

Sustainable Mobility Information System (SMIS)

A tool for decision-making.

1. Introduction

1.1 Contexte

En 1992, s'est tenue à Rio la Conférence des Nations Unies pour l'Environnement et le Développement. L'un des principaux documents adoptés au cours de cette conférence est l'Agenda 21¹. Ce document traite des actions à mener par les différents acteurs de la Société en vue d'assurer un développement socio-économique à l'ensemble des pays et qui ne mettent pas en péril celui des générations futures. Le dernier volet de ce document traite des modalités pratiques de mise en oeuvre des actions à mener, en particulier le chapitre 40 intitulé "Information for Decision-Making"². qui concerne le sujet que je vais développer ci-après.

Les deux grands objectifs de recherche à mener dans ce domaine d'action sont:

- combler le manque de données;
- augmenter la disponibilité de l'information et la rendre directement utile.

Ces deux actions passent par le développement de moyens comme:

- la standardisation;
- l'utilisation des ressources électroniques;
- toute autre méthode permettant une dissémination rapide et efficace des informations pertinentes.

Ces deux axes de recherche sont les deux conditions nécessaires à la construction d'un système d'information performant, mais celui-ci doit être construit en ayant comme autres priorités la simplicité et la rapidité des recherches d'information.

A la suite de la conférence de Rio, le gouvernement fédéral belge a mis en place en mars 1996, au travers des Services Scientifiques Techniques et Culturels du Premier Ministre (SSTC), le "Plan d'Appui Scientifique à une politique de Développement Durable"^{3 4}. Notre travail de recherche s'inscrit dans le volet "Mobilité" de ce programme de recherche.

1.2 Objectifs

Compte tenu de ce qui précède, les objectifs du travail sont les suivants:

- recherche des informations pertinentes en rapport avec la mobilité durable en Belgique;
- rendre ces informations accessibles et utilisables;
- servir de modèle pour le développement d'un système d'information plus large couvrant le développement durable en général;
- démontrer l'importance de l'interaction entre acteurs habituellement séparés pour permettre la mise sur pied de politique efficaces pour le Développement Durable.

Un des principaux problèmes auxquels les chercheurs et les preneurs de décision sont confrontés est la recherche d'informations pertinentes en liaison avec leurs travaux. Ce problème peut constituer une perte de temps, d'où l'idée de centraliser les sources de données et de références.

D'un autre côté certains détenteurs d'informations ne sont pas toujours en mesure de faire connaître leurs travaux d'où une perte potentielle d'informations. Notre but est donc de proposer un système de catalogue permettant une localisation rapide des données et, en parallèle, d'offrir la possibilité aux détenteurs d'informations de mettre à disposition les ressources en leur possession.

Une fois les informations collectées, il faut encore assurer la diffusion de celles-ci. L'émergence des technologies de l'information et d'Internet en particulier sont une opportunité dont il convient d'exploiter toutes les possibilités. L'encodage des données devra se faire en respectant des règles de standardisation afin d'assurer la compatibilité du système avec d'autres systèmes déjà existants ou émergents².

Notre travail sert de modèle au projet plus large de banque de données pour le développement durable. Il s'agit en effet, vu l'étendue du domaine à couvrir, de bien cerner les problèmes potentiels auxquels la construction d'un tel outil sera confrontée. L'intérêt de notre travail réside dans le caractère plus restreint de la mobilité, tout en offrant un champ d'investigation suffisamment large pour permettre d'envisager, d'étudier et de trouver des solutions à ces problèmes potentiels.

Un objectif scientifique du SIMS est aussi de démontrer pratiquement qu'une interconnexion étroite entre les politiques, les disciplines et les acteurs sera un élément déterminant du développement durable. Un système d'information standardisé croisant les différentes disciplines impliquées dans la mobilité soutenable (économique, institutionnelles, sociologique, environnementale) est un premier pas vers la construction des indispensables ponts entre des disciplines et acteurs habituellement séparés (chercheurs, administrateurs, public).

2. Métadonnées

2.1 Concept de métadonnées

Actuellement les recherches d'informations, notamment via les services électroniques, montrent une très grande hétérogénéité au niveau de la structure des réponses faites aux demandes. Les causes de cette variabilité sont diverses:

- disciplines en rapport avec le thème de recherche (sciences exactes, ingénierie, ressources humaines, économie,...);
- sources d'information (gouvernemental, privé, académique,...);
- types de ressources (documents, personnes, sites web,...);
- processus de création des données proposées (standards utilisés, langages, software,...);
- facteurs géographiques (langues, jeux de caractères, contexte socioculturel,...)
- ...

Tout ces facteurs influencent les processus de recherche d'information et la qualité des résultats proposés à la fin de la requête. Vu la quantité d'information actuellement disponible, l'Agenda 21 au chapitre 40.24¹ propose: "the organs and organisations of the United Nations system, as well as other governmental and other non governmental organisations, should document and share information about the sources of available information in their respective organisations" ². Ce processus de documentation des sources d'information revient à créer ce que l'on appelle des métadonnées.

Le terme métadonnées peut être défini comme "données sur les données". Une métadonnée constitue en une série d'éléments décrivant la source de données sans en proposer directement l'intégralité. Cette façon de procéder permet de réduire la quantité d'information que les outils de recherche vont devoir parcourir lors de la réalisation de requêtes. La métadonnée de base sera une série de mots clés susceptible d'être utilisés dans requête.

Une collection de métadonnées constitue une métabase de données ou une banque de métadonnées. Le rôle de ces métabases est d'accélérer les processus de recherche.

Il est à noter qu'une métadonnée ne doit pas obligatoirement présenter un caractère électronique. En effet, un fichier de bibliothèque n'est ni plus ni moins qu'une métabase de données, dans la mesure où chacune des fiches en papier contient une description en termes de titre, d'auteur, de mots clés ou encore de description. Et la consultation de ces fichiers permet au lecteur de localiser rapidement l'ouvrage désiré et d'y accéder.

La création de métadonnées présente donc un intérêt réel en terme de facilité et de rapidité des recherches. Mais plus on s'oriente vers des sujets variés tels que la description de: bases de données, organisations, personnes, échantillons, documents, ... plus il devient nécessaire d'utiliser des systèmes de description complexes. Et les dimensions des problèmes du langage et de l'appartenance culturelle peuvent venir ajouter une composante supplémentaire à ces difficultés de choix de champs descriptifs⁵.

Il convient donc de procéder à une standardisation des métadonnées. En effet si, indépendamment de la dimension du langage, les différentes sources de données se mettent d'accord sur un jeu d'éléments descriptifs communs et adaptés à la description d'une source de donnée, par exemple on utilisera les champs descriptifs: auteur, titre, nom, adresse, mots clés et description en vue de décrire un article scientifique. Une fois un standard défini au niveau de la structure et de la façon d'encoder les métadonnées, on peut alors établir des systèmes de traduction automatique ou de recherche multilingue permettant de contourner le problème du langage.

Mais la définition d'un standard doit également se faire en pensant à limiter la quantité d'informations que l'on propose. Pour résumer on pourrait dire que trop d'information tue l'information. En effet si on ne limite pas le nombre de champs descriptifs, la durée des requêtes ira donc en croissant et le risque de mauvaise interprétation du contenu des champs également. Cela conduira naturellement à des résultats sans rapport avec les résultats attendus par l'utilisateur, c'est le paradoxe de la recherche d'information⁶.

Pour résoudre ces problèmes, différentes organisations ont développé des standards de métadonnées, notamment:

- le langage HTML (HyperText Markup Language), le standard d'encodage utilisé pour la création des documents proposés sur le world wide web; ce standard⁷ prévoit différents champs qui ne sont pas visualisés automatiquement mais qui sont consultés par les moteurs de recherche; les champs sont notamment: nom d'auteur, date de création, description, mots clés.
- le Dublin Core⁸, qui est un standard de description à 15 éléments développé pour faciliter les recherches de ressources électroniques;
- Le Catalogue des Sources de Données (CDS)⁹, développé par l'Agence Européenne de l'Environnement (EEA), pour indiquer qui possède quelles informations en Europe, sous quelle forme et où et comment accéder à ces informations, il possède par ailleurs la particularité d'être multilingue.

2.2 Développement du standard de métadonnées propre à notre système d'information

Au vu de ce qui précède, nous avons développé un système de métadonnées adapté aux objectifs de notre recherche et compatible avec le caractère multilingue de la Belgique (les deux principales

langues nationales que sont le français et le néerlandais, ainsi que l'anglais, langue de référence pour de nombreux documents à caractère scientifique), ce qui apporte une dimension supplémentaire au problème¹⁰.

Notre standard a été également conçu pour être autant que possible avec d'autres standards utilisés notamment dans d'autres systèmes d'information à caractère environnemental en vue d'assurer la possibilité de transfert d'informations, notamment le "Catalogue of Data Sources"⁹ et le "Dublin Core"⁸ cités précédemment, le Federal Geographic Data Committee (FGDC)¹¹ standard développé pour les informations à caractère géographique, celui de "The U.S. Environmental Protection Agency (US EPA)"¹², différentes normes de l'Organisation Internationale de Normalisation (ISO)¹³ pour les champs de dates, de jeux de caractères, etc.

En vue de pouvoir identifier de façon unique chaque ressource décrite, nous avons d'abord défini le champ "Identifiant". Ensuite, nous avons défini les grands types de données et de sources de données que nous allons prendre en considération. Ce type d'information apparaissant dans le champ défini comme "Object_Type", les entrées possibles, pouvant se subdiviser en sous-catégories en vue d'affiner la description, sont:

- Organisation: publique, gouvernementale, académique, privée,...
- Personne: expert, promoteur, chercheur, documentaliste,...
- Activité: programme, projet, inventaire, conférence,...
- Produit: données statistiques, publication, rapport, software,...

Figure 1. Types de champ dans le standard de métadonnées du SMIS

Ensuite, nous avons introduit un champ "Name", ce champ prenant différents formats selon le type de ressource décrite, avec des références à différents champs tels que "Address", "Family_Name", "Phone", "Fax", "URL",... et qui permettent de localiser la ressource.

Différents champs doivent permettre à l'utilisateur de connaître le statut de la métadonnée proposée, tels que: "Metadata_Status", "Metadata_Date", "Metadata_Language",...

D'autres champs permettent à l'utilisateur de prendre connaissance des contraintes d'accès (prix, horaires, équipement nécessaire, confidentialité,...) aux données décrites.

Enfin un champ crucial¹⁴ pour les recherches est celui intitulé "Keywords", actuellement le système fonctionne avec un jeu de mots clés tiré du "General European Multilingual Environmental Thesaurus" (GEMET)¹⁵ développé par l'EEA. L'utilisation d'un thésaurus prédéfini permet aux utilisateurs de connaître les mots à utiliser en vue d'effectuer des recherches efficaces au sein du système d'information. L'un des autres avantages du GEMET est son caractère multilingue, ce qui rejoint un des prérequis de notre système.

3. Application du concept de métadonnées, construction du système d'information

3.1 Exigences propres au système et solutions retenues

Compte tenu de ce qui précède, les exigences à satisfaire en vue de construire d'information performant ont été recherchées. Elles sont les suivantes:

- rapidité, simplicité et accessibilité à un nombre maximal d'utilisateurs;
- pour limiter la quantité de données, le système ne doit pas être le détenteur de l'information, mais un outil d'aide à la localisation rapide des données recherchées. Cette approche permet

également de contourner les problèmes de droit d'auteur dans la mesure où le système d'information ne fait que signaler l'existence de la ressource, la décrire et donner les modalités d'accès à celle-ci;

- coût modéré en regard des autres exigences du projet (caractère multi plate-forme, accessibilité non restreinte en temps, gratuité du service,...);
- caractère multilingue au vu des exigences propres à la Belgique (français, néerlandais et anglais).

Les solutions retenues pour satisfaire à ces exigences sont:

- utilisation de métadonnées permet de donner à l'utilisateur l'information nécessaire à la localisation de la ressource;
- utilisations des technologies "Internet" permet d'assurer une dispersion maximale de l'information à un coût modéré, sans limite de temps ou de nombre d'utilisateurs;
- travail par partenariat entre deux équipes de recherches appartenant à deux communautés linguistiques différentes permet d'assurer le caractère multilingue français/néerlandais/anglais de l'outil.

3.2 Développement du Système d'Information sur la Mobilité Soutenable (SMIS)

Avant d'aborder la construction du système d'information, il fallait définir les relations entre les différents acteurs qui allaient être réunis dans et autour de celui-ci. La Figure 2 résume ces relations.

Il y a deux types de relations:

- la consultation des métadonnées par les utilisateurs qui peuvent être des administrations, des chercheurs, des membres de la société civile;
- la création de métadonnées par les sources de données, le SMIS doit offrir à terme la possibilité aux sources de données de créer leurs propres métadonnées au sein du système (institutions, organisations, experts,...); il est à noter que certains experts ont un rôle particulier puisqu'ils appartiennent au groupe d'utilisateur constitué en vue de contrôler la qualité des métadonnées proposées dans le système d'information;
- la possibilité d'interaction par l'intermédiaire du forum de discussion électronique développé au sein du système.

Le partenariat dont il est fait mention dans nos objectifs apparaît ici clairement. Les acteurs impliqués dans la mobilité durable en Belgique et habituellement séparés, se retrouvent réunis dans un système commun de sources de données où ils peuvent interagir et entrer en contact les uns avec les autres notamment par le biais d'un forum de discussion électronique consacré au thème de la recherche. Par ailleurs le fait que le développement du système soit assuré par les membres des deux principales communautés linguistiques du pays est un autre facteur de réussite.

Figure 2. Relations au sein du SMIS

3.2.1 Phase préliminaire, mise sur pied d'un outil de base¹⁶

Notre plan de travail prévoyait que, dans une phase préliminaire, le SMIS soit proposé sous la forme d'un ensemble de pages HTML reliées entre elles par liens hypertexte. La recherche de métadonnées se focalisa d'abord sur une réalisation de l'inventaire des grandes sources de données

et de métadonnées en place, notamment au niveau des instances officielles et académiques. L'accès au métadonnées se faisait au travers d'index alphabétiques thématiques.

Cette façon de procéder présentait une grande facilité de création de documents et une grande souplesse d'adaptation dans la mesure où le standard de métadonnées était encore sujet à des petites modifications. Mais à mesure que le nombre de métadonnées allait en croissant, la gestion de ces index devenait de plus en plus lourde et source de pertes de temps. La limite de gestion d'un tel système étant de l'ordre d'un millier de documents. Il était donc nécessaire de passer à des outils de gestion automatique des métadonnées de type base de donnée relationnelle. Ce type d'outil présentant l'avantage d'offrir des possibilités de recherche beaucoup plus évoluées, notamment par mots clés ou recherche textuelle au sein des informations, ce qui du point de vue de l'utilisateur est un énorme avantage en regard de la lourdeur des index thématiques.

3.2.2 Phase de développement, création du site évolué avec moteur de recherche¹⁷

La quantité de métadonnées actuellement collectées est de l'ordre de 4000 enregistrements qui étaient consultables selon les modalités de l'ancienne version. Ces enregistrements ont été importés dans une base de données relationnelle Access, qui est consultable via Internet au travers d'une interface spécialement développée pour notre système.

Les avantages d'un tel système sont nombreux dans la mesure où la gestion de l'outil n'est pas limitée par le nombre d'enregistrements, la mise à jour est quasi-automatique. Son défaut est le long temps de développement de l'outil, c'est d'ailleurs pour cette raison que nous avons développé le site préliminaire. Le tableau 1 résume les principales caractéristiques des deux approches.

	Site Préliminaire	Site Final
Type de support	Pages HTML	Base de données relationnelle
Nombre maximal d'enregistrements	1000	illimité (compte tenu des limites physique du matériel)
Traitement des métadonnées	Manuel, ou semi-automatique selon les capacités de l'administrateur	Automatique
Fréquence de mise à jour	max. une fois par semaine	Automatique/journalière
Facilité d'utilisation	bonne	très bonne
Recherche	manuelle, textuelle	assistée par moteurs de recherche

Tableau 1. Comparaison des deux approches du SMIS

3.3 Perspectives

Dans la suite de notre projet différentes tâches, seront réalisées en vue d'améliorer les performances du système, notamment:

- création de sites miroirs; le rôle d'un site miroir est de proposer exactement la même information que le site d'origine mais sur un autre serveur; l'intérêt de ce procédé est d'éviter l'encombrement du site et donc d'assurer de bonnes performances en terme de rapidité d'accès au métadonnées;
- automatisation de procédures de créations d'index thématiques pour répondre aux demandes de certains utilisateurs; le succès de certains moteurs de recherches tels que Yahoo¹⁸ montre la préférence que certains utilisateurs du web accordent aux index thématiques, pour satisfaire cette catégorie d'utilisateurs ainsi que les utilisateurs ne pouvant utiliser les moteurs de recherche pour cause d'incompatibilité de leurs browsers avec les exigences de l'interface de la base de données relationnelle;

- automatisation des procédures de collectes de métadonnées, notamment par l'utilisation de "robots" d'indexation tels que ceux utilisés par la majorité des moteurs de recherche commerciaux; ces "robots" sont en fait des programmes qui parcourent le réseau Internet et cataloguent automatiquement les informations qu'ils sont chargés de trouver;
- procédures d'aide à la création de métadonnées permettant aux utilisateurs qui le souhaitent de créer leurs propres métadonnées et de les insérer dans le système d'information; ces procédures doivent être développées de façon à garantir l'intégrité des données et éviter les malveillances ou les erreurs de manipulation.

Le but de notre travail est de construire un outil qui a terme sera susceptible de fonctionner de façon quasi autonome et ne nécessitant qu'un faible investissement en capital humain. Par la suite notre travail sera repris et inséré au sein du système d'information pour le Développement Durable que les SSTC développent sur base de l'expérience acquise au cours de notre programme de recherche.

4. Conclusions

L'essor des technologies de l'information sont opportunité et un facteur déterminant pour le développement durable. Les processus de prise de décision peuvent être améliorés grâce à la rapidité d'accès aux informations nécessaires à la mise sur pied de politiques efficaces et propices à favoriser les objectifs définis dans l'Agenda 21. Ces technologies permettent plus de transparence et une meilleure information du citoyen et des personnes concernées par les décisions prises ou à prendre.

Si l'on se place dans une optique mondiale, il importe de procéder à des standardisations en vue d'assurer la qualité de l'information et la rapidité de la transmission de celle-ci. L'intérêt de procéder à des standardisations réside aussi dans la minimisation des barrières culturelles et linguistiques, notamment via le développement des technologies de traduction automatique.

Cependant les technologies de l'information ne sont, bien sûr, qu'un outil d'aide à la décision et au Développement Durable et ne doivent pas être perçues comme une solution aux problèmes que l'on cherche à résoudre.

¹ UNEP & DPCSD 1992, "Agenda 21", United Nations Conference on Environment and Development, Rio de Janeiro 3 - 14 June 1992, at <http://kaos.erin.gov.au/portfolio/esd/nsesd/Agenda21.html>

² UNEP & DPCSD 1992, "Agenda 21", United Nations Conference on Environment and Development, Rio de Janeiro 3 - 14 June 1992, Summary of the chapter 40 at <http://kaos.erin.gov.au/portfolio/esd/nsesd/a21summ.html#ch40>

³ History of the SSTC, summary, at http://www.belspo.be/belspo/ostc/histo_uk.htm

⁴ Plan d'Appui Scientifique à une politique de Développement Durable, appel à propositions, at http://www.belspo.be/belspo/ostc/announ/pages/gc_fr.htm

⁵ DGXIII - E/4 Second Metadata Workshop, 26 June 1998, at <http://www2.echo.lu/libraries/en/metadata2.html>

⁶ B. Kestemont and W. Hecq, IATAFI 96, Information technology tools for sustainable development, at <http://www.ulb.ac.be/ceese/PAPERS/IATAFI/iatafi.htm>

⁷ World Wide Web consortium, HTML 4.0 Specification, W3C Recommendation, revised on 24-Apr-1998, at <http://www.w3.org/TR/PR-html40/struct/global.html#h-7.4.4>

⁸ Dublin Core, at http://purl.oclc.org/metadata/dublin_core/

⁹ European Environment Agency, Catalog of Data Sources, at <http://www.mu.niedersachsen.de/cds/>

¹⁰ Matthew Stiff, Multilingual Issues, in DGXIII - E/4 Second Metadata Workshop, 26 June 1998, at <http://www2.echo.lu/libraries/en/metadata2.html>

¹¹ Federal Geographic Data Committee, FGDC, at <http://fgdc.er.usgs.gov/>

¹² U.S. Environmental Protection Agency (EPA), at <http://www.lbl.gov/~olken/epaintro.html>

¹³ International Organization for Standardization, at <http://www.iso.ch/>

¹⁴ Paul Miller, Controlled Vocabulary, in DGXIII - E/4 Second Metadata Workshop, 26 June 1998, at <http://www2.echo.lu/libraries/en/metadata2.html>

¹⁵ General European Multilingual Environmental Thesaurus (GEMET), at

http://www.mu.niedersachsen.de/cds/objectives_products.html

¹⁶ Le site web où se trouve cette version est consultable à l'URL: <http://www.ulb.ac.be/ceese/SMIS/themes.htm>

¹⁷ Le site web où se trouve cette version est consultable à l'URL: <http://oder.agr.kuleuven.ac.be/smis/>

¹⁸ <http://www.yahoo.com>