

Numériser « Nos racines » : des principes et des techniques

Éléments de réflexion sur la création de collections virtuelles

Guy Teasdale, MBSI
Spécialiste en ressources documentaires électroniques
Bibliothèque de l'Université Laval
14 mai 2002, 10 h 50

(verbatim accompagnant le fichier Power Point)

Bonjour à tous,

C'est avec beaucoup de plaisir que je viens vous donner mon point de vue sur la question des publications numériques dans nos bibliothèques de recherche universitaires. Comme plusieurs d'entre vous le savent, la problématique des collections numériques et de leur gestion me passionne depuis plusieurs années déjà.

J'ai choisi de vous parler d'un des trois projets de collections numériques que je coordonne à l'Université Laval : le projet *Nos Racines*. *Nos racines* est une réalisation conjointe des Presses de l'Université de Calgary et de la Bibliothèque de l'Université Laval. Étant donné que la première phase de ce projet a été complétée le 19 avril dernier et que le lancement national aura lieu le 1^{er} juillet prochain, j'ai pensé que le présent colloque, situé entre ces deux dates était l'occasion idéale pour vous faire profiter d'une avant-première en levant le voile sur la genèse de *Nos racines*.

Bien que l'URL n'est pas encore publique, que notre contrôle de qualité n'est pas terminé et que le design n'est pas final, *Nos racines* me servira de modèle pour illustrer les pratiques courantes dans le développement de ce type de collections numériques.

[diapo 2]

Le temps qui nous est alloué pour cette présentation est très court mais je m'en voudrais de ne pas souligner le travail exceptionnel effectué par l'équipe de *Nos racines* dans des délais extrêmement difficiles; nous avons réalisé en 4 mois un plan de travail de 12 mois. Loubna Ghaouti, une diplômée récente de l'EBSI, était responsable de l'informatique documentaire et de la banque de données de gestion. France Parent est une historienne qui s'est chargée de la sélection des ouvrages à numériser en collaboration avec Monique Mailloux, conseillère à la documentation et responsable du développement de la collection en histoire à l'Université Laval. Enfin, Marie-Josée Verreault est

une sociologue ayant une bonne pratique des maisons d'éditions. Elle était chargée de la libération des droits d'auteurs pour les ouvrages qui n'étaient pas dans le domaine public. Nous reviendrons plus tard sur les différents comités.

[diapo 3]

Nos racines c'est maintenant, après quatre mois de travail intense à Québec et à Calgary, 300 000 pages d'ouvrages sur les histoires locales et régionales du Canada. Par « histoires locales », on entend des ouvrages qui décrivent des localités et les gens qui y vivent, à une période donnée. Il peut s'agir, par exemple, de l'histoire d'une paroisse publiée à l'occasion de son centenaire, de l'histoire de personnages, de familles, de lieux, d'édifices ou d'organismes importants pour une région. On peut penser à Alphonse Desjardins à Lévis, aux Ursulines à Québec, ou encore à l'Université Laval qui célèbre son cent-cinquantenaire cette année.

Mon objectif initial était de créer un corpus d'ouvrages sur l'histoire de Québec et de sa région en vue des Fêtes de Québec 2008. J'ai présenté deux demandes au Fonds de l'autoroute de l'information en 1999 et 2000 qui n'ont malheureusement eu qu'un succès ...d'estime.

Un appel de proposition de Patrimoine Canada en mai 2001 nous a permis de recycler notre projet en le greffant à un projet similaire alors en cours à l'Université de Calgary. En effet, les presses de l'Université de Calgary avaient déjà entrepris de numériser les histoires locales de l'Ouest canadien et possédaient une chaîne de production et une base de données fonctionnelle. Les critères d'attribution du ministère du Patrimoine canadien imposant un contenu français à 50%, nous devenions des partenaires incontournables.

Ce projet pancanadien est prévu sur 3 ans. Nous ajouterons des partenaires à chaque année avec l'objectif de numériser toutes les histoires locales canadiennes. Nous attendons d'un jour à l'autre le renouvellement de la subvention pour la 2^e année. Si c'est positif, nous accueillerons alors des partenaires de l'Université de Toronto, de Simon Fraser et de Victoria. De notre côté, nous nous sommes assurés d'une collaboration de la Bibliothèque nationale du Québec, de la Bibliothèque de l'assemblée nationale du Québec, de l'Institut canadien de Québec, du Centre d'études acadiennes de Moncton et de différents organismes et ministères canadiens et québécois.

Aux 300 000 pages déjà en ligne, nous prévoyons, avec l'aide de ces nouveaux partenaires, ajouter 750 000 pages cette années et 750 000 autres l'année prochaine. Au total 12 000 ouvrages représentant près de 2 millions de pages auront été numérisés à la fin de la 3^e année.

Ici nous pourrions faire une visite guidée et je vous montrerais nos plus belles pages numérisées. Mais nous réserverons cela pour la fin si le temps nous le permet. De toute façon, si le seul fait de mener à bien un projet de collection numérique pouvait être en soi un événement il y a quelques années, c'est une activité assez courante de nos jours et des projets de numérisation, il en pleut! Autant vous donner l' URL et vous inviter à y aller voir par vous mêmes!

[diapo 4]

Mentionnons simplement que l'interface permet la recherche dans les métadonnées auteur et titre pour l'instant mais que bientôt nous ajouterons également la date, les vedettes matières etc. La recherche en texte intégral dans chacune des pages du texte est aussi possible car chacune des 300 000 pages images numérisées a fait l'objet d'une opération de reconnaissance optique des caractères.

Principes à suivre pour construire ces collections

[diapo 5]

C'est un document récent qui a fait l'objet d'une communication en mars 2002 lors de la 3^e conférence sur les bibliothèques et les musées dans le monde numérique et dont les actes ont été publiés dans l'excellente revue *First Monday* le 6 mai dernier qui me permettra de faire d'une pierre deux coups, soit vous présenter indirectement « Nos racines » mais surtout, le faire sous l'angle des « problématiques et enjeux » annoncés dans le titre du présent colloque.

[diapo 6]

Le document en question s'intitule *Framework of Guidance for Building Good Digital Collections*. Ou, en traduction libre, *Canevas rassemblant des conseils et des principes permettant de construire de bonnes collections numériques*¹. Il témoigne de la volonté grandissante d'organisation et de normalisation les activités de production numérique dans nos bibliothèques. Dans les quelques minutes qui suivent j'essaierai donc de vous présenter les principes exposés dans ce document tout en vous indiquant de quelle façon, si c'est le cas, le projet *Nos racines* les a appliqués.

Quelques mots d'abord au sujet du *Framework*. Ce rapport a été développé dans le cadre des *Digital Library Forum* de la Digital Library Federation au cours de l'année 2001 principalement par l'Institute of Museum and Library

¹ Digital Library Forum. *A Framework of Guidance for Building Good Digital Collections* Washington, DC, Institute of Museum and Library Services, 2002 (<http://www.imls.gov/pubs/forumframework.htm>).

Services (IMLS). Le rapport a été publié sur le site de l'IMLS, pour commentaires, le 6 novembre 2001 et a été récemment entériné par la Digital Library Federation. Le rapport ne préconise pas une série de normes, de directives ou de meilleures pratiques spécifiques mais énonce plutôt **des principes de base**, généraux et englobants, pour la planification, la mise en œuvre et l'évaluation de projets de numérisation. On y trouve, par contre, une très riche série d'hyperliens vers des projets exemplaires de numérisation illustrant les principes de base. Par la diffusion de ces principes, les auteurs du rapport veulent favoriser le développement harmonieux, pérenne et interopérable des collections numériques en tant qu'éléments constitutifs de nos bibliothèques virtuelles au même titre que les journaux, les livres et les périodiques papiers ont été les constituants essentiels de nos bibliothèques traditionnelles.

Ce rapport vise deux auditoires : D'abord les personnes engagées dans la planification et le développement de collections numériques et ensuite les organismes subventionnaires. En effet, de plus en plus ces organismes veulent rentabiliser leurs investissements en s'assurant que l'on utilise les meilleures pratiques et que les coûteux objets numériques dont on finance la création ne disparaîtront pas avec la fin de la subvention et seront conservés afin d'être utilisés et réutilisés.

C'est également une préoccupation au Canada et le ministère du Patrimoine a publié un guide de meilleures pratiques intitulé « Normes et lignes directrices pour les projets de numérisation ». *Nos racines* se conforme à ces normes. Mentionnons au passage que ces normes canadiennes sont directement inspirées des travaux britanniques du Joint Information Systems Committee (JISC) (comité adviseur stratégique en TI) qui met en place un Distributed National Electronic Resource (DNER) et de l'UK Office for Library and Information Networking (UKOLN). Compte tenu des ressources énormes qu'ils y consacrent, ce n'est pas surprenant. Le framework (américain) fait d'ailleurs amplement référence aux travaux des britanniques.

[diapo 7]

Après cette longue entrée en matière, voyons maintenant le contenu du Framework et comment on l'a appliqué dans *Nos racines*.

Tout comme les bibliothèques traditionnelles sont plus qu'un assemblage épars de livres, les collections numériques sont plus qu'un assemblage d'objets numérisés; Ça implique sélection et organisation de même qu'un catalogage sous la forme de métadonnées descriptives, structurelles et administratives

Les métadonnées doivent être au niveau de la collection et au niveau de l'objet numérique. *Nos racines* possède une bonne description catalographique de chaque ouvrage numérisé. Nous dérivons, à partir de la notice MARC les métadonnées nécessaires en les convertissant en format Dublin Core pour chacun de nos objets numériques. Au niveau de la collection, on voit se développer des normes de description qui permettront une utilisation optimale et internationale des collections, notamment avec les travaux de l'Open Archives Initiative.

[diapo 8]

Principes d'une bonne collection numérique

- Politique explicite de développement des collections, approuvée et documentée avant le début de la numérisation. Nos racines a formé un comité éditorial national et local, composé d'historiens spécialisés en histoires locales, de bibliothécaires et d'archivistes pour raffiner et élaborer la politique énoncée dans la demande de subvention. Les partenaires s'engagent à respecter cette politique, notamment dans son concept de totalité.
- Les collections doivent être décrites afin que l'utilisateur puisse découvrir les caractéristiques importantes de cette collection telles que l'envergure (scope), le format, les restrictions d'accès, la propriété intellectuelle et toute autre information importante pour déterminer l'authenticité, l'intégrité et l'interprétation. Nos racines ne comporte aucune restriction d'accès, les formats sont normalisés grâce à des normes techniques déterminées par un comité technique.
- Les collections doivent être durables et pérennes. Nous avons utilisé des formats standards tels que le format image Tiff à haute résolution qui nous sert de format d'archivage. Pour ce qui est de la pérennité, nous croyons que les bibliothèques ont fait leur preuves avec les livres papiers et ne demandent qu'à les faire avec les livres électroniques comme le démontrent les nombreux colloques et les innombrables recherches sur la conservation des fichiers électroniques.
- Accessibles sans entraves de partout et sans restrictions d'usage. Idéalement accessible aux personnes handicapées et ouvertes aux technologies adaptées. L'accessibilité de *Nos racines* est très large. Pour l'instant, compte tenu du fait que l'on diffuse des images, notre projet est inaccessible aux technologies adaptées telles que les plages tactiles utilisées par les aveugles. Les fichiers de reconnaissance optique des caractères sont non corrigés, ce qui nous empêche aussi de les rendre accessibles aux handicapés.
- Respect de la propriété intellectuelle, maintien d'un registre des droits et autorisations. Cet aspect est très lourd à gérer en terme de temps et

- d'énergies mais nous avons eu l'agréable surprise d'un accueil enthousiaste des détenteurs des droits et des éditeurs
- Compiler la mesure de l'usage sous forme agrégée et maintenue dans le temps afin de voir l'évolution. Le projet n'étant pas encore officiellement lancé, nous n'avons pas encore choisi notre logiciel d'analyse de fichiers de logs
 - Se situer dans le contexte national et international d'initiatives similaires. Nos racines est un projet historique tout comme le projet « Notre mémoire en ligne » auquel nous avons participé. Nous avons négocié avec l'ICMH, organisme responsable de « Notre mémoire en ligne » la possibilité de numériser à leur place les microfiches correspondant à nos critères d'inclusion. Même si les technologies utilisées diffèrent, nous prenons toutes les précautions pour éviter de renumériser ce qui a été fait par eux et nous tenterons de relier les deux banques au moyen des métadonnées

[diapo 9]

Principes pour de bons objets numériques

- Sont produits afin de supporter les priorités de la collection. On s'est doté d'une politique de sélection à cet effet. De plus, la priorité c'est la recherche dans le texte intégral et non l'impression, ce qui rassure les éditeurs pour qui notre projet devient moins « menaçant ».
- Est persistant. L'institution prendra soin de le maintenir accessible à travers le temps même si les technologies changent. Le Format TIFF, et la base de données SQL représentent toutefois des normes assez stables
- Faciliter l'utilisation courante et future. Pour le courant on a dérivé 3 fichiers en format JPEG afin de diffuser sur le Web. Si un autre format s'imposait, il sera toujours possible de le dériver grâce à l'original à haute résolution en format TIFF
- Nomenclature solide et évolutive. Nous prévoyons nommer facilement les 12 millions de fichiers produits dans le cadre de ce projet grâce à notre politique de nomenclature des fichiers..
- Authentification. Les métadonnées élaborées permettent de déterminer exactement la référence bibliographique de l'original papier du document. D'autre part, les bibliothèques ne mettraient pas leur réputation en jeu en falsifiant des documents. Ça irait à l'encontre du principe de totalité
- Métadonnées descriptives et structurelles. Les métadonnées structurelles permettent la navigation dans la table des matières et le bouquinage page par page.

[diapo 10]

Principes pour de bonnes métadonnées

- Qui sont appropriées. En plus des métadonnées structurelles, nous utilisons le Dublin Core qui est normalisé au niveau international (norme Z39.85)
- Qui sont interopérables. Le Dublin Core sert de base à l'Open Archives Initiative. Il pourra donc nous permettre de créer des notices conformes à l'OAI et de faciliter le repérage de nos documents
- Qui ont un vocabulaire contrôlé. *Nos racines* intègre les Vedettes matières de l'Université Laval et l'équivalent LCSH, ce qui permettra un bouquinage bilingue dans nos collections. Comme nos notices sont dérivées de catalogues de bibliothèques, nous profitons d'un contrôle d'autorités de qualité
- Qui énoncent clairement les termes et conditions d'usage de l'objet numérique. Tous les objets numériques sont soit dans le domaine public, soit diffusables avec l'accord des ayants droits.
- Qui sont elles-mêmes des objets numériques. Nos métadonnées sont conservées dans des fichiers Excel, c'est une des faiblesses actuelles de la chaîne. Toutefois, il serait possible de programmer des sorties formatées en XML afin de les intégrer dans éventuelle base de données qui serait entièrement en XML. (xindice par ex.)
- Qui permettent la gestion à long terme des objets numériques. Chaque fichier est accessible par la base de données grâce à la nomenclature

Le Framework termine en mentionnant des aspects souvent négligés, parce que les projets sont souvent des efforts ponctuels limités dans le temps. mais qui sont tout aussi importants que les principes que nous venons d'énoncer sommairement.

- Le design. Nous venons d'ailleurs de lancer un concours de design. Cet aspect est essentiel. Il ne faut jamais oublier que nos usagers sont à un clic de souris d'un ailleurs meilleur dans Internet. Nous visons différents publics et nous prévoyons implanter plusieurs couches d'interfaces entre les documents et les usagers. Je ne crois pas qu'un adolescent de quinze ans naviguerait très longtemps dans une microfiche numérisée en mode bitonal. Le chercheur en histoire de son côté y verra un saut quantique qualitatif par rapport à la consultation sur microfiches et pourra nous pardonner un design austère
- L'évaluation. Nous formerons des groupes focus auprès justement de diverses catégories d'usagers. Nous ne nous fierons pas à des chercheurs en histoire pour nous dire que notre projet est très

70^e Congrès de l'ACFAS -- Session C-420

Publications et lectures numériques : problématiques et enjeux

intéressant. De plus les technologies évoluent tellement rapidement qu'il faut maintenir une veille constante.

- Le rapport de fin de projet. Qui servira aux suivants pour aller plus loin.

[diapo 11]

Je vous laisse sur la vision du caricaturiste du Soleil sur le congrès de l'ACFAS. Il paraît que le titre de la conférence est véridique

Démonstration et/ou questions