, mais ce choix compliquerait la maintenance du parc informatique. Des capacités graphiques et multimédia sont utiles à la mise en page et pour pouvoir suivre par la suite l'évolution des technologies Internet les plus avancées. L'ordinateur sera donc idéalement équipé en conséquence (CD-ROM, écran 17 », carte vidéo, carte son de qualité moyenne. Budget à prévoir: 100-150000 FB.

L'achat d'un nouveau PC peut être différé de quelques mois (maximum), un simple PC-DX2-66 sous Windows 3.1 pouvant faire l'affaire lors de la mise en place du système, non sans inconvénients quant à la gestion des données (vitesse insuffisante, gestion insuffisante des ressources système, qui obligera à redémarrer Windows plusieurs fois par jour).

La connexion à Internet (mail, WWW, FTP) est indispensable.

### Accès au serveur WWW

Le gestionnaire aura accès au serveur WWW sur lequel il aura l'autorisation de placer et éditer des documents, ainsi que d'effectuer des scripts UNIX ou cgi par exemple. L'accès se fera suivant le protocole FTP (pour la mise à jour des documents) ou en mode terminal UNIX.

### Changements futurs de configuration

Le SSTC envisage de se doter de serveurs sous système Windows-NT. A partir de ce moment, la base de données pourra être directement gérée sur un gestionnaire de bases de données comme MS-Access et éventuellement être interrogée directement par les utilisateurs, les fiches envoyées en ligne étant construites 'au vol'. Ne perdons pas de vue que la plupart des champs peuvent être répétés plusieurs fois, et d'autre part sont de longueur variable. En cas d'utilisation d'un gestionnaire de bases de données relationnelles comme Access, il faudra en tenir compte, ce qui risque d'exiger un modèle de données fort complexe (fiche de base ne comportant pratiquement que le numéro d'identification, soit l'adresse URL de la fiche textuelle, et autant d'index que de champs, avec une référence aux fiches décrites - relations 1-n). En réalité, l'utilisation d'une base de données relationnelle n'est pas recommandée, sauf pour la gestion éventuelle des fichiers d'index.

(Référence de ce chapitre: discussion avec Jean-Luc Delberghe; méthode utilisée pour le prototype sur le Changement Climatique et pour la base de donnée FEDRA).

## **Standards et programmes pouvant être utilisés dans certains domaines spécifiques**

La fiche-type est prévue comme un format d'échange à l'intersection de plusieurs standards existants. Cette caractéristique permet d'utiliser les outils existants pour des catalogues thématiques, ce qui peut dans certains cas favoriser la récolte de données. Les standards et outils existants ont tous leurs défauts, mais leur utilisation peut s'avérer intéressante pour la récolte de nouvelles données non disponibles autrement. Le choix d'un outil extérieur particulier dépendra de la force de persuasion que cet outil ou standard peut apporter pour qu'un groupe cible accepte de remplir des formulaires. Cependant, pour tout travail d'encodage devant être réalisé par le SSTC (formulaires papier), seuls la fiche type (ou son adaptation) sont utilisables. L'utilisation d'autres standards ne se justifie que si elle permet de s'affranchir d'encodage de fiches, partie la plus onéreuse de la constitution de la métabase de données.

Voici pour certains domaines d'application une liste d'outils utilisables immédiatement ou dans un futur proche:

### **Global change, Earth Observation**

Si des interlocuteurs désirent remplir des fiches directement sous forme digitale, plusieurs outils sont disponibles dans le cadre du CEOS. Cependant, ces fiches ne pourront être remplies qu'en anglais. Comme spécifié plus haut, une série d'outils (formulaires et scripts WWW, formulaires pour Windows) permettent d'enregistrer des données au format DIF. Les adresses sont les suivantes:

Un créateur d'entrées DIF est accessible à l'URL <http://gcmd.gsfc.nasa.gov/difweb.html>.

Un logiciel d'entrée DIF pour Windows peut être obtenu à l'URL [www.smithsys.co.uk/dif](http://www.smithsys.co.uk/dif)

Le thesaurus utilisé pour le DIF est bien adapté à la problématique de l'observation de la terre, de l'imagerie satellite et du changement global.

Les données utilisées par ce biais seront ensuite transférées telles quelles dans la métabase de données, moyennant la traduction des noms de champs et l'ajout du qualificatif 'SCHEME='DIF'' pour les mots-clés. D'autre part, ces données pourront être directement envoyées au Global Change Master Directory de la NASA comme contribution à cette métabanque de données très populaire au niveau mondial. Une autre manière de procéder est de conseiller aux acteurs du Global Change de se renseigner dans le Master Directory, puis

d'effectuer des extraits de cette base de données pour les importer dans la metabase de données du SSTC.

### Earth Observation, outil futur

Dans quelques années, le CEO (<http://www.ceo.org>) sera en mesure de fournir des outils adaptés à la communauté de l'observation de la terre. Il sera alors préférable d'encourager les fournisseurs belges de metadonnées à utiliser cet outil de préférence à celui du CEOS (voir ci-dessus) et d'abandonner les outils DIF.

### Environnement, outils futurs

Il faut suivre de très près ce qui se passe au niveau de l'Agence Européenne de l'Environnement, qui proposera bientôt des outils opérationnels sur Web et sur Windows. Le standard utilisé est par essence proche de la fiche-type, et les données ainsi récoltées seront aisément récupérables. Les adresses à contacter sont jointes dans un document annexé. L'URL de base est <http://www.mu.uni-hannover.de/cds>. En temps utile, ces outils seront certainement promus par le point focal belge de l'Agence, CELINE (voir adresse à <http://www.eea.dk>). Cependant, en attendant, il importe d'utiliser la fiche-type.

### Suivi des initiatives analogues

Le gestionnaire de la metabase de données suivra de près les développements des grosses initiatives suivantes, de manière à pouvoir proposer des outils adaptés pour les groupes-cible concernés:

-CEOS-DIF (<http://gcmd.gsfc.nasa.gov/difweb.html>)

-CEO (<http://www.ceo.org> ou <http://ewse.ceo.org>)

et surtout:

-AEE/CDS (<http://www.mu.uni-hannover.de/cds> ou <http://www.eea/dk>)

## **Limitations pour le passage d'un standard à un autre et problèmes d'exportation des données vers d'autres standards.**



Le tableau 2 propose des noms de champs provenant de standards de référence spécialisés dans la question posée. Dans la mesure du possible, la méthodologie par défaut et les limitations pour le contenu de ces champs (reprises dans la fiche technique présentée plus haut) s'inspireront du standard d'origine, sauf spécification contraire. Cependant, la possibilité de répéter le champ permet de le renseigner suivant différents standards, moyennant une qualification du standard utilisé (SCHEME ...).

L'importation en provenance de ces standards en est ainsi facilitée, sans perte d'information.

L'éditeur de la base de données opérera des conversions d'unités et de formats du contenu des champs importés en fonction des nécessités - en conservant néanmoins la version d'origine - de manière à rendre la base de données la plus intègre possible au sein des champs utilisés pour construire les indexes de recherche.

Une exportation éventuelle vers d'autres standards sera en principe facilitée à condition de connaître les règles de chacun de ces standards. Seuls les champs renseignés d'origine suivant le standard d'exportation seront conservés indemnes. Les standards les plus probables d'exportation étant DIF, FGDC, CDS ou WWW (Dublin), les champs correspondants seront conservés (le plus souvent seulement leur version anglaise), alors que les autres champs subiront des amputations ou ne seront pas exportables sans intervention humaine. Dans la plupart des cas, une exportation éventuelle entraînera inévitablement la perte de données, les standards extérieurs étant généralement plus limitatifs que le standard d'échange de la fiche-type.

## **Où trouver le thesaurus GEMET, et les autres standards?**

Les thesauri existants utilisés par défaut pour les mots-clés seront de préférence multilingues. Seuls les thesauri spécifiques seront acceptés pour les disciplines particulières (DIF pour le Global Change par exemple, listes d'espèces, ...). Même dans ce cas, il est vivement recommandé d'utiliser des mots clés ou descripteurs provenant d'un thesaurus général (GEMET) et simultanément de remplir un second champ 'Keywords' avec des descripteurs provenant d'un thesaurus spécialisé (à renseigner par les spécialistes belges de chaque discipline), avec cette fois une référence au thesaurus (Keywords SCHEME='DIF' par exemple).

Le thesaurus GEMET peut être obtenu via [felluga@relay.itbm.rm.cnr.it](mailto:felluga@relay.itbm.rm.cnr.it) ou [Wolf-Dieter.Batschi@uba.de](mailto:Wolf-Dieter.Batschi@uba.de). Une version préliminaire anglaise a été fournie dans le cadre de l'étude de faisabilité, au format MS-Word. Une autorisation éventuelle d'utilisation (par exemple sous

prétexte de 'test') peut être obtenue à l'Agence Européenne de l'Environnement via les mêmes auteurs.

En cas de problème pour l'utilisation du GEMET, il est toujours possible d'utiliser une version publiée plus ancienne, qu'on peut obtenir moyennant paiement sous forme papier chez Laurens de Lavieter, TNO, Schoemakerstraat 97, PO Box 6013, 2600 JA Delft, The Netherlands (phone +31 15 61 31 86) ou sur CD-ROM chez Bruno Felluga (felluga@relay.itbm.rm.cnr.it).

Quant aux adresses des autres standards cités dans le document, elles figurent à différents endroits de ce document et du rapport de l'étude de faisabilité (bibliographie, ...), ainsi que sur une page html attitrée:

Metadata standards directories sur <http://www.ulb.ac.be/ceese/meta.html>.

## **Comment structurer les documents-maîtres?**

Les documents maîtres sont sous le format (textuel) qui convient le mieux au gestionnaire de la metabanque de données en fonction de son expérience propre. Le plus facile serait sans doute de conserver les documents maîtres sous format html, sans fioritures particulière (donc en pratique en code ASCII avec retours chariots de type <P>. Mais, ces documents maîtres pourraient très bien être de simples fichiers ASCII (ou Word en format 'texte seulement' par exemple). Quel que soit le format de conservation des documents maîtres, des macros les convertiront automatiquement vers les fiches html correspondantes, y compris une mise en page éventuelle.

Une série d'éléments importants sont à observer: des délimiteurs et sous-délimiteurs clairement identifiables doivent organiser l'information. D'une part, les labels (libellés de champs) doivent être cohérents et comporter des éléments d'identification non ambigus (par exemple la suite de caractère suivante: RetourChariot, NomDeChamp, colonne, espace) et de préférence basés sur des caractères ASCII (bien qu'on puisse utiliser des séparateurs propres aux programmes utilisés, comme tabulation pour Excel ou 'champ' dans les traitements de texte).

En fonction de la taille de la base de données, les fichiers maîtres peuvent être séparés par fiche ou à raison d'un ensemble de fiches par fichier. La première méthode nécessite de pouvoir réaliser des scripts ou macros travaillant sur des fichiers successifs. Si la vitesse de calcul

l'exige, il peut être plus pratique de regrouper plusieurs fiches sur un même fichier. Dans ce cas, un délimiteur supplémentaire devra identifier chaque fiche (ou 'enregistrement') afin de permettre la publication posthume sous forme de fiches individuelles.

## **Création d'indexes**

A partir des documents-maîtres, des macros peuvent fonctionner d'une des manières suivantes:

1) Agréger les fiches dans un seul fichier (en prenant soin d'ajouter un délimiteur de fiches).

Trier les fiches suivant les critères voulus (ce qui peut se faire sous Word, Excel etc.)

Eliminer les informations superflues et ne garder que le titre et l'URL (URI, c'ad l'adresse html interne au système) des fiches correspondant au critère.

2) Faire une recherche au sein de chaque fiche et créer un index ne reprenant que le nom et l'URI des fiches correspondant au critère de recherche.

3) Sur base d'un seul fichier (sous Word) reprenant une fiche par page, indexer l'ensemble de la base de données à partir d'un thesaurus connu et créer l'index suivant les outils de MS-Word. Remplacer ensuite les numéros de pages par l'URI ou numéro de fiche correspondant. On peut travailler sur les fiches complètes (l'indexation est alors aveugle) ou sur des fiches réduites seulement à certains champs (champ Keywords).

4) Après une succession de tris et de remplacements (à conserver dans des macros pour usages ultérieurs), mettre en page la base de données de sorte qu'elle puisse être importée vers un gestionnaire de base de données. Les pertes d'informations dans les champs textuels longs ont moins d'importance dans la mesure où ces champs textuels ne serviront sans doute pas à la création des indexes. De même, les répétitions de certains champs devront être traitées en les remplaçant par de nouveaux champs ou en éliminant les parties non essentielles.

On peut de cette manière utiliser un gestionnaire de bases de données pour gérer les indexes et les recherches complexes en ligne, les documents-maîtres (contenant toute l'information) restant toutefois sous forme textuelle.

## **Création de meta-tags**

A partir de la base de données, on peut imaginer de progressivement remplacer l'URI (interne au système) par des URI décentralisés, conservés par des partenaires. Dans ce cas, il peut être utile de les aider à générer des champs META dans le HEADER de leurs documents (home-page en format libre par exemple).

Il est même conseillé de prévoir le remplissage des champs META suivants au sein de chaque fiche, déjà standard de facto sur le WWW, et reconnaissables par des moteurs de recherche comme Excite:

KEYWORDS, DESCRIPTION, AUTHOR etc.

Voir à ce sujet le W3C et le Dublin Core ([http://purl.org/metadata/dublin\\_core\\_elements](http://purl.org/metadata/dublin_core_elements)).

Le fait de remplir ces champs de metadonnées permettra à l'information d'atteindre toute personne recherchant des informations via l'Internet, même sans connaître l'existence du SSTC, ce qui est un avantage supplémentaire pour les fournisseurs de metadonnées.

## INDEX

### —C—

CELINE 6; 50  
champs 4; 5; 7; 12; 13; 21; 22; 23; 25; 26; 27; 28; 31;  
32; 34; 36; 42; 43; 44; 46; 47; 48; 49; 50; 51; 52;  
53  
CORDIS 4; 8; 28; 30

### —D—

DIF 5; 11; 22; 26; 27; 28; 29; 31; 33; 35; 36; 42; 43;  
49; 50; 51

### —F—

FEDRA 4; 8; 29; 48  
FGDC 22; 30; 36; 51

### —G—

GEMET 4; 28; 33; 51  
GILS 22

### —I—

IAFA 22; 30  
IDN 5  
ISO 5; 11; 22; 29; 34; 41; 45  
IWETO 4; 46

### —J—

jeu de caractère 45

### —L—

langue 5; 8; 9; 25; 27; 29; 46; 47

### —M—

metadonnées 4; 5; 6; 18; 21; 22; 25; 28; 29; 43; 50;  
53  
multilingue 4

### —O—

OCR 4

### —P—

PC 4; 45; 47

### —S—

source 7; 8; 9; 11; 13; 14; 15; 18; 19; 20; 21; 25; 26;  
27; 28; 30; 32; 34; 35; 37; 38; 39; 41; 43; 44; 45;  
46; 47  
SSTC 1; 4; 5; 8; 25; 27; 28; 29; 45; 46; 48; 49; 53  
standard 4; 5; 6; 7; 11; 22; 25; 26; 30; 34; 42; 43; 46;  
49; 50; 51; 53

### —T—

thesaurus 25; 26; 28; 33; 46; 49; 51; 52

### —U—

URL 5; 8; 10; 14; 15; 18; 20; 26; 27; 32; 38; 39; 43;  
48; 49; 50; 52

### —W—

WWW 4; 5; 6; 22; 33; 39; 45; 47; 49; 51; 53