

La création de métadonnées pour les Collections numériques des Bibliothèques de l'Université du Texas du Nord: une étude de cas

French translation of the original paper: "A Case Study of Metadata Creation in the University of North Texas Libraries' Digital Collections".

Translated by: Françoise Leresche, Bibliothèque nationale de France, Paris, France.

Le texte de ce document est une traduction en français et peut présenter des différences par rapport au texte original. La présente traduction n'est fournie qu'à titre de référence.

Hannah Tarver

University of North Texas Libraries, Digital Projects Unit, Denton, USA.

E-mail address: hannah.tarver@unt.edu

Oksana Zavalina

University of North Texas College of Information, Denton, USA.

E-mail address: oksana.zavalina@unt.edu

Mark Phillips

University of North Texas Libraries, Digital Libraries Division, Denton, USA.

E-mail address: mark.phillips@unt.edu



This is a French translation of "A Case Study of Metadata Creation in the University of North Texas Libraries' Digital Collections" copyright © 2016 by Françoise Leresche. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License:

<http://creativecommons.org/licenses/by/4.0>

Résumé :

Cette communication présente une étude de cas à partir du travail effectué par l'Université du Texas du Nord (University of North Texas, UNT) et l'intersection de la modélisation des métadonnées, des règles de saisie des métadonnées et leur documentation, des évaluations de la qualité des métadonnées, et de la technologie pour habiliter les éditeurs de métadonnées à créer des métadonnées de haute qualité. De plus, les exemples illustrent comment chacune de ces briques a façonné le développement des autres.

L'objectif de cette communication est de fournir une vue d'ensemble cohérente sur la manière dont la documentation, l'organisation de l'information, les métadonnées et les infrastructures technologiques fonctionnent ensemble dans les grandes collections numériques pour répondre à des besoins multiples.

Mots-clés : Création de métadonnées, Documentation sur les métadonnées, Spécifications techniques, Collections numériques

INTRODUCTION

Les Bibliothèques de l'Université du Texas du Nord (University of North Texas, UNT) ont commencé à recueillir et à générer activement des contenus numériques autour de 2002 et ont mis en ligne des ressources disponibles dans le domaine public dès 2004. Leurs Collections numériques ont ainsi plus de dix ans. Actuellement, les Collections numériques comprennent plus de 1,5 million d'objets qui sont hébergés dans une infrastructure technique unique, accessibles par trois interfaces utilisateur distinctes :

- le Portail sur l'histoire du Texas (Portal to Texas History) : un site collaboratif hébergeant les ressources numériques de plus de 300 partenaires, institutions ou collectionneurs privés, à travers le Texas ;
- la Bibliothèque numérique de l'UNT (UNT Digital Library) : un entrepôt de documents gouvernementaux, ainsi que des ressources possédées par l'UNT ou créées par la communauté de l'UNT, y compris la production savante ; et
- le Portail sur l'histoire de l'Oklahoma (Gateway to Oklahoma History) : une collection de ressources provenant de la Société d'histoire de l'Oklahoma (Oklahoma Historical Society), principalement des journaux et des photographies de presse.

En 2009, au terme de la subvention pour l'optimisation de l'interface pour les généalogistes (Interface Optimization for Genealogists, IOGENE)¹ financée par l'Institut des services aux musées et aux bibliothèques (Institute of Museum and Library Services, IMLS), le Portail sur l'histoire du Texas a lancé une nouvelle mouture de l'interface publique qui a été utilisée ensuite pour l'ensemble des Collections numériques. Une équipe de programmeurs a développé en interne le logiciel en utilisant des briques logicielles au code source ouvert (*open source*). Le système avait deux pièces principales : l'entrepôt Coda où les fichiers maîtres originaux sont archivés, reproduits et conservés ; et le système d'accès Aubrey qui organise et affiche les copies web dérivées et les métadonnées descriptives correspondantes. Les Bibliothèques de l'UNT utilisent les identifiants ARK (Archival Resource Key, Clé de ressource dans une archive) pour une identification univoque et comme un élément de la stratégie d'identification pérenne pour les ressources ajoutées aux Collections numériques.

Les Bibliothèques de l'UNT ont développé elles-mêmes les logiciels permettant d'assurer toute la gamme des services exigés dans une bibliothèque numérique moderne. En 2014, les Bibliothèques de l'UNT ont défini une Politique de logiciels au code source ouvert (*open source*)² qui encourage et expose les différentes manières dont nos programmeurs peuvent publier et contribuer à des projets logiciels. En outre, l'Équipe de développement des logiciels de la Division des bibliothèques numériques a commencé à publier et à développer davantage de leurs logiciels sur un entrepôt GitHub public³. Les Bibliothèques de l'UNT ont aussi développé une politique désignant clairement toutes les métadonnées et notices catalographiques créées par les Bibliothèques de l'UNT comme appartenant au domaine public et disponibles sous la licence CC0⁴.

À l'UNT, nous adhérons à l'idée de « version bêta perpétuelle » et nous déployons de nouveaux services et de nouvelles fonctions selon nos besoins, ou à mesure de l'achèvement et de la mise en production de composantes logicielles. Actuellement (au printemps 2016), nous sommes en train de tester une version bêta de l'interface utilisateur du Portail sur l'histoire du Texas ; c'est sa première refonte complète depuis 2009. Les Collections numériques de l'UNT continuent de mûrir et de croître, et plusieurs facteurs (entre les modèles sous-jacents, la documentation et les technologies) y contribuent, en fournissant un cadre stable pour la description des contenus et une large gamme de développements

technologiques. Les sections qui suivent détaillent les décisions, les outils et les pratiques de l'UNT en matière de métadonnées.

LES METADONNEES A L'UNT

Dans les Collections numériques, les métadonnées reposent sur du Dublin Core qualifié selon un profil d'application local, avec l'ajout de plusieurs champs de métadonnées descriptives pour répondre aux besoins locaux (voir Tableau 1, où les champs propres à l'UNT apparaissent en gras). Bien que nous permettions une certaine souplesse dans l'utilisation de ces champs, toutes les notices décrivant des ressources contiennent les mêmes 21 champs possibles, quels que soient la collection, le type de ressource ou la source. Un champ supplémentaire d'administration – Meta – documente l'information capturée automatiquement par le système ou des rubriques spéciales (comme le ou les système(s) dans lequel la ressource est visible), mais il n'est généralement pas éditable. Ces 22 champs constituent le format local « UNTL » (UNT Libraries, Bibliothèques de l'UNT) de métadonnées autour duquel nous construisons les systèmes et les fonctionnalités.

Puisque le contenu des Collections numériques est constitué grâce à des relations de coopération à l'intérieur de l'université comme à travers tout l'état, les métadonnées sont créées et mises à jour par des utilisateurs variés de l'UNT et des institutions partenaires. La création de nouvelles notices est réalisée pour la plus grande part par des étudiants recrutés sur le campus de l'UNT. La mise à jour des notices¹ concerne en revanche tout un éventail d'intervenants – des professionnels confirmés des bibliothèques, des étudiants et des membres du personnel formés localement, et des bénévoles sans aucune expérience préalable.

Pour traiter les questions de qualité, de cohérence et le besoin de conseils à plusieurs niveaux (pour les intervenants expérimentés comme pour les nouveaux), les gestionnaires de métadonnées ont créé une documentation approfondie destinée à l'usage interne comme à une diffusion publique⁵. Nous commençons avec le concept général de complétude, en établissant les valeurs pour les huit champs qui sont requis pour toute ressource afin de remplir le critère de notice « minimale viable » dans les Collections numériques⁶. Pour certains de ces champs, comme les champs Titre et Description, une information qualifiée spécifique est exigée (par exemple, Titre principal et Description du contenu), même si des informations complémentaires sont disponibles ou pertinentes pour la ressource (par exemple, Titre de périodique ou Description physique). Dans le Portail sur l'histoire du Texas et dans le Portail sur l'histoire de l'Oklahoma, chaque notice doit avoir au moins un mot-matière en provenance du répertoire des Bibliothèques de l'UNT pour la navigation par sujets (UNT Libraries Browse Subjects, UNTL-BS), qui est un vocabulaire hiérarchique, créé et maintenu localement.

¹ NdT : Dans la suite de la communication l'expression « édition des métadonnées » est utilisée pour désigner de façon générique la création, la mise à jour et l'enrichissement des métadonnées, et l'expression « éditeur de métadonnées » pour désigner les personnes qui réalisent les différents aspects de ce travail.

<i>Champ</i>	<i>Qualifié</i>	<i>Repérable</i>	<i>Vocabulaires contrôlés</i>	<i>Description</i>
Titre	Oui	Oui	Qualificatif uniquement	Titre publié, ou un bref titre forgé identifiant la ressource
Créateur	Oui	Oui	Type, Fonction	Entité ayant une responsabilité principale dans la création de la ressource
Contributeur	Oui	Oui	Type, Fonction	Entité ayant une responsabilité secondaire dans la création de la ressource, ou une relation de provenance
Éditeur	Non	Oui	N/A	Nom et lieu d'une entité qui a formellement publié la ressource
Langue	Non	Oui	Liste contrôlée	Langue de tout texte visible ou audible dans le contenu de la ressource (peut être « Pas de langue »)
Date	Oui	Oui	Qualificatif uniquement	Date ou plage de dates, connue ou approximative, à laquelle la ressource a été créée, moissonnée (sites web), soumise ou acceptée (applications de brevet), ou à laquelle un embargo sur une ressource numérique prendra fin
Description	Oui	Oui	Qualificatif uniquement	Attributs de l'objet physique (par exemple, matériau ou taille) et du contenu de la ressource
Sujet	Oui	Oui	Qualificatif, UNTL-BS	Termes qui décrivent le contenu de la ressource, provenant de vocabulaires contrôlés et/ou mots-clés
Source primaire	Non	Non	Liste contrôlée	Indication si la ressource est considérée comme une source primaire (par exemple, une photographie ou une lettre originale) ou non (par exemple, un livre écrit en utilisant d'autres sources)
Couverture	Oui	Oui	Qualificatif, Période chronologique	Information sur l'étendue géographique, géospatiale ou temporelle du contenu de la ressource
Source	Oui	Non	Qualificatif uniquement	Une ressource ou un événement plus vaste dont la ressource est extraite ou dérivée.
Citation	Oui	Oui	Qualificatif, Revue savante	Parties composantes de la citation formelle établie pour la ressource, telles que volume, fascicule, ou page de début et de fin
Relation	Oui	Oui	Qualificatif uniquement	Explication et lien vers une autre ressource qui a un type particulier de relation avec la ressource décrite
Collection	Non	Oui	Liste contrôlée	Une catégorie large rassemblant des ressources similaires, sur la base d'une collection archivistique, d'une source de financement, d'un sujet ou d'autres critères
Institution	Non	Non	Liste contrôlée	Département ou institution partenaire qui possède et gère les documents physiques
Droits	Oui	Oui	Qualificatif, Accès, Licence	Information sur les privilèges d'accès, les licences, et autres mentions pertinentes pour la ressource
Type de ressource	Non	Non	Liste contrôlée	Classification du type de support ou de contenu de la ressource originale
Format	Non	Non	Liste contrôlée	Classification du type de support de la ressource originale
Identifiant	Oui	Oui	Qualificatif uniquement	Tout code alphanumérique utilisé pour identifier la ressource (attribué par un éditeur, une institution partenaire ou une autre entité)
Diplôme	Oui	Oui	Qualificatif uniquement	Information identifiant essentiellement les aspects du diplôme associé à une thèse ou à un mémoire ; également utilisé pour les départements de l'UNT où la ressource a été créée
Note	Oui	Oui	Qualificatif uniquement	Information supplémentaire sur le contexte, l'administration ou la ressource qui n'est pas renseignée ailleurs dans la notice.

Tableau 1. Liste des champs à renseigner dans les métadonnées UNTL
(les champs obligatoires sont surlignés ;
les noms des champs définis localement apparaissent en gras).

Pour introduire les tout-nouveaux collaborateurs à la création de métadonnées, un « Guide de démarrage rapide sur les métadonnées » est disponible en ligne et expose les utilisations les plus courantes de chaque champ⁷. Des pages web complémentaires fournissent des descriptions approfondies des exigences et de l'utilisation appropriée pour chaque champ, décomposées en différents points d'une liste à puces, avec des exemples de valeurs couvrant autant de scénarios que possible, depuis le scénario de base jusqu'au plus rare⁸. Ces pages permettent aussi une maintenance relativement aisée pour documenter de nouveaux exemples et des précédents selon les besoins.

Pour certains documents très spécialisés, nous avons aussi créé des recommandations pour les métadonnées propres à une collection. Ces recommandations définissent les critères requis pour chaque champ et son utilisation pour la description de ces ressources⁹. Bien que cette approche nécessite de la maintenance supplémentaire, elle gagne du temps et encourage un surcroît de cohérence pour les vastes collections ouvertes.

SYSTEME / DEVELOPPEMENTS TECHNIQUES

Le succès local des Collections numériques des Bibliothèques de l'UNT est dû en grande part aux modèles de métadonnées qui combinent les approches conceptuelles et organisationnelles, ainsi qu'à des fondements techniques solides sur lesquels construire des services de métadonnées pour les utilisateurs, tant internes qu'externes. Bon nombre des composantes techniques exploitent aussi le fait que tous les objets numériques utilisent les mêmes champs possibles et sont donc construits selon la même structure.

Le format de métadonnées des Bibliothèques de l'UNT (format UNTL) a une sérialisation normalisée en XML qui est définie par un schéma XML¹⁰. Sur la base de ce format UNTL, il existe une bibliothèque Python pour lire, écrire et accomplir les opérations courantes sur ces fichiers de métadonnées, appelée « pyuntl »¹¹. En outre, « pyuntl » comprend plusieurs autres composantes, telles que : les programmes implémentant notre mesure de « complétude », les conversions depuis le format UNTL vers d'autres formats comme les notices en Dublin Core standard et les balises de métadonnées de HighWire Press utilisées par Google Scholar.

Au sens large, la structure des métadonnées renforce aussi la capacité de partage. Un des principaux mérites et objectifs que les Bibliothèques de l'UNT ont pris en considération pour le développement de l'infrastructure de leur bibliothèque numérique et pour les divers systèmes d'accès, de livraison et de préservation est que les partenaires qui travaillent avec le Portail sur l'histoire du Texas, la Bibliothèque numérique de l'UNT ou le Portail sur l'histoire de l'Oklahoma devraient toujours avoir accès aux métadonnées complètes pour leurs collections. Toutes les métadonnées du système sont mises à disposition pour être moissonnées à travers un serveur OAI-PMH, soit en bloc pour chaque système d'accès, soit au niveau d'un partenaire ou d'une collection selon les besoins. Pour la notice de chaque ressource, l'entrepôt OAI-PMH met à disposition les métadonnées UNTL brutes, les métadonnées en OAI_DC (Dublin Core simple, correspondant à la norme ISO) et les métadonnées en RDF/XML (Resource Description Framework/eXtensible Markup Language, Cadre de description de ressources/Langage de balisage extensible). Divers autres formats de métadonnées sont aussi disponibles à travers une API standard vers les objets dans le système de livraison de contenus Aubrey.

À l'intérieur des notices de métadonnées, des spécifications techniques peuvent contrôler les valeurs des champs. Chaque fois que c'est possible, nous gérons l'information avec un nombre limité d'options selon un vocabulaire contrôlé pour les métadonnées descriptives (voir Tableau 1), les métadonnées de préservation et d'autres activités du système. Ces vocabulaires sont gérés dans un système des Bibliothèques de l'UNT¹² et documentés par l'application des vocabulaires contrôlés de l'UNTL (UNTL Controlled Vocabularies App)¹³. Chaque vocabulaire est disponible dans divers formats, y compris JSON (JavaScript Object Notation, Notation Objet issue de JavaScript), les structures de données Python et RDF/XML. Les termes des vocabulaires ont des identifiants uniques qui répondent aux exigences des données liées et ouvertes ; la bibliothèque « pyuntl » peut convertir de manière normalisée des représentations sous forme de chaînes de caractères dans les qualificatifs des métadonnées UNTL en identifiants uniques pour chaque terme.

En plus des divers vocabulaires contrôlés, les Collections numériques utilisent l'application de gestion des noms de l'UNT (UNT Name App)^{14,15} pour effectuer un contrôle d'autorité sur les noms¹⁶ pour un certain nombre de collections¹⁷, y compris l'entrepôt des œuvres savantes de l'UNT¹⁸ et la collection des thèses et mémoires de l'UNT¹⁹. Actuellement, les membres du personnel ajoutent manuellement les nouveaux noms dans les fichiers d'autorité, aussi ce travail se concentre-t-il prioritairement sur les noms qui ne sont pas contrôlés ailleurs et qui apportent une amélioration des accès locaux (comme les noms des membres des facultés de l'UNT). Les fichiers d'autorité pour les noms dans l'UNT Name App ont des identifiants uniques et un champ pour enregistrer d'autres identifiants possibles, ce qui permet d'établir des liens avec d'autres notices d'autorité établies lorsqu'elles sont disponibles, par exemple : VIAF (Virtual International Authority File, Fichier d'autorité international virtuel), ORCID (Open Research and Contributor ID, Identifiant ouvert des chercheurs et des contributeurs), identifiants ResearchGate, ISNI (International Standard Name Identifier, Identifiant international normalisé des noms), profils des facultés de l'UNT et d'autres.

Ces divers services permettent la création d'outils de gestion des métadonnées et fournissent les mécanismes pour manipuler les métadonnées sur un plan technique.

RÉALISER DES CONNECTIONS

Il y a plusieurs façons dont l'infrastructure technique, les schémas de métadonnées et les interfaces utilisateur fonctionnent ensemble. Ces intersections améliorent les expériences utilisateur, à la fois pour les éditeurs de métadonnées et pour les utilisateurs finaux du grand public.

L'interface d'édition des métadonnées

Pour rendre la création et la mise à jour des métadonnées plus facile et plus cohérente pour des créateurs ayant différents niveaux d'expérience, l'interface d'édition des métadonnées utilise les spécifications techniques pour fournir des informations contextuelles. Des liens depuis chaque champ permettent à un intervenant de regarder toutes les recommandations écrites et les exemples, et la mise en page du formulaire de métadonnées fournit aussi des signaux visuels. Par exemple, une liste des champs sur le côté droit de l'écran permet la navigation parmi les champs et utilise également des codes de couleur pour montrer les champs obligatoires qui ne sont pas remplis (voir Annexe A).

Au cours de ces dernières années, l'interface utilisateur pour le système d'édition des métadonnées a connu des changements pour y incorporer des fonctions supplémentaires qui encouragent des pratiques conformes à nos normes. Par exemple, à l'UNT, nous avons adopté le format étendu de la date et de l'heure (Extended Date/Time Format, EDTF)²⁰ pour mieux représenter des dates complexes de manière lisible par une machine²¹. L'EDTF donne la possibilité de représenter dans un formatage normalisé des concepts qui sont courants pour les ressources patrimoniales, comme création « pendant l'été 2001 » (2001-22) ou « Mars 1903 » (1903-03). Afin de fournir un retour aux créateurs de métadonnées qui saisissent des chaînes de caractères en EDTF, les champs de date sont validés par rapport à la norme EDTF à travers un web service JSON²² qui utilise des applications appelées « django-edtf »²³ et « edtf-validate »²⁴. Si les éditeurs de métadonnées entrent des dates non conformes à EDTF, le formulaire de métadonnées affiche un avertissement et surligne le champ. En outre, des notes apparaissent pour certains champs quand des éléments d'information manquent (par exemple, quand les éditeurs de métadonnées entrent des valeurs sans qualificatifs).

Les vocabulaires contrôlés gérés à l'intérieur du système sont aussi connectés directement au formulaire d'entrée des métadonnées. Tous les qualificatifs et les listes contrôlées s'affichent comme des menus déroulants, de sorte que les éditeurs de métadonnées soient obligés de choisir un terme dans la liste, plutôt que d'entrer les termes à la main. De même, un outil UNTL-BS apparaît dans une fenêtre modale surgissante (*pop-up*) quand le choix du qualificatif s'effectue dans le champ Sujet. Bien qu'auparavant la fenêtre modale ait autorisé les utilisateurs à entrer les termes à la main, cela a été modifié pour que seuls des termes contrôlés valides puissent être ajoutés à la notice à partir de la liste actuelle.

La pratique de développer des outils locaux pour gérer les diverses activités dans notre bibliothèque numérique apporte de la souplesse pour intégrer des modules dans le système d'édition des métadonnées selon les besoins. L'application UNT Name App est reliée aux champs de nom désignés (créateur, contributeur et éditeur) pour fournir aux éditeurs de métadonnées une fonctionnalité de frappe anticipée, en sorte qu'il y ait intégration immédiate dans le système d'édition des métadonnées des noms ajoutés à l'application Name App. Un éditeur de métadonnées peut soit choisir un nom depuis la liste autorisée, soit entrer un autre nom puisque l'application UNT Name App n'est utilisée que pour des collections choisies.

Tous ces outils rendent la création de métadonnées plus facile pour les éditeurs et réduisent les erreurs autant que possible.

Apporter de la souplesse aux métadonnées

Au niveau le plus élémentaire, la façon dont les champs sont codés dans le système garantit que nous n'ajoutons pas de nouveaux champs à moins que ce ne soit absolument nécessaire, et encourage une utilisation souple des champs comme une alternative. Dans certains cas, l'infrastructure technique a eu des répercussions directes sur la manière dont nous adaptons les métadonnées pour répondre à des besoins divers. Par exemple, comme les titres de collections et de périodiques sont indexés pour fournir en option des facettes qui permettent de restreindre les résultats d'une recherche, nous utilisons souvent les titres de collections comme un moyen de réunir des groupes de ressources à l'intérieur d'une collection plus vaste. Cela fournit une alternative raisonnable à la création de nouvelles collections pour tout regroupement de ressources qui a besoin d'être identifié pour répondre aux besoins des partenaires ou des utilisateurs.

Une autre option est d'utiliser l'indexation plein-texte des métadonnées qui fait déjà partie du système, en exploitant des expressions formatées présentes dans les descriptions du contenu de toutes les notices appartenant à une collection ou à un groupe de ressources. Certains partenaires ont des besoins spécifiques ou des envies sur la manière dont ils seront capables d'utiliser ou de rechercher leurs propres collections dans le Portail sur l'histoire du Texas ou la Bibliothèque numérique, et nous pouvons répondre à leurs exigences avec cette technique. Une institution partenaire avait une collection de négatifs sur verre représentant des portraits, principalement de personnes seules ou de petits groupes, aussi souhaitaient-ils pouvoir rechercher par le nombre de personnes visibles sur chaque image. Pour faciliter cela, nous avons aménagé nos recommandations habituelles pour commencer chaque description par la mention « Photographie de [#] personne(s) ». Comme la formulation est identique dans chaque notice, quelqu'un qui sait que cette mention est indiquée dans les notices peut facilement rechercher un nombre particulier de personnes. Pour un projet subventionné par l'UNT, nous avons utilisé une approche similaire en travaillant avec des spécialistes pour développer une description du contenu qui liste toutes les caractéristiques physiques de spécimens biologiques de moules sous la forme de formules de manière à ce que les chercheurs puissent facilement trouver des sous-ensembles de types particuliers de moules.

Dans certains cas, le besoin d'exprimer l'information plus clairement (en particulier à travers diverses sortes de collections) a conduit à des changements dans l'implémentation technique. Il y a plusieurs années, nous avons ajouté un qualificatif au champ Source qui auparavant était un simple champ de texte. Un éditeur de métadonnées peut désormais préciser la nature de la « source » à partir d'une liste classée depuis les objets plus vastes dont fait partie la ressource (par exemple, un livre, une revue, un journal, etc.) jusqu'à des événements où la ressource a pris naissance ou a été présentée (par exemple, un congrès, un cours universitaire, une série de lectures, etc.).

Les métadonnées comme données

Un aspect important du système d'édition des métadonnées des Collections numériques est la gestion des versions des notices de métadonnées, introduite dans l'infrastructure durant l'automne 2009. Toutes les versions de chaque notice ont été sauvegardées depuis son implémentation, en sorte qu'il y a toujours un moyen de revenir en arrière pour consulter chaque version (voir Annexe B), ou pour retourner à une version antérieure si nécessaire. Une autre composante, déployée en octobre 2013, documente chaque événement d'édition associé au compte de l'éditeur de métadonnées (prenant note de la durée de l'édition, de la notice concernée, ainsi que de la complétude et de la visibilité de la ressource à la fin de l'édition) et rassemble toutes les statistiques d'édition (voir Annexe C).

Sauvegarder les versions et les données d'évènements offre une occasion d'analyser divers aspects de la création des métadonnées et de l'activité des éditeurs de métadonnées. Nous avons commencé les études initiales pour déterminer les informations que nous pouvons apprendre en regardant les versions des métadonnées²⁵, en comparant les modifications apportées aux notices²⁶ et en évaluant les temps de travail des éditeurs²⁷. Dans le futur, nous espérons utiliser ces données pour en apprendre plus sur les mesures déterminant la qualité des métadonnées et sur une analyse coûts-bénéfices appliquée à la création des métadonnées. Si d'autres institutions ont des données similaires, cela permettra aussi à des études comparatives d'en apprendre plus sur la création des métadonnées entre différents systèmes.

DISCUSSION

Une infrastructure conceptuelle et technique solide pour la création des métadonnées, soutenue par des politiques qui encouragent les développements ouverts et le partage des données, sont une fondation essentielle des Collections numériques des Bibliothèques de l'UNT. Cela constitue un socle solide à partir duquel développer les services existants et construire de nouveaux services pour les utilisateurs, internes et externes, des métadonnées et des contenus de la bibliothèque numérique.

De fait, beaucoup des fonctions implémentées dans l'interface utilisateur au cours de la refonte de 2009 ont été directement connectées aux valeurs des métadonnées en exploitant les données existantes. La plupart des facettes disponibles pour restreindre les résultats d'une recherche sont tirées de champs avec des vocabulaires contrôlés ou qui ont un formatage normalisé (par exemple, les dates et les noms de lieux). Les catégories pour la navigation par sujets sont fondées sur des termes du répertoire UNTL-BS, aussi seules les notices utilisant ces sujets sont retrouvables à travers cette option de navigation. Ajouter ces fonctions a aussi créé une boucle de rétroaction pour les éditeurs de métadonnées, dans la mesure où cela mettait en évidence beaucoup de zones où les métadonnées étaient moins cohérentes que ce qui était attendu, ou bien là où des erreurs avaient été faites dans des notices isolées.

De ces expériences à l'UNT, on peut tirer quelques recommandations générales qui peuvent être utiles à d'autres institutions. Une considération concerne le nombre de champs à inclure dans un schéma de métadonnées, et combien celui-ci est ouvert à l'ajout de nouveaux champs ou à des usages différents à travers une collection. Maintenir de la cohérence à l'intérieur du schéma de métadonnées et dans les utilisations à travers nos collections a apporté plusieurs bénéfices, non seulement de la fonctionnalité dans les interfaces des Collections numériques, mais aussi des connections entre collections, des facilités de partage à travers des initiatives comme la Bibliothèque numérique publique d'Amérique (Digital Public Library of America, DPLA), et un jeu de données plus uniforme pour les analyses.

Quels que soient les champs, la cohérence peut être la pratique générale la plus utile dans la mesure où elle ouvre la possibilité de services fondés sur les métadonnées, mais aussi facilite la migration ou de futurs changements, si nécessaire. Des institutions peuvent vouloir rechercher des champs gérés pour tout ou partie par des vocabulaires contrôlés ; cela renforce la cohérence et rend aussi la création des métadonnées plus facile pour les éditeurs de métadonnées. Partout où c'est possible, les vocabulaires contrôlés et tout autre contrôle d'autorité devraient s'aligner sur les données ouvertes et liées « cinq étoiles »²⁸ et se conformer au niveau le plus élevé possible. Dans certains cas, il peut y avoir des mesures relativement faciles pour faire progresser des vocabulaires contrôlés établis dans le continuum des données liées s'ils ne répondent pas aux exigences des données « cinq étoiles ».

Bien que le contrôle d'autorité et le développement de schémas puissent être complexes et sembler écrasants, il existe beaucoup de vocabulaires établis et de normes publiquement disponibles. Les institutions peuvent facilement réutiliser n'importe quelle de ces normes qui s'appliquent ou les utiliser comme point de départ pour développer localement un vocabulaire ou un référentiel. Dans les Collections numériques, plusieurs de nos vocabulaires sont fondés sur des listes établies en provenance d'autres organisations auxquelles nous avons apporté des modifications locales. Par exemple, le référentiel des « fonctions »²⁹ pour les créateurs et les contributeurs utilise les valeurs applicables des codes de fonction MARC³⁰ et inclut des

termes supplémentaires (comme « Artisan » ou « Détaillant ») qui figurent pas dans la liste originale ; le référentiel des types de relations³¹ est fondé sur la liste des raffinements des relations du Dublin Core³², encore une fois avec des ajouts locaux (par exemple, « A pour traduction » et « Est une transcription de »).

Enfin, la documentation écrite s'est avérée inestimable pour gérer les métadonnées dans nos Collections numériques, pour les aspects techniques comme pour la saisie des valeurs. Garantir que tous les éditeurs de métadonnées et le personnel technique aient accès à la documentation complète sur les métadonnées, même si elle n'est pas publiquement disponible, promeut la cohérence et ouvre un dialogue qui tient compte des outils et des discussions sur la façon d'optimiser l'efficacité des métadonnées. Dans notre système, tous les aspects de la création et de la gestion des métadonnées ont été développés en interdépendance, avec parfois des côtés positifs inattendus.

Nous espérons que partager des études de cas, telles que nos expériences à l'UNT, puisse donner des idées à d'autres institutions qui développent et gèrent des métadonnées ou qui envisagent des changements de système.

RÉFÉRENCES

- ¹ Hartman, C.N. & K.R. Murray. (February 2010). Optimizing the user experience in a rapid development framework: final project report, February 2010. <http://digital.library.unt.edu/ark:/67531/metadc32974/>
- ² <http://www.library.unt.edu/policies/other/open-source-software-policy>
- ³ <https://github.com/unt-libraries/>
- ⁴ <http://www.library.unt.edu/policies/other/locally-created-cataloging-and-metadata-records-rights-policy>
- ⁵ Tarver, H. (2010). Better guidelines, better functionality: how metadata supports the cycle of system improvement at the University of North Texas. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, pp. 165-176. <http://digital.library.unt.edu/ark:/67531/metadc29323/>
- ⁶ <http://www.library.unt.edu/digital-projects-unit/minimally-viable-records>
- ⁷ <http://www.library.unt.edu/digital-projects-unit/quick-start-metadata-guide>
- ⁸ <http://www.library.unt.edu/digital-projects-unit/input-guidelines-descriptive-metadata>
- ⁹ <http://www.library.unt.edu/digital-projects-unit/project-specific-guidelines-and-documents>
- ¹⁰ <http://digital2.library.unt.edu/untl.xsd>
- ¹¹ <https://github.com/unt-libraries/pyuntl>
- ¹² <http://digital2.library.unt.edu/vocabularies/>
- ¹³ <https://github.com/unt-libraries/django-controlled-vocabularies>
- ¹⁴ <http://digital2.library.unt.edu/name/>
- ¹⁵ <https://github.com/unt-libraries/django-name>
- ¹⁶ Tarver, H., L. Waugh, M.E. Phillips, & W. Hicks. (2013). Implementing name authority control into institutional repositories: a staged approach. <http://digital.library.unt.edu/ark:/67531/metadc172365/>
- ¹⁷ Waugh, L., H. Tarver, & M.E. Phillips. (2014). Introducing Name Authority into an ETD Collection. *Library Management*, 35(4/5), pp. 271-283. <http://digital.library.unt.edu/ark:/67531/metadc279706/>
- ¹⁸ <http://digital.library.unt.edu/ScholarlyWorks/>
- ¹⁹ <http://digital.library.unt.edu/explore/collections/UNTETD/>
- ²⁰ <http://www.loc.gov/standards/datetime/>
- ²¹ Tarver, H. & M.E. Phillips. (July 2013). Lessons learned in implementing the Extended Date/Time Format in a large digital library. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, pp. 60-70. <http://digital.library.unt.edu/ark:/67531/metadc174739/>
- ²² <http://digital2.library.unt.edu/edtf/>
- ²³ <https://github.com/unt-libraries/django-edtf>
- ²⁴ <https://github.com/unt-libraries/edtf-validate>
- ²⁵ Zavalina, O.L., & P. Kizhakkethil. (2015). Exploration of metadata change in a digital repository. *Proceedings of iConference 2015*, Newport Beach, California. <http://digital.library.unt.edu/ark:/67531/metadc503265/>

-
- ²⁶ Zavalina, O.L., P. Kizhakkethil, D. Alemneh, M. Phillips, & H. Tarver. (2015). Building a framework of metadata change to support knowledge management. *Journal of Information and Knowledge Management*, 14 (1), 1-16. <http://digital.library.unt.edu/ark:/67531/metadc505014/>
- ²⁷ Tarver, H., O.L. Zavalina, M. Phillips, D. Alemneh, & S. Shakeri. (2014). How descriptive metadata changes in the UNT Libraries' Collections: a case study. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, pp. 43-52. <http://digital.library.unt.edu/ark:/67531/metadc406345/>
- ²⁸ Berners-Lee, Tim. (July 7, 2006) "Linked Data," W3C. <http://www.w3.org/DesignIssues/LinkedData.html>
- ²⁹ <http://digital2.library.unt.edu/vocabularies/agent-qualifiers/>
- ³⁰ <http://www.loc.gov/marc/relators/relators.html>
- ³¹ <http://digital2.library.unt.edu/vocabularies/relation-qualifiers/>
- ³² <http://dublincore.org/documents/dcmi-terms/>

ANNEXE A : Le système d'édition de l'UNT

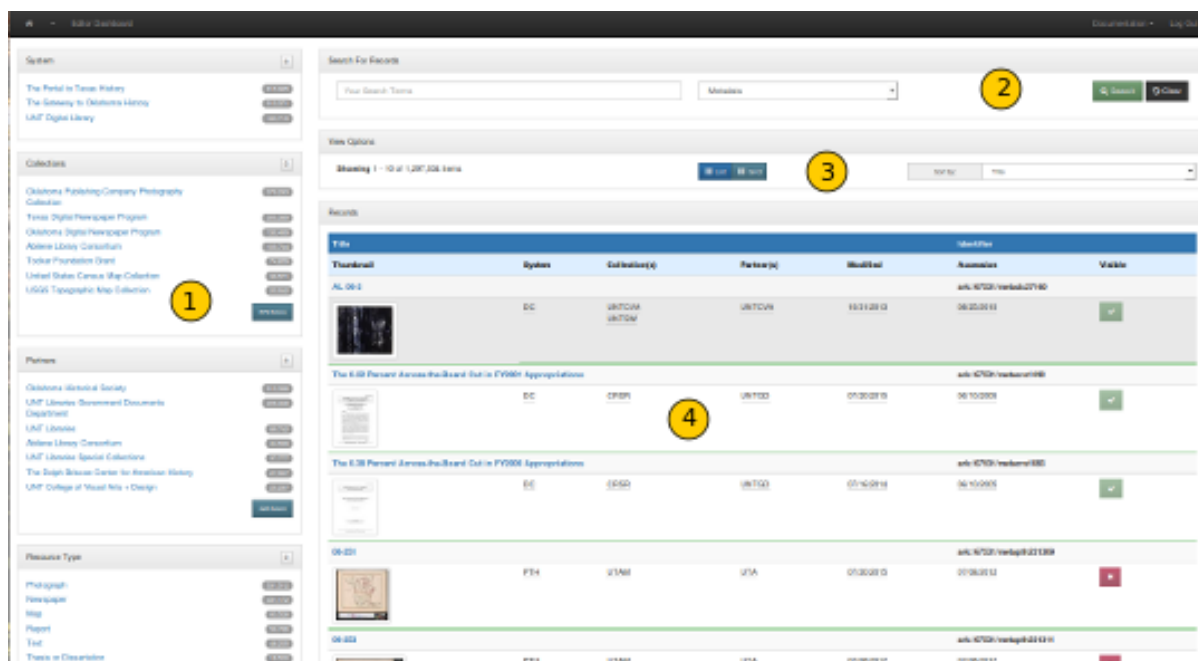


Figure 1. Capture d'écran du tableau de bord de l'utilisateur dans le système d'édition

Voici une vue du tableau de bord qu'un utilisateur voit après s'être connecté au système d'édition pour les Collections numériques. Chaque éditeur de métadonnées voit la liste des notices auxquelles il a accès. Les notices individuelles peuvent être ouvertes en cliquant sur le titre ou sur la vignette (voir Figure 2).

Caractéristiques du tableau de bord :

1. Des facettes permettent aux éditeurs de métadonnées de restreindre la liste de notices selon les critères applicables, comme l'interface système, la collection, le partenaire, le type de ressource et la visibilité par le public.
2. Une barre de recherche pour trouver des valeurs n'importe où dans les métadonnées ou dans un champ spécifique.
3. Des options d'affichage pour voir les ressources sous la forme de liste (ce qui est montré ici) ou de grille, et un menu déroulant pour trier par titres, par dates de création ou de modification des notices, par date de création des ressources ou par identifiants uniques ARK.
4. Liste des notices décrivant des ressources, avec le titre, la vignette, le système, le partenaire, la collection, la date de création et de modification, l'identifiant ARK et le statut de visibilité de la ressource.

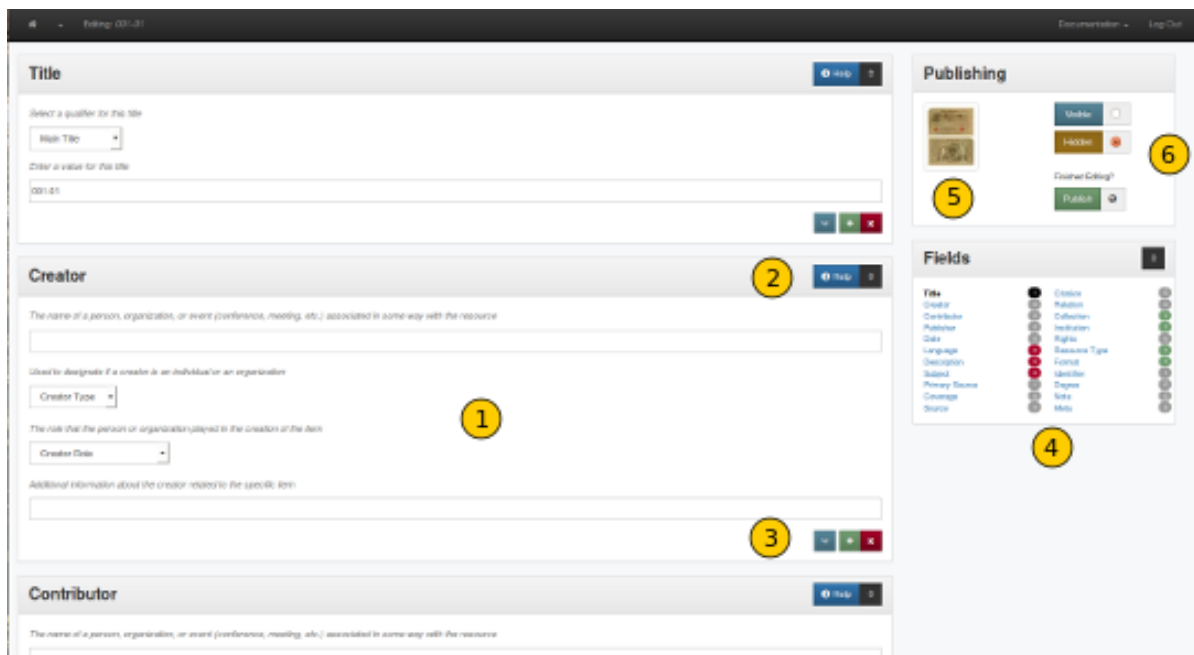


Figure 2. Capture d'écran d'une notice décrivant une ressource dans le système d'édition

L'image en Figure 2 montre le formulaire de métadonnées pour une notice décrivant une ressource ; seules certaines valeurs sont renseignées dans cette notice.

Champs de la notice

1. Chaque champ est limité dans un cadre et comporte une (ou des) boîte(s) appropriée(s) de texte et/ou des menu(s) déroulant(s) pour les valeurs de métadonnées.
2. Un lien « Aide » dans la barre de titre pour chaque zone permet de se référer aux recommandations (qui s'ouvrent dans une fenêtre modale surgissante (*pop-up*)) ; une icône pour replier le champ figure aussi dans le coin supérieur droit de chaque champ.
3. Des boutons dans le coin inférieur droit de chaque champ permettent aux utilisateurs d'insérer des symboles et caractères spéciaux, d'ajouter et de supprimer des entrées.

Navigation

4. Une liste de champs sur le côté droit de l'écran affiche le nombre d'entrées pour chaque champ et offre une liste cliquable pour permettre à un éditeur de métadonnées d'accéder à un champ spécifique. Les nombres d'entrées utilisent des codes de couleur pour les champs obligatoires (rouge = pas de valeur renseignée, vert = valeur renseignée) et pour signaler des dates non valides ou des sujets insuffisants (jaune).
5. La vignette est cliquable et ouvre dans un autre onglet ou une autre fenêtre une version de toutes les pages ou images associées à la ressource, avec la possibilité de zoomer.
6. Des boutons dans la section supérieure droite contrôlent la visibilité par le public de la ressource (visible ou cachée) et permet à l'éditeur de métadonnées de « Publier » (sauvegarder) la version de la notice

ANNEXE B : Historique d'une notice décrivant une ressource

History: ark:/67531/metaph638535 Documentation - Log Out

ark:/67531/metaph638535

Created by: mphilips on 2015-10-09, 16:44:39

UNTL History:

Metadata modification date	Metadata modifier	Metadata file name	Change	Content length	Status	Completeness
2016-05-03, 13:45:19	jlepper	metaph638535.untl.xml	+1094	3,120 bytes	Visible	1.00
2016-03-04, 13:00:37	hharvor	metaph638535.untl.xml.2	+1428	2,026 bytes	Hidden	1.00
2016-03-02, 14:47:32	hharvor	metaph638535.untl.xml.1	+0	568 bytes	Hidden	0.98
None	None	metaph638535.untl.xml.0		560 bytes	Hidden	0.98

METS History:

Metadata creation date	Metadata file name	Record length	Metadata modification date	METS metadata creator
2015-10-09T16:44:39Z	metaph638535.mets.xml	702,253 bytes	2015-10-09T16:44:39Z	UNT Libraries: Digital Projects Unit

Figure 3 : Historique d'une notice décrivant une ressource, comprenant les dates de modification, les noms des éditeurs de métadonnées, la version de la notice, l'importance des changements (par nombre de caractères), la taille de la notice, le statut de visibilité de la ressource et le taux de complétude.

ANNEXE C : Système de l'UNT pour le suivi des événements d'édition



Figure 4. Capture d'écran du tableau de bord du système de suivi des événements d'édition

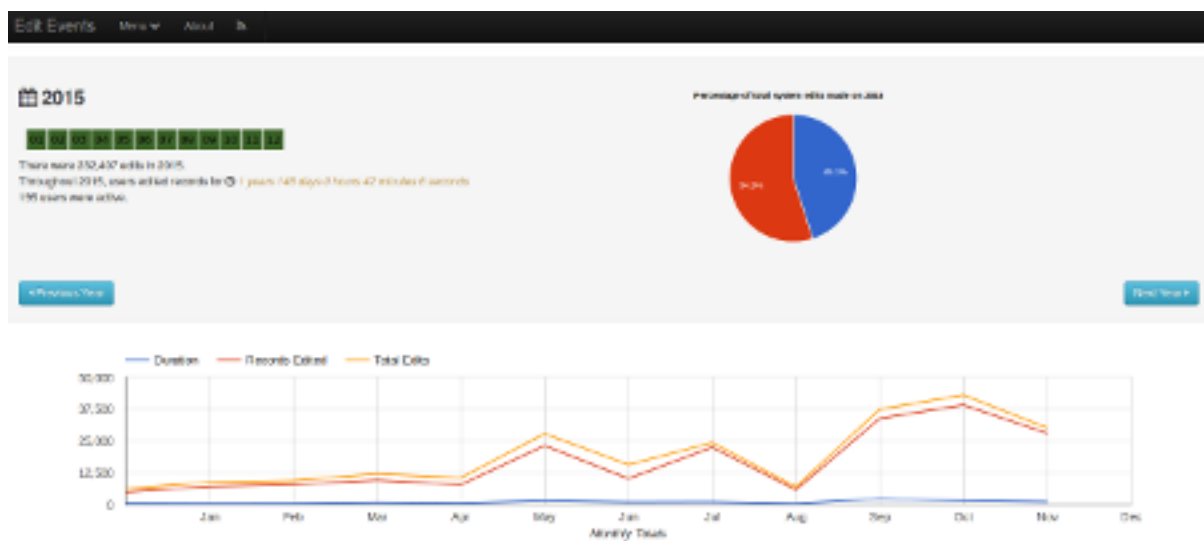


Figure 5. Capture d'écran de la page de statistiques pour l'année civile 2015