

Une bibliographie en ligne des Lettres : un modèle à code source libre

Le cas de www.CompCanLit.ca

An Online Bibliography About Literature: An Open Source Model, the Case of www.CompCanLit.ca

Una bibliografía en línea sobre las Humanidades: modelo de códigos abiertos, el caso de www.CompCanLit.ca

John Taylor-Johnston

Volume 50, Number 2, April–June 2004

URI: <https://id.erudit.org/iderudit/1030082ar>

DOI: <https://doi.org/10.7202/1030082ar>

[See table of contents](#)

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)

2291-8949 (digital)

[Explore this journal](#)

Cite this article

Taylor-Johnston, J. (2004). Une bibliographie en ligne des Lettres : un modèle à code source libre : le cas de www.CompCanLit.ca. *Documentation et bibliothèques*, 50(2), 177–188. <https://doi.org/10.7202/1030082ar>

Article abstract

When the Bibliographie d'études comparées des littératures canadienne, québécoise et étrangères / Bibliography of Comparative Studies in Canadian, Québec and Foreign Literatures (CCL Bibliography) was first published on-line in 1995 at the Université de Sherbrooke, the researchers were unable to handle the necessary technology. Consequently, the first on-line attempts fell short of the target. This state of affairs changed however in 2002 when the CCL Bibliography enabled access to its data using the PHP-MySQL script and the Apache Web server, using a research model based on open source code and a new approach to the on-line publication of a bibliography about literature.

Even though the research has stayed the same, since 2002 the work process and publication methods have changed. The researchers at CCL currently develop the electronic bibliography using phpMyAdmin, an on-line MySQL interface, thereby eliminating considerable technological requirements. The electronic publication allows for the publication of paper formats using MySQL. This article constitutes a preliminary report and is an example of the amplification of Open Source in the transition from hard copy to the electronic format for the CCL project.

Une bibliographie en ligne des Lettres : un modèle à code source libre

Le cas de www.CompCanLit.ca

JOHN TAYLOR-JOHNSTON

Université de Sherbrooke
taylorjo@collegesherbrooke.qc.ca

RÉSUMÉ | ABSTRACTS | RESUMEN

Lorsque la *Bibliographie d'études comparées des littératures canadienne, québécoise et étrangères/Bibliography of Comparative Studies in Canadian, Québec and Foreign Literatures (Bibliographie CCL)* a été publiée en ligne pour la première fois, en 1995, à l'Université de Sherbrooke, les chercheurs n'étaient pas prêts à composer avec la technologie requise. Par conséquent, ses premières tentatives en ligne sont tombées à plat. Cet état de fait a toutefois changé, en 2002, lorsque la *Bibliographie CCL* permettait l'accès à ses données en utilisant un script PHP-MySQL et un serveur Apache Web, adoptant un modèle de recherche à code source libre (*Open Source Initiative*) et une nouvelle approche à la publication en ligne d'une bibliographie des Lettres.

Depuis 2002, bien que la recherche soit restée la même, les procédures de travail et les méthodes de publications ont changé. Les chercheurs de la CCL développent maintenant leur bibliographie électroniquement en utilisant *phpMyAdmin*, une interface MySQL en ligne, éliminant ainsi des besoins technologiques considérables. La publication électronique offre aujourd'hui la possibilité d'utiliser MySQL pour la compilation des formats papier. Cet article se veut un rapport intérimaire et une exemplification *Open Source* de la transition de la copie papier vers l'électronique pour le projet de la CCL.

An Online Bibliography About Literature : An Open Source Model, the Case of www.CompCanLit.ca

When the *Bibliographie d'études comparées des littératures canadienne, québécoise et étrangères / Bibliography of Comparative Studies in Canadian, Québec and Foreign Literatures (CCL Bibliography)* was first published on-line in 1995 at the Université de Sherbrooke, the researchers were unable to handle the necessary technology. Consequently, the first on-line attempts fell short of the target. This state of affairs changed however in 2002 when the CCL Bibliography enabled access to its data using the PHP-MySQL script and the Apache Web server, using a research model based on open source code and a new approach to the on-line publication of a bibliography about literature.

Even though the research has stayed the same, since 2002 the work process and publication methods have changed. The researchers at CCL currently develop the electronic bibliography using *phpMyAdmin*, an on-line MySQL interface, thereby eliminating considerable technological requirements. The electronic publication allows for the publication of paper formats using MySQL. This article constitutes a preliminary report and is an example of the amplification of *Open Source* in the transition from hard copy to the electronic format for the CCL project.

Una bibliografía en línea sobre las Humanidades : modelo de códigos abiertos, el caso de www.CompCanLit.ca

Cuando en 1995 se publicó en línea por primera vez la *Bibliographie d'études comparées des littératures canadienne, québécoise et étrangères / Bibliography of Comparative Studies in Canadian, Québec and Foreign Literatures (Bibliographie CCL)*, en la Universidad de Sherbrooke, los investigadores no estaban

preparados para lidiar con la tecnología requerida. En consecuencia sus primeras tentativas en línea se vinieron abajo. No obstante, este estado de hecho cambió en 2002 cuando la *bibliografía CCL* permitió el acceso a estos datos utilizando el tipo de programa (script) PHP-MySQL y un servidor Apache Web, que adopta un modelo de búsqueda por código abierto (*open source*) y un nuevo enfoque para la publicación en línea de una bibliografía para las Humanidades.

Desde 2002, si bien la búsqueda no se ha modificado, los procedimientos de trabajo y los métodos de publicación sí cambiaron. Los que buscan en la CCL desarrollan ahora su bibliografía electrónicamente utilizando *phpMyAdmin*, una interfase de MySQL en línea que elimina necesidades tecnológicas considerables. La publicación electrónica ofrece actualmente la posibilidad de utilizar MySQL para la compilación de formatos impresos. Este artículo es un informe provisional y un ejemplo de amplificación del código abierto de la transición de la copia impresa hacia la electrónica para el proyecto de la CCL.

L'analyse de l'évolution de la *Bibliographie d'études comparées des littératures canadienne, québécoise et étrangères/Bibliography of Comparative Studies in Canadian, Québec and Foreign Literatures* à l'Université de Sherbrooke peut avoir des incidences importantes sur les autres projets de recherche en Lettres qui voudraient utiliser la technologie des bases de données en ligne. Déjà, en 1995, les principaux chercheurs du projet de bibliographie de la Littérature canadienne comparée (CCL) publiaient leurs recherches en ligne. Cependant, durant les six premières années du projet, la surabondance de changements technologiques a lourdement entravé leurs efforts de publication en ligne. Ce n'est qu'en 2001, lorsque l'équipe de la CCL a enregistré www.CompCanLit.ca et qu'elle a porté ses données sur un serveur de bases de données en ligne, que le projet de bibliographie en ligne a commencé à se développer de manière significative. Utilisant un serveur à code source libre, (la combinaison PHP, MySQL et *phpMyAdmin*) le groupe CCL pouvait dorénavant maintenir, mettre à jour et administrer sa *Bibliographie* entièrement en ligne. Le seul logiciel dont le chercheur a maintenant besoin est un simple navigateur Internet.

Ainsi, le groupe CCL a-t-il non seulement adopté une nouvelle technologie, mais aussi un nouveau modèle de recherche, celui de l'*Open Source Initiative* (OSI). Retraçant ses origines auprès d'une communauté diversifiée de scientifiques, de programmeurs et

d'universitaires, les fondateurs de l'OSI affirment qu'Internet a vu le jour en tant que projet scientifique et académique et qu'il doit donc rester en tout temps un outil collectif ouvert, librement accessible. Il ne doit pas être une entreprise commerciale fermée et propriétaire. Dans cette optique, notre article s'insère dans la bibliographie en ligne d'*Open Source* selon la « Déclaration des droits » de l'OSI, telle que présentée ci-dessous :

« Le droit de copier ce programme et d'en distribuer les copies [;]

Le droit d'accéder au code source du logiciel... [;]

Le droit d'apporter des améliorations au programme. » (Perens, p. 172)

Le maintien du projet CCL en tant que modèle « scientifique » à code source libre exige non seulement que nous partagions la source de programmation, afin que ce modèle soit reproductible, mais également que le « processus de découverte et le processus de justification » soient expliqués (DiBona, Ockman et Stone, p. 2).

À l'instar d'une méthode scientifique de recherche, ce rapport :

- ▶ présentera l'*Open Source Initiative*;
- ▶ résumera les six ans qui ont mené à l'adoption d'un serveur à code source libre;
- ▶ exposera l'ensemble des logiciels que le projet de Bibliographie CCL a mis sur pied;
- ▶ servira à illustrer une bibliographie *Open Source* en ligne;
- ▶ traitera des avantages de l'informatisation; et
- ▶ étudiera comment le modèle *Open Source* a changé la manière dont le groupe CCL a compilé sa Bibliographie.

L'OPEN SOURCE INITIATIVE

Le mouvement ou « révolution » *Open Source*, selon DiBona, Ockman et Stone, a commencé dès 1984, lorsqu'un chercheur du MIT, Richard Stallman, a créé la *GNU General Public Licence* (GPL)¹. Le rêve de l'informaticien Stallman était de créer une « communauté de partage libre de logiciels » (le projet GNU) basée sur le principe que « le savoir scientifique ... doit être partagé et distribué ».

« ... la méthode scientifique repose sur un processus de découverte et sur un processus de justification. Pour que des résultats scientifiques soient justifiés, ils doivent être reproductibles. Cette reproductibilité n'est possible que si la source est partagée : l'hypothèse, les conditions de contrôle et les résultats. Le processus de découverte peut emprunter plusieurs avenues et, à certains moments, les découver-

tes scientifiques peuvent même naître dans l'isolement. Mais au bout du compte, le partage ouvert des résultats scientifiques facilite les découvertes, permettant ainsi à d'autres d'y contribuer. » (DiBona, Ockman & Stone, p. 2)

Le projet de Stallman, www.gnu.org, est basé sur la notion que les logiciels doivent être « libres » d'accès. Il insiste sur le fait que les programmeurs doivent partager le code source de leur logiciel et le redistribuer librement. Stallman considère que ce « principe de base » est moralement essentiel pour que « continue l'innovation » (DiBona, Ockman & Stone, p. 2). Pour plus de précision, Stallman le définit en tant que « libre, comme dans liberté » (p. 56). Stallman n'a rien contre la vente de collections de logiciels libres sur Cd-rom, en autant que cela serve à « lever des fonds pour le développement des logiciels libres » et que ces logiciels puissent être librement inclus (p. 56) ou « [démontés]... pour bâtir de nouveaux logiciels » (p. 53). Toutefois, les adeptes de www.gnu.org sont des puristes qui préfèrent que leur code source ne soit pas mêlé ou associé à des logiciels commerciaux.

À l'autre extrémité de la gamme, on retrouve OpenSource.Org, une organisation fondée en 1997 (DiBona, Ockman & Stone, p. 3). Ses adeptes ont créé un terrain commun entre les puristes du GNU et les programmeurs commerciaux propriétaires. La « Déclaration des droits » du code source libre est nécessairement plus libérale. Malgré de solides liens avec la communauté académique, les projets de code source libre comme *PHP*, *MySQL* et *Apache* sont fréquemment, quoique parfois inopportunément, appelés « applications de e-Commerce ». Il est vrai que les fondateurs de OpenSource.Org, comme Tim O'Reilly, considèrent les sites Web tels Amazon.com, Yahoo.com (voire www.CompCanLit.ca), comme des « applications d'information », ou « infoware » (O'Reilly, p. 189). Les sites Web qui gèrent les tâches sont donc considérés comme des logiciels « infoware » (p. 189). Parce qu'Internet s'adresse commercialement au grand public, l'*infoware* peut être de nature commerciale, mais pas obligatoirement. Bref, contrairement aux projets GNU, le projet *Open Source* entremêle souvent sa recherche avec celle d'autres produits ou organismes commerciaux, ce qui, du même coup, aide à développer et à faire progresser le projet *Open Source* initial.

Mettons les licences *Open Source* et GPL en perspective : Rasmus Lerdorf, l'inventeur du *PHP*, distribue ses scripts de langage avec une licence *Open Source* (*The Approved Licenses*), parce qu'elle est moins « restrictive » qu'une licence GPL (Lerdorf). À l'instar de Stallman et de son idéal, Lerdorf insiste sur le fait

1. D'après www.gnu.org, GNU est un acronyme « récursif » pour « GNU's Not Unix » (GNU n'est pas Unix). <<http://www.gnu.org/>>.

L'accès aux forums de discussion Open Source a été la clé de la croissance rapide de la bibliographie en ligne de la CCL.

qu'une licence *Open Source* est cruciale pour l'avancement de la recherche *PHP*. Dans son opposition à Stallman, Lerdorf est inflexible : « plus il y aura de distributeurs de logiciels commerciaux qui adoptent le *PHP*, mieux ce sera [pour faire avancer le *PHP*] » (Lerdorf). Toutefois, Lerdorf souligne qu'une licence GNU fournit une importante protection pour les programmeurs qui ne veulent pas que des logiciels commerciaux propriétaires s'approprient leur code. Néanmoins, « [cette] même protection restreint le choix du type d'utilisateur du logiciel et la manière dont il peut être utilisé, et ce n'est pas ce qui nous préoccupe dans cette question. » (Lerdorf)

Perens voit « la définition de l'Open Source comme une Déclaration des droits de l'utilisateur » (p. 171). Elle est fondée sur cinq principes de base :

- 1) l'*infoware Open Source* doit être gratuit, sans licence ou dérivation supplémentaire à la licence gratuite;
- 2) on doit être capable d'accéder au code source et de le modifier;
- 3) on doit avoir le droit de modifier le code source de programmation selon ses besoins;
- 4) on doit avoir le droit de redistribuer librement sa version du logiciel « *openware* » original;
- 5) on ne doit exercer aucune discrimination à l'endroit d'une personne ou d'un groupe utilisant son programme, ni restreindre l'utilisation de celui-ci (*The Open Source Definition*). C'est ce dernier point qui différencie www.gnu.org de www.opensource.org.

Tim O'Reilly, le cofondateur de www.opensource.org, souligne qu'il y a eu un « glissement du pouvoir » (p. 193), du logiciel propriétaire vers l'*Open Source*. « Le logiciel [fermé], ajoute-t-il, n'est plus le moteur central... de l'industrie de l'informatique » (p. 193). O'Reilly présente une chronologie du mouvement en commençant par la dominance du matériel avec IBM, en passant par l'ère Microsoft, jusqu'à l'ère *Open Source* où l'élaboration d'une version sur mesure de son logiciel favori est possible. « Il y a de la place pour l'expérimentation ... l'évolution n'engendre pas un gagnant unique, mais bien la diversité ... le logiciel *Open Source* vous permet de "gratter là où ça démange" » (p. 194).

Pour des projets comme la Bibliographie CCL, la révolution *Open Source* a été novatrice, spécialement à cause du système de soutien de sa communauté et de la manière dont elle a rendu opportuns la recherche et le développement. L'accès aux forums de discussion

Open Source a été la clé de la croissance rapide de la bibliographie en ligne de la CCL. Les solutions ponctuelles et utiles à des problèmes mineurs de programmation sont presque toujours trouvées par la « communauté ». Les réponses arrivent en quelques heures voire parfois en quelques minutes.

Par contre, la culture de soutien et l'esprit communautaire sont différents. Les programmeurs propriétaires hésitent à partager totalement. Ils donnent parfois des indices, mais partagent rarement le code lui-même. L'apprentissage par l'exemple est impossible, obligeant le programmeur novice à se débrouiller seul. Par contraste, [news://news.php.net/php.general](http://news.php.net/php.general) a su me fournir suffisamment de codes de langage *PHP*, en novembre 2001, pour me permettre de construire un modèle, de le présenter, de le ratifier et de transformer www.compcanlit.ca en un serveur *Open Source* entièrement fonctionnel dès janvier 2002 (*Minutes of Biblio Meeting, 21 janvier 2002*). Qui plus est, la combinaison *PHP-MySQL* a pu résoudre plusieurs problèmes qui étaient demeurés sans réponse pour la CCL avec ses serveurs de 1995, 1997 et 2001 « programmés sur une base propriétaire ».

LA PÉRIODE 1989-2001²

Durant la période allant de 1989 à 2001, le nombre d'entrées dans la base de données de Sirois et al. est passé de 934 à 1624. Plus particulièrement, la bibliographie sur papier a été mise à jour jusqu'en 1995. Toutefois, les trois premières versions électroniques de la Bibliographie CCL se sont avérées inadéquates. Les sites Web de 1995, 1997 et 2001 étaient gênés par des problèmes de gestion du contenu et des problèmes techniques qui, pour les chercheurs de la CCL, devenaient compliqués et prenaient énormément de leur temps. En particulier, le moteur de recherche booléenne ne fonctionnait pas et les chercheurs n'avaient pas le soutien technique nécessaire pour mettre à jour la base de données en ligne.

Serveur 1995 : <http://www.usherb.ca/CCL/>
Serveur 1997 : <http://www.si.usherb.ca/CCL/>
Serveur 2001 : <http://webdev.si.usherb.ca/CCL/>
Serveur 2002 : <http://CompCanLit.ca/>

Les sites Web de 1995 et 1997 étaient, à l'origine, des projets scolaires de Christopher Enright, un professeur d'informatique qui utilisait *Active Server Pages* (ASP) et *Access* de Microsoft pour emmagasiner et produire en ligne les données bibliographiques du groupe CCL. Bien que les sites Web de 1995 et 1997 aient été à la fine pointe de la technologie, les chercheurs n'étaient pas prêts à utiliser une technologie en ligne. Ceux-ci mettaient sur pied la Bibliographie à l'aide de *NotaBene for DOS* (NBDOS). Leurs besoins

2. Gregory J. Reid utilise l'expression « interstice » dans Sirois et al. (i).

technologiques immédiats dépendaient de *Ibidem*, un outil de gestion bibliographique compris dans *NotaBene*, pour produire leurs publications sur papier. Lorsqu'en 1997, Enright a laissé le projet, les chercheurs ne possédaient pas le savoir technologique et les ressources humaines requises pour maintenir et mettre à jour la banque de données en ligne. Parce que leur perspective et leurs besoins étaient concentrés ailleurs, l'utilisation d'Internet leur semblait inaccessible. Malgré l'aide du Service des technologies de l'information (STI), la banque de données en ligne a cessé d'être mise à jour jusqu'en 2001.

Vers la fin de 1999, le projet de recherche CCL engageait un technicien/analyste afin de réorganiser sa tentative de 1997 sur Internet. Cependant, parce que l'emphase était mise sur la préparation de la publication de 2001 sur papier, on a dû résoudre plusieurs problèmes techniques, y compris une technologie de base de données périmée, avant que la bibliographie ne soit imprimée. En 2001, les chercheurs utilisaient une version MS-DOS de *NotaBene* (un logiciel de publication bibliographique) antérieure à Windows 3.1. Au départ, seule une mise à niveau de *NBDOS* à *NBWin* (*NotaBene for Windows*) était requise. *NBDOS* s'est avéré instable; il était congestionné par plus d'entrées que sa structure ne le permettait et avait été installé sur des ordinateurs qui n'avaient pas suffisamment de mémoire RAM. Par conséquent, *NBDOS* tombait fréquemment en panne, malgré les mises à niveau recommandées de la mémoire RAM et des disques durs. *NBWin* n'a pas non plus été à la hauteur; il était incapable de reproduire les fonctions d'indexation qui avaient rendu *NBDOS* essentiel. Alors, la Bibliographie complète a été transposée sur *WordPerfect* afin de produire les index des savants, des auteurs sous étude, des régions géographiques et particulièrement des titres bibliographiques. Toutefois, après la publication, le passage à *WordPerfect* se traduisait par des mois de travail pour mettre à jour la base de données *NBWin* à cause des changements apportés dans *WordPerfect*. Finalement, en janvier 2001, j'ai pu concentrer mon attention sur le site Web de 1997 qui n'avait toujours pas été mis à jour.

Ma première impression des méthodes de recherche du projet CCL a été le constat que trop de technologies avaient déjà été utilisées pour produire la version papier. Qui plus est, lorsque vint le temps de réorganiser le site de 1997, on était encore loin d'avoir une base de données utilisable à portée de la main. Il n'y avait que deux options possibles quant à la procédure. D'une part, je pouvais reconstruire les données *NBWin* existantes dans le format *NBDOS* qui avait été utilisé en 1995 et 1997. Il m'aurait néanmoins fallu transcoder les données dans un format que *MS Access* pouvait utiliser. Ceci avait déjà été fait en utilisant un outil intermédiaire de programmation appelé Perl³.

Malheureusement, les STI avaient perdu le code Perl, ce qui signifiait qu'il fallait repartir à zéro.

Par ailleurs, *NBWin* pouvait exporter les données vers un « format tableau » que *MS Access* pouvait lire, mais *NBWin* n'a pas pu suffire à la tâche. La base de données de la CCL comprenait 27 champs, desquels *NBWin* pouvait exporter 12 champs à la fois (Image 1). Pour mener à bien ce travail, l'exportation devait se faire en trois fois et le script Perl devait quand même être programmé pour enchaîner ces trois exportations afin qu'elles puissent fonctionner dans *MS Access*.

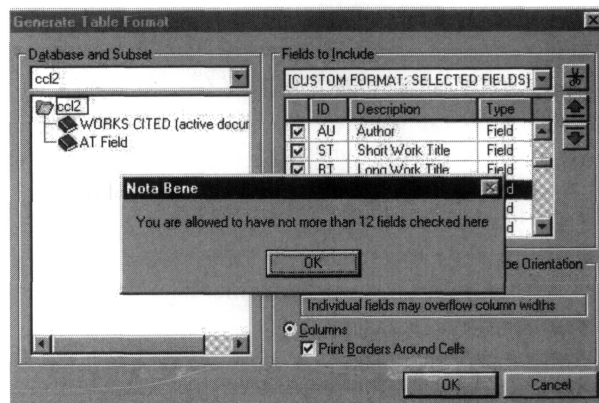


Image 1 : Format Tableau d'exportation de *NBWin*

Ces deux procédures prenaient beaucoup trop de temps et se sont avérées finalement peu efficaces. Pour que chacune de ces possibilités puisse fonctionner, on devait malgré tout modifier manuellement le code de la base de données pour la rendre utilisable. Cette deuxième procédure a toutefois été utilisée comme mesure palliative pour le site Web de 2001. Il était clair que le projet de la CCL exigeait une solution qui puisse être à la portée du coordonnateur de la recherche. De plus, il ne fallait pas que le projet dépende trop d'un technicien/analyste pour produire des résultats. Une autre méthodologie était nécessaire pour que le site Web de 2002 puisse permettre un travail d'équipe.

Deux problèmes administratifs devaient également être résolus en janvier 2001. La mission initiale n'était pas seulement de faire en sorte que le site de 1997 soit fonctionnel, mais on se devait également de créer un pont entre la technologie, les chercheurs CCL et les STI. Une mesure d'autonomie était également souhaitable. Les données du site de 1997 étaient emmagasinées sur un site Web de développement (WebDev). Seuls les techniciens des STI avaient le droit de transférer les données mises à jour du WebDev au site Web de la CCL. Dès la création de *www.CompCanLit.ca*, le site de 2001 se trouvait directement lié à la base de données WebDev, économisant ainsi le temps nécessaire aux STI pour effectuer des

3. Perl est un langage de scripts UNIX, qui fut le précurseur de *PHP*. Perl est l'acronyme de *Practical Extraction and Report Language*. On peut obtenir plus de renseignements sur Perl au http://www.perl.org/press/fast_facts.html.

mises à jour à la base de données principale. Quoique fonctionnelle, cette mesure n'était que palliative. Il restait encore à trouver une solution au problème de la procédure de mise à jour des données qui était très difficile au plan technologique pour les chercheurs.

La deuxième difficulté consistait à enregistrer le nom de domaine www.CompCanLit.ca lui-même. La Bibliographie désirait donner à sa publication une adresse Internet distincte. Les politiques de l'Université ne permettent pas l'enregistrement d'un nom de domaine extérieur à la structure de l'établissement. C'était une dure leçon et ce problème n'est pas encore résolu. Pour conserver le nom de domaine original, une firme externe doit gérer et rediriger, moyennant des frais annuels, l'adresse www.CompCanLit.ca vers CompCanLit.USherbrooke.ca.

« Si [l'informatique] a permis des gains considérables, il ne faut pas perdre de vue, toutefois, qu'elle est également à la source de nouvelles attentes. Y répondre et combler les lacunes existantes s'avèrent par ailleurs difficilement pensable... » (Masson, p. 2)

De plus, plusieurs nouvelles demandes ont été faites concernant le site Web de 2001, entre autres, celle d'une méthode de recherche booléenne. Les données devaient aussi être accessibles dans un format que le site Web et le coordonnateur de la recherche pouvaient utiliser. *NBWin*, un outil bibliographique essentiel, n'était pas compatible avec *MS Access*, malgré son format « tableau d'exportation ». Il était possible de sauvegarder les données exclusivement en ligne, mais il fallait créer un outil d'administration en ligne pour le serveur *ASP*. Ceci allait exiger des centaines d'heures de programmation, de recherche et de développement dont ne disposait pas le projet. Qui plus est, la perspective d'une édition papier en 2005 impliquait encore la création d'index, ce qui allait demander trop de temps. Ce processus allongé de mise à jour du site Web n'était désormais plus faisable au plan technologique ou administratif. On se devait donc de trouver d'autres solutions.

LA SOLUTION DU PROGICIEL OPEN SOURCE

En septembre 2001, alors que je travaillais sur un autre projet académique, j'ai découvert la solution *Open Source* qui allait résoudre le problème du projet CCL, tant sur le plan technique qu'administratif. Regroupé en un ensemble sur un serveur *Windows*, *EasyPHP* (www.EasyPHP.org) comprenait déjà tous les éléments *Open Source* essentiels : un serveur Web *Apache*; *MySQL*, une base de données en ligne; *PHP*, un script de langage qui produit des pages Web à partir de *MySQL*; et le plus important, *phpMyAdmin*, une interface en ligne à utilisateurs multiples pour *MySQL*, construite à l'aide de *PHP*.

J'ai proposé cette technologie *Open Source* en décembre 2001 comme alternative viable à l'achat de mises à niveau pour notre serveur *Microsoft ASP/MS Access* et j'ai également proposé d'équiper le site Web avec de nouveaux outils de programmation. Toutefois, le facteur décisif n'a pas été une question de budget, mais plutôt de polyvalence. *EasyPHP* regroupait toutes les technologies nécessaires à la Bibliographie CCL en une seule méthodologie exécutable. Le projet CCL nécessitait un effort collectif de recherche; il n'était pas question qu'un seul chercheur fasse la totalité des entrées de données.

Au plan administratif, lorsque j'ai été engagé, les responsables de la Bibliographie CCL avaient exprimé leur besoin d'autonomie technologique par rapport aux STI. Leur premier geste a été d'installer le serveur de base de données en ligne *PHP-MySQL* dans le local de recherche de la CCL. Toutefois, à la suite d'un coûteux vol par effraction en mars 2002, le projet est retourné aux STI, transférant son site Web et toute sa base de données sur un serveur *RedHatUnix* offrant une meilleure sécurité. Ce progiciel de serveur *RedHat*, à l'instar de *EasyPHP*, était lui aussi un projet *Open Source* (jusqu'en 2004). Bien que le serveur et la base de données soient installés aux STI, le projet CCL a tout de même conservé son autonomie d'accès et continue de recevoir un soutien utile du personnel des STI.

Au niveau technique, les couches multiples de logiciels propriétaires, *NBDOS*, *NBWin*, *WordPerfect*, les macros *WordPerfect*, *Perl*, *MS ASP*, *MS Access*, *MS FrontPage*, ont été progressivement retirées. Au premier plan, les seuls logiciels dont ont besoin les chercheurs sont un navigateur Web et *phpMyAdmin* et l'interface en ligne pour la base de données *MySQL*. La technologie physique elle-même réside en arrière-plan. En fait, *phpMyAdmin* a supplanté *NBWin* lors d'une réunion du service : « *Latelier NBWin Ibidem n'a pas eu lieu, parce qu'il appert que nous passerons probablement à MySQL, ce qui rend inutile la connaissance de Ibidem* » (*Minutes of Biblio Meeting, January 21, 2002*). Au cours de la même réunion, l'adoption du serveur *EasyPHP* a été ratifiée; j'ai dû faire la démonstration que la base de données *MySQL* était compatible à rebours. « *Peut-on transférer MySQL de nouveau à NBWin pour la publication de copies papier?* » Ma réponse était « *oui* ». J'ai immédiatement fait la démonstration en transférant des données à partir de *MySQL* en fichier texte, que j'ai ensuite annexées à *NBWin Ibidem*.

Les professeurs de la CCL utilisent encore *NotaBene* en tant qu'outil bibliographique essentiel. Cependant, les chercheurs se servent exclusivement de *phpMyAdmin* et d'un navigateur Web. En fait, l'interface *MySQL*, *phpMyAdmin* est devenue un outil essentiel pour la cueillette, l'entrée et la maintenance des données. Elle peut fonctionner avec n'importe quelle

plateforme (*Macintosh, Windows* ou *Unix/Linux*) et il en va de même pour le visiteur du site. Principalement, les chercheurs devraient pouvoir travailler indépendamment sans tenir compte de la technologie qui opère en arrière-plan.

Cependant, la technologie, elle aussi, devait faire la preuve qu'elle était « ouverte ». À l'instar de la méthodologie *Open Source*, le savoir, quant à la manière dont cette technologie fonctionne, ne doit pas être centralisé, mais plutôt être ouvert à tout le groupe. Si je devais quitter le projet, la Bibliographie ne devrait plus être tributaire d'un unique programmeur. Avant que ce modèle de recherche ne soit ratifié en janvier 2002, j'ai dû aussi démontrer que les STI pouvaient facilement gérer la base de données de la CCL et nommer mon successeur. J'ai démontré que les STI utilisaient déjà la même technologie pour d'autres projets.

UN EXEMPLE OPEN SOURCE

« *Donnez-moi quelque chose que je puisse utiliser demain matin!* »⁴

Cette partie de l'article a pour ambition de faire la démonstration d'un modèle téléchargeable de base de données bibliographiques en ligne, complète avec son code source pour :

- ▶ une page couverture HTML :
- ▶ une page *PHP* pour traiter la recherche bibliographique ;
- ▶ un échantillon de tableau de base de données.

Cette section reproduit⁵ (en moins détaillé) le code de programmation actuellement utilisé par www.CompCanLit.ca. Une personne ressource en informatique peut installer la programmation fournie. L'important est que le lecteur comprenne comment cette démonstration fonctionne. Les explications techniques ne sont offertes que pour favoriser l'intérêt et encourager les autres à bâtir sur cet exemple pour répondre aux besoins d'autres chercheurs.

La Plaque 1 fonctionne en n'utilisant que deux fichiers informatiques. Il y a tout d'abord la page couverture `index.html` (plaque 1⁶), qui est une page Web standard en HTML (*HyperText Mark-Up Language*). La capacité de traitement d'une banque de données bibliographique en ligne est démontrée « côté serveur » par un second fichier d'ordinateur. Lorsqu'on clique `submit`, `index.html` appelle `index.php`, qui à son tour traite la demande de recherche par le biais de *MySQL* et en affiche le résultat à l'écran. Tout le calcul est fait côté serveur par `index.php`.

On construit `index.php` (*sic*) en utilisant *PHP*, un « langage de scripts » (*PHP Manual, Chapter 1. Introduction*). Pour un fichier texte, on peut combiner *PHP* avec HTML et générer dynamiquement des pages Web. On utilise le terme « dynamiquement », parce

Plaque 1 : Page couverture
<<http://CompCanLit.ca/sample/>>

UNIVERSITÉ DE
SHERBROOKE
Sample Bibliography / Échantillon de
bibliographie
Département des lettres et communications - Université de
Sherbrooke

Margaret Atw. Search / Recherche

Advanced Search / Recherche avancée

© <http://www.CompCanLit.ca/> - Département des lettres et communications - Université de Sherbrooke
Programming/programmation : jt.johnston@usherbrooke.ca

Voir le code⁷ :

- ▶ <http://CompCanLit.ca/sample/index.html>
- ▶ <http://CompCanLit.ca/sample/index.php>

que le caractère et la nature de chaque page Web générée avec *PHP* dépendent des données que *MySQL* retourne lorsque l'on clique `submit`. Donc, `index.php` n'est pas une page Web statique comme `index.html`. Son contenu varie selon la demande de recherche entrée à `index.html`. De fait, si l'on ne fait que consulter `index.php` et qu'il n'y a pas eu de recherche effectuée, `index.php` est alors programmé pour afficher le contenu de `index.html`.

Les opérateurs de recherche MySQL 4.x

Ce qui fait qu'une base de données en ligne fonctionne bien repose sur la capacité de grammaire et de syntaxe de sa fonction de recherche. Une recherche *simple* de « Margaret Atwood » dans la plaque 1 ne trouvera pas « Atwood, Margaret » ou d'autres variations. Elle trouvera les exemples se rapportant seulement à la phrase exacte tapée; ce n'est pas une recherche suffisamment intelligente ou polyvalente pour chercher d'autres possibilités. C'est à ce chapitre que les fonctions de recherche des sites Web de 1995, 1997 et 2001 ont échoué, académiquement parlant. Pour

4. Mon premier professeur formateur, Carole Lanthier, m'a enseigné l'importance de ce précepte de base en démontrant quelque chose de nouveau.

5. `index.html` (*sic*) est la page principale. Dès qu'on clique `submit`, `index.html` appelle `index.php`. `index.php` (*sic*) appelle le serveur *MySQL*, qui possède les données que contient dans `sample.sql`. Selon la requête soumise, (Plaque #4), `index.php` rendra tout ou n'importe quel tableau d'archive qui corresponde à la requête. Si vous désirez reproduire cette expérience, tout le code source est contenu dans les liens ci-dessous. Vous n'avez donc qu'à copier-coller dans un fichier texte le code apparaissant à l'écran et à nommer ce fichier en conséquence.

<http://CompCanLit.ca/sample/index.html>
<http://CompCanLit.ca/sample/index.php>
<http://CompCanLit.ca/sample/generate.php>
<http://CompCanLit.ca/sample/sample.sql>

6. La syntaxe `http://CompCanLit.ca/sample/` est la même que `http://CompCanLit.ca/sample/index.html`. Si vous ne spécifiez pas une page à la fin de la syntaxe, le serveur Web Internet assignera une page à afficher, habituellement `index.html`. Cette page `index.html` est appelée *page DirectoryIndex* (page du répertoire). (http://http.apache.org/docs/mod/mod_dir.html#directoryindex). On utilisera typiquement la syntaxe `index.html`.

7. *Ibid.*

Plaque 2 : Opérateurs booléens

Extrait de : <http://CompCanLit.ca/mysql4.html>

Margaret Atwood

[Cette opération de recherche] trouvera des lignes contenant au moins un de ces mots. [Le résultat peut très bien être Atwood, Margaret.]

+Margaret +Atwood

... les deux mots.

+Margaret Atwood

... le mot « Margaret », mais le classer plus haut s'il contient aussi « Atwood ».

+Margaret -Atwood

... le mot « Margaret » mais sans « Atwood ».

+Margaret +(>Atwood <Lawrence)

... « Margaret » et « Atwood », ou « Margaret » et « Lawrence » (dans n'importe quel ordre), mais classer « Margaret Atwood » plus haut que « Margaret Lawrence ».

Marg*

... « Margaret », « Margret », « Margie », ou « Marge ».

« quelques mots »

... « quelques mots de sagesse », mais pas « quelques mots bruyants ».

Plaque 3 : Échantillon bibliographique

<http://CompCanLit.ca/sample/generate.php>

Ouvrages cités

1 Allen, Carolyn. *Failures of word, uses of silence* : Djuna Barnes, Adrienne Rich et Margaret Atwood. *Regionalism and the Female Imagination*. 4.1 (1978) : 1-7.

2 Atwood, Margaret. *Canadian-American relations : surviving the eighties*. *The Literary Criterion*. 20.1 (1985) : 227-246. *The Literary Criterion*.

3 Atwood, Margaret. (1972) : *Survival : a thematic guide to Canadian literature*. Toronto : Anansi, 1972.

4 Baer, Elizabeth Roberts. *Pilgrimage inward : quest and fairy tale motifs in Surfacing*. (1988) : 24-34. *Margaret Atwood : Vision and Forms*. Carbondale : So. Illinois University Press, 1988.

5 Colman, S.J.. Margaret Atwood, *Lucien Goldmann's « Pascal », and the meaning of Canada*. *University of Toronto Quarterly : A Canadian Journal of the Humanities*. 48 (1979) : 245-262.

code PHP incrusté dans cette page¹⁰ :

▶ <?php comprend (« échantillon/generate.php ») ;?> dont l'URL est :

▶ http://CompCanLit.ca/sample/generate.php

qu'une demande de recherche soit utile, elle nécessite un ensemble de règles de grammaire.

Le mode de recherche le plus reconnu aujourd'hui est la recherche booléenne. Nommée d'après un mathématicien français⁸, la recherche booléenne comprend habituellement les opérateurs relationnels «+ — * < >» et les opérateurs logiques «ET, OU, SANS», qui permettent à l'utilisateur de choisir comment se déroule la recherche. *MySQL* considère les opérateurs logiques périmés et ne les utilise pas. La Plaque 2 explique la grammaire de recherche booléenne, dans le cadre de la base de données de la CCL (*Comparative Canadian Literature*)⁹.

La Plaque 3 montre les données de l'échantillon bibliographique entré dans le tableau *MySQL* utilisé dans la Plaque 4. La fonction «generate.php» (*sic*) nous permet d'afficher à l'écran le contenu de la bibliographie, produite en format M.L.A.

La Plaque 1 démontre pourquoi la recherche booléenne est importante pour une recherche académique. La simple recherche de «Margaret Atwood» dans la Plaque 1 est insuffisante pour une recherche

Plaque 4 : Échantillon de recherches booléennes

<http://CompCanLit.ca/sample/>

Opération 1

Code du tableau *Open Source MySQL*¹¹ :

▶ http://CompCanLit.ca/sample/sample.sql appelé par :

▶ http://CompCanLit.ca/sample/index.php

académique. Le nombre d'entrées qu'elle peut rendre est trop vaste. Une telle recherche rend tous les exemples de «Margaret» et de «Atwood», même si «Margaret Lawrence» est mentionnée dans la Plaque 3, entrée 6. Par contre, la Plaque 4 utilise une grammaire de recherche relationnelle, réduisant ainsi le résultat de la recherche. Plus particulièrement, l'opération 2 réduit la recherche aux exemples où «Margaret Atwood» est mentionnée, mais pas où «Atwood, Margaret» est l'auteur.

Au plan académique, l'absence de recherche booléenne pour les versions 1995, 1997 et 2001 du site Web de la CCL présentait une sérieuse problématique. Pour l'équipe de recherche, la question était le besoin de minutie. Cependant, pour le programmeur, il

8. Renseignements sur George Boole au <http://www.texas.net/~square1/start2a.html>.

9. La Plaque 2 est dérivée de http://www.mysql.com/doc/en/Fulltext_Search.html et a été modifiée pour refléter le contenu de la base de données de la CCL.

10. *Ibid.* 5.

11. *Ibid.* 5.

s'agissait du nombre exponentiel de lignes de programmation requises pour effectuer l'analyse grammaticale de la requête et pour le code des opérateurs booléens. Pour les « amateurs qui construisent le Web » (O'Reilly, p. 192), cela représente un ordinogramme parfois compliqué. La recherche booléenne est complète, mais pour être précise, elle requiert une programmation méticuleuse. En effet, lorsque les gens de la CCL ont demandé aux programmeurs des STI de créer un moteur de recherche booléenne pour le site de 1997, programmé sur Microsoft ASP, ces derniers ont poliment refusé (*Re : Modifications apportées au site Web*).

Dès 2002, MySQL améliorait légèrement la question booléenne. La version 3.23 comprenait un « langage naturel de recherche » (*6.8 MySQL Full-text Search*). Toutefois, on était encore loin de la recherche booléenne. Sa construction représentait un travail de programmation titanesque, si l'on voulait que cette recherche soit à 100 % efficace. Lorsque www.MySQL.com a lancé une version de production de MySQL version 4.x en mars 2003, le projet de la CCL y a immédiatement intégré un « mode pleine recherche booléenne "texte complet" ». Aujourd'hui, parce que les opérateurs relationnels sont préprogrammés, cela se traduit par des avantages immédiats tant pour le programmeur que pour l'académicien. Une seule ligne de programmation en remplace maintenant des centaines. Le code de programmation derrière la Plaque 1 se lit comme suit :

```
SELECT AU, ST, BT, JR, VNum, INum, PL, PR,
YR, PG FROM Sample
WHERE MATCH (AU, ST, BT, JR, VNum, INum,
PL, PR, YR, PG)
AGAINST "Margaret Atwood" IN BOOLEAN
MODE)
```

Essentiellement, lorsqu'on clique *Submit* dans la Plaque 1, on demande à MySQL de choisir tous les champs (*SELECT [ALL]*) à partir (*FROM*) du tableau de la banque de données appelé *Échantillon (Sample)* où (*WHERE*) notre recherche sur « Margaret Atwood » est comparée (*MATCH*) à (*AGAINST*) tous (*[ALL]*) les champs.

Comment reproduire cet exemple sur votre serveur PHP-MySQL ?

Pour reproduire cette expérience, téléchargez <http://CompCanLit.ca/sample/sample.zip>. Ce fichier contient *index.html* et *index.php*, qui requièrent tous deux un serveur *PHP-MySQL*. Veuillez noter que *index.php* doit être configuré avec un nom d'utilisateur et un mot de passe sur le serveur *MySQL*¹². Les données bibliographiques sont sauvegardées dans un fichier texte tiers, *sample.sql*, qui doit être installé dans un tableau de banque de données *MySQL*, habituellement par un technicien, ou en utilisant l'inter-

face en ligne *MySQL — phpMyAdmin*. Pour des renseignements sur *phpMyAdmin* ou sur la manière de l'installer, veuillez consulter <http://www.phpMyAdmin.net/>.

La technologie

En 2002, la Bibliographie CCL a d'abord choisi *EasyPHP*, un logiciel de serveur sur Windows, pour développer leur premier serveur *Open Source*. C'était parce que *EasyPHP* regroupait *PHP*, *MySQL* et *phpMyAdmin* en un seul progiciel d'installation, plutôt que d'avoir à installer séparément chaque élément. Il a fallu environ 30 minutes pour télécharger, installer et configurer *EasyPHP*. Même s'il a fallu un technicien pour installer le premier serveur *EasyPHP*, il s'est avéré facile à opérer ainsi qu'à réinstaller sur d'autres machines. On peut obtenir une aide utile et gratuite sur *PHP* à partir de www.EasyPHP.org. Il était pratique et efficace au niveau des coûts d'utiliser une plateforme *Windows 98*. Cela permettait également de fournir une certaine mesure d'autonomie au technicien lors du développement du site Web de la Bibliographie, sans contrainte administrative. Toutefois, deux problèmes sont survenus lorsque le serveur a été installé dans le local de recherche de la CCL. Les panes inopportunes de courant exigeaient que je doive redémarrer le serveur en dehors des heures de travail. Au surplus, l'appareil lui-même, un ordinateur *Windows*, a été subtilisé en mars 2002 lors d'un vol par effraction ; il a fallu six semaines pour le remplacer. C'est à cause de cela que le groupe CCL a décidé d'investir dans un serveur dédié *RedHat Unix*. Par opposition à l'utilisation d'un appareil *Windows*, ce choix s'avérait plus coûteux et nécessitait le concours des services techniques des STI et l'installation de l'appareil dans leurs locaux. Les données étaient aisément transférables d'une plateforme *Windows* à un appareil *Unix* et des copies de sauvegarde étaient automatiquement générées.

LES AVANTAGES DE L'INFORMATISATION

« La bibliographie sert à mettre de l'ordre dans le chaos du savoir enregistré. » (*Hackman*)

Depuis au moins 1992, on a vu se développer une perspective académique différente de la bibliographie. En effet, la préparation d'une bibliographie en 2003 nécessite l'inscription informatique. Selon l'avis des savants, la bibliographie sur papier est tombée en défaveur. William Bowen, directeur fondateur de *l'Iter Renaissance Bibliography* de l'Université de Toronto, trace un aperçu des lacunes de la bibliographie papier :

12. L'*index.php* doit être composé à l'aide d'un éditeur de texte. *NotePad* ou *WordPad* feront l'affaire sur *Windows*. *SimpleText* servira aux utilisateurs *Macintosh*.

La page électronique peut réarranger et afficher dynamique- ment les données « à une vitesse inégalée par l'imprimé ».

«... les outils de recherche imprimés avaient des défauts communs que Iiter a été conçue pour éviter. Par exemple, la plupart des bibliographies étaient produites en tant que volumes annuels qui étaient habituellement quelques années en retard par rapport à la date de publication des articles qu'elles présentaient.

Aussi, elles offraient des possibilités de recherche limitées, surtout à cause de leur indexation rudimentaire par sujet et à cause des restrictions pratiques inhérentes aux bibliographies imprimées. De plus, en concevant les bibliographies comme des rapports annuels plutôt que cumulatifs, il était impossible pour les savants de faire une recherche approfondie sans devoir consulter chaque volume individuellement. Ce format avait aussi un inconvénient supplémentaire lorsque les omissions d'une année n'étaient pas corrigées dans les volumes subséquents; la couverture inégale de la documentation était encore plus apparente dans le cas de bibliographies qui dépendaient de fragiles réseaux bénévoles de savants et d'instituts de recherche à travers le monde. Et finalement, compte tenu de la qualité inégale des produits, les bibliographies offertes semblaient onéreuses. » (Bowen, p. 1)

La discussion savante en faveur d'une bibliographie en ligne va bien au-delà de ses avantages évidents. Dès 1992, George Shirinian, le coordonnateur des services techniques du réseau des bibliothèques publiques de la ville de New York, décrivait la structure académique et théorique de la bibliographie en ligne :

«La préparation d'une bibliographie est remplie de problèmes ardues, tant sur le plan technique que théorique. L'ordinateur est un excellent outil pour faciliter la tâche du bibliographe, mais il ajoute ses propres problèmes au processus. » (Shirinian, *Problems of Bibliographic Control in Armenian Studies: Computers and Armenian Bibliography*, p. 83)

Shirinian fait référence à six paramètres pour la bibliographie : la cueillette, l'entrée et la maintenance. En discutant de l'âge de l'électronique, Shirinian y ajoute l'accessibilité, la facilité de recherche et l'agencement des données en tant qu'éléments théoriques nécessaires (p. 83-84). Les deux derniers sont d'importants « avantages » de l'espace électronique (*Problems*, p. 89). Selon Shirinian, les opérateurs de recherche logiques et relationnels mettent en valeur l'espace électronique, parce que l'information dynamique qui est produite, « lorsque la page glacée disparaît » (Sidler), met de l'ordre dans le chaos (Hackman). Les opérateurs booléens en sont une conséquence opportune. « Si une recherche donne des résultats insatisfai-

sants, il est facile et rapide de tenter une approche différente » (Shirinian, *Problems*, p. 89). « L'élément [théorique] » qui soutient la discussion de Shirinian « met l'accent sur le concept de l'agencement » (p. 89). Les banques de données relationnelles, comme MySQL, permettent un « agencement virtuel » (p. 91) des données

bibliographiques qui est impossible avec la page imprimée : « Étant donné que la bibliographie imprimée est statique et linéaire de par son agencement, tandis que le savoir et l'information enregistrés sont dynamiques et multidimensionnels, la bibliographie imprimée n'est pas assez flexible pour accepter les ajouts d'information ni la signification changeante des relations entre les publications. » (p. 90-91)

La page électronique peut réarranger et afficher dynamiquement les données « à une vitesse inégalée par l'imprimé » (p. 89). Dans le cas de la base de données de la CCL, une banque de données relationnelle était cruciale pour travailler avec un système complexe de sujets bibliographiques. Alors que ces sujets jouaient un rôle de premier plan dans la publication de 2001, cette façon d'agencer les données bibliographiques n'a pas été oubliée lorsque le site Web de 2002 a été créé. Pour le « spécialiste du sujet » (p. 90), la bibliographie multidimensionnelle permet de voir les données d'une manière substantiellement différente : « L'agencement logique des entrées sous des vedettes-matières et l'agencement logique de ces vedettes-matières elles-mêmes, illustrent bien ce système de relations. L'agencement par sujet aide l'utilisateur de bibliographie en regroupant les documents apparentés, lui permet de voir d'un simple coup d'œil la documentation qui a été produite sur un sujet donné et lui donne un aperçu des forces et faiblesses des études et des recherches sur ce sujet. » (p. 90)

Finalement, la dimension théorique la plus avantageuse de la bibliographie en ligne « centralisée » est que « plusieurs personnes peuvent y avoir accès, à grande distance, à l'aide d'un modem » (p. 89). Bien qu'une telle affirmation puisse sembler banale en 2003, l'idée de Shirinian sur la connectivité doit être considérée dans le contexte de 1992. Le fait que depuis 1992 Internet ait provoqué une explosion de l'information à l'échelle mondiale sert à appuyer la perspective de Shirinian. Car jusqu'alors, la connexion « par modem » à des bases de données en ligne se traduisait par des « tableaux d'affichage privés et des catalogues de bibliothèque » (Shirinian, 22 avril 2003). Jusqu'alors réservé aux universités, le CERN : Laboratoire européen de physique des particules (<http://www.cern.ch/>) ne lançait le World Wide Web publiquement qu'en 1992. L'utilisation du Web pour disséminer des quantités imprévues d'information, à l'aide d'une technologie

qu'O'Reilly appelle « *infoware* », était encore au stade embryonnaire. Notons que 1992 marquait plusieurs premières dans l'évolution de l'*infoware*. Pizza Hut, qui commençait en 1992 à vendre ses pizzas sur le Web et First Virtual, la toute première cyberbanque (Kristula), étaient des pionniers technologiques, les précurseurs des projets *Open Source* d'aujourd'hui.

En conclusion partielle, la publication d'une bibliographie en ligne comporte aussi des embûches. Shirinian souligne que « *les bibliographies sur le Web s'écroulent lorsque des versions [en ligne] périmées mènent à un manque général d'autorité* » (For George Shirinian, 15 avril 2003). La transformation en ligne du travail savant présente la même problématique pour Dyck et Siemans qui parlent de « *la légitimité de la publication virtuelle* » dans la revue en ligne *Early Modern Literary Studies* de l'Université d'Alberta (en ligne, p. 2). Shirinian admet que la bibliographie imprimée est encore nécessaire, si « *elle fait preuve d'un certain niveau de qualité et d'autorité* » (For George Shirinian, 15 avril 2003). Bien que plusieurs « *publications à compte d'auteur* » existent, Shirinian souligne qu'elles sont « *facilement reconnaissables* » et n'annulent pas l'importance des bibliographies imprimées « *de grande qualité... certifiées* » (Shirinian).

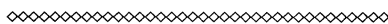
LE MODÈLE OPEN SOURCE : UNE NOUVELLE MENTALITÉ DE RECHERCHE

« *Afin de rendre l'information aussi utile que possible, l'idéal serait que la bibliographie soit accessible à distance, par modem, à un large groupe d'utilisateurs.* » (Shirinian, Problems, p. 97)

L'adoption d'un progiciel *Open Source* en ligne a créé un paradigme de recherche entièrement nouveau pour l'équipe de la Bibliographie CCL. L'interface *MySQL* en ligne *phpMyAdmin*, est devenue l'outil bibliographique pour la collecte, l'entrée et la maintenance des données. Cette méthodologie est drastiquement différente des procédures de travail des années 1988 à 2001. À cette époque, les chercheurs entraient les données directement dans la base de données principale *NBDOS* sous la supervision de la bibliothécaire du projet et sur son propre ordinateur. L'entrée des données était surveillée, non pas pour des raisons de sécurité, mais plutôt parce que la banque de données *DOS* tombait souvent en panne. Malgré les mises à niveau des disques et de la mémoire, *NBDOS* n'avait pas été conçu pour supporter la quantité de données

L'interface

phpMyAdmin permet un système hiérarchique des comptes d'utilisateurs. À mesure qu'avancait le projet de la CCL, une pyramide inversée de sécurité s'est développée.



qu'il contenait dès 2001. Par contraste, l'utilisation d'une banque de données en ligne créait une nouvelle mentalité et de nouvelles procédures de travail.

Depuis que le serveur en ligne *Open Source* est entré en service en 2002, les chercheurs travaillent indépendamment. On leur a alloué des comptes *phpMyAdmin* individuels et ils entrent les données depuis leur propre poste de travail. Lorsque le bibliothécaire en chef approuve une entrée, elle est ajoutée à la base de données principale et si elle est rejetée, elle est envoyée dans la base de données des rejets. À notre avis, le premier avantage à pouvoir accéder à une banque de données en ligne a été de fournir une certaine mobilité. Les chercheurs peuvent maintenant entrer les données directement de la bibliothèque ou de leur domicile. Cependant, le concept de l'accès en ligne a également élargi nos horizons de recherche sur le plan national. L'utilisation de *phpMyAdmin* et *MySQL* permettait un nouveau partenariat avec l'Université Simon Fraser, en Colombie-Britannique. En octobre 2001, le docteur Kathy Mezei se joignait à la Bibliographie CCL comme l'un de ses chercheurs principaux (*Minutes of Biblio Meeting, October 1, 2001*). Ce nouvel apport à l'équipe de la CCL signifiait que l'interface *phpMyAdmin* serait de nouveau testée. Les chercheurs en Colombie-Britannique pourraient-ils utiliser *phpMyAdmin* avec un minimum de supervision? La réponse était « oui ». Dès juin 2002, après quelques explications rudimentaires sur la façon dont fonctionnait *phpMyAdmin*, le docteur Mezei et son assistant ont pu commencer à fournir des données à partir de leur compte *MySQL*.

L'interface *phpMyAdmin* permet un système hiérarchique des comptes d'utilisateurs. À mesure qu'avancait le projet de la CCL, une pyramide inversée de sécurité s'est développée. Au premier niveau, les chercheurs ont accès en lecture seule à la base de donnée principale ainsi qu'aux comptes des autres chercheurs. Au deuxième niveau, les chercheurs principaux ont accès en lecture seule aux comptes de tous les chercheurs ainsi qu'à la banque de données principale. Au troisième niveau, le bibliothécaire en chef a accès en lecture/écriture à tous les comptes. À la racine, l'analyste/technicien a le plein accès et gère le système de permission d'accès. Pour assurer la continuité, ses homologues des STI peuvent, si nécessaire, remplacer le technicien de la CCL.

Du côté technique du projet, le technicien de la CCL se voyait lui aussi adopter une nouvelle mentalité de recherche. Son travail s'élargissait à celui d'analyste. Non seulement il devenait un programmeur *PHP*, mais il était également l'administrateur de système du

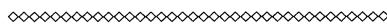
site Web, en plus d'être le premier administrateur de la base de données MySQL. Aussi, lorsque les STI ont installé le serveur Internet RedHat, le technicien de la CCL avait également la responsabilité de fournir des comptes d'utilisateurs pour le serveur principal de la CCL. Au début de 2003, certains membres du personnel de la CCL ont manifesté le désir de prendre un cours de programmation en PHP et MySQL, ce qui a nécessité un espace de travail sur www.CompCanLit.ca.

Si l'installation d'un serveur *Open Source* allait remplacer une myriade de technologies variées, elle créait également de nouvelles attentes, comme Masson nous le rappelle (p. 2). On s'attendait à ce que cet investissement technologique puisse résoudre d'autres problèmes. Au départ, les chercheurs avaient besoin d'un filtre de base de données (<http://ccl.flsh.usherb.ca/db/tools/filter/>) pour importer les données d'un autre format provenant d'autres bases de données disponibles en ligne à partir de la bibliothèque de recherche de l'Université. De plus, les chercheurs demandaient une fonction améliorée d'impression d'écran (<http://ccl.flsh.usherb.ca/print/>) pour rechercher et imprimer les entrées de leur base de données individuelle. Une telle caractéristique n'est pas disponible dans *phpMyAdmin*. Le bibliothécaire de la CCL avait également besoin de listes d'experts (http://ccl.flsh.usherb.ca/db/tools/index_scholars.php). Nous avons déjà mentionné que les chercheurs principaux avaient exigé une fonction sur mesure pour exporter les données MySQL vers *NBWin* (<http://ccl.flsh.usherb.ca/db/tools/nb/>) (*Minutes of Biblio Meeting, January 21, 2001*).

ATTENTES FUTURES

Si le développement d'un serveur *PHP-MySQL* pouvait remplacer les fonctions d'autres logiciels, il devait également être en mesure de produire en 2005 une édition papier en format « prêt-à-photographier ». La publication 2001 de Sirois et al. nous a donné des leçons inestimables quant aux pertes de temps attribuables au double emploi. La publication 2001 était éditée en *WordPerfect*, ce qui impliquait une réédition en *NBWin*, puis en *MySQL* pour que la base de données principale soit à jour avec la publication papier. De plus, les champs de *MySQL* devaient être restructurés pour pouvoir reproduire les index générés par *WordPerfect*. Avec la perspective d'une nouvelle publication en 2005, l'équipe de la Bibliographie est en train d'étudier diverses possibilités d'utiliser *MySQL* pour générer un produit « prêt-à-photographier ».

Le projet de la CCL possède une structure Open Source qui a su faire ses preuves en tant que solide modèle pour la recherche et pour la gestion du temps, en plus d'offrir une courbe d'apprentissage acceptable.



L'objectif pour la publication 2005 est de pouvoir faire tout le travail d'édition dans *MySQL* et d'utiliser *PHP* pour faire la composition. En fait, *PHP* peut être programmé pour exporter les données en RTF (*Rich-Text Format*), un format universel de traitement de texte que *Word*, *WordPerfect* et tous les logiciels d'édition peuvent importer (<http://ccl.flsh.usherb.ca/db/tools/rtf/>). Pour 2005, la seule modalité sera de programmer *PHP* dynamiquement pour qu'il produise une bibliographie en format *MLA (Modern Languages Association)*.

Il est aussi simple d'utiliser le RTF que de générer des pages en HTML. Cependant, le RTF permet une meilleure manipulation des marges, des fontes et de l'espacement des lignes, ce qui est nécessaire à un produit « prêt-à-photographier ». Une autre possibilité serait d'exporter en format de lecture *Adobe Acrobat*. Dans les deux cas, il sera nécessaire de créer une base de savoir permettant de produire la Bibliographie 2005 en style *MLA* « sans erreur de syntaxe » à l'aide de *MySQL*.

Le débat sur le remplacement des technologies commerciales par *Open Source* se poursuit encore. Cependant, l'exemple du projet de la Bibliographie CCL a beaucoup de potentiel pour d'autres projets en Lettres. Le choix d'un serveur *Open Source* a su centraliser notre technologie, tout en lui permettant d'être polyvalente. Présenté en décembre 2001, démontré en janvier 2002 et ratifié le mois suivant, le projet de la CCL possède une structure *Open Source* qui a su faire ses preuves en tant que solide modèle pour la recherche et pour la gestion du temps, en plus d'offrir une courbe d'apprentissage acceptable. Qui plus est, cette technologie offre un potentiel de publication. Les fruits de l'*OSI* nous ont permis de remplacer une quantité de logiciels commerciaux par une interface *Open Source* en ligne *MySQL*, avec un minimum de gadgets Internet.

Nul doute que le modèle *Open Source* soit devenu une solution viable pour le projet de la Bibliographie CCL. Toutefois, la présence d'un technicien, ne serait-ce que sur une base consultative, est une nécessité malgré une technologie facilement utilisable par le chercheur. En effet, le modèle de travail actuel confine la technologie à un rôle d'arrière-plan : il s'agit d'un avantage pour les chercheurs qui préfèrent ne pas s'encombrer de gadgets technologiques. Leur seul outil est devenu *phpMyAdmin* et le navigateur Web leur seul logiciel. Pour le projet de la CCL, le facteur décisif quant à l'adoption d'une méthodologie *Open Source* ne se traduisait pas en termes budgétaires, mais plutôt en termes de potentiel académique.

Il existe beaucoup de logiciels commerciaux prêts à utiliser pour qui désire emprunter une autre voie que celle du PHP, mais leur réseau de soutien n'appartient pas à la communauté des développeurs. Chaque jour, non seulement les utilisateurs d'*Open Source* dépendent de l'effort académique collectif, mais ils y contribuent aussi. À leur tour, leurs projets de recherche aident ceux de milliers d'autres programmeurs, technologues et bien sûr, de chercheurs dans le domaine des Lettres. ♦

SOURCES CONSULTÉES

- 6.8 MySQL Full-Text Search. MySQL Reference Manual. 15 avril 2003. <http://www.mysql.com/doc/en/Fulltext_Search.html>.
- The Approved Licenses. 15 avril 2003. <<http://www.opensource.org/licenses/>>.
- Bowen, William R. *Iter : Where Does the Path Lead? Early Modern Literary Studies* 5.3. Special Issue 4 : 2.1-26. 15 avril 2003. <<http://www.shu.ac.uk/emls/05-3/bowiter.html>>.
- Castell, Tracy. 1997. *Maintaining Web-Based Bibliographies : A Case Study of Iter, the Bibliography of Renaissance Europe*. Présenté à l'American Society for Information Science, (nov). 15 avril 2003. <<http://www.asis.org/annual-97/castell.htm>>.
- Chapter 1. Introduction. PHP Manual. 15 avril 2003. <<http://www.php.net/manual/en/introduction.php>>.
- Chapter 2. A Simple Tutorial. PHP Manual. 15 avril 2003. <<http://www.php.net/manual/en/tutorial.php>>.
- DiBona Chris, Sam Ockman, et Mark Stone. 1999. *Introduction. Sam Ockman DiBona Chris et Mark Stone éd. Open Sources : Voices from the Open Source Revolution*. Sebastopol, CA : O'Reilly, 1-18. 15 avril 2003. <<http://www.oreilly.com/catalog/opensources/book/intro.html>>.
- Dyck, Paul, et R.G. Siemens. *The Janus-Face of Early Modern Literary Studies : Negotiating the Boundaries of Interactivity in an Electronic Journal for the Humanities*. Early Modern Literary Studies 5.3.Special Issue 4 : 4.1-20. 15 avril 2003. <<http://purl.oclc.org/emls/05-3/dslwemls.html>>.
- The GNU Project. 15 avril 2003. <<http://www.gnu.org/gnu/thegnu-project.html>>.
- Kristula, Dave. *The History of the Internet*. 15 avril 2003. <<http://www.davesite.com/webstation/net-history.shtml>>.
- Lerdorf, Rasmus. *Re : PHP Licensing*. Courriel à John Taylor-Johnston. 7 avril 2003.
- Hackman, Martha L.. *The Practical Bibliographer*. Cité dans Shirinian, George. *Problems of Bibliographic Control in Armenian Studies : Computers and Armenian Bibliography*. Journal of the Society for Armenian Studies 6 (1992-1993) : 83-98. Englewood Cliffs, NJ : Prentice, 1970.
- Minutes of Biblio Meeting, January 21 2002. Comparative Canadian Literature Bibliography Group. 15 avril 2003. <<http://ccl.flsh.usherb.ca/minutes/Biblio20020121.htm>>.
- Minutes of Biblio Meeting, October 1, 2001. Comparative Canadian Literature Bibliography Group. 15 avril 2003. <<http://ccl.flsh.usherb.ca/minutes/Biblio20011001.htm>>.
- Netcraft Survey. 15 avril 2003. <<http://www.netcraft.com/survey/>>.
- The Open Source Definition. 15 avril 2003. <<http://www.opensource.org/docs/definition.php>>.
- O'Reilly, Tim. *Hardware, Software, and Infoware. Sam Ockman DiBona Chris et Mark Stone éd. Open Sources : Voices from the Open Source Revolution*. Sebastopol, CA : O'Reilly, 1999. 189-96. 15 avril 2003. <<http://www.oreilly.com/catalog/opensources/book/tim.html>>.
- Perens, Bruce. 1999. *Hardware, Software, and Infoware. Sam Ockman DiBona Chris et Mark Stone éd. Open Sources : Voices from the Open Source Revolution*. Sebastopol, CA : O'Reilly, 171-88. 15 avril 2003. <<http://www.oreilly.com/catalog/opensources/book/perens.html>>.
- Search Options. 15 avril 2003. <<http://ca.search.yahoo.com/search/ca/options>>.
- Shirinian, George. *Problems of Bibliographic Control in Armenian Studies : Computers and Armenian Bibliography*. Journal of the Society for Armenian Studies 6 (1992-93).
- _____. *Re : For George Shirinian*. Courriel à John Taylor-Johnston. 21 avril 2003.
- _____. *Re : For George Shirinian*. Courriel à John Taylor-Johnston. 15 avril 2003.
- Sirois, Antoine et al. 2001. *Introduction*. Introduction par Gregory J. Reid et Maria van Sundert. *Bibliography of Comparative Studies in Canadian, Québec and Foreign Literatures/Bibliographie d'études comparées des littératures canadienne, québécoise et étrangères 1930-1995*. Sherbrooke : Productions G.G.C. 293 p.
- Van Sundert, Maria. *Re : Modifications apportées au site Web*. Courriel à Maria van Sundert. 2 juillet 1997.