Technical manual for the implementation of a metadatabase on sustainable development

Version 1,01

Federal services of Scientific, Techniques and Cultural Affairs (SSTC).

Ir. Bruno Kestemont* and Ir. Dirk Le Roy**
under the direction of:
Prof. Walter Hecq* and Dr. Ir. Paul Vanhaecke**

*CESSE-ULB, 44 avenue Jeanne, CP 124, B-1050 Brussels
**ECOLAS N.V., Lange Nieuwstraat 43, B-2000 Antwerp

TABLE OF CONTENTS

Introduction

This manual gives the technical specification of the interchange format for the Metadatabase on Sustainable development. All existing metadata should be downloadable to this open textual-based format. Also textual-based questionnaires can be used to collect new metadata from data providers. Exportation to other metadatabases is facilitated by the fact that the standards used are as close as possible to the dominant standards in each field on the information marked.

Representation mode of the type-form.

The type-form (interchange standard) is presented under the form of a usable questionnaire as such. However, in practice, this form will be adapted in function of :
-the type of source (delete irrelevant questions);
-the proposed theme (provide only relevant keywords).

The following format is used:

''Section title underlined'': only aim is to present fields by logical groups in order to increase the readability of the questionnaire.
''Remarks in italic'' : assist the questionnaire user.
''QUESTION'': questions asked (corresponding most often to a field in the database). By default, all questions may have several answers, any length, but short in preference.
''* QUESTIONS MARKED WITH A STAR'' : one answer only (the field can not be repeated).
''QUESTIONS IN BOLD CHARACTERS'': minimal answer to constitute a separate form.
    box to be ticked followed by '' : '' : one can tick and suggest an under-descriptor;
    simple box to tick: tick only (notes will not be encoded)

Questions are temporarily numbered for the use of the database administrator to return field names to the table presented further on.

Under-separators to be used in case of several answers to one same field should be clearly identifiable: semicolons to separate keywords, authors or names, line return and hyphen '-' to separate list elements.

Codification of the metadatabase.

Names or codes given to descriptive fields have no thematic interest and can be given by the database administrator once it will be have been set-up. However, it is interesting to use names or codes included in existing standards, notably because they can facilitate ulterior exchangers. The rule is therefore to make reference to at least one standard whenever possible.

The method to follow can be inspired from annexe 8 (example of meta-information in the header of a html page).

Rule for field names (META NAME)

We propose to adopt the following rule:
Metadata standards emerging on the WWW (Dublin core, IAFA, html Meta-Tags) should have priority.
If the field is not standardised at this level, the following standards will be adopted in the order of: CDS-EEA when it will exist, DIF (last version), FGDC (or ISO when it will exist), GILS, US-MARC.
Provisionally, a free field name can be used, for example in the case of data import, but this is not recommended.

Table 2 shows a proposal of standard field names.

Numbers corresponding to questions of the «type-form» questionnaire mentioned above for legibility reasons. These numbers do not belong to the standard. With the utilisation of an existing standard, all types of fields can be added, similarly fields marked with * can be repeated.

Table 2: Proposed field names for the main questions of the type-form.

Table 2 (continuation):

Specification of the language of the field content

Current standards suppose the utilisation of English (or of a code) in the field content. To represent another language for information provided in a field, it is necessary to specify by using one of the following three methods:
- add a suffix to the field name (absence of suffix indicating English by default), for example: «Name fr:» if the name is given in French;
- add a qualificative (possible in html), for example:
META NAME= «Keywords» LANGUAGE= «French» CONTENT= «list in French»
-use the field "Language of the form" as the main language for all the metadata. In this case, it is imperative to construct a separate form for each translation. This solution will be retained if there is no possibility of an international agreement for a multilingualism management method.

Codification of keywords used for the fields themselves.

The rule is the following:
-priority to codes used in the standard chosen for the field name; however, other lists can be used with the condition of giving the reference;
-for the general thesaurus-subject : General Multilingual Thesaurus for the Environment when it will exist (draft version in the meantime) should be used in priority; the other thesaurus' or code lists which can also be used are INFOTERRA for environmental matters, the Thesaurus Library of Congress (US-MARC) or the EUROVOC Thesaurus for matters uncovered by the precedents. Systems of more specific keywords are also authorised with the condition of giving the reference (CASE number for chemical products, lists of species, NUTS codes etc.). Finally, specific keywords to a particular under-group of the SSTC are admitted if found to be indispensable by user groups. A reference to the SSTC will have to, in this case, accompany them.
All keywords where the source has not been specified are considered as free keywords.

During the utilisation of databases under textual format, it is possible to give keywords making a link to several reference systems if necessary. The following nomenclature is, for example, used on Internet (but the html standard is not yet fixed):

META NAME= «keywords» CONTENT= «list of keywords»
for any free keywords (by default), or
META NAME= «keywords» SCHEME= «US-MARC» CONTENT = list of keywords or codes
to give the reference to an existing thesaurus (here Library of Congress).

Technical remarks for each field (thesaurus used,...)

This section gives, for each field, technical characteristics, such as formats and possible field lengths, as well as the thesaurus by default or recommended.

General considerations.

Remember that each field can be repeated as many times as necessary, in the image of the DIF interchange standard and most of the documentary management software. In case of multiple values, the DIF standard recommends the repetition of the field names, and to make it be succeeded by only one value. For example:
Keyword: water
Keyword: quality
However, for fields containing, in general, multiple values, we recommend the use of a unique field in plural , followed by a series of separate values by a determined delimitator (comma for keywords). For example:
Keywords: water, quality
It is widely recommended not to exceed 31 characters for names and acronyms expressed in unique fields (marked with a *), this is to facilitate a possible export to the DIF format.
The DIF format is based on lines of 80 characters maximum. In case of export to the DIF format, fields exceeding 80 characters will be automatically cut all 80 characters, the name of the field being repeated at the beginning of each line.

In every part of the form, one can specify a hypertext link towards an existing document on internet or towards a form already filled in this inventory (example: in 'parent data source', one can put the URL of a description of this source).

This remark is important and should be applied as much as possible to the extent where the database is conceived as a series of forms connected together by hypertext links.

Questions in bold characters are mandatory.

This rule is more advice than an obligation. It is possible, in practice, to describe or make reference to an object without giving a name or title, but this method is not elegant. In practice, the existence of a form begins from when it contains something legible and a reference allowing a connection (URL address, for example), the name of the object described can always be added or replaced afterwards. In case of exportation towards other interchange formats (DIF, for example), forms without names will be ignored.

For the other questions, only answer if they apply to the source described .

Some fields have no reason to be for certain types of objects. They will be omitted in the form describing the object, and when it is possible, in the questionnaire itself. Inventories are, indeed, most often directed towards a particular type of object (for example, databases). In this case, it is useless to display specific fields to people (NAME, FIRST NAME, TITLE). If one desires only a comprehensive list (Internet sites, for example) only some fields will be preserved (NAME/TITLE, URL, KEYWORDS and DESCRIPTION, for example) to obtain quickly all the data looked after. If however the depth of information is looked after (description of databases, for example), a maximum of questions will be asked. Each specific inventory will adapt to its own objectives and to the best strategy of information research .

*Questions marked * can only have one answer (without counting translations, of course).*

In some cases, one reply only is imperative to avoid all ambiguity in the database. However, translations are allowed in the following cases:
-translation in another language (for example, official name given in several languages).
In this case, fields will carry the mark of the language used: Metadata_Language field has absolutely to be filled in (if it is not informed, the considered system by default than the language used is English), and the totality of the form is filled in the same language, unless otherwise indicated (fields translated in English have, in this case, to carry the mark 'en', for example «TITLE en:»). Then, in all cases where a field is filled in another language than the language by default of the form (generally English), it is necessary to create a specific field name for the language used. Different translations of the source name can therefore coexist in a same form, for example, if the language of the form is not specified and is therefore supposed to be English:
Acronym: OSTC
Acronym nl: DWTC
Acronym fr: SSTC
Example if the language of the form is specified as French:
Metadata_Language: fr
Acronym: SSTC
Acronym nl: DWTC
Acronym en: OSTC
Another method would be to use attributes to characterise the content of a field:
Acronym (Language= en): OSTC
or<meta Name=«acronym» Language=«French» content=«SSTC»
It is necessary to follow the appearance of standards on this subject on Internet.

In the future, one can imagine other translation levels, such as translations for different publics-targets (see Kestemont and al, IATAFI 1996). This case is especially applied to descriptive fields like keywords or the summary. Once more, it is the use of attributes for field names that allows us to envisage this possibility, for example:
Keywords (Scheme= GEMET): water, air
Keywords (Scheme= DIF): GLOBAL_CHANGES> CLIMATE
Description (Users= expert; Language= French): Description précisé.
Description (Users= novice; Language=French): Description vulgarisée.

All the other questions can have several answers (for example AUTHOR is singular but one can list several authors in their logical order of publication).

It is an application of the repeatability characteristic of each field, particularly useful if the metadata comes from different sources (possible attributes of the field can, in this case, vary).

Questions or multiple choice ending by two points: indicating that one can specify by a free text with regard to choice.

This point is very important and happens to be used in the DIF format. It allows the evolution of thesaurus' provided by default in function of the group of real users. Compared to a free keyword list, it offers the advantage of allowing the user to situate keywords in a comprehensible context from others. For example, the keyword «Brussels» entails confusion between several cities existing in Europe and on the American continent, but if this keyword is informed as an under-entity of a thesaurus informing the country «Belgium», the ambiguity disappears.

Do not hesitate to provide supplementary information with regard to each question, annexe or via a present or non present reference on Internet.
If you have already described this source in other similar questionnaires, sent in thirds (CORDIS, AEE, etc.), it is useless to restart: simply return a copy of this form to us, and only answer questions that seem to bring you complementary information (for example, the translation of an official name or the summary in another language).

The proposed system has to allow a maximum of suppleness. The administrator will see to the transfer in a common interchange format. Indeed, this metadatabase has not, in theory, a system of pressure to oblige suppliers information to conform to its standards. Being a cross-roads of information, stress is put on the possibility of reusing the existing metadata, without losing the wealth of description in other systems.

Guidelines for each frield or question

<u>Technical data:</u>
1.1.* NUMBER ORDER (code to which the respondent can make reference to possibly annexed documents: translations, illustrations, copies of other forms. It is recommended to use a short character chain allowing the identification of only the object described, like a series of acronyms in hierarchical order, for example: ULB_CEESE_CDS or SSTC_FEDRA).

Ideally, this number order should not exceed 31 characters (DIF). By default, we will use the DIF methodology .

1.2.* LANGUAGE OF FORM *(language used by default in the answers)*
The ISO-2 code will be applied in preference. The name of the language in English can be chosen according to the evolution of the most well known standards.
As a reminder, the ISO-2 code is written in small for the name of language, with possibly a an addition representing the name of the country for an individual idiom. For example «nl BE» for Flemish and «nl NL» for the Dutch from the Netherlands, nl for ABN. In the frame of the metadatabase, it is recommended not to use national linguistic particularities and only the general code in two characters will be used.

Some language code and their equivalent in English:

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

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

<u>Source type :</u>
2.1. SOURCE TYPE
Reference: This list is the product of source types used in numerous inventories. Sources: EEA-CDS, UNEP-HEM and CORDIS.
The list can be developed by using specific standards (IAFA for Internet documents or FGDC for geographical databases). Several source types can be used (field repetition). For example:
Object_Type: institution
Object_Type: academic

Here are some types of objects being able to be used in a standard manner (ref.: EEA, 1994):

Main categories (highly recommended utilisation):

Institution
Person
Activity
Product

Secondary categories (utilisation recommended in addition to the main category, the code being that of the AEE, it is necessary to specify Scheme=AEE if one uses this code. One can, on the other hand, use the proposed English terms.)

Sort of institution:
PU Public (or public power dominant in the managing board)
PA Parapublic
PR Private (not specified)
AS Private association/foundation
NA Private non associate

Type of institution:
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-governmental
PO Political responsibles
ME Media, e.g. press
RL Religious

Type of person:
Expert
Documentalist
Promoter
Researcher
etc.

Type of activity:
Program
Project
Monitoring network
Inventory
Statistical enquiry

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

A more complete codification or designation in English of other types of sources can be inspired from what is proposed for the field FORM, CONTENT. Another alternative is to use the codification or designations of US-MARC which is also very detailed on this subject.

Title:
* 3.1. NAME OR TITLE:
Maximum of 80 characters and, in preference, less than 31 characters (DIF).
In the case of one person, this field is built up from the following fields, beginning, in preference, with the surname to facilitate sorting. Example: Dupont, John.

(if one person:
3.1.1. TITLE: * 3.1.2. **SURNAME**:   * 3.1. (FIRST NAME)
One can add a field Middle_Name for some foreign names.
The first name is entirely given by default or in the form of an initial (example J.L. for John-Luc). If it is necessary to put several names, join them in the same field.

3.2. ACRONYM:
By default, the letters are, in preference, in capitals and without points. Example: OECD. Other representations are admitted if it concerns official representations. Examples: UN/DPCSD, ifen, gsf, AIrBr. One can, if needed, put several synonyms.
One can also make reference to a small logo (URL in GIF image).

3.3. PARENT DATA SOURCE: *(database, series, program of which this is possibly a part). One can fill a minimal form for this reference (see later).*
Call the source by its official name (if possible following the defined format by NAME/TITLE) or by a non ambiguous acronym, and/or give a precise reference (URL or joint questionnaire).

3.4. POSSIBLE REMARKS ON THE INTERNAL ORGANISATION OF THE DATA
SOURCE :

Text and/or reference (URL,...) to an explanatory external document (image, organigramme,...).

3.5. LINK WITH ANOTHER DATA SOURCE :
Specify the type of link and name of the source, as well as an identification element, such as the URL.
Use, in preference, the following formulation (respect the case and spaces):
Link_type: name of the link (and URL)
Example:
Relationship: Promoter: Prof. André Berger
Relationship fr: Promoteur: Prof. André Berger

However, a more complex sentence can be used, with, if possible, a hypertext reference towards linked elements (use the html tag).

Objectives:
Except in import cases from an existent base, remain brief!

4.1. OBJECTIVE:
4.2. GENERAL DESCRIPTION (SUMMARY):
4.3. UTILISATION OR SERVICE RENDERED:

Content:
5.1. GENERAL PROCESSED AREAS, KEYWORDS:
(General keywords from a current Multilingual Thesaurus are presented by default according to the specific inventory, but the respondent can use keywords (descriptors) coming from another classification system, that must be then specified). Minimum example:

environment:
economy:
sociology:
institutions:
other:

Use the GEMET of the European Agency for the Environment, and in a first stage its provisional versions (see document provided in the annexe).
One can also use the DIF thesaurus «disciplines», to load onto the WWW.
For specific forms, the administrator will have to operate a selection of appropriate words before shortening the forms (the complete listing of the GEMET takes 70 pages!).
Even in case of using a specific vocabulary at the interface-user level (for example CO_2 instead of dioxide of carbon), the administrator will insure that the closest synonym of the GEMET will, in practice, be encoded.

5.2. PARAMETERS, UNITS AND METHODS:
(brief description of the product, especially taking into account detailed elements best describing the quantity and quality of data as well as their disintegration. If possible, in case of classification, list classes. Join an illustration to the annexe) :
[example: Emissions of 4 air pollutants- NOX, SO2, N2O and CO2-in tons per year by 44 economic sectors and 33 fuels in 2 countries (The Netherlands and Belgium) over a period of 10 years (1980-1990). Calculated by multiplying specific emission factors by the utilisation of final energy. Follow the list of sectors and fuels: ...].

This information can also be obtained by using the «Parameters» DIF thesaurus. In this case, it is better to use a new (DIF) field name :
Parameter

All specific lists can be informed in a Parameter field (list of species, etc.) or in a separate keyword list in a supplementary Keywords field (or in Keyword fields)

5.3. ANNEXED ILLUSTRATION (give a reference):
(illustration or descriptive brochure for the product, such as map, graph, example of form, model of data, formulae, photograph, etc. under the digital form in an interchange standard GIF, TXT or HTML or maximum 2 A4 pages. For an illustration provided on digital support, specify the access path and filename)
If necessary, the administrator could scan (digitaliser) documents provided on paper, and include them in GIF format. However, it is only necessary to process, in priority, images stocked on other servers.

5.4. USED STANDARD :
Each standard can eventually be described on a new form, with a source type : product, database, publication: standard
or in practice
Object_Type: product
Object_Type: document
Object_Type: standard
or only
Object_Type: standard

Geographical coverage:
6.1. GEOGRAPHICAL COVERAGE: *(to which the source relates)*
Standard descriptors are proposed, any keywords can be added.

Administrative:
> *specify (use ISO and NUTS codes if there isn't any space)*

International:
National:
Regional:
Local:

Non administrative:

Course of water:
Plans of water:
Oceans and seas:
Coasts:
Basins of drainage:
Urban zones:
Industrial zones:
Rural zones:
Others: (see DIF location keywords, by discipline: astronomy, sciences of the earth, planetary sciences, solar physics, physics of space).

Geographical zone covered *(rectangle, volume, or central point)*
Data that follows delimits a volume, a surface or a point on the worldmap, in which the specialisation area of the source described is situated. As an example, the maxima and minima are confused.
6.2. WESTERN MOST LONGITUDE ddd mm ss E/W:
6.3. EASTERN MOST LONGITUDE ddd mm ss E/W:
6.4. SOUTHERN MOST LATITUDE ddd mm ss N/S:
6.5. NORTHERN MOST LATITUDE ddd mm ss N/S:
6.6. ALTITUDE MIN (meters, positive above sea level, negative for depth):
6.7. ALTITUDE MAX (meters, positive above sea level, negative for depth):
6.8. MIN THICKNESS (meters):
6.9. MAX THICKNESS (meters):
6.10. ALTITUDE/STRATUM (keywords):

6.11. GEOGRAPHICAL ZONE COVERED (FREE FORMAT):
One can also specify a more complex zone by a series of object co-ordinates (points, lines, polygons, volumes) under the following format (units: degrees, meters): LAT11, LONG11, ALT11) (LAT2, LONG2, ALT2)… (LAT1n, LONG1n, ALT1n)… (LAT11, LONG11, ALT11). It is necessary to close the lid by repeating the first value if it acts as a polygon.

6.12. SPACIAL RESOLUTION OR SCALE:
Free format [ex: '1:1000000', '25 ha/pixel', 'districts', 'country', 'cities of more than 10000 inhabitants', 'units of more than 2 ha', 'resolution of 10 meters', etc.]

Period covered and frequency of data (if applicable):
(format YYYYMMDD hhmmss, GMT hour, or specify the hour ???)

7.1. ACQUISITION TIME *(for example, for an image)*:
7.2. PERIOD FROM: 7.3. TO:
(For a series of data, two times are given. In case of multiple periods, it is necessary to repeat these two fields by pairs)

7.4. PERIOD *(free format)*:

7.5. DATA FREQUENCY

This question necessitates several fields, or a strictly defined format. One can alternately use US-MARC or DIF, but these standards are less precise. Notice that it is possible to use only a numerical field by using the next convention (or all other equivalents):

Questionnaire		Code
		YYYY/MM/ DD hh: mm: ss
continuous		0
regular, every…………………	seconds	0.0000...1 to 58.99999... (not 0)
(number)	minutes	1:00 to 59:00
	hours	1:00:00 to 23:00:00
	days	1 00:00:00 to 31 00:00:00
	months	1/ 00 00:00:00 to 12/ 00 00:00:00
	years	1/ 00/ 00 00:00:00 and more
irregular		9999/ 12/ 31 23:59:59
other:		9999/ 12/ 31 23:59:58

example: regular every 3 days, 20 hours 5 minutes: 3 20:05:00

Quality:
8. QUALITY:
Free textual format (DIF). More precise fields exist in FGDC.

Possible reference of the product, or copyright:
9.1. AUTHOR (repeat fields if several):
9.2. ORIGIN (organisation):
9.3. PUBLICATION DATE :
9.4. SERIES:
9.5. PUBLISHER:
9.6. PLACE OF PUBLICATION:

Versions
* 10.1. CREATION DATE :
10.2. FREQUENCY OF UPDATE:
(Same format as data frequency)

* 10.3. LAST UPDATE:
10.4. NEXT UPDATE:
10.5. STATUS: under construction: fixed:
(is the data source already or more enriched or is-it fixed?)

10.6. VERSION:

11.0.1.
Unique identificator of reference for the source
This section is destined to identify the 'mother-source', unique reference in databases or on internet, where the original source or its official description are found. Several unique reference types are possible on different supports. Only give one reference by type, the other copies being able to be indicated in the next question. Other identificators can be unique identificators in the existent metadatabases.

If need be, indicate remarks (ex: login, password, configuration)

Address
* 11.1.1. STREET/N°/POSTAL BOX:
* 11.2.1. ZIP CODE:
* 11.3.1. CITY:
* 11.4.1. COUNTRY:
* 11.5.1. LOCATION *(longitude-latitude)*:
* 11.6.1. ACCESS MAP *(join map or file)*:

Telephone
Enter the complete code, including the international code.
Example: 32-2-650 35 88
* 12.1.1. TELEPHONE (voice):
* 12.2.1. FAX:
* 12.3.1. MODEM:

* 13.1.1. URL:
enter the complete URL, including the protocol.
Example
http://www.ulb.ac.be and not www.ulb.ac.be
* 13.2.1. E-MAIL:
Example info@ulb.ac.be
13.3.1. OTHER INTERNET (specify, one address only by supplementary type):
others
* 14.1.1. ISBN:
* 14.2.1. ISSN:
15.1.1. OTHER (specify, one address only by supplementary type):

Direct access to the source or to its description (mirror sites)
Indicate here the different contact points where the source can be directly consulted (several answers possible by type).

Same format as the above mentioned. The sequence can be repeated as many times as necessary, finally to the continuation of the field corresponding to the above mentioned (this first field giving access to the original).

Construction mode, sources or references used
16.1. SOURCE
Use a URL if possible.
16.2. CONSTRUCTION METHOD :

16.3. FORM, CONTENT (tick one or several options and specify):
General types (recommended as minimum to inform in any event, even if more precise indications are given in a repetition of the same field). It could be an opportunity to align on the WWW standards for this subject, which is not the case for the list proposed here.

Questionnaire	English term (code)
fact data, tables, diagrams	Facts
texts	Texts
graphs, drawings, images, maps	Pictures
films, video, animation	Animations
samples, collection pieces, models,...	Samples
sounds:	Sounds

A more precise representation can get inspiration from the reference below.

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

Code Name

Type of product not specifying the support (content of systems):

FA Facts/tables
TX 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*

Specifying the support type :
(.P=printed;.D= digital;.A=analogic;.L= Actual*; M= Manual*; O=Others*)
examples:

Alps	EEA1994	Name
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)

Real support (combining the support type and the real support asked for later : P= Paper*; L= Line; T= tape*; D= Disk(ette)*; R= CD-ROM*; H= Hard; O= Other)

PP printed/paper
DL on-line version
DT digit-(magnetic)tape
DD digit-disk(ette)
DHdigit hard disk
DR CD-ROM (read-only memory)
DW optical disk (re-writable)
DO other digital medium
OD other kind of disk

Example of complete codes for products, combining support and content:

code	name
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. AVAILABLE LANGUAGE *(several possible answers)*
In the database, use the English designation or, in preference, the ISO-2 code (see ''Source type'').

16.5. CHARACTER SET:
ASCII
ANSI
ISO Latin-1   (used by html)

are the most well known, but certainly not the only ones !

16.6. SUPPORT (tick and specify available formats)
Questionnaire	Databases (English)
digital:	digital
diskette:	diskette
CD-ROM:	CD-ROM

magnetic tape:	digital tape
on-line (Internet, ...):	on-line
hard disk:	hard disk
analogic:	analogic
paper (manuscript, printed, photograph):	paper
disk:	analogic disk
film(negative, microfilm, film, ...):	film
magnét tape:	analogic tape
'on Line'' (fax, téléphone, TV, ...):	analogic-line
waves (TV, radio, mobile...)	hertz
sample:	
model, ...:	
other:	

In practice, at the least two fields are used. One summarises a standard appellation (above mentioned), the other specifies the format in free textual format.

For formats, one can use the lists provided on Internet in general, or the following list (EEA):

Formats are presented in function of product types (see earlier)

Type of product	System	Formats
.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, ...
PI.D:	DOS	Raster, BMP, TIF, ...
.DT:		CCT, 8mm, 1/4"
.DD:		5", 3", ...

Distribution policy :
17.1. DISTRIBUTION

Questionnaire	Code (non standard)
internal use only :	internal
external use:	external
free:	free
for sale:	cost
limited:	limited

All that remains is to look for a standard code (see DIF or US-MARC)

Deontology:
18.1. HAVE YOU DEFINED RULES?
        for the collection of data
        for data processing
        for access to data
        for the publication and the dissemination of data
        for payment of information or services

others
Look for a standard code or propose a code.

18.2. ANNEX TEXTS *(if yes, give a reference)*

Access time :
19. ACCESS TIMETABLE *(in local time, or specify)*:
Format: free text.

Approximate volume of the source:
20. APPROXIMATE VOLUME (number of recordings, pages, Kbs, values, members of personnel of a company, annual budget,...)

One can use a free format.
One can also use specific fields foreseen to this effect notably in DIF.

Organisation or person responsible for management of the source
21. RESPOSIBLE:
(give a URL)
Reference to a questionnaire already completed :

If need be, fill in a questionnaire for the company in question. See below an example of a minimum questionnaire for companies .

Other links can be defined in various places of the form.

Source of the metadata (who filled in this questionnaire?)
Anticipate a URL link or a separate minimum form for the documentalist and/or his company, or for the metadata source already published (see minimal form below).

22.1. DOCUMENTALIST:
Person, company, metadatabase, catalogue or/and document having served as source of information provided in this questionnaire... Give elements allowing the identification of the forms describing the mentioned sources.

22.2. DATE:
Last update date of this form.

22.3. REMARKS ON THE SOURCE OF METADATA:

22.4. OTHER METAINFORMATION:

Remarks and others:
23. REMARKS AND OTHER INFORMATION *(you can recopy here, or join extracts of other questionnaires to which you have replied for the same source).*
Reuse imported fields as they are, by specifying, if possible, their origin by the attribute Scheme=...).

(END of basic form.)

Limitations for the passage of one standard to another and export problems of data to other standards.

Table 2 proposes field names coming from reference standards specialised in the question asked. Whenever possible, the methodology by default and limitations for the content of these fields (taken into account in the technical form presented earlier) will get inspiration from the original standard, unless otherwise specified. However, the possibility to repeat the field allows to inform it following different standards, in average a qualification of the standard used (SCHEME...).
Importation of these standards is thus facilitated, without loss of information.
The publisher of the database will operate unit conversions and formats of the field contents imported in function of necessity - by preserving, nevertheless, the original version - in such a way as to make the database as much as possible integrated in the heart of the fields used to construct research indexes.

A possible export towards other standards will be, in principle, facilitated with the condition of knowing the rules for each of these standards. Only informed fields of origin following the export standards will be preserved and unharmed. Most probable export standards such as DIF, FGDC, CDS or WWW (Dublin), corresponding fields will be preserved (most often only their English version), whereas other fields will undergo cuts or will not be exportable without human intervention. In most cases, a possible export will inevitably entail the loss of data, external standards being generally more limiting than the interchangestandard of the form-type.

Where to find the GEMET thesaurus, and the other standards?

The existent thesaurus' used by default for keywords will be, in preference, multilingual. Only specific thesaurus' will be accepted for particular disciplines (DIF for Global Change, for example, lists of species,...). Even in this case, it is widely recommended to use keywords or descriptors coming from a general thesaurus (GEMET) and simultaneously fill in a second «Keywords» field with descriptors coming from a specialised thesaurus (to be informed by Belgian specialists from each discipline), with, this time, a reference to the thesaurus (Keywords SCHEME= 'DIF' for example).
The GEMET thesaurus can be obtained via felluga@relay.itbm.rm.cnr.it or Wolf-Dieter.Batschi@uba.de. A preliminary English version has been provided in the frame of a feasibility study, in MS-WORD format. A possible utilisation authorisation (for example, under the pretext of 'test') can be obtained at the European Agency for the Environment via the same authors.
In case of utilisation problems of GEMET, it is always possible to use an older published version. One can obtain this in paper form at Laurens de Lavieter, TNO, Schoemakerstraat 97, PO BOX 6013, 2600 JA Delft, The Netherlands (phone + 31 15 61 31 86) or on CD-ROM at Bruno Felluga (felluga@relay.itbm.rm.cnr.it).

As for the addresses of other standards quoted in the document, they appear in different places of this document and the report of the feasibility study (bibliography,...), as well as on an accredited html page:
Metadata standards directories on http://www.ulb.ac.be/ceese/meta.html.

Creation of meta-tags

From the database, one can imagine to gradually replace the URL (intern for the system) by decentralised URL, preserved by partners. In this case, it could be useful to help them to

generate META fields in the HEADER of their documents (home-page in free format, for example).

It is even advisable to anticipate the filling in of following META fields within each form, already *de facto* standard on the WWW, and recognisable by research engines : KEYWORDS, DESCRIPTION, AUTHOR etc.
See W3C and the Dublin Core (http://purl.org/metadata/dublin_core_elements).

The fact of filling in these metadata fields will allow information to reach all people researching information via the Internet, even without knowing of the existence of the SSTC, which is a supplementary advantage for metadata suppliers.