Summary Report

Datamanagement and Indicators of Sustainable Development

Standard Documents and Information Exchange

A Contribution to the Information System for Sustainable Development (ISSD)

W. De Jonge, S. Deconinck

Centre for Sustainable Development, University of Ghent
http://cdonet.ugent.be

Financed by the Belgian Federal Office for Scientific, Technical and Cultural Services (OSTC) - Researchcontract AS/O2/019 (Researchplan 'Scientific Support for a Sustainable Development Policy' - Supporting Actions) - http://www.belspo.be

1. Data about Indicators for Sustainable Development (ISD)

During the first decade that followed the UN Summit of Environment and Development (UNSED, Brazil, 1992) we saw a fast growing number of initiatives focussing on Indicators for Sustainable Development. Worldwide, on different geographical scales (city, region, country,…) collections of indicators were selected and often presented as a multicriteria evaluation aid to support Sustainable Development Policy.
Agenda 21, one of the main outputs of UNSED, stated that *"Indicators of sustainable development need to be developed to provide solid bases for decision-making at all levels and to contribute to a self-regulating sustainability of integrated environment and development systems"*(1).

Supporting decision-making (ex-post or ex-ante evaluation of policy options) remained the main argument, cited by many publications on this issue (2). Another argument is the need to operationalize the complex concept 'sustainable development', defined in many vague definitions (using vague concepts like needs, equity, natural capital, etc...). In many publications we can find a detailed interpretation of individuals or consensus seeking groups.
If we look for indicators (measurable parameters) however, we are forced to be very clear about the meaning of sustainability. One can only measure something, if you know what should be measured. Sustainability, being complex, is then split up in themes and subthemes (popular complementary themes are: social, economic and ecological, the latter often split up in soil, water and air). The fuzzy ambition 'sustainability' is translated into a list of concerns for which measurable parameters exist or could be developed. This analytical approach is a well known procedure in multicriteria evaluation.
Some researchers developed so called indices: a set of parameters combined in a formula that produces a more balanced view on a certain issue. The Human Development Index (HDI) is a well known example. Many alternatives for GNP are formulated, alternatives that take external social costs of economic growth into account. However, these indices do not completely cover the issue 'sustainability'. Aggregation always leads to methodological discussions, and can even be counterproductive. If we have one number, ranking becomes easy, and we have a simple message: Belgium scores worse than...!! The press might like it, but experts often do not. Sustainability is not at all that simple. Furthermore, one number often leads to the assumption that only one actor is responsible for the results: government. If we look at the data that are aggregated however, more actors become visible and responsible: consumers, industry, ...
Indicators for Sustainable Development are generally presented as a set of complementary parameters, which are all insufficient on their own, but they add a detail to the whole picture which is given by the full indicator set.

It is a common practice to produce for each indicator a document, describing what has been measured, by whom and when, but also why exactly this has been measured and how it should be interpreted. Generally these documents describe the relations between a certain concern (for example: poverty) and a measurable parameter that is observed (for example: number of people below a certain income level).

A concern like air-quality might be covered with more than one parameter (concentration of small particles, ozone concentration,…). The description of the relation between concern and observed parameters, might focus on the shortcomings of the latter one (not covering the concern completely). Selection of a parameter can be explained by practical means (data are available and updated regularly).
Some documents give a full description of observation method, responsible authority, quality of the data, etcetera. Sometimes short and/or long term targets are added, or a time series with an evaluation of the developments during one, five or more years.
Some documents include (optional) policy responses, others only refer to other documents that focus on the same concern.
There is no standard document for the description of Indicators for Sustainable Development, and perhaps it is not even possible to develop a de facto standard. Five different projects on ISD, in which the Centre for Sustainable Development of the University of Ghent (CSD-UG), played a central role, ended up with five different documents. Of course there was an important overlap, but budgets and preferences of the contractor can lead to differences. A 'first exercise' might lead to rather brief description of the indicators, but some projects end up with detailed data. The type of reader can have an influence: is it the broad public, or the professional policy maker, or both.

Sometimes one document is made for each parameter that is measured. In other cases, one document is made for each concern, issue or subtheme like poverty, which might then include one, two or more parameters.
A (sub)group of indicator can be introduced by a separate document that discusses the complementarity of the indicators in the group and perhaps the whole selection procedure. This means that some data, which are discussing the indicator, are not necessaraly part of the document that describes the indicator. A correct interpretation of an indicator however might only be possible if the context is read also.

Al these observations give us enough arguments to conclude that one should not seek to develop a standardized description of Indicators for Sustainable Development, that is a standardized record with fields like: name of indicator, definition, observed data, etcetera . Such standards will never become 'de facto' standards.

Secondly, if we discuss the (international) exchange of data about indicators for Sustainable Development, we should not see those data as isolated documents, each describing one indicator. These documents are part of a broader context of information, and consulting them, out of this context, might lead to the wrong conclusions. If someone looks for information about a certain indicator, he or she should be able to consult the context of this indicator too.

2. Metadata

If we cannot standardize the description of ISD - the data - we can still try to standardize the metadata. Metadata are data that describe data. In a library, all the books can contain different types of data, but in the system used by the visitors who look for books about a certain subject, each book is described with the same set of data: title, keywords, author, publisher, publication date, etcetera. These are metadata, and indeed: metadata are often used in search-facilities.

This project builds on an earlier project that launched the Belgian Information System for Sustainable Development. This ISSD contains metadata about projects, publications, people and organisations that focus on Sustainable Development (3).
The metadata for organisations include: name, type (academic, administration,...), acronym, description, keywords, address, contactperson, projects (that are described by other metadata), etcetera,...
For some metadata-elements a datafield is provided for Dutch, French and/or English. (Dutch and French are the main languages spoken in Belgium).
The types of metadata used by the ISSD are not just copied from existing metadata-initiatives. The selection of metadata is an answer to local needs as perceived, mainly by researchers in the domain of Sustainable Development. The main function of the metadata is to support search facilities. The ISSD helps people to find (Belgian) projects, publications, people and organisations and the links

(relations) between them.

During the development of ISSD, the question has been raised: could we use existing metadata-schemes? One of these existing sets is known as the Dublin Core Metadata (DC).
It is important to notice that at that time the DC was not yet as popular as it is today. The development of the DC and applications of it were in their early stages and one could only speculate about the future success of the Dublin Core Metadata Initiative. This has changed a lot during the last years. The DC has become a standard used in important international initiatives that focus on the exchange of metadata over the World Wide Web (4). Furthermore it has become clear that the Dublin Core Metadata should be used as a basic set for international and cross-domain exchange. For specific or local needs, one can always add some metadata (5). The Dublin Core Initiative provides options for extension, which give us some flexibility for adaptation to local needs. Last but not least, the Dublin Core is perceived as the best option, when we look for a general metadata-standard to be used for the wide variety of resources on the Word Wide Web. In fact, this was the first major goal of the Dublin Core Initiative: to develop a simple, flexible set of metadata that can be used on the World Wide Web, where the information crosses all domains, from poetry to mathematics, from sports to philosophy , where the creators and users of data are children, as well as professional librarians. Already in 1998 the Nordic Metadata Project concluded, after a thorough evaluation of existing metadata-initiatives that: "*the Dublin Core metadata Element Set has gained a strong status as the format for description of document-like Internet objects. At the moment is does not have serious rivals - if we do not take domain specific formats such as MARC as competitors with the more generic Dublin Core. We do not see the relationship the relationship of Dublin Core and domain specific formats as a competitive one. On the contrary, due to its flexibility the Dublin Core is an ideal exchange format, which can enable us to pass bibliographic information from one community to another...*"(6)
This viewpoint is now shared by many others and the question now becomes: how should we use the DC in 'our' context? First we will demonstrate the 'flexible' use of the Dublin Core.

The basis Dublin Core Metadata Set has 15 elements (7):

- **dc.title**: the name given to the resource
- **dc.subject**: keywords, key phrases or classification codes
- **dc.description**: an abstract or table of contents
- **dc.type**: the nature or genre of the content of the resource
- **dc.source**: a reference to a resource from which the present resource is derived
- **dc.relation**: a reference to a related resource
- **dc.coverage**: the extent or scope of the content of the resource; spatial location (place name or geograhical coordinates), temporal period (date or date range) or jurisdiction
- **dc. creator**: an entity (person, organisation) primarily responsible for making the content of the resource
- **dc. publisher**: the entity responsible for making the resource available
- **dc.contributor**: an entity for making contributions tot the content of the resource
- **dc.rights**: intellectual property rights, copyrights
- **dc.date**: a date associated with an event in the life cycle of the resource (generally date of publication)
- **dc.format**: physical or digital manifestation of the resource (media-type: print, CD-ROM, pdf,...)
- **dc.identifier**: an unambiguous reference tot the resource (often this is the URL, the web address (http://...) or the International Standard Book Number (ISBN))
- **dc.language**: the language of the intellectual content of the resource.

To give an example we produce some metadata of this article:

dc.title: Summary Report: Datamanagement and Indicators of Sustainable Development; Standard Documents and Information Exchange; A Contribution to the Information System for Sustainable Development (ISSD)
dc.creator: W. De Jonge
dc.creator: S. Deconinck
dc.date: 05-12-2003

dc.language: en
dc.publisher: Centre for Sustainable Development, University of Ghent - http://cdonet.ugent.be (adress: Poel 16, 9000 Gent, Belgium - email: cdo@ugent.be - tel: +32(0)9 264.82.10)
dc.source: De Jonge W., Deconinck S. (2003) Gegevensbeheer inzake indicatoren voor duurzame ontwikkeling; Standaardfiches en Informatiestromen; een bijdrage aan het Informatie-systeem voor Duurzame Ontwikkeling (ISDO), Centrum voor Duurzame Ontwikkeling - Universiteit Gent
dc.subject: sustainable development, metadata, xml, indicators for sustainable development, datamanagement, archive, Dublin Core - http://cdonet.ugent.be/issd/rapport.html
dc.contributor: financing of project: Belgian Federal Office for Scientific, Technical and Cultural Services (OSTC) - Researchcontract AS/O2/019 (Researchplan 'Scientific Support for a Sustainable Development Policy' - Supporting Actions) - http://www.belspo.be
dc.rights: summary can be reproduced and distributed free , copy of whole or parts should however be distributed with references to the original resource, the authors and the publisher.
dc.format: pdf
dc.identifier: http://cdonet.ugent.be/issd/summary.html
dc.relation: Information System for Sustainable Development (ISSD) - http://www.belspo.be/issd/

First one might notice that not all 15 elements are used. Elements are in fact always options and if no relevant metadata are available, one should not invent them, just to fill gaps. One should serve users of metadata with a minimum of metadata, but enough to help the user decide whether this resource is useful or not. In our example, one might suggest that a dc.description (an abstract) is missing. Making links with related resources is however important.

We've used dc.creator twice. Another option is to use it only once:

dc.creator: W. De Jonge; S. Deconinck

But our example with two separate elements is generally considered as best practice. Furthermore it is not forbidden to ad an email-address or telephone number to it. We've noticed that metadata-initiatives are often to restrictive on this, but metadata are produced to serve people. If a reader wants to know who the creator is, perhaps he or she also wants to know how to contact tat creator.

3. Relational Database management System

If the actual number of elements can be lower or greater than 15, how can I manage this with a databasesystem. If metadata are stored in a relational database management system like MySQL, one often starts to build records with one field for each metadata-element. But then of course, we have a problem. Should we give only one dc.creator field. How many dc.relation fields should we provide?

These questions are solved simply by a different construction, where we use one record for each metadata-element.

The basic fields of such a record should be:

id-key: this key is the internal unique number of the record in our databasesystem, used for cross references (common practice in relational databases)
name: name of metadata-element
content: metadata for this element
scheme: specifies specific encoding-standards or controlled vocabularies used to describe content
language: language used in content-field
resource-id: a unique identifier that links all records that describe a certain resource

With our previous example we will produce records like this:

id-key: 1
name: dc.title
content: Summary Report: Datamanagement and Indicators of Sustainable Development; Standard

Documents and Information Exchange; A Contribution to the Information System for Sustainable Development (ISSD)
scheme:
language: en
resource-id: 1

id-key: 2
name: dc.creator
content: W. De Jonge
scheme:
language: nl
resource-id: 1

id-key: 3
name: dc.creator
content: S. Deconinck
scheme:
language: nl
resource-id: 1

id-key: 4
name: dc.date
content: 05-07-2004
scheme: W3CDTF
language:
resource-id: 1

etcetera ...

All records share the resource-id value 1, which is the id-key of the first record that we've produced here. This id links all metadata of one resource.
Language here is not the language used in the resource that we are describing, but the language used in the metadata-content. The names of the creators should be pronounced as Dutch names, and therefore we use the code nl (Nederlands).
DC.date is encoded using the W3CDTF-scheme (Date and Time Format of the Word Wide Web Consortium), that is: day (two digits) - month (two digits) - year (four digits), in short: dd-mm-yyyy.

For dc.language we can us the ISO639 scheme (two letter codes for the representation of languages) Scheme can also be used for identification of controlled vocabularies. Suppose that we want to use a dc.subject, where the keywords are selected from a standard list, than we can refer to this standard in scheme.

If a metadata-provider wants to give the name of an organisation in more than one language, he or she just adds a second or third record with another language-value. For example :

key-id: 5
name: dc.publisher
content: Centrum voor Duurzame Ontwikkeling - Universiteit Gent
language: nl
scheme:
resource-id: 1

key-id: 6
name: dc.publisher
content: Centre for Sustainable Development - University of Ghent
language: en
scheme:
resource-id: 1

Now we are in a flexible position. We can use a relational database management system to manage our data (robust software is available, also as free software), and we can add as many metadata-elements as we want, provide content in many languages, using an unlimited list of schemes, etcetera. If certain metadata-providers feel that they should add new metadata-elements, the database does not need to be reorganised or changed . Users just add new elements.

What if we want to add metadata-elements that are not provided by the Dublin Core Metadata Initiative. Education Network Australia for example, preferred to add some information about the audience, a category of users for whom the resource is intended. So they added **edna.audience** to the list. The prefix edna indicates that this element is not defined by the Dublin Core Metadata Initiative, but by EdNa (5).

The Dublin Core Metadata Initiative has added refinements to the basic 15 elements. For example dc.title can be refined as dc.title.alternative. If for example the resource-title is the name of a project, it happens that the project is also known by it's acronym. The acronym can become the content of an element dc.title.alternative.

dc.description can be refined as: dc.description.abstract and dc.description.TableOfContents
dc.date also has refinements: dc.date.created, dc.date.valid, dc.date.available, dc.date.issued, dc.date.modified.
dc.relation has many refinements: IsVersionOf, HasVersion, IsReplacedBy, Replaces, IsRequiredBy, Rquires, IsPartOf, HasPart, IsReferencedBy, References, IsFormatOf, HasFormat.
dc.coverage can be refined with 'spatial' and 'temporal'.

Instead of writing dc.title.alternative, we could also write dcterms.alternative, and dc.coverage.spatial can be written as dcterms.spatial. Using prefix dcterms for refinements might be the best option for the future. It seems that dc becomes the prefix that is reserved for the basic 15 elements.

Some people might wonder. If we speak of the 'name' of the project and not it's 'title', why should I use dc.title? Because dc.title is defined as "the name by which the information resource is known". dc.title is just a label for exchange of metadata between metadata-archives. Using this label for data-exchange does not necessaraly mean that it should be used in every communication.

Suppose we have following metadata stored in our database

One resource with following records:

id-key: 0067
name: dc.title
content: Consumers and Energy
language: en
scheme:
id-resource: 0067

id-key: 0068
name: dc.type
content: project
language: en
scheme: ISSD
id-resource: 0067

Another resource with following records:

id-key: 0088
name: dc.title
content: Consumers and Energy, Final Report: conclusions
language: en

scheme:
id-resource: 0067

id-key: 0068
name: dc.type
content: publication
language: en
scheme: ISSD
id-resource: 0088

We've used dc.type to differentiate between different types of resources, in our example: project and publication (the value of scheme indicates that these terms belong to a controlled vocabulary managed by ISSD).

If our database is consulted by a person who looks for information resources that have the word 'energy' in their title (or name), both resources will be selected.
But we can change the output on screen, depending on the value of dc.type:

Name of project: Consumers and Energy
Title of publication: Consumers and Energy, Final Report: conclusions

The same goes for creator. The output on screen may be 'author' if that is less confusing for end-users that do not have to bother about international metadata-standards for data-exchange.

Generally: the labels used for forms (data-input) and responses of search-facilities, do not necessaraly have to be the same as the labels used in international communication. In Belgium we will use French or Dutch labels for communication with end-users, and not the English terms, like coverage.temporal.

Concerning the existing ISSD (Information System for Sustainable Development) it is possible to change from the actual system to a system that uses the Dublin Core with some refinements. However the construction of records in the database should follow the procedure as described in this paper, with use of language-values and schemes that refer to some local controlled vocabularies. This will cause no problems for metadata about projects and publications. For organisations and persons we propose a mixed use of Dublin Core and the vCard-standard (an electronic visiting card that has gained popularity in e-mail). (8)
In any case, it is necessary to open metadata-archives like ISSD for international exchange using protocols like those of the Open Archives Initiative.

What about metadata about Indicators for Sustainable Development? The Dublin Core can support many of our basic needs. Some additional elements could be suggested, although we fear that too much detail might become counterproductive.

There is an important difference between a publication and a set of Indicators. Writing metadata about a publication takes 20 minutes at most. Producing metadata about 20, 50, 100 or more indicators is another task. We therefore looked for a procedure that might facilitate the production of metadata for ISD.

4. Extensible Markup Language (XML)

The Extensible Markup Language (XML) has become a de facto standard for the exchange of data on the Internet. XML is a standard, set by the World Wide Web Consortium (W3C, http://www.w3c.org) , the organisation that sets most standards used for data exchange over the Internet.

If we want to produce Dublin Core metadata in XML-format, our document could be structured like this:

<dc>

<title>Summary Report: Datamanagement and Indicators of Sustainable Development; Standard Documents and Information Exchange**</title>**
<creator>W. De Jonge**</creator>**
<creator>S. Deconinck**</creator>**
<date scheme="W3CDTF">05-07-2004**</date>**
<language scheme="ISO639">en**</language>**
<publisher>Centre for Sustainable Development, University of Ghent - http://cdonet.ugent.be (adress: Poel 16, 9000 Gent, Belgium - email: cdo@ugent.be - tel: +32(0)9 264.82.10)**</publisher>**
<subject lang="en">sustainable development, metadata, xml, indicators for sustainable development, datamanagement, archive, dublin core - http://cdonet.ugent.be/issd/rapport.html**</subject>**
</dc>

All XML documents are formed by elements. Elements are delimited by start and end tags and may contain other elements and text. In our example we have an element 'title' and the start tag is <title>, the end tag is </title>. Between the tags we write the content of the element.
In the start tag we can include one or more attributes. For the date and language-elements we've added an attribute scheme. Attributes are generally used to define the type of content between the tags.

In our example we have one element <dc> that contains all the other elements, a must in XML-documents.

If we had to mix metadata-elements from the Dublin Core, with metadata-elements from EdNA, we could use another structure:

<metadata>
<dc>
<title>Summary Report: Datamanagement and Indicators of Sustainable Development; Standard Documents and Information Exchange**</title>**
<creator>W. De Jonge; S. Deconinck**</creator>**
</dc>

<edna>
<audiance>broad public**</audiance>**
</edna>
</metadata>

We could invent many other structures to support our metadata-exchange. The point is, that each piece of information (element) is wrapped between tags. A group of elements can be part of a 'parent'-element.

XML has many advantages. First of all, we can produce XML-documents with simple editors. No special software is needed. You can produce and read XML-documents on any platform (Windows, Mac, Linux, ...) without any license.
Secondly, as our examples already demonstrated, XML is flexible. One can ad new elements if needed. A group of people or organisations can decide to use a predefined structure. It is very easy to transform one structure into another. There is even a standard from W3C to define a transformation (XSLT).
If we agree to exchange data worldwide in XML-format, we can still store these data in a database to support search-facilities. Today, most database management systems provide software to export or import data in XML-format.

At the Centre for Sustainable Development we've used XML to store data about indicators (not metadata, but the description of indicators). In fact, these documents were produced with a word-processor, which allowed us to use colours or other markers to indicate changes between different versions. So, during editing we used MS-Word doc-format, and this facilitated the communication between the people involved in the content-production (in Belgium almost everyone use MS-Word,

today we could use OpenOffice.org as well).
Once the document was finalized, it was saved as plain text and so we had XML.

If we had tot store the data in a database, we transformed the XML-documents (with XSLT) into SQL-commands (the output of a transformation can be PDF, LateX, HTML or any other format). If we needed to produce metadata we used XSLT to extract the data from the indicator-documents.

Producing an XML-document is easy and everyone can learn the basics in 30 minutes. However, transformations with XSLT needs more expertise. If an organisation produces documents with descriptions of Indicators for Sustainable Development, one of the final outputs is generally a website with these documents formatted as HTML. The Hypertext Markup Language is in fact related to XML (using tags like for paragraph, for bold text, etcetera). HTML-documents are viewed with browers, and can include links to other HTML-documents. That's the basic structure of the World Wide Web.

We can produce HTML-documents that are conformant to the XML-standard (known as the XHTML-standard, also from W3C). These documents are ready for publication on the Internet, but at the same time we can use them as resources for data-collection and extraction, using XML-technology.

Suppose we have organisation X that publishes a set of indicators on the web. A HTML-document that describes an indicator could be structured like this (we use an example from Calgary, Canada):

<h1>Crime Rate and Rate of Victimization**</h1>**
** **The facts: **** The estimated person crime rate per 100,000 people for 1999 in Calgary was 1,015. The estimated property crime rate was 6,724. ** **
** **Definition: ****These statistics come from the Calgary Police Service Annual Statistical Report for 1995-99.....** **
** **Trend: **** Since 1995 the rate of property crime has decreased consistently .....** **
** **Importance: **** A community in which citizens do not feel safe, or are not in fact safe, is not sustainable.... ** **
....

In this case, the tags that are used are common in (X)HTML.
h1-elements are headers (in this case level 1)
p-elements are paragraphs
b-elements are used for bold text.

When viewed in a browser we'll see this:

Crime Rate and Rate of Victimization

The facts: The estimated person crime rate per 100,000 people for 1999 in Calgary was 1,015. The estimated property crime rate was 6,724.
Definition: These statistics come from the Calgary Police Service Annual Satistical Report for 1995-99.....
Trend: Since 1995 the rate of property crime has decreased consitently .....
Importance: A community in which citizens do not feel safe, or are not in fact safe, is not sustainable....

A webmaster can manipulate the style of presentation (colour, typography,...) with a Cascading Style Sheet (CSS). For example, if he/she wants the header to be red, the HTML-document can include a CSS with the command:

h1 {color: red}

A browser should execute this command and anything between h1-tags will be red on screen. CSS (also a W3C-standard) is a simple mechanism for adding style to HTML documents. With CSS, you

can specify such styles as size, colour and spacing of text.
If we decided, not to use bold text, we just had to give a command:

b {font-weight: normal}

We can even give commands that make certain elements invisible in the browser canvas:

h1 {display: none}

The h1-element and its content are still there, but you don't see them on screen.

5. Metadata in XML-format

Anticipating on the exchange of metadata we could use following XHTML-structure:

** **
** **cd.title** **
** **Crime Rate and Rate of Victimization** **
** **

** **
** **The facts: ** **
** **The estimated person crime rate per 100,000 people for 1999 in Calgary was 1,015. The estimated property crime rate was 6,724. ** **
** **

** **
** **Definition: ** **
** **dc.description** **
** **These statistics come from the Calgary Police Service Annual Satistical Report for 1995-99.....** **
** **

We've put each piece of information in a div-element (a general purpose XHTML-element) and added some hints for later use in metadata collections.

If a citizen from Calgary views this document in his browser, he can view it with a CSS that hides the metadata-information and sets the div 's of class=local_name in boldface.

div.dc {diplay: none}
div.local_name {font-weight: bold}

With a CSS we can produce the same output on screen, as with the original HTML-document, however we've included information that can be used for extraction of metadata.
It is indeed simple to produce a program that downloads the XHTML-document and selects the content of div-elements that contain a div where class=dc.

Conclusion: one XHTML-document can serve many purposes: publication for broader public, and metadata collection. If the content is changed, the metadata change also, without special interventions. If the program that collects metadata, visits the websites regularly, updates are made automatically.

It is easy to demonstrate that we can add not only hints for DC-metadata, but also for other metadata-schemes.
A div-element can give a reference to a thesaurus of keywords:

** **GEMET** **
** **en** **

** **energy, sustainability, Agenda 21** **

Multiple languages can be mixed, and one could give a visitor of the website the option to choose a language (which just means a change in displayed and hidden elements).

7. Harvesting metadata

Before we develop a procedure to collect metadata, we will focus on a common practice in the development of Indicators for Sustainable Development.

As we have mentioned earlier, descriptions of indicators are often not just a loose collection of documents, but often need to be interpreted in a broader context.

Fortunately the broader context in which indicators are placed have always a tree-like structure. The complex concept 'sustainability' is split up in themes and subthemes. It's like a tree were the indicators are the leaves that are attached to thematic branches.
It often happens that for each theme a standard set of observations is used (for example: pressure, state, response – the PSR-frame), but it remains a tree if we see P, S and R as smaller branches, attached to a thematic branch. If we have a two dimensional matrix, with a theme in each row and each column representing a certain aspect (like P, S,R), the columns can become the smaller branches, attached to the bigger branches - the rows.

We've seen reports that use the same indicator in more than one tree. For example: environmental indicators that are first presented in a thematic scheme (acidification, climate change,...) and a second tree that starts with sectoral branches: trafic, agriculture, housing,… But they are still trees.

The tree like structure is in fact the only 'standard' that we've discovered in data about ISD. The fact that sustainable development is generally evaluated by multicriteria analyses explains this. But there is also another argument: tree-like hierarchical structures are simple and therefore good for communication. Everyone is used to it: books are trees with chapters and sections as their branches; the filesystem on a computer is a tree.

Suppose we have a program that is able to download the XHTML-documents described earlier, and extracts the metadata using the hints given with XHTML-attributes. If an organisation has published a set of indicators in XHTML-format, and these are organised in a tree-structure, then we need to give an overview of this structure to our program that collects the metadata.

This structure can be described with an xml-document, which can be read by our harvesting software first. If we had only one indicator, the structure could be described like this:

```
<?xml version="1.0" ?>
<harvest>
  <folder>
     <indicator>http://cdonet.ugent.be/project_x/indicator_one.htm</indicator>
  </folder>
</harvest>
```

Our 'harvest' includes only one folder, containing one indicator which is referenced by its URL (link). Our software can download the page with that URL and extract the metadata.

A larger collection of indicators might be presented with following structure (in this case we have two folders, where each folder contains some indicators from two different projects. We've added some meta-elements that give us metadata about the folders.

```
<?xml version="1.0" ?>
<harvest>

<folder>
```

```
   <meta name="issd.resource" content="indicatorset" />
   <meta name="dc.title" content="IDO Project X" />
   <indicator>http://cdonet.ugent.be/project_x/indicator_1.htm</ indicator >
   <indicator>http://cdonet.ugent.be/project_x/indicator_2.htm</ indicator >
   <indicator>http://cdonet.ugent.be/project_x/indicator_3.htm</ indicator>
</folder>

<folder>
   <meta name="issd.resource" content="indicatorset" />
   <meta name="dc.title" content="IDO Project Y" />
   <indicator>http://cdonet.ugent.be/project_y/indicator_1.htm</indicator>
   <indicator>http://cdonet.ugent.be/project_y/indicator_2.htm</indicator>
   <indicator>http://cdonet.ugent.be/project_y/indicator_3.htm</indicator>
</folder>

</harvest>
```

We could allow a structure of folders, containing other folders, in order to construct a tree of themes and subthemes. We could also develop an XML-structure that allows to produce a mixture of metadata about indicators, projects, publications, organisations, etcetera.

Notes

(1) Agenda 21, Chapter 40: Information for Decision-making

(2) Gouzee N., Mazijn B., BillHarz S. (eds.) (1995); Indicators for Sustainable Development - Report of the Workshop Ghent, Belgium, 9-11 – 1- 1995, submitted to the UN Commission on Sustainable Development CSD), Federal Planning Office of Belgium, Brussels.
We mention this report, also to focus attention on the fact that the UN CSD was/is very active on this point and the fact that an institute like the Federal Planning Office, traditionally working with socio-economic indicators, supports this approach. Socio-economic policy-making and evaluation is supported by indicators for more than fifty years, environmental policy since the eighties.

(3) http://www.belspo.be/issd/

(4) One important project that uses the DC is the relatively young but fast growing Open Archives Initiative (OAI) - http://www. openarchives.org. Another important initiative that uses both DC and OAI-protocols is the European Library project (TEL) - http://www.europeanlibrary.org. Both initiatives demonstrate that major players and metadata-experts percieve the Dublin Core as a good basic metadataset, however to be used in a context that allows flexible extension.

(5) This is demonstrated by EdNA (Education Network Australia) - http://www.edna.edu.au/index.html (concerning metadata EdNA now participates in the AGLS Metadata Standard, an Australian standard for cross-domain resource description, that must improve visibility, accessibility and interoperability of government information and services (other participants are government, national archives and library). AGLS is based on the DC and the Australian standard is published on the National archives of Australia website: http://www.naa.gov.au/agls .
Another important user of DC is PRISM (Publishing Requirements for Industry Standard Metadata) - The PRISM- specification defines a standard for interoperable content description, interchange and reuse in both traditional and electronic publishing contexts (Among the PRISM Working Group members we find representatives from Time inc., McGraw-Hill, Reuters and other publishers and vendors) - http://www.prismstandard.org

(6) Nordic Metadata Project: http://www.lib.helsinki.fi/meta/index.html

(7) Dublin Core Metadata Initiative: http://dublincore.org

(8) vCard - structured values: http://www.imc.org/pdi/