

Un Dublin Core Culture pour accéder à des ressources patrimoniales hétérogènes

Présentation des travaux réalisés dans le cadre des projets Minerva¹ et 24 Hour Museum², à partir de textes de David Dawson (Resource), Richard Light et Damien Dudouit (System Simulation Ltd.)

« La norme de métadonnées du Dublin Core est un ensemble d'éléments simples mais efficaces pour décrire une grande variété de ressources en réseau. La norme du Dublin Core comprend 15 éléments dont la sémantique a été établie par un consensus international de professionnels provenant de diverses disciplines telles que la bibliothéconomie, l'informatique, le balisage de textes, la communauté muséologique et d'autres domaines connexes. »³.

Conçu pour l'ensemble des ressources électroniques, le Dublin Core est donc très général, tous les éléments sont facultatifs et peuvent être répétés sans limitation. Afin de décrire des ressources et effectuer des recherches sur des éléments plus précis, des groupes de travail ont défini, dans domaines tels que l'éducation ou les bibliothèques, des « qualificatifs » aux éléments du Dublin Core, qui permettent de les détailler, comme l'aire géographique sur laquelle porte une ressource : « DC.coverage.spatial »⁴.

Largement préconisé pour retrouver les ressources disponibles sur l'Internet, dans les balises méta (ex : <META NAME="DC.Creator" CONTENT="Foulonneau, Muriel">), il sert aussi de plus grand dénominateur commun pour connecter des ressources hétérogènes. Des équivalences entre des systèmes descriptifs plus complexes et le Dublin Core sont alors dessinées, c'est le cas par exemple du projet SCRAN pour les ressources patrimoniales écossaises⁵ ou du portail SBN-Online⁶ qui connecte des catalogues de bibliothèques, musées et archives via le protocole Z39.50 en se référant à du Dublin Core non qualifié. Il est important, dans ce cas, de différencier l'usage d'une balise pour décrire une ressource (par exemple renseigner sur les DROITS qui y sont attachés) et pour localiser une ressource (usage notamment des balises META). Dans le cadre de recherches simultanées dans plusieurs catalogues, le Dublin Core est en réalité destiné à un affichage d'informations minimales standard et surtout à servir de format pivot pour retrouver les ressources.

Utilisé comme une manière simple et standard de décrire des ressources, le Dublin Core ne permet donc pas de les définir de manière très complète et, en particulier celles qui sont déjà indexées selon des schémas complexes, telles que les ressources patrimoniales.

Le texte suivant a permis d'élaborer un schéma, le Dublin Core Culture, conçu pour localiser des ressources patrimoniales hétérogènes. Il sera implémenté dans le cadre du projet 24 Hour Museum.

24 Hour Museum est un portail d'accès aux ressources des musées qui fonctionnera avec le protocole de collecte des métadonnées de l'Initiative Archives ouvertes (OAI-PMH). Il va fédérer les ressources des musées britanniques en utilisant deux schémas : le Dublin Core Culture pour les descriptions et un schéma distinct, « résumé méta-recherches » pour l'affichage des enregistrements. Des recommandations ont été élaborées pour que les partenaires du projet créent des équivalences entre leur système descriptif et ces deux schémas⁷.

Le Dublin Core Culture répond à la nécessité de créer des points d'accès standards aux ressources patrimoniales, quelle qu'en soit la nature et l'origine, afin de faciliter la mise en œuvre de portails patrimoniaux.

Objets numériques – les problèmes du Dublin Core

Par David Dawson, Senior ICT officer, Resource⁸, david.dawson@resource.gov.uk, 2002⁹

¹ <http://www.minervaeurope.org/>

² <http://www.24hourmuseum.org.uk/>, projet d'extension du portail à la collecte des métadonnées

³ <http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>

⁴ <http://dublincore.org/documents/dcmi-terms/>

⁵ Scottish Cultural Resources Access Network, <http://www.scran.ac.uk>

⁶ <http://sbnonline.sbn.it/>

⁷ Document de Richard Light et Damien Dudouit, « 24 Hour Museum Metasearch project : schemas », 11 mars 2003

⁸ <http://www.peoplesnetwork.gov.uk>

⁹ Traduction Muriel Foulonneau

Il existe un large consensus sur le fait que le Dublin Core fournit un ensemble de métadonnées de base pour rechercher des ressources hétérogènes d'origine différente et portant sur des sujets divers, de films d'archive à des sites archéologiques, de références bibliographiques à des ressources d'informations communautaires. Le Dublin Core est un ensemble de métadonnées de base qui comprend 15 éléments:-

Titre : TITLE
Créateur : CREATOR
Sujet : SUBJECT
Description : DESCRIPTION
Editeur : PUBLISHER
Contributeur : CONTRIBUTOR
Date : DATE
Type : TYPE
Format : FORMAT
Identifiant : IDENTIFIER
Source : SOURCE
Langue : LANGUAGE
Relation : RELATION
Couverture : COVERAGE
Droits : RIGHTS

L'ensemble de métadonnées initial peut être sub-divisé en sous-éléments plus détaillés en utilisant des Qualificatifs – ainsi Couverture, par exemple, peut être qualifié en:-

dc.Coverage.geographic.coordinates
dc.Coverage.date

Comme le Dublin Core qualifié, des métadonnées plus riches contenues dans les bases de données peuvent être représentées, permettant des recherches plus fines dans un domaine particulier mais aussi des recherches croisées dans plusieurs domaines.

L'étude d'évaluation comparative¹⁰ n'a pas permis d'identifier le schéma ou les standards utilisés pour les métadonnées utilisées pour décrire les ressources patrimoniales, ceux utilisés par les institutions culturelles sont souvent spécifiques à chaque pays, ou liées à des logiciels ou à des systèmes propres aux institutions nationales. Certains ensembles de métadonnées sont largement acceptés et diffusés au niveau international, tels que l'*Encoded Archival Description* (EAD)¹¹ pour les archives, le *CIDOC Reference Model*¹², les recommandations du *RLG Cultural Materials*¹³ et le MARC (*Machine Readable Catalogues*) pour les bibliothèques. Malgré l'intérêt international pour des standards tels que SPECTRUM¹⁴ (musées) et MIDAS¹⁵ (environnements historiques), ceux-ci ne sont pas diffusés de manière aussi large. Bien que la situation ait évolué, l'enquête sur les standards menée par l'Institut Européen d'Information sur les Musées (*European Museum Information Institute*¹⁶) illustre bien cet aspect.

Ces standards spécifiques à chacun des secteurs patrimoniaux sont plus largement diffusés lorsqu'ils sont adoptés par des fournisseurs de logiciels spécialisés et une volonté générale des fournisseurs de logiciels est apparue pour accepter et promouvoir l'usage de standards. Mais, à quelques exceptions près, les produits développés par l'industrie logicielle sont spécifiques à chaque secteur patrimonial.

Ces différences entre secteurs deviennent une question essentielle pour la mise en place de fonctionnalités de recherche croisée dans des bases de données. L'expérience du projet Aquarelle¹⁷ a démontré que créer un système d'équivalences entre 11 bases de données connectées grâce au protocole Z39.50 représentait un effort très lourd. Ce travail fut d'autant plus difficile que les

¹⁰ Menée dans le cadre du projet Minerva sur les projets de numérisation en Europe

¹¹ <http://www.loc.gov/ead/>

¹² http://cidoc.ics.forth.gr/what_is_crm.html

¹³ <http://www.rlg.org/culturalres/>

¹⁴ <http://www.mda.org.uk/spectrum.htm>

¹⁵ voir <http://www.britarch.ac.uk/HEIRNET/rep1apps.html#4> pour la description des monuments britanniques

¹⁶ www.emii.org

¹⁷ <http://aquarelle.inria.fr/aquarelle/welcome.html>

ressources couvraient des domaines différents et dans un premier temps, des équivalences furent réalisées avec le profil du CIMI¹⁸, conçu d'abord pour les objets de musées. Le profil a été étendu pour couvrir les besoins de l'archéologie et des environnements historiques, mais l'une des grandes leçons du projet a été que pour la plupart des utilisateurs, les structures de données complexes pouvaient être simplifiées en :

- qui – personnes ou institutions
- quoi – sujets, noms d'objets, d'évènements, titres
- où – information géographique textuelle
- quand – dates, périodes

Le même constat se retrouve dans le rapport dans le rapport DIGICULT¹⁹:-

*Les vocabulaires contrôlés utilisés actuellement sont conçus pour la communauté académique la plus spécialisée et avertie, avec pour conséquence que, – s'ils sont fournis en ligne – les index sont rarement utilisés. Selon l'expérience tirée de Sandy Buchanan, Resource Manager à SCRAN, Royaume-Uni, "80% de nos utilisateurs utilisent les recherches textuelles, et seulement 20% se servent des recherches structurées telles que les index. Nous avons besoin d'outils qui soient confortables pour les gens. Il ne s'agit pas d'adapter les utilisateurs à l'Internet, mais l'inverse."*²⁰

Le SCRAN²¹ a mis en place l'approche 'Qui, Quoi, Où et Quand', et a constaté qu'elle est utile à la majeure partie des 20% d'utilisateurs qui se sert des options de recherche plus avancées.

Ces quatre éléments ont été incorporés au profil du CIMI – comme les points d'accès de haut niveau CIMI / Aquarelle. Le profil du CIMI comprend un ensemble beaucoup plus large et plus fin de points d'accès – environ 40 points d'accès au total.

Ainsi qu'il a été démontré précédemment, ce besoin des utilisateurs n'est pas pris en compte dans les seules métadonnées du Dublin Core, en particulier la différence entre l'espace et le temps, les deux étant inclus dans l'élément dc.Coverage.

Points d'accès de haut niveau CIMI / Aquarelle	Éléments du DC
Qui	CREATEUR EDITEUR CONTRIBUTEUR DROITS COUVERTURE
Quoi	TITRE SUJET DESCRIPTION TYPE FORMAT IDENTIFIANT SOURCE LANGUE RELATION COUVERTURE
Où	COUVERTURE
Quand	COUVERTURE DATE

COUVERTURE se réfère aux 4 points d'accès de haut niveau CIMI / Aquarelle. La définition de cet élément indique les questions posées pour l'information culturelle:-

¹⁸ www.cimi.org

¹⁹ "Technological Landscapes for Tomorrow's Cultural Economy", <http://digicult.salzburgresearch.at/>

²⁰ DigiCULT ERT, Amsterdam, September 25-26, 2001 in Mulrenin, A., Geser, Guntram & Others, 2001, p. 176.

²¹ Scottish Cultural Resources Access Network <http://www.scran.ac.uk/>

dc.coverage

Définition: La portée ou la couverture spatio-temporelle de la ressource.

Commentaire: La couverture inclut typiquement une position géographique (le nom d'un lieu ou ses coordonnées), une période de temps (le nom d'une période, une date ou un intervalle de temps) ou une juridiction (telle que le nom d'une entité administrative).

Cette problématique a été identifiée depuis plusieurs années au sein de la communauté de développement du Dublin Core, mais elle devient critique pour le déploiement de services interopérables. Créer des équivalences vers des profils plus complexes, tels que celui du CIMI permet de préciser une recherche, mais ce niveau de complexité est en général utilisé par une très faible proportion des utilisateurs. La large adoption du Dublin Core, en particulier telle qu'implémentée par de nombreux projets *Open Archives PMH* repose sur la collecte de métadonnées Dublin Core non qualifié.

Il peut s'avérer nécessaire de définir un Dublin Core Culture – utilisant les éléments Dublin Core de base, avec un nombre minimum d'extensions pour traiter ces questions. L'étude DIGICULT²² indique la nécessité de poursuivre les travaux sur ce problème.

La couverture géographique

Elle comporterait la définition d'un qualificatif pour les descriptions textuelles de la couverture géographique.

Mais, l'information géographique soulève un autre problème qui pourrait empêcher le développement de services interopérables. Pour utiliser de manière efficace l'information relative aux localisations, elles doivent être décrites, non pas en termes textuels, mais sous la forme de coordonnées qui peuvent être incluses dans un système de SIG²³. Afin d'implémenter ces fonctions dans un système de SIG, il faudrait définir une coordonnée X (*X-coordinate*), une coordonnée Y (*Y-coordinate*) et le système de référence utilisé. Les systèmes de SIG pourraient alors mettre en relation des informations issues de différents ensembles de données, pouvant couvrir plusieurs pays.

La couverture temporelle

Le même besoin de traiter des métadonnées de manière automatique s'étend à la couverture temporelle des ressources. La définition de la couverture temporelle comprend trois concepts distincts:

date – une seule date de calendrier

intervalle de dates – deux dates de calendrier définissant un intervalle entre elles

période – une formulation textuelle pour représenter un intervalle de dates

Les dates de calendrier devraient être traitées de sorte que des machines puissent fournir des résultats pertinents aux utilisateurs – en leur permettant d'effectuer une recherche sur une date, une année, ou une période. Ce niveau de métadonnées autoriserait également le développement de modes de navigation basés sur le temps, sous la forme de périodes et permettrait de générer des combinaisons innovantes d'informations spatiales et temporelles.

La couverture juridictionnelle

La définition de Juridiction n'est pas très précise et se réfère à des concepts larges tels que 'Empire Romain' ou 'Azèque' qui combinent le temps, l'espace et la culture humaine. Ce manque de définition suggère qu'il serait préférable d'utiliser les concepts de 'Culture' ou de 'Peuple'.

Ainsi, il est possible de créer des équivalences de COVERAGE.Culture avec chacun des quatre points d'accès d'Aquarelle, et un utilisateur peut raisonnablement espérer utiliser n'importe lequel de ces points d'accès pour recevoir des résultats pertinents.

²² (Mulrenin, A., Geser, Guntram & Others, 2001, p.174)

²³ Système d'Information Géographique

Peuples et Organisations

En Dublin Core, il n'existe pas d'élément spécifique pour les noms des peuples et des organisations qui sont le SUJET de la ressource, et non ceux associés à la création de la ressource. L'élément SUJET du Dublin Core est inévitablement un ensemble de termes généralisé et est défini comme

dc.subject

Définition: Le sujet du contenu de la ressource.

Commentaire: Typiquement, le sujet sera décrit par un ensemble de mots-clefs ou de phrases ou un code de classification qui précisent le sujet de la ressource. L'utilisation de vocabulaires contrôlés et de schémas formels de classification est encouragée.

En conséquence, il est difficile de percevoir comment une recherche sur des enregistrements Dublin Core non qualifié peut renvoyer des résultats pertinents dans une interrogation sur le nom d'une personne ou d'une organisation. Il serait possible de répondre à cette problématique en identifiant spécifiquement où le SUJET d'une ressource est le nom d'une personne. Cela équivaudrait au point d'accès Aquarelle de haut niveau 'QUI'.

DC.Culture

A ce stade, il est trop tôt pour être certain de l'ensemble des besoins, mais un DC.Culture devrait être discuté pour prendre en compte au moins les éléments suivants:-

Points d'accès de haut niveau CIMI / Aquarelle	Éléments DC
Qui	CREATOR PUBLISHER CONTRIBUTOR RIGHTS COVERAGE.Culture SUBJECT.Person
Quoi	TITLE SUBJECT DESCRIPTION TYPE FORMAT IDENTIFIER SOURCE LANGUAGE RELATION COVERAGE.Culture
Où	COVERAGE.Placename COVERAGE.CoordinateX COVERAGE.CoordinateY COVERAGE.CoordinateSystem
Quand	COVERAGE.Culture DATE.BeginDate Date.EndDate Date.Period

Cet ensemble de métadonnées pourrait être utilisé dans le cadre de services interopérables construits en utilisant les protocoles Z39.50²⁴, OAI-PMH²⁵ ou SOAP²⁶, ou la combinaison de plusieurs d'entre eux. Ces services devraient aussi pouvoir être compatibles avec d'autres projets basés sur du Dublin Core simple, avec un minimum de travail sur les structures de données et les schémas.

²⁴ <http://www.loc.gov/z3950/agency/>

²⁵ Open Archives Initiative Protocol for Metadata Harvesting <http://www.openarchives.org/>

²⁶ Simple Object Access Protocol <http://www.w3.org/TR/SOAP/>

Éléments du DC Culture utilisés pour le projet 24 Hour Museum²⁷

<i>Elements de haut niveau DC Culture</i>	<i>Correspondant aux éléments simples du Dublin Core</i>
Qui	Créateur Editeur Contributeur Droits
Quoi	Titre Description Type Format Identifiant Source Langue Relation
Où	
Quand	

²⁷ Traduction d'un extrait du document de Richard Light et Damien Dudouit