briefing paper

digital preservation europe

Un modèle de données pour les métadonnées de préservation

En juin 2003 OCLC et RLG ont mis sur pieds un groupe de travail international nommé PREMIS (PREservation Metadata Implementation Strategies) qui a fonctionné pendant deux ans. Ce groupe de travail comprenait plus de trente experts de cinq pays différents, représentant les bibliothèques, les musées, les archives, les administrations et le secteur privé. L'objectif était de définir un « ensemble de groupes 'essentiels' de métadonnées de préservation » utilisable par la communauté de la préservation numérique. En mai 2005, un rapport final était publié. Il comportait un modèle de données pour les métadonnées de la préservation et un dictionnaire des données version 1.0. Actuellement, l'activité d'implémentations est assurée par le groupe de maintenance PREMIS qui est en charge de la révision et de la maintenance du modèle et du dictionnaire des données.

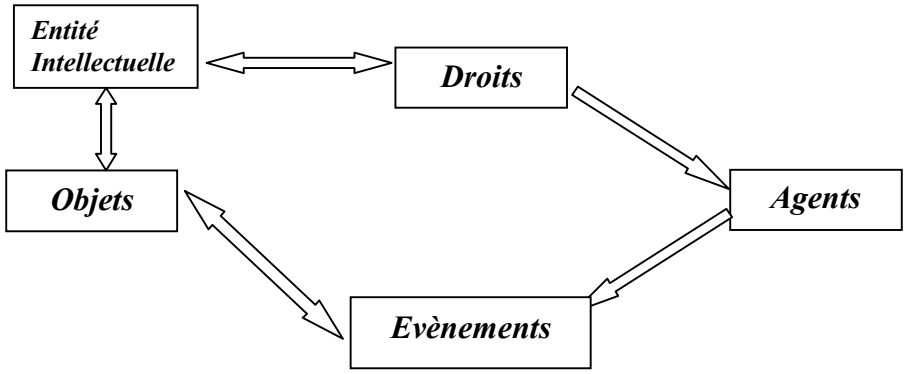
Les besoins

Le groupe de travail a identifié cinq domaines pertinents pour les métadonnées de la préservation :

• La provenance : Ce domaine devrait conserver les informations ayant trait à l'historique de la conservation de l'objet numérique, depuis sa création et en suivant les changements successifs dans la conservation physique et / ou de propriété.
• L'authenticité : Ce domaine devrait comprendre les informations nécessaires pour attester que l'objet numérique conservé est bien ce qu'il prétend être, qu'il n'a pas été altéré, soit intentionnellement, soit par inadvertance, sans documenter le fait.
• Les activités de préservation : Ce domaine devrait comporter la documentation des actions successives entreprises pour la préservation de l'objet numérique, et garder la trace des conséquences de ces actions sur la présentation, le rendu ou la fonctionnalité de cet objet.
• L'environnement technique : Il devrait décrire le matériel, le système d'exploitation et les progiciels nécessaires à l'activation et à l'utilisation de l'objet en l'état dans lequel il est conservé.
• La gestion des droits : Devraient être conservés tous les droits restrictifs de la propriété intellectuelle qui limitent le pouvoir du dépôt de conservation de prendre des dispositions pour la préservation de l'objet numérique et sa dissémination vers les utilisateurs présents ou futurs.

Résumé du modèle des données.

This Résumé du modèle des données.
Le modèle de données PREMIS est constitué d'entités, de relations et de propriétés qui sont appelées des unités sémantiques.



Les entités

• L'entité intellectuelle – un ensemble cohérent de contenu qui est raisonnablement décrit comme une unité, par exemple un livre, une carte, une photo, ou une base de données. Parce que cette entité est bien décrite dans les métadonnées descriptives, elle est considérée comme en dehors du champ du dictionnaire des données.
• Objet ou objet numérique – une unité discrète d'information dans la forme numérique.
• Evènement – une action qui implique au moins un objet ou un agent connu du dépôt de préservation.
• Agent – une personne, un organisme ou un progiciel associé aux évènements de préservation durant la durée de vie d'un objet.
• Les droits – les assertions liées à un ou plusieurs droits ou permissions appartenant à un objet et/ou un agent.

Les relations

Les relations sont des formulations d'associations entre des instances d'entités. « Relation » peut être interprétée au sens large ou restreint et toute relation de fait peut être formulée de beaucoup de manières différentes. Les relations entre objets sont des variantes de trois catégories de base :
• Les relations de structures montrent les relations entre les objets et leurs parties. Les relations de structures entre les fichiers qui constituent une représentation d'une entité intellectuelle, sont très clairement des métadonnées essentielles de préservation. Si dans un dépôt de préservation, les parties d'un objet numérique ne peuvent être remontées ensembles, alors cet objet n'a pas été préservé.
• Les relations de dérivation résultent de la réplication ou de la transformation d'un objet. Le contenu intellectuel de l'objet résultant est le même, mais l'instanciation de l'objet et son format, sans doute, sont différents. De nombreux objets numériques sont complexes et les informations de structures et de dérivation peuvent changer dans le temps, selon les activités de la préservation.
• La relation de dépendance existe lorsque un objet en requière un autre pour renforcer sa fonction, ses résultats ou la cohérence de son contenu. L'objet de complément ou d'aide ne fait pas formellement partie de l'objet lui-même mais est nécessaire à sa présentation ou son fonctionnement.

Les propriétés.

Les unités sémantiques sont les propriétés d'une entité. Dans certain cas, une unité sémantique peut être un contenant qui regroupe un ensemble d'unités sémantiques de la même famille ; les sous unités regroupées sont alors appelées des composants sémantiques de l'unité sémantique.

Le principe de 1:1

Pour les métadonnées, le principe de 1 :1 pose que chaque description ne décrit qu'une et seulement une ressource. Appliqué aux métadonnées de PREMIS, chaque objet détenu dans un dépôt de préservation (fichier, chaîne de bits, représentation) est alors décrit comme une chaîne de bits statique. Il n'est pas possible de modifier cet ensemble, mais seulement d'en créer un nouveau qui est relié à l'objet source par une relation de dérivation. En effet, le dictionnaire des données a une unité sémantique uniquement pour la date de modification d'un objet puisque, par définition, un objet ne peut être modifié.

Du modèle au dictionnaire

Le modèle de données a un dictionnaire de données associé qui comprend toutes les unités sémantiques pertinentes décrivant les quatre entités couvertes par le dictionnaire (objets, agents, évènements, droits). Dans le modèle de données, les relations entre entités de types différents sont matérialisées par des flèches.

Le dictionnaire de données les traite en tant qu'information de liaison, incluant dans l'information de l'entité A un pointer apparenté à l'entité B. Dans le modèle de données, chaque entité possède un identifiant unique qui peut être utilisé comme pointer. Par exemple, l'entité objet a des flèches qui pointent vers les entités Intellectuelle et Evènements. Celles-ci sont implémentées dans le dictionnaire des données par les unités sémantiques « identifiant de liaison entité Intellectuelle » et « identifiant de liaison Evènement ».

Les avantages et les inconvénients

Le dictionnaire de données PREMIS est le résultat d'un processus consensuel de construction, inter domaines et international ; ce qui augmente ses chances d'être très largement utilisable par toutes sortes d'institutions, de contextes de préservation numérique et d'implémentations dans différents systèmes informatiques. En tant que schéma de métadonnées de préservation, il devrait pouvoir couvrir la totalité des besoins actuels ; il est orienté vers les implémentations concrètes et l'interopérabilité afin de faciliter les manipulations d'objets et remplir sa fonction primordiale : documenter les objets numériques à travers le temps et les rendre ainsi accessibles sur le long terme. L'inconvénient du dictionnaire de données PREMIS est que personne ne peut dire avec certitude que son efficacité durera. L'impacte des développements futurs dans l'utilisation des objets est difficile à prévoir.

Conclusions

Le dictionnaire des données est une pièce cruciale de l'infrastructure de la préservation numérique ; c'est aussi une composante avec laquelle on peut implémenter des stratégies durables et efficaces de préservation numérique.

Informations et ressources complémentaires

PREMIS (PREservation Metadata: Implementation Strategies) Resources.
http://www.oclc.org/research/projects/pmwg/resources.htm

Data Dictionary for Preservation Metadata: Final Report
of the PREMIS Working Group.
http://www.oclc.org/research/projects/pmwg/premisfinal.pdf

Brian Lavoie, Richard Gartner, Preservation Metadata.
DPC Technology Watch Report No. 05-01: September 2005
http://www.dpconline.org/docs/reports/dpctw05-01.pdf

PREMIS Working Group.
http://www.oclc.org/research/projects/pmwg/

PREMIS Maintenance Activity.
http://www.loc.gov/standards/premis/

Auteur : Angela Di Iorio, *Fondazione Rinascimento Digitale*
angeladiiorio@gmail.com