

Metadata : nouveau concept de description de ressources ?

AMAROUALI Youcef

Les metadonnées (ou metadata) peuvent être définies comme étant des données relatives à d'autres données (data about data : Donnée sur des données). Par conséquent, une notice catalographique classique peut-être considérée comme une métadonnée. Le terme "metadata" est surtout utilisé pour désigner l'information "lisible par machine" concernant des fichiers de données "lisibles par machine", donc ce terme désigne en quelques sortes une information référentielle sur des électroniques.

Utilisée dans le contexte de l'information numérique géospatiale, une metadonnée sera perçue comme l'information de fond qui décrit le contenu, la qualité, les conditions et autres propriétés et caractéristiques des données.

Tim BARNERS L. du W3C définit les metadonnées comme : "... une information compréhensible (interprétable) par machine (ordinateur) sur des ressources d'information du Web ou d'autres choses (autres sources de données). L'expression "compréhensible par machine" est la clef. Nous parlons ici l'information que des logiciels traitent pour nous faciliter la vie, nous assurer que nous obéissons à nos principes, à la loi, vérifier que nous pouvons croire en ce que nous faisons, et exécuter tous les travaux régulièrement et rapidement. Metadonnée définit bien la sémantique et la structure. Metadonnée a été appelée ainsi à sa création; et est encore actuellement une information sur des ressources du Web, ou des données sur des données. Dans l'avenir lorsque les metadonnées, les langages, les moteurs de recherches seront plus évolués, ils formeront une base plus forte pour l'information "compréhensible par machine" du Web, sur n'importe quoi : les gens, les choses, les concepts et les idées. Gardons dans notre esprit ce modèle (préconisé), bien que la première étape soit de réaliser un système d'information sur l'information."

Historique :

Première époque :

Le premier standard de metadata, le format MARC (Machine Readable Cataloging) fût développé en 1965 par la bibliothèque du Congrès américain à des fins d'échanges de notices bibliographiques sur bandes magnétiques. A côté du format USMARK vers lequel évolua le format MARC, apparurent rapidement d'autres formats dérivés de ce dernier. Ce fut le cas d'abord en 1968 de UKMARC qui fut établi par la British National Bibliography en Grande Bretagne. Universitaire de GRENOBLE (France) définit sur la même lancée, un format analogue en 1970 : Monocle. A partir de là, ce fut le déferlement : AUSMARC en AUSTRALIE CANMARC au Canada, DANMARC au DANEMARK, et finalement INTERMARC pour l'EUROPE. Ce foisonnement de formats nationaux risqua de devenir un sérieux obstacle aux échanges d'information bibliographique entre pays. Ce qui amena l'IFLA à développer un format d'échange international : UNIMARC. La première version de ce format apparut en 1977. Elle fut révisée et complétée en 1987 et en 1994. Actuellement une instance de l'IFLA, le *Permanent Unimarc Committee* est chargé de son évolution et de la promotion de son utilisation dans le monde...

Quant au terme metadata, il fut employé pour la première fois par un informaticien américain Jack E. Meyers, en définissant des architectures informatiques appelées plus tard metamodèles, au début de l'été 1969. Il fit d'ailleurs des recherches dans des publications et des bases de

données pour s'assurer que ce mot n'a pas été utilisé avant lui. Après confirmation, il décida en 1986 d'enregistrer le mot « metadata » aux USA comme une marque déposée d'une compagnie de services informatiques. Cela n'a évidemment aucun lien avec le sens que nous connaissons au mot metadata, aujourd'hui.

En novembre 1987, un projet de norme d'échange de documents électroniques (TEI pour Text Encoding Initiative) fût lancé, lors d'une conférence préparatoire qui s'est tenue au VASSAR Collège de New York. Les objectifs généraux de la TEI ont été définis par la résolution finale, d'ailleurs connu sous le nom de « principes de Poughkeepsie » (du nom de l'endroit où s'est tenue la conférence). La TEI connut deux cycles de développement :

- Le premier cycle (1988/1990) déboucha sur la publication d'une ébauche des recommandations de la TEI (document TEI P1).
- Le deuxième cycle (1990/1994) permit la publication de la version officielle des recommandations de la TEI, en mai 1994.

De par leur structure, les entêtes TEI peuvent être perçues comme des métadonnées.

Deuxième époque :

Nous situons le début de cette deuxième époque au moment où des agences gouvernementales Américaines, grandes utilisatrices des systèmes d'informations géographiques ont ressenti le besoin d'organiser et de cataloguer l'information géospatiale ; ce qui était devenue une nécessité pour faciliter l'accès à cette dernière.

Dès le 19 octobre 1990, une circulaire présidentielle (A 16) américaine établit des missions pour une structure fédérale le FGDC (*federal Geographic Data Committee*) de développement d'un standard de catalogage de l'information géospatiale. Cela permit le lancement d'un standard de metadata : le *GSDGM* (Content Standard Data for Geospatial Metadata) qui fut basé sur le SDTS (Spatial Data Transfer standard) et l'*USMARC*.

Le *CGDGM* fut rapidement développé et implanté au niveau des agences gouvernementales entre 1993 et 1994.

En parallèle, d'autres standards de metadata se développaient aux USA, dont le plus connus :

- *LAFS* (Internet Anonymes FTP Archives) standard conçu par l'IETF (Internet Engineering Task Force) pour cataloguer les archives des sites FTP.
- Le *GILS* (Government Information Locator Service) élaboré pour cataloguer et ainsi faciliter l'accès à l'information détenue par les organismes gouvernementaux. En décembre 1994, le *GILS* fut adopté comme norme fédérale à laquelle devait se conformer tous les départements d'état.

Ce même standard a été adopté aussi par le Canada en 1995 pour cataloguer toute l'information des organismes gouvernementaux.

- Les standards développés par les agences américaines de l'environnement :
- + *SMSP* (Scientific Metadata Standards Projects) en 1994 par l'EPA (US Environment Protection Agency) ;
- *CERES* (California Environmental Resources Evaluation System) en 1995.
- *MRLC* (Multiple Resolution Landscape Characteristics) développé par le NBS (National Biological Survey) structure dépendant du ministère américain de l'intérieur.

Troisième époque :

Nous situons le début de cette dernière époque en 1995, c'est à dire l'année où, sous la houlette d'OCLC (Online Computer Library Center), un colloque international, réunissant des professionnels des bibliothèques et de l'informatique, fût organisé à DUBLIN, dans l'état de

l'OHIO (USA). Il permit d'aboutir au lancement d'un standard simple et ouvert : le Dublin Core. Celui-ci est supporté par une série de workshops qui lui assurent un état d'évolution permanent. Dès le deuxième workshop, qui s'est tenu à l'université de Warwick, en avril 1996 au Royaume-Uni, une syntaxe particulière est proposée pour le Dublin Core ; syntaxe approuvée rapidement par le W3C. Cela raccroche ce standard à des développements majeurs du W3C, que sont la RDF (Resource Description Framework) et XML. Une autre conférence d'OCLC sur les metadata, relative aux images accessibles en réseau, qui s'est tenue en septembre 1996, a rendu possible l'utilisation du Dublin Core dans le domaine des images.

Enfin le quatrième workshop qui a eu lieu à Canberra en Australie en mars 1997 fit l'unanimité sur une proposition d'élever le Dublin Core à un niveau de standard international.

Plus d'une quarantaine de projets, se basant sur ce standard commencent à s'implanter un peu partout dans le monde. Des standards dérivés ont vu aussi le jour tel standard de l'EDNA (Educational Network Australy) ou le standard du secteur de l'éducation américaine : GEM metadata (Gateway to Educational Materials).

Il faut signaler qu'en octobre 1995 fut fondée aussi la « Metadata Coalition », association regroupant alors 53 compagnies d'informatique (dont IBM et Microsoft) et d'autres organismes, dans le but de définir un ensemble de spécifications standards pour l'interchangeabilité et la prise en charge des metadonnées par des outils logiciels. Cela aboutit à un projet dénommé : *MDIS* (Metadata Interchange Spécification Initiative).

Devant ce début de prolifération de standards de metadonnée, et sur les recommandations de l'ISO (International Standardisation Organisation), s'est tenue une conférence du 8 au 11 juillet 1997, à BERKELEY (CALIFORNIE / USA). Celle-ci est sponsorisée par l'Agence de Protection de l'Environnement U.S., l'Online Computer Library Center et la *Metadata Coalition*, avait pour but de dégager des recommandations ou des normes la mise en place et le développement de standards de metadonnées afin d'assurer pour ces derniers :

- L'interopérabilité entre eux ;
- Le mapping ;
- L'évolutivité ;
- La flexibilité ;

Objectifs de description de ressources :

Tous les standards de metadonnée sont orientés principalement vers la description de ressources et c'est même l'objectif majeur qui a prévalu à leur développement. Pour la plupart, ils conservent leur orientation de départ. Pour le Dublin Core, nous assistons à une étonnante évolution des objectifs de description de ressources. Défini au départ, au premier workshop en 1995, comme un standard de metadonnée pouvant permettre à l'utilisateur final de décrire son propre document ou objet-document (DLO), le Dublin Core change, aujourd'hui au gré des propositions des membres actifs du groupe en charge de son développement ; membres issus de diverses communautés scientifiques. Actuellement un bouillonnant débat anime ce groupe suite à une proposition de la communauté des musées pour introduire dans le Dublin Core, la description d'événements. Par exemple, nous ne décrivons plus tout simplement un Document (compte-rendu, appel à la communication, ou programme) relatif à une conférence ; mais nous décrivons celle-ci, comme un événement, avec des liens dans le champ relation vers d'autres événements. Evidemment, la définition de certains champs sera nécessairement revue et étendue à cette nouvelle spécificité du Dublin Core. Nous pourrions imaginer aussi que d'autres communautés ; tels les gens du cinéma vont proposer que le

Dublin Core soit étendu à la description d'actions (pour définir des plans de vue dans un scénario) ; les gens des radios et les musicologues proposeront la prise en compte de la description de sons peut-être (pour mettre en valeur une différenciation de tonalités).

Que le Dublin core soit un standard de metadonnée souple, extensible et modulable que chaque communauté peut adapter à ses propres besoins de description, dans le cadre d'un projet spécifique est certainement un avantage qui a entraîné un véritable engouement autour de lui : plus de quarante projets de description de ressources ont été développés sur la base du Dublin Core, par des communautés différentes, en l'espace de deux ans. Mais peut-on concilier des propriétés simples, définies à sa création ; et qui ont permis son essor, avec une incessante tentation de structuration à chaque fois qu'une proposition est lancée ? Peut-on satisfaire l'ambition de structurer un standard au point de prendre en charge tous les besoins de descriptions aussi différents et spécifiques soient-ils ? Nous avons suivi l'évolution du Dublin Core au cours des différents workshops et avons eu une légère impression de mise à l'écart ou d'oubli (temporaire ?) à chaque nouvelle étape de certaines spécificités définies à l'étape précédente c'est ainsi que l'expression, objet - document (Document - Like Objet ou DLO) introduite au premier workshop pour définir une ressource sur le Web, au lieu d'être plus clarifiée aux autres workshops, fût écartée des discussions. De même qu'une structure (warwick framework) définie au second workshop qui s'est tenu à Warwick, en Grande-Bretagne ; et qui apportait au Dublin Core une possibilité d'interopérabilité avec d'autres standards de metadonnées ; et par conséquent une adaptabilité pour tout besoin de description de ressources et par toutes communautés, fût aussi peu développée par la suite. Il est vrai que les développements du Dublin Core, étant pris en charge par une communauté subdivisée en deux sous-groupes : les minimalistes et les structuralistes, peuvent ne pas avoir l'approbation de tout le monde. Les minimalistes prônent un Dublin Core simple (sans qualificatifs pour les champs de description) ; alors que les structuralistes militent pour un Dublin Core très structuré. Jusqu'au quatrième workshop, nous pensions que les deux tendances trouvaient une voie médiane et proposeraient, en bout de chemin un standard de metadata simple ; et pouvant constituer la base à tout autre standard de description de ressources, propre à une communauté donnée. Il est clair qu'une description de ressources sur Internet est une description d'objets documents, qui peuvent être des pages Web, des graphiques, des images, des séquences vidéo, des diapositives (ou pourquoi pas une présentation entière de type Powerpoint, genre exclu au premier workshop en tant qu'objet document). Un événement est un fait qui peut être décrit entièrement dans un objet document ; lequel est décrit par des éléments de metadata. Vouloir décrire un événement est certes plausible et correspond, sémantiquement à juxtaposer des éléments d'informations ou plutôt des éléments de repères pour le situer dans le temps et l'espace. Nous ne pourrions plus appeler cela une description de ressources selon le sens bibliothéconomique de l'expression ; mais tout simplement une description référentielle de faits. Toutefois ce type de description est tout à fait conforme à la définition générale des metadonnées (toute information sur l'information).

Nous avons évoqué au début de cet écrit d'autres standards de metadata tels que le Content Standard Data for Geospatial Metadata (CSDGM), pour l'information géospatiale, qui a pour objectif une description des ressources naturelles (essentiellement des lieux géographiques).

Nous pourrions évoquer d'autres objectifs de description de ressources :

- Information administrative pour le GILS (Government Information Locator Service) ;
- Information d'archives FTP pour IFA Templates ;
- Information catalographique pour les entêtes TEI (Text Encoding Initiative) et le standard de metadata SOIF (Summary Object Interchange Format) ;
- Informations climatologique et écologique pour le standard de metadata développé en Allemagne : CERA (Climate and Environmental Data Retrieval and Archive) ;

- Informations sur l'environnement pour le standard de metadata CERES (california Environmental Resource Evaluation System) ;
- Information écologique pour NEDRES (National Environmental Data Referral Service) ;
- Information de codage électronique des matériels tels les bandes optiques et magnétiques, pour le standard FSMS (File Storage Management System) développé par l'Association for Information and Image Management International (AIIM) ;
- Information non géospatiale pour le Content Standard Data for Non Geospatial Metadata (CSDNGM) développé par le National Biological Survey (NBS) pour le compte du National Biological Information Infrastructure (NBII).

Dans le monde des bibliothèques, l'expression « description de ressources » sous-entend directement « description de ressources documentaires ». Nous avons de mettre en relief les nuances qui peuvent exister avec des description d'autres types de ressources, lesquelles sont apparues avec l'avènement de l'information en réseaux, notamment sur Internet. Nous pensons que dans les pareils contextes, il est plutôt question d'une information de type multimédia et nous parlerons dans ce cas, tout simplement de description de ressources d'informations (ou informationnelles). Décrire un livre ou un document constitue une description d'une ressource documentaire (classique) ; qui est un cas particulier de description de ressources d'informations. Décrire un événement constitue une description d'une ressource informationnelle « abstraite ». Cette description nous apporte une multitude d'information sur cet événement, tel la date, la durée, le lieu du déroulement ; éventuellement la manière ; les attentes ou les retombées ; les participants ou les acteurs, etc...

Tout événement peut générer de l'information et par conséquent se prêter à une description comme une ressource d'informations quelconque, mais non ne pouvons l'assimiler à une ressource documentaire, selon le point de vue bibliothécaire. Par contre, il constitue bien une ressource informationnelle.

Nous avons commencé cet écrit en donnant une définition générale des metadonnées l'historique nous a permis de saisir la genèse des différents standards de metadata et de comprendre qu'à chaque fois, il y a eu le besoin de description d'une ressource d'information. L'évolution dans la description de ressources d'information a été surtout amorcée par le Dublin Core. Conçu au départ comme un descriptif bibliographique, il rapidement suivi les besoins de descriptions de ressources multiformes existantes sur le réseau Internet. C'est tout à fait un autre concept dans la description de ressources qui est inauguré, et qui correspond à un type d'information mise en réseau, propre à des communautés scientifiques très diversifiées. Cette idée ou cet été de fait ne peut certainement plus être remis en Cause, mais la question qui demeure est : Peut-on structurer un standard de metadonnées, en l'occurrence ici le Dublin Core, au point de prendre en charge tout type de description d'information, sans que cela soit au détriment de la pertinence des éléments de description ? N'est-il pas plus judicieux de proposer un standard de base que toute communauté scientifique peut moduler selon ses propres spécificités de description, tout en respectant des gabarits prédéfinis de façon à assurer l'interopérabilité des standard ? Quelqu'un me dira sûrement : c'était le borne objectif du Dublin Core, avant qu'elle ne soit dépassée par un excès d'accélération. Toujours est-il que le Dublin Core s'impose de plus en plus et est pris en compte par tous les développements qui tournent autour d'Internet. Plusieurs systèmes de recherches d'informations l'ont déjà adoptés et nous sommes complètement convaincus qu'à terme la plupart des outils de recherche d'informations sur le réseau Internet bascront l'indexation des documents qu'ils mettent en ligne sur lui. Les workshops périodiques qui se tiennent sur le Dublin Core sont devenus une véritable institution internationale où se

décident les nouvelles orientations et les nouveaux développements de ce dernier ; en fonction, évidemment des nouvelles spécificités et besoins de l'information en réseau . C'est une nouvelle ère, dans la description de ressources, qui commence.

1. Articles de périodiques :

- (1998) M. Youcef Amerouali, S. Laine-Cruzel
Prise en compte du système descriptif de documents Profildoc
Dans le cadre de metadata (Dublin (Core)
Mémoire de DEA en Science de l'Information et de la Communication
ENSSIB, juillet 1998.
- (1997) Roman S. Panchyshyn et France Bouthillier
Cataloguer le cyberspace : le défi des ressources électroniques
Documentation et bibliothèques, juillet:septembre 1997
- (1997) Younger J.A.
Resource description in the digital age
Dans : "Library trends", 45.3, pp. 462-487
- (1996) Jan Smits
Digital Metadata, Standards for Communication and Preservation
European Research Librarians Cooperation : The Librarian Quarterly
6(1996), 383-406
- (1996) Dempsey, L.
ROADS to Desire : Some UK and Other European Metadata
And Resource Discovery Projects dans : "D-lib Magazine", july/august 1996.
- (1996) Smith, Terence R.
The Meta-information Environment of Digital libraries
Dans : "D-lib Magazine", july/August 1996
- (1996) Weibel, S. et al
Advances in metadata. (1996)
Dans. "D-lib Magazine, july/august 1996"
- (1995) Mangan Elizabeth U.
The making of a standard
Information technology and libraries, 2(14), 1995
- (1994) S. Laine-Cruzel
Vers de nouveaux systèmes d'information prenant en compte le profil des utilisateurs.
Documentaliste-sciences de l'information, 1994, vol. 31, n°3, pp.143-147.
- (1994) Riekert Wf
Manager of data and services for environmental applications.
Environnement knowledge organisation and information management :
Bratislava, 14-16 septembre 1994
Knowledge Organisation in Subject Areas, DEU 1994.

2. Webographie

(1) Dempsy , Lorcan

ROADS to Desire : Some UK and Other European Metadata and Resource Discovery Projects

D. Lib Magazine, juillet/août 1996

<http://www.dlib.org/>

(2) IEEE.

The metadata and Data Management Information Page.

http://www.llnl.gov/liv_comp/metadata/metadat.html

(3) UKOLN

Metadata

<http://www.ukoln.ac.uk/metadata/interoperability/>

(4) UKOLUN

Metadata : Mapping between metadata formats.

<http://www.ukoln.ac.uk/metadata/interoperability/>

(5) EPA (Environmental Protection Agency)

Scientific Metadata Standards Project.

<http://www.lbl.gov/~olken/epa.html>

(6) weibl, S., Kunze, J. and Lagoze, C.

Dublin Core Metadata for Simple Resource Description.

ftp://ds.Internic.Net/internet-drafts/

Draft-kunze-dc-01. Txt.

(7) Gill, Tony, Grout, Catherine and Smith, Louise.

Visual Arts, Museums and Cultural Heritage Information

Standards : a domain specific review of relevant standards for networked information discovery.

<http://vads.ahds.ac.uk/standards.html>.

(8) Lagoze, Carl.

The Warwick Framework: A Container Architecture for Diverse Sets of Metadata. Juillet/août 1996.

<http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>.

EPA (Environmental Protection Agency)

(9) Library of Congress.

Metadata, Dublin Core and USMARC:

A Review of Current Efforts.

[Gopher://marvel.loc.gov/00/.listarch/usmarc/dp99.doc](http://gopher.marvel.loc.gov/00/.listarch/usmarc/dp99.doc)

(10) Miller, Paul.

Metadata the masses.for

Ariadne, Issue 5, septembre 1996.

<http://www.ukoln.ac.uk/ariadne/issues5/metadata-masses/>

American Society for Testing and Materials. (11)

ASTM Section D18.01.05 Draft Specification Content Specification for Digital Geospatial Metadata

<http://info.er.usgs.gov/research/gis/standard/index.htm>

(12) Federal Geographic Data Committee (FGDC)

<http://fgdc.er.usgs.gov/fgdc2.html>

(13) Federal Geological Data Committee.

Metadata Standards Development

<http://www.fgdc.gov/Metadata/metahome.html>

(14) U.S. Geological Survey.

Government Information Locator Service.

<http://www.usgs.gov/public/gils/>

(15) American Society for Testing and Materials.
ASTM Section D18.01.05 Draft Specification content
<http://info.er.usgs.gov/research/gis/standard/index.html>

(16) Federal Geographic Data Committee.
Metadata Standards Development.
<http://www.fgdc.gov/Metadata/metahome.html>

(17) U.S. Defense Information Technical Center.
Technology Transfer Center GILS Toolbox.
<http://skydive.ncsa.uiuc.edu/toolbox/>