

LE LIVRE UNIVERSITAIRE NUMÉRIQUE EN SCIENCES HUMAINES ET SOCIALES :

POUR UN FOYER D'EXPERTISE
ET UNE INFRASTRUCTURE DE STATURE MONDIALE

Gérard Boismenu,
Émilie Paquin,
Martin Boucher,
Dominic Forest,

avec la collaboration de
Michael Eberle-Sinatra

Pour le
Conseil de recherche en sciences humaines et sociales du Canada

1^{er} décembre 2010

TABLE DES MATIÈRES

Introduction	7
Les contours de la question	7
Concevoir l'édition numérique de livres	7
Penser aux attentes du lectorat.....	8
Traiter de la volonté et des contraintes des éditeurs	8
Notre démarche	8
1- Tirer partie du numérique.....	11
Le numérique au service des sciences humaines	11
Et si on avait déjà un passé ! 10 ans de livres numériques	12
2- Les modèles analysés.....	13
La recherche d'une certaine exemplarité.....	13
L'agrégateur et l'entrepôt numérique	13
Offre de livres numériques : les commerciaux en tête	14
Des projets éditoriaux variés.....	14
Diversité des modèles d'affaires et des modalités d'accès	15
Modèle d'affaires de netLibrary.....	16
Modèle d'affaires de MyiLibrary	17
Modèle d'affaires de Cairn	17
Modèle d'affaires de Humanities E-Book	17
Modèle d'affaires de BiblioVault.....	18
Modèle d'affaires de OAPEN	18
Processus d'édition numérique : le grand bazar	18
Services des plateformes	20
Conclusion	20

3- Ce que veulent bibliothécaires, chercheurs et étudiants.....	23
JISC National E-books Observatory Project	23
HighWire Press Librarian eBook Survey and Analysis	24
How Readers Navigate to Scholarly Content.....	24
Conclusion	24
4- Enquête auprès des presses universitaires	27
Les presses universitaires canadiennes : une offre numérique abondante	27
Production de livres numériques : variété des pratiques.....	27
Élargir les marchés grâce aux canaux de diffusion déjà existants.....	28
Conclusion	28
5- La chaîne de traitement du livre numérique	31
La pluralité des formats de diffusion et des supports de lecture.....	31
Plusieurs formats, plusieurs niveaux de services.....	32
La qualité des métadonnées, un enjeu pour la diffusion et la distribution.....	33
XML, un format orienté services	35
Conclusion	37
6- Le processus d'édition numérique	39
Les types de traitement du livre numérique	39
Une analyse des coûts de production : à traitements différents, coûts différents.....	40
Traitement en lot (numérisation rétrospective)	41
Traitement minimal et complet, du simple au triple	41
XML, entre l'intérêt et la réalité des canaux de distribution.....	42
Conclusion	42

7- Les systèmes documentaires numériques comme laboratoires de recherche : enjeux et développements	45
Les thématiques.....	46
Les modèles d'imbrication.....	47
La technologie comme outil	47
La technologie comme objet d'étude	48
La technologie comme moyen d'expression	48
La technologie comme laboratoire exploratoire de recherche	48
Conclusion	49
8- Remarques finales et propositions	51
Il y a d'abord eu les revues.....	51
Il y a numérique et numérique	51
Les éléments d'une énigme	52
De la modestie en prime	54
Une entreprise concertée	55
Arrimer une dimension recherche	55
Le partenariat comme démarche stratégique.....	56
Annexe 1 : Agrégateurs et entrepôt	61
Entrepôt numérique.....	76
Agrégateurs	78
Entrepôt numérique.....	78
Annexe 2. : Canevas d'entretien avec les dirigeants des presses universitaires canadiennes...	79
Annexe 3 : Projets de recherche issus de différentes institutions	81
Annexe 4 : An Annotated Review of Centers and Large Projects listed in the CenterNet website	102

Introduction

Au moment où le numérique occupe de plus en plus d'espace dans la communication scientifique, il est utile de s'interroger sur la capacité d'implantation (formes organisationnelles et modèles d'affaires) d'un système de publication numérique issu de la communauté universitaire en mesure d'éditer, de mettre en valeur et d'assurer la présence internationale du livre universitaire en sciences humaines et sociales.

Les contours de la question

Au Canada, la publication numérique de revues savantes et leur diffusion sur des plateformes a trouvé des réponses porteuses et dont les effets sont structurants (pensons à <www.erudit.org> et à <www.synergiescanada.org>), aussi bien pour le fonds publié que pour leur diffusion et consultation internationales.

Le livre, comme genre éditorial, est cependant peu exploité en dehors de grands agrégateurs commerciaux (comme *EBSCO*). Pourtant, le livre numérique est devenu un « objet » voué à de grands développements, surtout depuis la commercialisation de tablettes de lecture performantes et l'intérêt décuplé des acheteurs institutionnels pour le livre numérique, telles les bibliothèques.

L'heure a sonné pour la mise en place d'une infrastructure de qualité en mesure de soutenir le livre universitaire numérique et touchant toutes les étapes de la chaîne de l'édition, de la publication à la diffusion.

Cette question n'est pas particulière au Canada, mais elle mérite un traitement qui tienne compte de l'état des lieux au Canada, des expertises disponibles, de la volonté des éditeurs, de la capacité organisationnelle et des conditions d'implantation d'une infrastructure

pertinente. En raison de l'accélération de projets dédiés au livre aussi bien aux États-Unis qu'en Europe, tout autant que de la diffusion des équipements d'appropriation des contenus numériques (à commencer par les tablettes de lecture), il est impératif que la communauté universitaire canadienne se dote des moyens pour occuper une place dans la diffusion du savoir, sinon le champ va être occupé et structuré de telle manière que les barrières à l'entrée deviendront infranchissables.

Toute initiative doit pouvoir s'appuyer sur une bonne compréhension des règles de l'art en édition numérique, des acteurs impliqués, des conditions de réussite et de la capacité de réunir les moyens d'implantation. Une synthèse des connaissances de l'économie numérique du secteur du livre universitaire est indispensable. C'est cette question qui est au cœur de ce texte.

Concevoir l'édition numérique de livres

L'édition dans l'environnement numérique suppose une série de choix. Le paradigme numérique va de pair avec plusieurs types d'appareils et services. Le lecteur, qu'il soit individuel ou institutionnel, peut privilégier l'un ou l'autre de ces modes d'accès au texte. Cette diversité est appelée à se confirmer. Le livre numérique devra vraisemblablement prendre plusieurs formes de matérialisation pour rejoindre son lecteur.

L'édition du livre numérique doit porter une attention aux formats, d'abord en pensant aux formats de restitution, mais aussi aux formats d'encodage en considérant la diversité des médias, la conservation à long terme, l'interopérabilité entre plateformes et l'adaptation aux différents types d'écran (*reflowable text*, image), mais aussi aux supports matériel que sont les diverses tablettes de lecture. Ajoutons que la qualité des

métadonnées pour décrire l'unité documentaire est cruciale. Il faut penser à ces éléments constitutifs de l'édition numérique de livres et faire les arbitrages qui s'imposent en mesurant leur portée.

Penser aux attentes du lectorat

Les différents services ou plateformes qui diffusent des livres numériques ont reproduit largement des pratiques qui étaient liées à la mise en circulation de l'imprimé en cherchant à limiter la multiplication des usagers pour la même licence d'utilisation.

D'où une série de restrictions dans la consultation et l'utilisation des livres numériques. Même des projets issus de la communauté universitaire ou apparentés s'inscrivent dans cette démarche.¹ Pourtant l'opinion des utilisateurs va à l'encontre de ces restrictions. Il est utile d'en prendre acte.

S'il y a un écart appréciable entre les pratiques des éditeurs et les attentes des lecteurs (individuels et institutionnels), la question se pose : comment combler cet écart tout en permettant aux éditeurs de développer un modèle d'affaires « soutenable » ?

Traiter de la volonté et des contraintes des éditeurs

Les éditeurs universitaires ou apparentés (*responsible publishers*) conduisent leurs activités avec la préoccupation d'assurer un recouvrement des coûts et même de développer une marge bénéficiaire. La position défensive concernant l'utilisation des livres numériques est-elle la seule qui peut

permettre une rentabilité minimale ? Il est vrai que nous sommes dans une période de transition et cette dernière passe par un processus d'hybridation, alors même que nous sommes toujours dans un environnement marqué par l'imprimé.

Avec un certain recul, ce n'est pas sans rappeler certaines pratiques que l'on a voulu mettre en place au moment de l'avènement des revues savantes numériques. Or, dans ce domaine, un modèle d'affaires a été trouvé sans que l'on s'en tienne aux restrictions de tout genre. Aujourd'hui, les revues savantes pourraient servir de référence à plusieurs égards pour les livres.

Notre démarche

Nous traitons de quatre enjeux fondamentaux : la constitution (conditions et contenus) de la collection de livres numériques et son évolution ; le processus de publication numérique adapté au livre ; le développement d'un processus de publication de haute qualité mais à faible coût d'opération ; l'élaboration d'un modèle économique viable situé dans une période de transition.

Pour apporter des propositions répondant à ces enjeux, quatre volets ont structuré notre étude.

D'abord, nous procéderons à un inventaire des principaux projets porteurs (États-Unis, Europe) pour le livre numérique universitaire. L'objectif est de tirer le maximum de renseignements portant sur la constitution du corpus et son mode de sélection, sur le modèle organisationnel et l'unité institutionnelle, sur les caractéristiques (choix) du processus d'édition numérique et sur le modèle économique.

Il importe également de faire enquête auprès des principaux éditeurs universitaires

¹ Pensons aux plateformes suivantes : <http://www.humanitiesebook.org>, <http://www.oxfordscholarship.com>, http://www.cdlib.org/services/publishing/ucpress_ebooks.html ; ou même à <http://lib.mylibrary.com/>.

canadiens (*University of Toronto Press, McGill-Queen's University Press, UBC Press, Presses de l'Université de Montréal et Presses de l'Université Laval*) concernant leur appréciation sur l'avenir du livre numérique, les expériences qui ont été menées, la place du livre numérique dans leur plan d'activités, leur expertise en édition numérique, les conditions économiques à réunir pour un engagement dans ce champ, leur capacité de participer à un projet multiéditeurs (consortium ou autre), la volonté de faire eux-mêmes ou de déléguer leurs livres numériques, etc. Cela se fera par des entrevues approfondies.

Ensuite, en s'appuyant sur les bonnes pratiques en matière d'édition numérique et sur les standards assurant à la fois la polyvalence recherchée et une adaptabilité évolutive, il est possible de définir les paramètres d'une chaîne de publication numérique du livre qui se situe dans un environnement de rareté de ressources. En d'autres termes, trouver les bons arbitrages entre les caractéristiques les plus porteuses en termes de qualité et d'adaptabilité et la recherche d'un processus de publication dont la structure de coût n'est pas trop élevée (coût d'apprentissage et coût d'opération, notamment).

Pour terminer, le développement de plateformes diffusant un corpus universitaire ne se limite pas à la diffusion : c'est en soi une condition de développement d'une infrastructure de recherche, comprise en termes de laboratoire supportant de nouvelles recherches et de nouveaux protocoles. Cela est particulièrement central dans le domaine des sciences humaines, il faut donc apprécier cette dimension.

1- Tirer partie du numérique

Le nouveau monde numérique bouleverse les pratiques dans tous les domaines de la vie. Le milieu universitaire et la diffusion de la connaissance n'y échappent pas. Si cela peut être vu comme une menace par quelques-uns, il est sans doute plus approprié d'y voir une occasion de changer les pratiques et les façons de travailler pour l'avancement de la connaissance et pour l'amplification de son impact. Les sciences humaines et sociales ne sont pas à l'écart d'un tel mouvement. Le numérique est appelé à s'investir dans toutes les ramifications de leur avancement et rayonnement, et il faut en saisir les retombées, notamment dans le domaine de la production et la diffusion de la connaissance.

Le numérique au service des sciences humaines

La capacité de rayonnement des sciences humaines et sociales n'est pas en baisse en raison du numérique. Voilà plus de 40 ans que les ventes des livres en sciences humaines et sociales sont en perte de vitesse. Mais comme l'évoque avec humour Colin Steele, il semble bien que le livre savant ait été, tout comme la dynastie des Habsbourg, en déclin depuis toujours.²

Hausse des prix des abonnements aux périodiques numériques, augmentation continue du nombre de titres publiés, évolution rapide des technologies de l'information, pressions exercées sur les presses universitaires par les institutions d'enseignement, toutes ces raisons sont évoquées pour expliquer la délicate situation

de l'édition du livre en sciences humaines et sociales.³

Il n'en reste pas moins que la publication de livres demeure un étalon important du parcours universitaire en sciences humaines et sociales, et que le capital symbolique entourant les maisons d'édition pèse encore de tout son poids dans le processus de reconnaissance institutionnelle qui mène de la permanence à la titularisation. Dans ce contexte, le recours au numérique apparaît comme un moyen à explorer pour accroître le rayonnement des publications, diminuer les coûts de production et de gestion des stocks, et libérer le livre des limites de l'impression sur papier.

La rapidité avec laquelle les projets dédiés au livre numérique apparaissent, aux États-Unis aussi bien qu'en Europe, qu'ils soient tarifés ou en accès libre, témoigne éloquemment de cette nouvelle réalité. De même, l'engouement pour les nouvelles tablettes de lecture, *iPad* en tête, laisse supposer que l'heure du livre numérique est arrivée. Entreprises commerciales et sociétés sans but lucratif déjà engagées dans la diffusion de contenus numériques universitaires offrent toutes, ou le feront prochainement, des bouquets de livres numériques auxquels sont associés des services plus ou moins sophistiqués. L'offre de contenus numérisés se diversifie et le livre savant en fait résolument partie.

Bien que le livre numérique ne semble pas encore en voie de remplacer définitivement l'imprimé, les forces à l'œuvre dans le champ de l'édition savante exigent de la communauté universitaire canadienne qu'elle se dote des

² Steele, Colin, « Scholarly Monograph in the 21st Century : The Future More Than Ever Should Be An Open Book », *The Journal of Electronic Publishing*, vol. 11, no 2, Spring 2008.

³ Williams, Peter et al., « The Role and Future of the Monograph in Arts and Humanities Research », *Aslib Proceedings : New Information Perspectives*, vol. 61, no 1, 2009, pp. 67-82.

moyens nécessaires à la diffusion numérique du savoir. À défaut de quoi, le champ sera occupé et structuré de telle manière que les barrières à l'entrée deviendront infranchissables dans quelques années

C'est dans cet esprit que notre étude a été entreprise, afin de circonscrire les grandes lignes d'une stratégie numérique canadienne en matière de publication de livres en sciences humaines et sociales.

Et si on avait déjà un passé ! 10 ans de livres numériques

Les premières initiatives en matière de publication et de diffusion de livres numériques remontent à la fin des années 1990.

L'air du temps était alors marqué par un optimisme certain et les entreprises commerciales flairant le bon coup offraient aux éditeurs universitaires des services de numérisation gratuits afin d'inciter les plus réticents à opérer un passage vers le numérique.⁴ L'éclatement de la bulle technologique vint rapidement mettre fin à ces généreuses politiques, mais c'est tout de même ainsi que certaines plateformes, dont *netLibrary*, ont pu constituer au tournant des

années 2000 une imposante collection de livres numériques à l'intention d'une clientèle institutionnelle.

Le succès de ces premières initiatives fut relatif, en raison essentiellement du peu de nouveautés numérisées et du redoublement des collections numériques et imprimées. Ce n'est que plus récemment que nous avons vu l'écllosion d'une offre de livres numériques riche et variée.

Au cours des dernières années, l'adoption massive des périodiques numériques par la communauté universitaire, le développement rapide de l'informatique, d'Internet ainsi que des appareils de lecture ont préparé le terrain pour une évolution rapide aussi bien des collections de livrels que de leur utilisation.

Cela est particulièrement vrai dans le domaine de la lecture savante, puisque, comme le souligne Hubert Guillaud « le livre papier n'est pas fait pour travailler. Je ne peux rien faire avec : ni prendre de notes (sans recopier), ni chercher dans son contenu (retrouver une définition...). » Ainsi, toujours selon Hubert Guillaud, « [à] l'âge de l'accès, on ne va plus avoir accès au produit, mais au contenu. »⁵

L'intérêt du livre numérique réside, en ce sens, dans les fonctionnalités des plateformes depuis lesquelles nous y avons accès. Ce qu'ont très bien compris les gros joueurs de l'édition numérique, au rang desquels figure, par exemple, *EBSCO*, qui rivalisent en ce moment d'ingéniosité dans l'objectif d'offrir des services novateurs, au plus près des besoins particuliers des différentes clientèles.

⁴Thompson, John B., *Books in the Digital Age. The Transformation of Academic and Higher Education Publishing in Britain and the United States*. Cambridge, Polity Press, 2005, p.335. « The temptation to experiment was accentuated by the proliferation of third-party organizations which were springing up in the publishing field throughout the second half of the 1990's. Well funded by venture capitalists who were knocking on the doors of publishers in search of content which they could license for use in their prospective web-based businesses. The approaches were often linked with financial incentives to participate – for example, a significant amount of content would be digitized free of charge if publishers were willing to commit to an as yet untested business venture.»

⁵ Guillaud, Hubert, « Qu'est-ce qu'un livre numérique? », in *Read/Write Book. Le livre inscriptible*, sous la dir. de arin Dacos. <<http://cleo.revues.org/147>>

2- Les modèles analysés

Les plateformes retenues pour étude présentent une offre de livres numériques substantielle et intéressante en sciences humaines et sociales ainsi qu'en arts. Les initiatives trop récentes ou spécialisées en sciences, technologie et médecine (STM) ont été dès le départ écartées.

La recherche d'une certaine exemplarité

Afin d'assurer une représentativité des modèles, nous avons exploré des projets à vocation commerciale (*netLibrary* et *MyiLibrary*) et d'autres sans but lucratif, pilotés par des regroupements de sociétés savantes (*Humanities E-Book*), d'éditeurs indépendants (*Cairn*) ou issus de la communauté universitaire (*BiblioVault*). Une importante initiative européenne en accès libre a également été évaluée (*OAPEN - Open Access Publishing in European Networks*).

Compte tenu de la nature du marché universitaire et des caractéristiques de ses clientèles, nous n'avons pas cru bon de recenser les services des librairies en ligne ni les structures regroupant des ouvrages d'édition générale (la plateforme française *Numilog* par exemple).

Notre choix s'est porté essentiellement sur des projets de type « agrégateur » (*netLibrary*, *MyiLibrary*, *Cairn*, *Humanities E-Book*, *OAPEN*) et « entrepôt numérique » (*BiblioVault*). Cinq critères d'analyse ont balisé notre recherche : le modèle organisationnel, le projet éditorial, le modèle d'affaires, le processus d'édition numérique ainsi que les services des plateformes.

L'agrégateur et l'entrepôt numérique

D'emblée, quelques précisions s'imposent sur la mission et le statut de deux types d'organisation qui portent le livre numérique.

L'*agrégateur* regroupe des documents numériques produits par ailleurs afin d'en vendre l'accès, par abonnement ou licence, à une clientèle institutionnelle depuis une infrastructure informatique complexe offrant différents services de recherche et d'appropriation du contenu (exportation de notices, annotation des PDF, création d'un panier, etc.) Une redevance est versée aux éditeurs participants, suivant différentes formules dont les modalités restent à la discrétion des entreprises. La plupart des agrégateurs offrent des services de numérisation et de production de fichiers numériques.

L'investissement financier de l'éditeur est, dans ce modèle, réduite à sa plus simple expression, et bien que des frais soient chargés pour la création des fichiers, elle peut même s'avérer relativement lucrative. Évidemment, l'éditeur n'a pas le contrôle de l'environnement dans lequel apparaissent ses titres pas plus que sur la présentation des livres diffusés, mais leur rayonnement s'en trouve considérablement élargi de même que leur lectorat potentiel.

Les entreprises proposant des services d'*entrepôt numérique* produisent et hébergent des fichiers numériques, ils gèrent les droits numériques, distribuent les livrels et offrent, pour certains, de l'impression à la demande ainsi qu'une structure de distribution aux libraires. Les clients des entrepôts numériques sont les éditeurs, les librairies en lignes, les bibliothèques et les agrégateurs de contenus. Le modèle de l'entrepôt permet aux éditeurs d'avoir accès à une infrastructure technologique de pointe et à une main d'œuvre hautement qualifiée. Cela leur permet

d'intégrer leurs contenus éditoriaux dans les principaux canaux de diffusion selon les normes en vigueur, afin de respecter une interopérabilité maximale.

L'entreprise opérant un service d'agrégateur, tout comme celle opérant un entrepôt numérique, ne dispose pas des droits des livres diffusés et ne possède pas non plus d'exclusivité de diffusion.⁶

Offre de livres numériques : les commerciaux en tête

Comme l'illustre le tableau ci-dessous, les modèles organisationnels des plateformes de production et de diffusion de livres numériques varient fortement, bien que les regroupements d'acteurs soient, dans le secteur sans but lucratif ou indépendant, une formule largement répandue.

L'offre actuelle est principalement composée de livres de langue anglaise, bien que le projet européen *OAPEN* diffuse des documents en allemand, français, italien, néerlandais, etc., et que *Cairn* propose une collection d'ouvrages en français. L'offre de livres numériques est également beaucoup plus abondante sur les plateformes commerciales que sur celles sans but lucratif. Les plateformes commerciales ont aussi une politique de mise en ligne plus dynamique, marquée par des ajouts importants à chaque mois (5000 nouveaux titres par mois pour *MyiLibrary*) et une offre diversifiée (*netLibrary* diffuse des milliers de livres sonores).

Cela révèle l'avantage des joueurs déjà impliqués dans l'édition numérique, qui bénéficient de l'expérience acquise avec la publication de périodiques numériques, des installations technologiques nécessaires à l'édition de contenu électronique, sans compter une clientèle déjà constituée.

Des projets éditoriaux variés

À côté des ambitions généralistes des agrégateurs commerciaux (*netLibrary*, *MyiLibrary*), qui diffusent des nouveautés, des livres disponibles à la vente, des livres épuisés, passés dans le domaine public et des rééditions, les projets éditoriaux des plateformes analysées varient sensiblement.

La collection de livres numériques de *Humanities E-Book* est composée à 87 % de livres se trouvant en librairie. Le reste de la collection est représenté par des ouvrages épuisés ou passés dans le domaine public, le livre le plus ancien datant de 1820. La plateforme *Cairn* a par ailleurs décidé de spécialiser son offre en numérisant des ouvrages de références (*Que sais-je ?*, *l'État du monde*, etc.) et des collectifs disponibles en librairie. Cette stratégie a pour effet d'offrir aux éditeurs intéressés un espace de diffusion pour des documents nécessitant des mises à jour fréquentes et pour des ouvrages difficiles à commercialiser.

BiblioVault collabore avec plus de soixante-dix presses universitaires afin de produire, d'héberger et commercialiser des livres disponibles à la vente exclusivement, pour lesquels une version imprimée est vendue en librairie. Aucun livre exclusivement numérique n'est produit par cet entrepôt. Enfin, créé en 1996 par le *National Academies Press*, *OAPEN* diffuse, depuis son lancement officiel en octobre 2010, des nouveautés et des livres numérisés en accès libre (la publication du plus ancien remonte à 1969) en s'appuyant sur

⁶ Voir pour plus d'information sur ces deux modèles l'étude dirigée par Guylaine Beaudry pour le compte de l'Association nationale des éditeurs de livres. *Les enjeux de l'édition du livre dans le monde numérique*. Montréal, Association nationale des éditeurs de livres, 2008, 228 p.

un réseau de 35 partenaires.⁷ La collection ainsi constituée est évaluée par les pairs selon différents critères déterminés par les membres.

Diversité des modèles d'affaires et des modalités d'accès

De façon schématique, nous pouvons identifier deux principales modalités d'achat des collections de livres numériques : l'achat à l'unité et la constitution de collections, aussi appelées panier ou bouquet. La tarification peut s'effectuer sur le mode de l'achat pérenne avec accès illimité, sur la base d'un abonnement annuel ou d'une licence.

Toutefois, contrairement au modèle économique en vigueur dans le domaine des périodiques numériques, où les restrictions d'accès sont nulles, dans le secteur des livrets, il n'est pas rare que le nombre d'utilisateurs simultanés soit limité ou encore qu'un plafond annuel soit fixé au nombre de téléchargements.

Acheter un titre est généralement plus cher que de s'abonner à une collection, augmenter le nombre d'utilisateurs simultanés augmente proportionnellement les coûts d'abonnement, tout comme augmenter le nombre de téléchargements annuel.⁸ De plus, les institutions desservant une vaste population payent, de façon générale, plus cher leurs collections de livres numériques que celles qui ont une communauté étudiante et un corps professoral plus limités.⁹

Voilà pourquoi les informations concernant les modèles d'affaires sont si difficiles à obtenir : elles varient le plus souvent en fonction des ententes conclues entre les fournisseurs de services et les institutions documentaires.

Un rapide coup d'œil à l'offre existante révèle que les modalités d'accès ne sont pas soumises à une norme commune et que l'application de restrictions peut varier à l'intérieur d'une même plateforme, selon la volonté des acteurs impliqués. Cette situation complique la vie des bibliothécaires, qui ont pour mission de guider leurs usagers dans leur utilisation des livres numériques, et elle irrite les lecteurs qui se voient imposer des limitations sans savoir pourquoi certains documents en comportent et d'autres non.¹⁰

De façon générale, il apparaît que les modalités d'accès aux livres diffusés sur les plateformes issues de la communauté universitaire ou des maisons d'éditions indépendantes sont plus abordables, moins compliquées d'utilisation et plus transparentes que celles élaborées par les plateformes commerciales.

⁷ Pour la liste complète des institutions partenaires du réseau : http://project.oopen.org/partner_network.asp

⁸ Grigson, Anna, « Evaluating Business Models for E-Books Through Usage Data Analysis : A Case Study from the University of Westminster », *Journal of Electronic Resources Librarianship*, vol. 21, no 1, pp. 62-63.

⁹ Il est à noter que cette formule de tarification peut s'avérer inéquitable puisque la taille de la population

étudiante d'une université ne correspond pas forcément à celle des facultés et départements. Ainsi, une grande université avec un petit département de théologie se verra désavantagée lors de l'achat de documents dans ce domaine par rapport à une université moyenne disposant d'un département de théologie plus important.

¹⁰ Voir à ce propos les études de *HighWire Press* <http://highwire.stanford.edu/PR/HighWireEBookSurvey2010.pdf> et de *JISC* <http://www.jiscebooksproject.org/reports/finalreport>

Projets	Modèle organisationnel	Collections
<i>netLibrary</i>	Division d'EBSCO Publishing	§ 260 000 titres § Plus de 6000 compatibles avec <i>iPod</i> § 16 000 livres audios
<i>MyiLibrary</i>	Division d'Ingram Content Group	§ 250 000 titres § Ajout de 5000 titres par mois § 18 000 éditeurs représentés
<i>Cairn</i>	Regroupement d'éditeurs : La Découverte, De Boeck, Belin, Érès Bibliothèque nationale de France Centre national du livre Différents investisseurs institutionnels, dont Gesval	§ 1400 titres
<i>Humanities E-Book (HEB)</i>	Regroupement de plus de 20 sociétés savantes de l'American Council of Learned Societies (ACLS) University of Michigan's Scholarly Publishing Office	§ 2800 titres § Ajout de 500 titres par année
<i>BiblioVault</i>	University of Chicago Press Chicago Distribution Center Chicago Distribution Center	§ 20 000 titres
<i>OAPEN</i>	Presses de l'Université d'Amsterdam, de Leiden, de Manchester, de Lyon, de Firenze Presses du Museum Tusulanum du Danemark. Université d'Amsterdam et Université de Leiden	§ 724 titres

Modèle d'affaires de netLibrary

Les livres offerts dans *netLibrary* sont disponibles sous abonnement et en acquisition permanente (paiement unique), ce qui est appréciable lorsque les budgets d'acquisitions de livres numériques ne sont pas récurrents. Les fonctionnalités disponibles sur *netLibrary* dépendent des conditions d'achat de chaque

institution abonnée ainsi que des restrictions propres à chaque éditeur. Une très faible portion de la collection est en accès libre (3461 titres) et en français (2600 titres).

L'interface est toutefois disponible en plusieurs langues (dont le français) et la recherche pourra être effectuée depuis l'agrégateur *EBSCOhost* prochainement. Les documents sont consultables à distance ou

téléchargeables avec *Adobe Digital Editions* pour les utilisateurs de la bibliothèque.

Les éditeurs peuvent bénéficier d'un service de protection des données (DRM) avec l'application *Adobe Digital Editions*. L'impression est limitée à quinze pages par heure si l'utilisateur est connecté. De même, le copier/coller est disponible mais suivant certains critères dans un souci de respect du droit d'auteur. Un message d'avertissement apparaît lorsque le système détecte un abus possible.

Modèle d'affaires de MyiLibrary

Les options d'achat sont également flexibles sur *MyiLibrary*. Il est possible de constituer une collection par acquisition de titres seuls ou par composition d'un bouquet. Les livres peuvent être achetés avec un accès pérenne ou selon une formule d'abonnement. Cette structure commerciale comprend des offres particulières telles que des prêts interbibliothèques et des options de paiement multiples, par éditeurs ou par disciplines notamment.

Des outils simples et efficaces sont disponibles pour automatiser l'acquisition d'un nouveau contenu. Plusieurs niveaux d'accès à la plateforme sont offerts par l'entremise d'une grande variété de méthodes d'authentification et d'accès contrôlés. La formule DRM est appliquée à l'ensemble du contenu qui peut être téléchargé à raison de dix pages à la fois uniquement. L'impression est elle aussi limitée à quatre-vingts pages à la fois et un utilisateur ne peut consulter plus de quatre livres simultanément. Le copier/coller est disponible de même que les annotations, tout autant que l'ajustement à la vue et la recherche en plein texte. Un dictionnaire est intégré.

Modèle d'affaires de Cairn

Cairn offre aux institutions documentaires des bouquets thématiques ou interdisciplinaires sous forme de licences d'accès forfaitaire (le nombre de consultations étant alors illimité). Des abonnements individuels sont également proposés, mais les conditions commerciales de cette formule d'abonnement ne sont pas présentées. Il est possible d'acheter des chapitres à la pièce, ce qui coûte entre 3 et 5 €.

Les utilisateurs peuvent consulter ou imprimer les chapitres, et, éventuellement, les reproduire, pour autant évidemment que cela s'effectue dans les limites de la réglementation en vigueur (nous n'avons pas accès à plus de précisions). Les livres ne semblent pas être soumis à une gestion des droits d'accès (DRM). Il est possible de copier/coller, mais les fonctions d'annotation et de surlignage ne sont pas supportées à l'exception des termes recherchés dans les résultats de recherche. L'ajustement à la vue est offert ainsi que la recherche en plein texte.

Modèle d'affaires de Humanities E-Book

HEB ne charge aucun frais aux éditeurs pour la production de leurs fichiers numériques. Ceux-ci n'ont qu'à transmettre une copie du livre de l'ouvrage, qu'elle soit imprimée ou en format PDF. 50 % des revenus d'abonnements sont divisés entre les détenteurs de droits en fonction du nombre de requêtes que chaque titre reçoit.

Les prix sont établis en fonction de la taille de l'institution et la collection n'est pas offerte en panier personnalisable. Il est possible d'acheter une licence d'une durée de 10 ans pour tous les titres actuels et à venir. Les abonnements individuels ne sont offerts qu'aux membres de l'une des soixante-dix sociétés savantes de l'ACLS. Un service de

prêt est assuré et aucune restriction d'accès aux titres n'est imposée. Il est possible d'enregistrer et d'imprimer certains ouvrages (390 sur 2800) localement trois pages à la fois. Il est possible de copier/coller, mais pas d'annoter ni de surligner.

Un ajustement à la vue est possible selon certains pourcentages (50% à 400%) pour les pages « image » uniquement. La recherche dans le texte intégral d'un ouvrage est supportée et des liens vers les critiques du livre sur d'autres plateformes sont associés aux contenus éditoriaux (*JSTOR*, *Project MUSE*, *History Cooperative* notamment). L'impression et le téléchargement sont soumis au « *Scholarly Fair Use rights & restrictions* ».

Modèle d'affaires de BiblioVault

L'accès aux livres numériques depuis *BiblioVault* est possible sous condition d'achat uniquement. Les modalités de fixation des prix sont variables, mais le plus souvent la version numérique est équivalente ou un peu moins cher que la version imprimée. Il est parfois possible d'emprunter le livre pour une durée déterminée, à prix moindre. L'achat par téléchargement est offert sur le site des presses concernées avec consultation partielle de certaines pages du livre. La quantité de pages disponibles pour consultation varie beaucoup d'une presse et d'un livre à l'autre.

Un dispositif de protection des données (DRM) est disponible avec les applications *Adobe Digital Editions* (pdf, epub) et *MobiPocket Reader* (MOBI, PRC). L'impression des documents est interdite, ou permise, généralement en partie, selon les droits mis en place et le type d'accès choisi (achat ou emprunt). Les fonctionnalités de surlignage et

d'annotation sont tributaires des formats de restitution, du lecteur, et donc des choix que la presse a effectués.

Modèle d'affaires de OAPEN

Le projet européen *OAPEN* diffuse des livres numériques suivant le principe de l'accès libre. Les droits entourant les publications diffusées sur la plateforme varient en fonction des politiques éditoriales des intervenants impliqués et sont soumis, pour certains, à des licences *Creative Commons*.

OAPEN fait ainsi figure d'agrégateur moissonnant les publications et les fonds des institutions et presses universitaires européennes le désirant. Les documents doivent être soumis à un processus de révision par les pairs. Les PDF peuvent être imprimés en entier pour autant que les droits annoncés pour le livre soient respectés. Le copier/coller et les autres fonctionnalités propres au format PDF sont disponibles.

Processus d'édition numérique : le grand bazar

L'évolution rapide des technologies de l'information et l'apparition récente des nombreux appareils de lecture expliquent en partie la coexistence des différentes pratiques éditoriales observées. Comme nous pouvons le constater dans le tableau 2, certaines plateformes offrent des formats adaptés aux nouveaux supports de lecture, mais ce n'est pas encore généralisé. Il semble que les commerciaux seront plus rapides à procéder aux changements nécessaires à la prise en charge des nouveaux supports, en raison de leurs capitaux d'investissement principalement.

Projets	Formats de restitution	Particularités
<i>netLibrary</i>	PDF, DjVu, WMA protégé, MP3	<p>§ Supporte les appareils de lecture nook (B&N), Sony Digital Readers, COOL-ER (Interead) et tous ceux qui prennent en charge l'Adobe ContentServer</p> <p>§ Métadonnées balisées : titre de l'ouvrage, titre des chapitres, auteur, éditeur, lieu de publication, année de publication, langue, ISBNs, sujets, ID du produit</p>
<i>MyLibrary</i>	PDF, OeB (ancien ePub)	<p>§ Pas d'adaptation aux différents types appareils portatifs et d'écrans (reflowable text)</p> <p>§ PDF en format image uniquement</p> <p>§ Métadonnées balisées : titre de l'ouvrage, titre des chapitres, auteur, éditeur, EAN/ISBN [hardcover, paper, electronic, MIL], LCC, BIC, Dewey, lectorat visé, résumé</p>
<i>Cairn</i>	PDF, HTML	<p>§ DTD Erudit Article 3.0.0</p> <p>§ PDF en format image uniquement</p> <p>§ Métadonnées balisées : titre de l'ouvrage, rédacteurs en chef, titre du chapitre, auteur, affiliation, pagination, résumé [ou premières lignes], intertitres du chapitre [plan de l'article], collection, éditeur, lieu de publication, année de publication, ISBN</p>
<i>Humanities E-Book (HEB)</i>	PDF, HTML	<p>§ Texte OCR (recherche en plein texte dans PDF)</p> <p>§ TEI</p> <p>§ Images en GIF et PNG</p> <p>§ DTD aclshebook.dtd version 1.5 (78 titres en XML seulement)</p>
<i>BiblioVault</i>	PDF, ePub, HTML, PRC, MOBI	<p>§ PDF en format image uniquement</p> <p>§ Service de préparation des livres à l'intention des libraires et agrégateurs (<i>Amazon, B&N, netLibrary</i>, etc.)</p> <p>§ Dépôt à long terme des fichiers numériques</p> <p>§ Adaptation aux différents types d'écrans et aux appareils mobiles</p> <p>§ Métadonnées balisées : page couverture, titre de l'ouvrage, éditeur, année de parution, ISBN-10, ISBN-13, extraits, critiques, titres des chapitres, notes biographiques</p>
<i>OAPEN</i>	PDF, HTML	<p>§ PDF texte avec polices incorporées et texte/image pour les plus anciens</p> <p>§ TEI</p> <p>§ Métadonnées basées sur le Dublin Core, avec 19 champs disponibles, dont 10 obligatoires</p>

Par ailleurs, pour ce qu'il est possible d'en juger, le XML semble utilisé par toutes les

plateformes, comme le révèle l'affichage des métadonnées des livres numériques. Cela

étant, il n'est pas possible de déterminer l'étendue du balisage effectué par les plateformes ni de savoir comment les documents ont été produits, avec quels outils, suivant quelles étapes et à quel prix. Une étude plus approfondie permettrait peut-être de trouver des réponses à ces questions, qui nécessiterait une enquête auprès des plateformes.

Une chose est certaine toutefois : le PDF demeure le format de restitution le plus utilisé à ce jour, sur les plateformes sans but lucratif aussi bien que commerciales, mais cette situation risque fort d'évoluer rapidement dans les prochaines années.

Services des plateformes

Tous les projets analysés offrent des services de recherche simple et avancée avec différents filtres de résultats et différentes fonctionnalités de navigation dans les collections. Tous également, à l'exception de *OAPEN* (mais peut-être que ces services ne sont tout simplement pas encore annoncés), offrent des notices *MARC* à télécharger gratuitement, des outils de consultation des statistiques d'utilisation et d'exportation des notices vers des logiciels de gestion des références bibliographiques.

Les plateformes offrent leurs services à tous les éditeurs intéressés dont la production correspond aux normes établies (processus d'évaluation par les pairs notamment) et prennent en charge la commercialisation des contenus diffusés sans imposer d'exclusivité toutefois. Notons par ailleurs que *Cairn* dispose d'un outil de feuilletage et qu'il permet aux éditeurs qui le souhaitent de recourir à un système de tatouage numérique.¹¹ Notons par

la même occasion que les services aux éditeurs développés par *BiblioVault* sont plus raffinés et comprennent, entre autres, l'ajout de ISBN à treize chiffres pour les publications antérieures à 2007, des services d'impression de courts tirages (*short-run digital printing*) et la préparation de copies numériques requises dans le cadre de cours universitaires particuliers ou à l'intention d'une clientèle handicapée.

Conclusion

- L'offre de livres numériques est principalement développée en anglais (États-Unis et Grande-Bretagne) et diffusée par des entreprises commerciales.
- Les entreprises commerciales offrent une variété plus grande de formats de restitution et semblent plus à même d'offrir des documents adaptés aux nouveaux appareils de lecture.
- Le PDF reste à ce jour le format de restitution le plus employé.
- L'accessibilité aux ressources est mieux respectée par les structures sans but lucratif pas ou peu de DRM, de contraintes d'impression ou de téléchargement).
- La clientèle naturelle du livre numérique est encore institutionnelle en raison de la culture de « gratuité » du Web.
- La formule de l'entrepôt offre des services plus développés aux éditeurs.
- La formule de l'accès libre, bien qu'intéressante, en est encore à ses

¹¹ Le tatouage numérique (*watermarking*) consiste à imprimer en filigrane d'un document le nom de la

plateforme ou de l'éditeur afin de décourager les utilisateurs de le faire circuler.

débuts comme le montre le projet
OAPEN, lancé en octobre 2010
après plusieurs années d'élaboration.
Elle exige également un
investissement massif des
institutions publiques et, par là,
suppose une redéfinition imposante
des structures subventionnaires.

3- Ce que veulent bibliothécaires, chercheurs et étudiants

Trois études d'utilisateurs ont été retenues afin de prendre le pouls des attentes et besoins des utilisateurs de livres numériques savants.

Une première, menée entre 2008 et 2009 en Grande Bretagne par le *Joint Information System Committee (JISC)* auprès de sept universités, cent vingt-sept bibliothèques universitaires et deux agrégateurs de livres numériques, cherche à identifier, par une méthode dite de *deep log analysis*,¹² comment les livres numériques sont utilisés par les membres de la communauté universitaire.

Une deuxième, conduite par *HighWire Press* en 2009 auprès cent trente-huit bibliothécaires provenant de treize pays différents, a pour objectif de sonder les professionnels de l'information concernant les attitudes et pratiques reliées au livre numérique.

Et une dernière, réalisée en 2008 par *Annual Reviews, PNAS, MetaPress et Nature Publishing Group*, vise à illustrer comment les utilisateurs accèdent aux documents numériques universitaires et y naviguent.

JISC National E-books Observatory Project

Les conclusions de l'étude de *JISC* révèlent que plus de la moitié des membres de la communauté universitaire de Grande-

Bretagne a déjà consulté un livre numérique dans le cadre de leurs activités (recherche ou loisirs), principalement par l'entremise d'une bibliothèque universitaire. L'analyse de l'utilisation des livres numériques démontre qu'ils sont consultés sur de courtes périodes, pour de l'extraction de contenu et de données depuis le campus ou via un service de type Proxy, à toute heure du jour ou de la nuit, et tous les jours de la semaine.

Il est apparu que la constitution d'une offre de livres numériques adaptée aux besoins de la communauté universitaire pouvait aider à diminuer la pression exercée sur la collection imprimée des bibliothèques, en période de pointe notamment (mi-session et fin de session).

Trois remarques sont ressorties des données colligées concernant, plus généralement, le marché du livre numérique : la quantité de modèles d'affaires coexistants complexifie le secteur, les restrictions d'accès limitent l'usage des livres numériques et l'offre telle quelle est commercialisée actuellement n'est pas centrée sur les utilisateurs, qui voudraient une plus grande uniformité des structures de diffusion. Le catalogue de bibliothèque des institutions documentaires et les liens contenus sur le site web des institutions documentaires restent par ailleurs des points d'accès privilégiés aux contenus numériques. Enfin, durant le laps de temps qu'a duré l'enquête, aucun impact négatif n'a été enregistré sur les ventes de livres imprimés, ce qui est probablement dû aux prix élevés des monographies en Grande-Bretagne et aux faibles ventes de ce secteur.¹³

¹² La *deep log analysis* est une méthodologie par laquelle des données brutes d'interactions enregistrées sur un serveur sont colligées ce qui permet notamment de décrire de façon précise comment les utilisateurs interagissent avec le système (longueur de la session, nombre de pages vues, termes recherchés, etc.) L'information ainsi recueillie est combinée à des enquêtes qualitatives (focus group, sondage, etc.) afin de dresser un portrait global illustrant pourquoi les utilisateurs agissent comme ils le font.

¹³ *JISC National E-Books Observatory Project. Key Findings and Recommendations.*
<http://www.jiscebooksproject.org/reports/finalreport>

HighWire Press Librarian eBook Survey and Analysis

L'ensemble des participants à l'étude de *HighWire* prévoit une hausse significative des budgets consacrés aux livres numériques dans les cinq prochaines années, la majorité des institutions sondées consacrant déjà une partie plus ou moins importante de leurs ressources aux livres numériques sous forme de budgets récurrents ou non renouvelables.

La simplicité et la facilité d'accès et d'usage sont les composantes des livres numériques les plus appréciées des bibliothécaires sondés alors que les sources traditionnelles d'accès aux livres continuent d'être importantes dans le monde numérique puisque les bibliothécaires acquièrent et s'informent des livres numériques depuis les mêmes canaux que dans le monde imprimé, via les distributeurs et diffuseurs de livres.

Le catalogue de bibliothèque et Internet restent des voies d'accès privilégiées au livre numérique selon les sujets de l'étude, qui indiquent également que les utilisateurs préfèrent encore le format PDF, bien que l'évolution des supports de lecture risque de modifier cette tendance.

Enfin, il est apparu que les systèmes de gestion des droits d'accès (DRM) sont le plus important irritant pour les usagers de bibliothèque et que l'achat avec accès pérenne est considéré comme le modèle d'affaires le plus intéressant par 83 % des répondants bien qu'un nombre significatif d'entre eux ait indiqué que d'autres modèles d'affaires sont également équitables (les modèles préférés ne sont malheureusement pas identifiés).¹⁴

¹⁴<http://highwire.stanford.edu/PR/HighWireEBookSurvey2010.pdf>

How Readers Navigate to Scholarly Content

L'étude, réalisée en Grande-Bretagne pour le compte de *Annual Reviews*, *PNAS*, *MetaPress* et *Nature Publishing Group*, a permis d'identifier que, plus que jamais, les lecteurs accèdent aux ressources numériques au niveau de l'article ou du résumé, en raison de l'indexation des collections sur Internet notamment. Ainsi, certaines fonctionnalités des plateformes de diffusion deviennent invisibles ou inaccessibles alors que d'autres, offertes à un plus haut niveau dans la structure des plateformes, sont moins pertinentes qu'auparavant.

La recherche a permis de dévoiler également que la fonctionnalité la plus prisée est la création d'alertes et non pas la possibilité de personnaliser l'interface de recherche ni le degré de sophistication des options de recherche.

Il est ressorti enfin qu'il est essentiel d'utiliser le plus de canaux de diffusion possible afin de rendre le contenu accessible, sans toutefois héberger des copies des documents sur différents serveurs. Ainsi les éditeurs devraient collaborer avec *Google*, publier en XML, être compatible avec la norme *OpenURL*, collaborer avec *CrossRef*, avoir une syntaxe d'URL prévisible, offrir des fils RSS, des alertes, etc.¹⁵

Conclusion

- La demande pour les livres numériques est forte et elle devrait s'accroître au cours des prochaines années.

¹⁵ Inger, Simon et Tracy Gardner, *How Readers Navigate to Scholarly Content*.
<<http://www.sic.ox14.com/howreadersnavigatetoscholarlycontent.pdf>>

- L'offre de livres numériques gagnerait à être adaptée aux besoins des communautés universitaires afin que la pression sur les collections imprimées soit réduite en période de pointe (livres qui se retrouvent fréquemment à la réserve par exemple).
- Les fonctionnalités entourant les livres numériques devraient prendre en considération le fait que les livres sont utilisés pour de courtes périodes et pour de l'extraction de contenu plutôt que pour une lecture intensive.
- Elles devraient aussi considérer le fait que les chercheurs accèdent au contenu numérique principalement au niveau de l'article et du résumé (faire en sorte que les fonctionnalités soient accessibles à tous les niveaux).
- Les acteurs du numérique devraient considérer des moyens alternatifs aux DRM pour limiter le piratage de leurs collections, parmi lesquels le tatouage numérique serait à considérer.
- Les documents numériques devraient être conçus de telle sorte que leur intégration à différents canaux de diffusion soit maximale (métadonnées riches et complètes) et de façon à prendre en charge les nouveaux supports de lecture (utilisation du XML plutôt que du PDF image).

4- Enquête auprès des presses universitaires

Tous les éditeurs sondés ont développé au cours des dernières années une offre de livres numériques importante. Pour l'essentiel, les livres numériques offerts existent également en version imprimée. Les éditeurs universitaires canadiens ne se sont pas encore lancés dans la production de livres exclusivement numériques.

Les presses universitaires canadiennes : une offre numérique abondante

Les *Presses de l'Université de Montréal (PUM)* comptent 50 % de leur catalogue en version numérique, les *Presses de l'Université Laval (PUL)* plus de 30 %, les *Presses de l'Université de Colombie-Britannique (UBC Press)* 75 %, alors que les *Presses de l'Université de Toronto (UTP)* ont converti en numérique plus de 1500 titres de leur catalogue et que *McGill-Queen's University Press (McGill-Queen's UP)* offrira d'ici six mois l'ensemble de leur collection en version numérique.

McGill-Queen's University Press est sans contredit la presse universitaire canadienne qui a élaboré la stratégie numérique la plus avant-gardiste. Ce passage précoce au numérique a permis à la structure éditoriale de bénéficier de retombées économiques considérables, grâce à la signature d'entente avec des agrégateurs canadiens et américains notamment, alors que l'offre numérique était moins abondante qu'aujourd'hui. *McGill-Queen's University Press* est d'ailleurs la seule presse sondée à avoir déclaré retirer des bénéfices importants de la vente de livres numériques. Les autres éditeurs ne voient pas encore de retombées financières significatives à leur engagement numérique.

UBC Press est par ailleurs la seule à avoir noté des ventes importantes réalisées depuis son site Internet, probablement grâce à la

politique de Google qui met le lien de l'éditeur sous les titres des livres diffusés. *UBC Press* affirme malgré cela devoir mettre fin prochainement à la publication de livres en grand format (hardcover), un secteur lucratif de leur production destiné aux bibliothèques, en raison de la demande des institutions pour le livre numérique, bien que les revenus provenant de la vente de livrels ne parviendront sûrement pas à compenser les pertes de revenus encourues.

Certains éditeurs (*PUM*, *PUL* et *UBC Press*) ont relevé à cet égard la nécessité de subventions pour soutenir leur investissement dans le numérique.

Production de livres numériques : variété des pratiques

Alors que les *Presses de l'Université de Toronto* et de *l'Université de Colombie-Britannique* travaillent à l'élaboration d'une chaîne de production éditoriale permettant de produire à l'interne des fichiers XML (pour le moment les éditeurs recourent à de l'aide extérieure) et que les *PUL* produisent elles-mêmes leurs fichiers numériques, *McGill-Queen's UP* et les *PUM* confient à des contractuels le soin de générer leurs fichiers de diffusion. Cette pratique s'insère dans une tendance plus générale qui voit les tâches reliées à la production déléguées à différents intervenants externes suivant la nature des projets entrepris. Cette indépendance semble chère aux éditeurs sondés qui veulent par là s'assurer de la qualité de leurs publications et, comme les *PUM*, concentrer leurs efforts sur la dimension intellectuelle de leur travail éditorial.

Les éditeurs semblent bien connaître les différentes pratiques en matière de livres numériques et paraissent soucieux de suivre l'évolution des standards. Les éditeurs entendent produire des livres numériques dans les formats et suivant les standards du marché,

afin de maximiser leur visibilité et favoriser l'essor de leurs ventes dans ce secteur. Il est à noter toutefois que *UBC Press* s'est montré sceptique en ce qui concerne la conservation des documents numériques et a affirmé l'importance d'une stratégie nationale visant à assurer la préservation des patrimoines scientifiques numériques.

Chez *McGill-Queen's University Press*, le choix des formats de fichiers de restitution et des canaux de diffusion s'inscrit dans la stratégie de mise en marché des titres. Ainsi, les livres très pointus ne figureront pas sur la plateforme *iTunes* d'*Apple*, qui force les éditeurs à pratiquer des prix très bas, alors que les ouvrages plus grand public, qui coûtent moins chers à produire, y seront disponibles. *UBC Press* a également souligné les conditions commerciales difficiles imposées aux éditeurs universitaires par *Amazon* et *Apple*, deux joueurs qui cherchent à dominer le marché par une politique de prix cassés.

Élargir les marchés grâce aux canaux de diffusion déjà existants

Les éditeurs sondés s'accordent tous également pour affirmer l'importance d'une diffusion large de leur fonds et de l'utilisation la plus ouverte des principaux canaux de diffusion, qu'il s'agisse d'agrégateurs, de moteurs de recherche ou de librairies en ligne. Ils ont massivement investi ces espaces de diffusion et de commercialisation, localisés aux États-Unis et en Europe principalement.

Aucun ne semble craindre de perdre le contrôle de sa production éditoriale ni de son capital symbolique par son inclusion dans des bases de données généralistes, bien que certains affirment (*UTP*) qu'il leur importe de garder le contrôle sur la fixation des prix de leurs ouvrages, sur la réputation de leur presse (assurée par la renommée des entreprises avec lesquelles ils travaillent), sur les dispositions de

gestion des droits d'accès et les capacités de commercialisation de leurs fonds par leurs partenaires.

Les éditeurs cherchent clairement à bénéficier de l'abolition des frontières qu'offre Internet et ont pour objectif d'investir les marchés américains et européens depuis les structures de diffusion déjà en place. *UBC Press* affirme malgré cela privilégier les initiatives universitaires (*BiblioVault* aux États-Unis par exemple) aux entreprises commerciales dans ses choix de partenaires d'affaires.

Toutefois, aucun n'affirme voir l'intérêt d'un nouveau projet canadien de production et de diffusion des livres numériques savants, bien que *McGill-Queen's University Press* ait affirmé qu'un consortium regroupant des presses universitaires canadiennes pourrait être une avenue intéressante pour la promotion des collections canadiennes auprès des institutions documentaires au pays en vue de la négociation de contrats de diffusion.

Conclusion

- Les éditeurs universitaires canadiens sont tous impliqués dans l'économie numérique et cette implication s'accroîtra au cours des prochaines années.
- Le livre numérique universitaire est toujours créé en fonction des paradigmes de l'impression papier. Les fonctionnalités propres au numérique ne sont pas encore pleinement exploitées.
- La question de la pérennité des fichiers produits se posera avec acuité lorsque les livres exclusivement numériques apparaîtront. Les coûts de migration des formats devraient faire pencher

les éditeurs en faveur de l'utilisation du XML.

contrats de diffusion du type *CRKN*.

- Le secteur du livre numérique universitaire n'est pas encore rentable, à l'exception de l'expérience particulière de *McGill-Queen's UP*. Les conditions commerciales imposées par les grands acteurs du numériques, qui veulent offrir les plus bas prix possibles, sont souvent peu avantageuses pour les éditeurs de livres savants.
- En raison de l'absence de structure canadienne de production et de diffusion de livres numériques universitaires, les éditeurs se sont massivement tournés vers des initiatives commerciales basées aux États-Unis et en Europe.
- La vente des collections numériques à une clientèle institutionnelle depuis des plateformes semble la formule la plus lucrative, malgré des conditions commerciales frileuses.
- Les éditeurs canadiens estiment avoir accès à de nombreux canaux de diffusion et veulent garder leur indépendance en refusant toute forme d'exclusivité de diffusion. Alors que certains éditeurs veulent produire leurs fichiers à l'interne (*UTP, PUL*), d'autres préfèrent recourir aux services de contractuels (*McGill-Queen's UP* et *PUM*). Aucun ne souhaite disposer d'un service de production centralisé pancanadien.
- Les éditeurs notent tous cependant l'importance des investissements requis pour le passage au numérique. Ils déclarent leur intérêt pour le déblocage de subventions ou pour la création d'un consortium ayant pour objectif la négociation de

5- La chaîne de traitement du livre numérique

Le processus d'édition du livre numérique s'intercale actuellement dans le processus d'édition du livre imprimé, faisant de l'objet numérique un dérivé de la chaîne de traitement du livre papier.

À l'exception des formats de restitution visant une technologie spécifique tels que les formats PDF, ePub ou PRC, les traitements visant à produire des fichiers XML, mieux adaptés à la conservation et à l'évolution des supports de lecture, requièrent toutefois un processus plus complexe de transformation des documents à partir de fichiers sources. Cela entraîne inévitablement des coûts plus élevés. L'expertise nécessaire pour produire de tels fichiers XML force les éditeurs à se tourner vers de formats moins coûteux ou à chercher à diminuer les coûts de production en faisant faire le travail par des sous-traitants.

La pluralité des formats de diffusion et des supports de lecture

À l'heure où certains formats de livres numériques tentent de s'imposer, il convient d'analyser leur pertinence et leur utilisation dans le contexte d'un processus d'édition numérique et de diffusion. Les formats PDF,¹⁶ ePub, PRC/MOBI et XHTML dominent actuellement le marché du livre numérique et leur utilisation semble se généraliser. Toutefois, dans le contexte du livre savant, on dénote une forte prédominance du format PDF.

Le format PDF est certainement celui qui se rapproche le plus de ce que l'éditeur connaît déjà, c'est-à-dire un format permettant un rendu où il a le contrôle absolu sur ce que

le lecteur aura sous les yeux. Depuis 1993, c'est donc le format de prédilection pour l'impression et l'affichage préservant la mise en forme, peu importe le lecteur ou la plateforme.¹⁷ De plus, c'est certainement le format le plus simple à produire et le moins coûteux.

Contrairement au format PDF, qui préserve toutes les caractéristiques de mise en page de l'éditeur, les formats ePub et PRC/MOBI permettent, quant à eux, l'affichage des contenus de manière ajustable (*reflowable text*) aux différentes tailles d'écran des appareils mobiles, portables ou tablettes de lecture. Ces deux formats sont surtout des contenants structurés pouvant recevoir des fichiers XHTML (un fichier par chapitre), des fichiers d'images (dans le cas du format ePub), des fichiers de style CSS ainsi que des fichiers de métadonnées.

Malgré l'apparent bénéfice pour le lecteur de pouvoir consulter un document s'adaptant à son appareil de lecture préféré, certains types de contenu ou certaines conventions de présentation n'ont pas encore trouvé d'équivalent dans ce mode présentation,¹⁸ d'où une adaptation inévitable de la part de l'éditeur ou de l'auteur.

Le format HTML (ou XHTML dans sa version la plus stricte) constitue probablement une exception à la règle chez les éditeurs universitaires. Pour ce cas de figure, le format de diffusion est généralement produit dynamiquement à partir de fichiers XML finement balisés, suivant des schémas de données standards ou développés à l'interne

¹⁶ http://fr.wikipedia.org/wiki/Portable_Document_Format

¹⁷ Dès sa deuxième version (2.3) sortie en novembre 2009, le Kindle d'Amazon permet la lecture native des fichiers PDF (http://fr.wikipedia.org/wiki/Amazon_Kindle)

¹⁸ https://webshare.uchicago.edu/users/kdavey/Public/eformats_aaup_lores.pdf

tels TEI (*BiblioVault*, *OAPEN*, *Humanities E-Book*), ACLS-HEBOOK-DOC (*Humanities E-Book*) ou même ÉruditArticle (*Cairn*). Même si le format HTML permet, en plus d'une consultation aisée sur tous les types de supports possédant un navigateur, l'ajout de fonctionnalités supplémentaires et, dans certains cas, interactives, on constate que seule une infime portion des collections recensées est disponible en format HTML.¹⁹

Enfin, notons que les formats de restitution sur liseuses ne semblent pas encore être totalement adéquats pour la réalité des chercheurs, comme le laisse supposer une étude préliminaire réalisée pour la plateforme *Humanities eBook* :

« HEB's initial findings in this study indicate that titles formatted for existing handheld devices are not yet adequate for scholarly use in terms of replicating either the benefits of online collections—cross-searchability, archiving, multifarious interactive components—nor certain aspects of print editions that users reported missing, such as being able to mark up and rapidly skim text. A turnaround is underway once a common and more robust format optimized for handheld readers is determined and devices themselves evolve, adding improved display options and better and more intuitive web-access, searching and other interactive use of content. »²⁰

Plusieurs formats, plusieurs niveaux de services

Le tableau suivant présente brièvement les principales caractéristiques des formats actuellement privilégiés par les éditeurs universitaires.

Il importe de mentionner que tous les schémas XML ne sont pas équivalents en terme de consignation des éléments sémantiques, notamment au niveau du corps du texte.

¹⁹ Dans le cas de *Humanities E-Book*, on dénombre (en novembre 2010) moins de 80 titres en XML sur l'ensemble de la collection qui en compte plus de 2800, soit moins de 3 %.

²⁰ <http://www.humanitiesebook.org/heb-whitepaper-3.html>

	PDF	EPub	PRC/MOBI	XML (HTML)
Préservation de la mise en forme du document imprimé	OUI	NON	NON	NON
Adaptation du texte à la taille de l'écran	NON	OUI	OUI	OUI
Multiplateforme	OUI	OUI	NON	OUI
Fonctions d'annotations (commentaires/surlignage)	OUI	OUI	OUI	NON*
Facilité de conversion vers d'autres formats de restitution	Moyenne	Moyenne	Moyenne	Élevée
Facilité d'extraction des métadonnées	Moyenne	Moyenne	Moyenne	Élevée
Facilité d'exportation de portions de texte (copier/coller)	Moyenne	Faible	Faible	Élevée

* Des applications tiers, telle que Zotero (www.zotero.org) et Mendeley (www.mendely.com), permettent de sauvegarder une copie du document et d'y superposer une couche d'annotations et de surlignage, à même la consultation dans le navigateur.

La qualité des métadonnées, un enjeu pour la diffusion et la distribution

À l'heure où le repérage, la diffusion et l'indexation des livres sont au cœur des préoccupations des auteurs et des éditeurs, l'importance de la qualité des métadonnées se pose comme un véritable enjeu. Il ne s'agit pas de mettre en ligne des documents de formats PDF ou ePub pour que ceux-ci soient automatiquement référencés. Il importe que les métadonnées de chaque document soient produites de manière rigoureuse et puissent être ajoutées aux fichiers de restitution ou transmises de manière automatique aux différentes plateformes de diffusion ou de distribution.

Actuellement, autant les formats PDF, ePub et PRC/MOBI permettent d'incorporer aux livres numériques des métadonnées en utilisant le standard Dublin Core. Le format PDF, à l'aide de son propre conteneur de

métadonnées appelé XMP,²¹ permet quant à lui de greffer d'autres formats de métadonnées. Par contre, dans ce dernier cas, le recours à des applications externes est indispensable.

La consignation rigoureuse des métadonnées dans un schéma XML à la fois enrichit le document lui-même et permet de reproduire les métadonnées dans une diversité de formats alternatifs, tels que MARC21 (pour les catalogues des bibliothèques) ou MARCXML, Dublin-Core, RSS, RDF ou même OPDS (Open Publication Distribution System)²². Ce dernier format est actuellement en émergence et est principalement utilisé pour la distribution, l'agrégation et l'acquisition de publications électroniques.

²¹ <http://www.adobe.com/devnet/xmp.html>

²² <http://fr.wikipedia.org/wiki/Bookserver> et <http://opds-spec.org/>

FORMAT	MÉTADONNÉES	NIVEAU DE GRANULARITÉ DES MÉTADONNÉES
PDF	Deux niveaux de consignation des métadonnées <ul style="list-style-type: none"> • Propriétés simples du document • Format XMP (eXtensible Metadata Platform), développé par <i>Adobe</i> qui incorpore plusieurs ensemble de formats dont le Dublin Core. La gestion en est faire à l'aide d'outils Adobe ou tiers 	Livre
ePub	Les métadonnées sont consignées dans un fichier spécifique (<i>content.opf</i>) du conteneur ePub en utilisant la norme Dublin Core	Livre
PRC/MOBI	La consignation des métadonnées est faite à l'aide des outils de publication tels que <i>MobiPocket Creator</i> ²³ et permet d'inscrire à même les fichiers les métadonnées les plus courantes (comme le Dublin Core) ainsi certaines données plus spécifiques à la commercialisation (image de couverture, fichiers démo, prix de détail suggéré, etc.)	Livre
XHTML	La consignation des métadonnées est faite directement dans le fichier XHTML, à l'aide des balises <meta>	Livre ou chapitre selon le découpage de la publication
XML	La consignation est faite à l'aide des outils des différentes chaînes de traitement et du schéma XML sous-jacent.	Livre, chapitre ou section, selon le schéma utilisé

²³ <http://www.mobipocket.com/dev/article.asp?BaseFolder=creatorpublisher&File=metadata.htm>

On constate que pour les formats PDF, ePub et PRC/MOBI, les métadonnées font partie intégrante du document de restitution et ne peuvent être facilement réutilisées, à condition de disposer des outils nécessaires à leur extraction. Dans le cas d'ouvrages collectifs, seules les métadonnées du livre sont repérables, ce qui n'est pas nécessairement le cas pour les ouvrages possédant un équivalent XML, découpés par chapitre.

Le découpage par chapitre assure aussi une exploitation maximale des contenus par l'utilisateur. Comme le font remarquer Michèle Lénart et Claudine Masse :

« La notion de chapitre devient beaucoup plus importante qu'avec le papier. Les lecteurs effectuent une lecture rapide : 20 % balaisent les pages, 80 % lisent des extraits. Ils utilisent les fonctionnalités de zoom, recherche dans le texte, annotation, mémorisation de la dernière page consultée. Les utilisateurs ont une forte demande d'accès immédiat, de téléchargement et de copier-coller. La notion de type de documents n'a pas d'importance pour eux, ce qu'ont bien compris les éditeurs qui proposent des plates-formes communes de livres et périodiques. »²⁴

Ce type d'usage est d'ailleurs confirmé par l'étude de JISC.²⁵

XML, un format orienté services

L'apparition rapide des nouveaux supports de lecture démontre une évolution sans précédent dans la définition de nouvelles

technologies d'affichage, rythme que les éditeurs universitaires ne semblent pas suivre pour l'instant, les livres en format PDF constituant l'essentiel des collections de ces derniers.

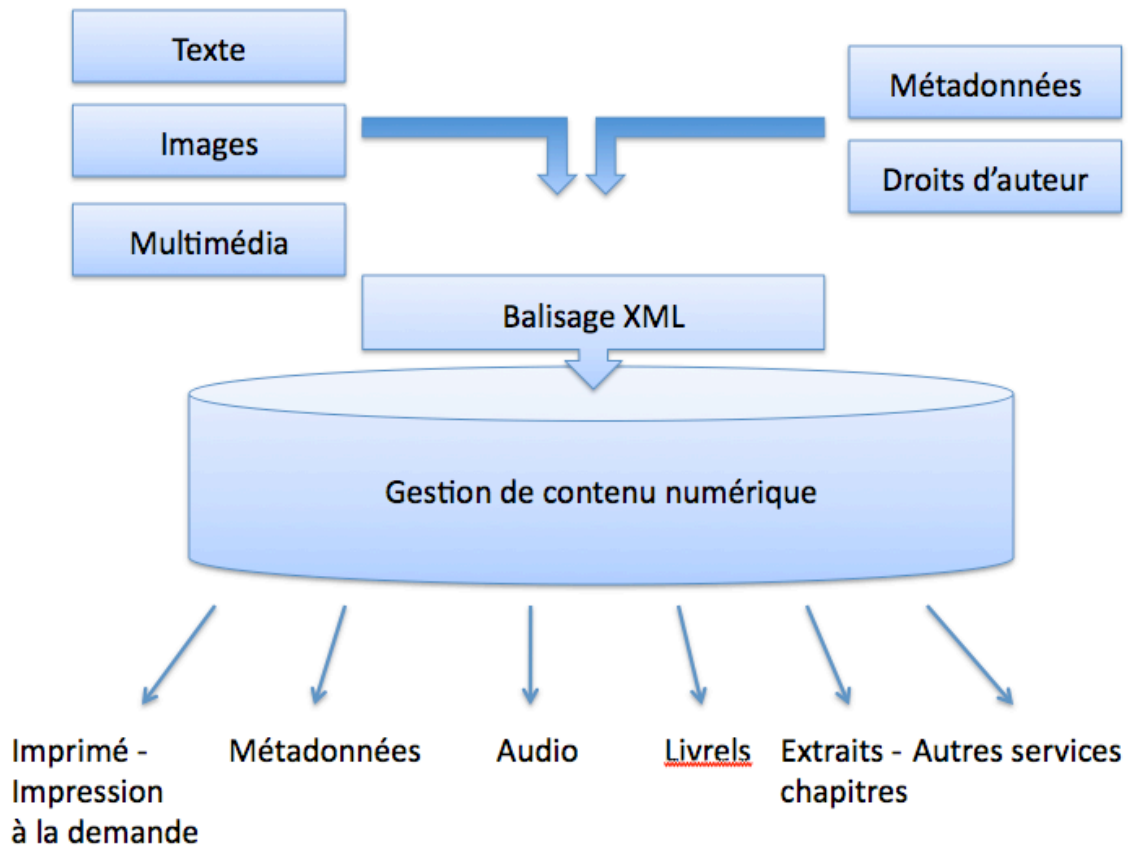
Toutefois, peu d'intérêt semble être porté sur la préservation et la pérennité des formats de diffusion. Bien que les formats HTML et PDF dominant actuellement le paysage numérique depuis plus de 15 ans, l'arrivée de nouvelles technologies telles que la version 5 du standard HTML ou les nouvelles tablettes de lecture forcera peut-être les éditeurs à revoir leur position.

Dans une perspective de pérennisation, il est important de s'assurer que les documents puissent être facilement convertis d'un format obsolète à un autre plus actuel. Pour l'instant, seul le format XML permet une préservation adéquate des contenus et des métadonnées et rend leur conversion aisée.

La plupart des schémas XML ne s'attachent pas à une technologie de restitution précise, le format d'affichage étant habituellement généré dynamiquement. Par contre, un tel affichage présuppose une plateforme de diffusion beaucoup plus développée, pouvant prendre en charge un ou plusieurs schémas de données XML et permettant de générer simultanément plusieurs formats de diffusion.

²⁴ Michèle Lénart et al. « Méthodes, techniques et outils », *Documentaliste-Sciences de l'Information* 1/2010 (Vol. 47), p. 14. <http://dx.doi.org/10.3917/docs.471.0014>.

²⁵ « JISC national e-Books observatory project, » sans date, <http://www.jiscebooksproject.org/reports/finalreport>, p. 17.



Parmi les services offerts par les plateformes XML, notons entre autres :

- L'accès aux tables de matières complètes et la navigation par chapitre ou section
- La numérotation des paragraphes
- Le référencement permanent
- La navigation aisée entre le texte et les notes ou références bibliographiques
- Des résumés parfois multilingues
- L'intégration aisée de contenu multimédia
- La production « à la volée » de plusieurs formats de restitution
- La production de métadonnées dans une panoplie formats et standards

Certaines plateformes offrent une partie de leur collection en format XML dont *Cairn*, *OAPEN*, et *Humanities E-Book (HEB)*. On y différencie deux types d'usage du XML, le premier sert principalement à consigner les métadonnées et le corps du texte sans balisage, en appui à un format de restitution PDF, et le second permet un rendu HTML/XHTML complet. À cet égard, *HEB* offre une plateforme où il est possible, même pour les livres en format PDF, de naviguer à travers les

tables des matières et d'accéder au texte intégral (en format texte brut)

Enfin, notons que la recherche fondamentale s'intéresse de plus en plus au contenu des documents numériques en tant qu'objet de recherche. Par contre, seul le format XML semble répondre actuellement aux besoins en matière de fouille de texte, de bibliométrie, etc.

En effet, l'extraction de contenu à valeur sémantique est très difficile, voire impossible à partir des autres formats de restitution, surtout lorsqu'elle est effectuée de manière automatique sur de larges collections. Le balisage XML acquiert encore ici une valeur indéniable pour une exploitation optimale par les chercheurs.

Conclusion

- Le PDF est actuellement le format de prédilection des éditeurs universitaires. Ce choix semble motivé tant par la portabilité du format, par sa facilité de production et de commercialisation que par la faiblesse des coûts associés à sa production, surtout dans le contexte du livre numérique dérivé du livre papier. Les formats qui s'adaptent aux nouveaux supports de lecture comme le ePub sont, pour l'instant, peu supportés par un grand nombre de plateformes commerciales.
- Les formats XML permettent une panoplie de services ne pouvant être offerts par les autres types de formats de restitution, notamment la possibilité de générer statiquement ou dynamiquement les métadonnées dans des formats variés. Malgré cela, et bien que les utilisateurs semblent y trouver un certain nombre d'avantages, le livre en format XML

ne constitue qu'une très faible proportion des collections actuelles.

- La consignation rigoureuse des métadonnées est un enjeu important pour la diffusion et la distribution des livres numériques. Actuellement, la norme Dublin Core est celle qui est la plus utilisée et intégrée dans la plupart des formats actuels, comme le PDF, ePub ou PRC/MOBI.
- Le format XML est le seul qui offre à ce jour un plein potentiel de pérennisation car il est facilement stockable et permet une conversion aisée et sans perte, ce que les formats tel que le PDF ne peuvent offrir actuellement.
- Seul le format XML semble répondre actuellement aux besoins de la recherche fondamentale.

6- Le processus d'édition numérique

Peu importe le format de restitution offert au lecteur, qu'il soit XHTML, PDF, ePub ou autre, le travail d'édition et de consignation des métadonnées s'effectue soit à partir de sources de données externes comme des fiches de catalogage en format MARC, soit directement à partir des fichiers de mise en page produits par l'éditeur.

Parmi les logiciels de mise en page les plus populaires, notons entre autres *Adobe InDesign*²⁶ qui permet de produire à la fois les formats de restitution PDF et ePub pratiquement sans effort. Pour ces deux formats, l'expertise nécessaire est donc réduite au minimum, confinée seulement à un travail minutieux de consignation des métadonnées à même le fichier de mise en page.

De même pour le format PRC/MOBI, en utilisant les outils mis à la disposition des éditeurs (*Mobipocket Creator* en l'occurrence), la consignation des métadonnées est simplifiée et ne dépend que d'une expertise limitée. Nul besoin de connaître une technologie autre que celle définie par la plateforme Mobi.

Par contre, pour la constitution de fichiers XML, l'extraction du texte et des métadonnées des fichiers de mise en page n'est pas trivial car les mises en formes complexes (appels de notes, colonnes multiples, etc.) ne sont pas toujours bien gérés par les outils d'exportation mis à la disposition des éditeurs numériques. Un travail manuel est donc inévitable. L'éditeur doit alors recourir à une expertise dédiée, qu'il possède ou non.

Les types de traitement du livre numérique

On distingue actuellement plusieurs types de traitement menant à la production de livres numériques, ces traitements font appel à des processus qui varient principalement en fonction des deux paramètres suivants:

- Le type de document source (images numérisées, PDF, documents de mise en page)
- Le type de document généré en sortie, généralement le format de restitution et/ou de préservation (PDF, ePub, XML, etc.)

Par exemple, pour réaliser son étude sur la conversion en format XML, *Humanities E-Book* a préféré recourir à des documents sources de type images numérisées déjà en sa possession pour produire en sortie des documents XML suffisamment balisés pour un rendu XHTML.²⁷

Dans le contexte d'une chaîne de traitement produisant des fichiers XML pour la conservation et la diffusion, nous distinguons plusieurs types de traitement qui dépendent à la fois du type de document source et du type de fichier XML résultant :

- Traitement **en lot** pour les projets de numérisation rétrospective : production de PDF texte/image et consignation des métadonnées avec identification sommaire du corps du texte pour indexation, le corps contenant le texte fourni par un processus de reconnaissance optique de caractères sans correction des erreurs (les erreurs de

²⁶ <http://www.adobe.com/products/indesign/>

²⁷ <http://www.humanitiesebook.org/heb-whitepaper-2.html>

reconnaissance sont laissées telles quelles dans le fichier XML) ;

- Traitement **minimal** : Production de PDF texte/image, consignation des métadonnées pouvant comprendre les références bibliographiques et identification sommaire du corps du texte pour indexation, le corps contenant le texte tel qu'il apparaît dans la version finale soumise par l'auteur ;
- Traitement **complet** : Production ou non d'un PDF texte/image, consignation des métadonnées et balisage de tous les éléments sémantiques du texte, à savoir les sections, sous-sections, notes, tableaux, images, etc., pouvant servir à un rendu XHTML.

Certains éditeurs numériques offrent toutefois la possibilité de convertir en format XML finement balisé des livres numérisés et ayant fait l'objet d'une reconnaissance optique de caractères. Dans ces cas, le travail est plus laborieux car il demande une vérification supplémentaire, caractère par caractère, entre le document numérisé et le document final. C'est le cas notamment de la chaîne de traitement utilisée par SPO (*The Scholarly Publishing Office of the University of Michigan Library*), le partenaire de *Humanities E-Book*.

Enfin, de plus en plus de services de conversion sont proposés aux éditeurs, permettant à ces derniers de bénéficier d'une expertise qui peut être parfois coûteuse à acquérir (nous reviendrons d'ailleurs sur ces coûts plus loin). Par exemple, la plateforme *Cairn* (ainsi que *Gallimard*) a confié une partie de son travail de production de fichiers XML à une firme externe, *Isako* (www.isako.com). De même, *Humanities E-Book* encourage les éditeurs à produire eux-mêmes leurs fichiers,

qu'ils soient PDF ou XML ou à impartir cette production.

Toutefois, la mise en place d'une expertise spécifique à une chaîne de traitement XML ouvre la voie à l'élargissement vers des types de contenus autres que textuels, comme l'image, l'audio ou la vidéo. De la même manière, le développement de collections englobant plus d'un genre scientifique (revues savantes, actes de colloques ou conférences, thèses, etc.) tend à s'étendre. Les portails *Cairn*, *Érudit* et bientôt *Project Muse* en sont de bons exemples.

Une analyse des coûts de production : à traitements différents, coûts différents

Les différents types de traitement mentionnés précédemment amènent un éventail de grilles de coûts. À un extrême de cette gamme, on retrouve des coûts relativement faibles pour la production de fichiers PDF ou ePub à partir des fichiers de mise en page, alors que la production de fichiers XML finement balisés peut se révéler coûteuse.

Une étude du MOTif²⁸ a évalué le coût de production d'une édition numérique en format ePub avec fichier XML à l'arrivée pour un essai de 320 pages. Ces coûts se chiffrent de la manière suivante :

- environ 2000 \$ (un peu plus de 1500 euros) si la conversion est faite à partir d'un exemplaire papier (comprenant numérisation et OCR)
- environ 1200 \$ pour une conversion faite à partir des fichiers de mise en page

²⁸ <http://www.lemotif.fr/fr/etudes-et-analyses/etudes-du-motif/cout-d-un-livre-numerique/>

Comme le sujet de cette étude ne concernait pas exclusivement le livre savant, une majoration de ces coûts en fonction de la complexité inhérente à ce type de publications (bibliographies, notes et appels de notes, index, etc.) semble nécessaire.

Dans son étude sur la conversion en format XML d'une vingtaine de titres de sa collection, *Humanities E-Book* abouti à des coûts nettement inférieurs, de l'ordre de 570 US \$ en moyenne. Toutefois, pour atteindre de tels coûts, *HEB* doit recourir à de la main d'œuvre beaucoup moins dispendieuse, en Inde en l'occurrence, et simplifier le balisage. Ainsi, dans une logique de production de masse, les coûts de production des fichiers XML n'étant pas négligeables, le recours à l'impartition séduit de plus en plus d'éditeurs. Le cas de *Humanities E-Book* n'est certainement pas marginal.

Traitement en lot (numérisation rétrospective)

Lorsqu'elle est bien développée et rodée, une chaîne de traitement performante peut permettre la production de fichiers de diffusion et de conservation à des coûts relativement bas, tout en satisfaisant à des exigences élevées en termes de qualité. Par exemple, des coûts moyens de l'ordre de 0,60 \$ à 0,70 \$ la page sont tout à fait possible, ces coûts comprenant à la fois la numérisation en lot, la reconnaissance optique de caractères (OCR), la consignation des métadonnées et la production des fichiers de diffusion/conservation de type "minimal", c'est-à-dire sans balisage sémantique détaillé du corps du texte.

Par contre, de tels tarifs ne permettent pas d'assumer les coûts de correction de toutes les erreurs provenant de l'OCR, qui peuvent être de l'ordre de 1 % pour la langue anglaise. De plus, le contenu des fichiers XML provenant

d'une numérisation rétrospective faite en lot ne pourrait servir à offrir l'éventail des services d'une plateforme de diffusion riche, à savoir les tables des matières complètes, des bibliographies, etc. Seules les métadonnées de base de chaque chapitre pourraient ainsi être préservées.

Enfin, le tarif "à la page" est considéré dans ce type de traitement essentiellement en raison des étapes de massicotage, de numérisation (scan) et d'OCR.

Traitement minimal et complet, du simple au triple

Pour ces deux types de traitement, l'expérience montre que le nombre de page n'intervient que de manière marginale dans le coût de production — même pour le traitement complet à partir des fichiers de mise en page —, puisque le texte est sensé être exempt d'erreur, ayant préalablement fait l'objet d'un processus d'édition sévère de la part de l'éditeur. Ainsi, lorsqu'il est extrait des fichiers sources ou de mise en page, peu de balisage supplémentaire serait requis, ce qui rend difficile l'élaboration d'une métrique de calcul de coût basée sur le nombre de pages.

Dans le cas du traitement complet, c'est la complexité des éléments de balisage qui influe de manière prépondérante sur les coûts de production. Le traitement des images, des tableaux, des équations ou de tout autre objet multimédia nécessite souvent des manipulations particulières. Le tableau ci-dessous présente des coûts moyens approximatifs (en \$ canadien) pour la chaîne de production de la plateforme d'édition du Centre d'édition numérique de l'Université de Montréal.

Type de traitement	Nbre pages	Nbre chapitres	Nbre images	Nbre tableaux	Coût approximatif
Numérisation en lot	300	10	-	-	195 \$ (0,65 \$/page)
Minimal	300	10	-	-	400-500 \$
Minimal	300	15	-	-	550-650 \$
Complet	300	10	-	-	1000-1200 \$
Complet	300	10	30	10	1300-1400 \$
Complet	300	15	-	-	1400-1500 \$
Complet	300	15	30	10	1600-1700 \$

Il est à noter que ces coûts couvrent l'ensemble du processus de production numérique, de la réception des fichiers en provenance de l'éditeur, à la diffusion et la distribution ainsi que le balisage fin de tous les éléments sémantiques du texte à travers un processus rigoureux. Mais ils ne tiennent pas compte de l'ensemble des coûts liés à l'infrastructure technologique, à savoir les coûts d'hébergement des fichiers, de préservation, de conversion, de gestion des droits d'auteurs (DRM), etc.

Enfin, notons que certaines plateformes comme *Humanities E-Book*, plutôt que d'offrir des services complets d'édition XML, proposent en fait un schéma XML et une intégration des contenus plus en aval dans la chaîne de traitement. L'éditeur fournit tous les contenus balisés et assume ainsi la pleine charge des coûts de production, laissant l'intégration et une partie du contrôle qualité à la charge de la plateforme.

XML, entre l'intérêt et la réalité des canaux de distribution

Malgré la pertinence du format XML pour la préservation et les différents services qui peuvent lui être rattachés, peu d'éditeurs semblent enclin à privilégier ce format. Lors de l'enquête menée auprès des éditeurs, UBC Press a confirmé son intérêt pour ce format mais il a été mentionné que les canaux de distribution pour cette technologie ne sont pas adaptés. Il est effectivement beaucoup plus facile d'acheminer des fichiers PDF aux différents distributeurs que des fichiers XML.

Conclusion

- Le processus d'édition numérique menant à une version XML comprend plusieurs niveaux de traitement, du plus simple menant essentiellement à la consignation des métadonnées, au plus complexe offrant un balisage de tous les éléments sémantiques du texte.

- Les formats XML permettent une panoplie de services ne pouvant être offerts par les autres types de formats de restitution, notamment la possibilité de générer statiquement ou dynamiquement les métadonnées dans des formats variés.
- La qualité et le type de balisage entraînent des variations significatives de coût pour la production des livres en format XML. Il importe toutefois de se questionner sur l'importance relative de ces coûts sur l'ensemble du processus d'édition et de commercialisation.
- L'utilisation du XML permet d'offrir d'avantages de services, si bien que la différence de coût entre un balisage minimal et un balisage complet demande à être évaluée dans un contexte beaucoup plus large.
- À ce jour, aucune plateforme de commerciale n'est adaptée à la réalité des livres en format XML, du moins en dehors du champ de l'expérimentation et de l'initiative marginale, comme en fait fois la très faible proportion des livres numériques ayant fait l'objet de ce type de traitement.

7- Les systèmes documentaires numériques comme laboratoires de recherche : enjeux et développements

La production scientifique en sciences humaines est influencée par l'émergence des technologies d'information et communication (TIC). De là procède la création d'un nouveau champ d'intervention et de réflexion scientifique. Certes, les technologies s'infiltrant dans les modes de production scientifique — en servant à la fois comme instrument de communication et de collaboration et en façonnant la production de connaissances —, mais également elles deviennent l'objet même de l'investigation scientifique. L'héritage culturel de notre société change d'artefact grâce au passage de l'analogique au numérique. Cette perspective s'exprime par la définition de ce nouveau champ d'intervention scientifique, dans le *Digital Humanities Manifesto* :

« Digital Humanities is not a unified field but an array of convergent practices that explore a universe in which: a) print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated; instead, print finds itself absorbed into new, multimedia configurations; and b) digital tools, techniques, and media have altered the production and dissemination of knowledge in the arts, human and social sciences ».²⁹

Cette compréhension ne pose pas les Sciences humaines numériques en champ unifié, mais plutôt en ensemble de pratiques convergentes qui ont la technologie comme

noyau commun. L'article *The Landscape of Digital Humanities* est éclairant à cet égard et suggère une voie d'analyse à travers l'identification de différents modèles paradigmatiques d'imbrication des sciences humaines et des technologies d'information ou le numérique. Il en découle que les modèles d'imbrication se traduisent par des approches où la technologie se trouve placée sous différentes perspectives :

- la technologie comme outil ;
- la technologie comme objet d'étude ;
- la technologie comme moyen d'expression ;
- la technologie comme laboratoire de recherche exploratoire ;
- et la technologie comme voie activiste.

Ces modèles ne sont pas mutuellement exclusifs ni radicalement différents, mais peuvent être vus comme des couches qui coexistent et qui sont interdépendantes.³⁰

Dans le court texte qui suit, nous avons conjugué les différents projets de recherches trouvés avec la typologie proposée par Svensson, tout en nous souciant des intersections entre les approches. Tout en mettant en relief les projets qui utilisent un système documentaire numérique comme laboratoire de recherche (qui est un modèle d'imbrication proposé par Svensson), nous avons identifié des chevauchements avec d'autres modèles d'imbrication épistémologique.

²⁹ Presner, Todd, et al. *The Digital Humanities Manifesto 2.0*. UCLA Mellon Seminar in Digital Humanities. http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf.

³⁰ Svensson, Patrick. *The Landscape of Digital Humanities*. Digital Humanities Quarterly, HUMlab, Umeå University, 2010. <http://digitalhumanities.org/dhq/vol/4/1/000080>

Les thématiques

Avant de mettre en relief la typologie des projets de recherche, qui reposent sur des systèmes documentaires numériques dans le domaine des sciences humaines numériques, il nous semble utile de faire un sommaire des sujets abordés par les projets analysés. Nous avons retenu une liste qui repère les vingt principaux projets de recherche issus de différentes institutions.

Cinq thématiques générales se dégagent des projets analysés :

- Gestion et étude de besoins ;
- Génération automatique de métadonnées ;
- Extraction de connaissances ;
- Création d'hyperliens ;
- Études linguistiques.

Vous trouverez une fiche descriptive de chaque projet en annexe de ce rapport.

Dans le volet, « Gestion et étude de besoins », se retrouvent les projets qui ont un intérêt marqué par le côté gestion des systèmes documentaires numériques. Parmi les principaux, nous pouvons citer ceux qui portent une attention particulière sur la façon de gérer les ressources électroniques et les rendre repérables par les utilisateurs

- soit par l'identification de méthodes efficaces, de traitement, préservation et diffusion, employées par les plateformes de publications électroniques,
- soit par l'étude des besoins et des statistiques d'utilisation de ces ressources.

S'y retrouvent aussi les projets qui cherchent à automatiser certaines tâches dans la gestion de grandes bases de données de ressources électroniques.

Pour ces projets, les systèmes documentaires numériques sont utilisés comme laboratoires dans la mesure où ils servent comme plateforme d'expérimentation pour la validation de données de ces recherches ou comme source de collecte d'information pour la recherche en tant que telle.

Dans la deuxième catégorie, « Génération automatique de métadonnées », on retrouve les projets qui cherchent de nouvelles formes de traitement documentaire et de création automatique de métadonnées. À cela s'ajoutent les méthodologies d'évaluation de la pertinence des métadonnées générées automatiquement ; pour ce dernier aspect, la recherche a un intérêt institutionnel dans la mesure où on cherche à justifier le retour sur l'investissement pour la mise en place et la maintenance de systèmes documentaires numériques. Les plateformes utilisées pour ce type de projet servent comme source de collecte et environnement d'expérimentation pour ces projets.

Il faut compter également sur les projets d'« extraction de connaissances » qui sont centrés sur l'analyse statistique de textes à travers des processus de fouille. Il s'agit de recherches visant à développer des requêtes à haute valeur ajoutée, dans des corpus volumineux offerts par de grandes plateformes de publication numérique. Ici la combinaison de différentes ressources électroniques permet la modélisation de présentations multidimensionnelles de l'information d'un corpus donné, par exemple, le projet *Emancipating Digital Data - Scanning and Image Analysis of the Lincoln Papers*.

De plus, d'autres projets traitent de la « Création d'hyperliens » entre les documents d'un système documentaire numérique, ce qui permet notamment la recherche sur la pratique de notation et le référencement entre diverses couches d'information dans les documents numériques en sciences humaines.

Finalement, le volet « Études linguistiques » regroupe les recherches (dont celles menées par LASLA dans la base Opera Latina) qui portent sur l'utilisation des technologies pour la reconnaissance de structures linguistiques, dans le but de développer le domaine de classification automatique de ressources électroniques.

Les modèles d'imbrication

Il convient de caractériser les modèles d'imbrication des sciences humaines et des technologies de l'information (identifiés par Svensson). Cela permet d'expliquer les projets de recherche qui leur sont associés.

La technologie comme outil

L'imbrication où la technologie a le rôle d'outil est plus ou moins évident : il s'agit de recherches qui se servent des fonctionnalités propres aux logiciels afin de créer des plateformes innovantes en termes de recherche, de visualisation de l'information, ou encore d'environnements pédagogiques. Suivant cette perspective, nous trouvons notamment le projet *Emancipating Digital Data - Scanning and Image Analysis of the Lincoln Papers* dont l'objectif est de construire une présentation multidimensionnelle pour permettre la consultation d'une grande collection numérisée d'écrits d'Abraham Lincoln. La recherche d'informations est ici combinée avec une interface dynamique composée de cartes géographiques (*Google Maps*).

Même si le projet est encadré dans un sens instrumental — en utilisant les technologies comme outil pédagogique —, il s'insère aussi dans le modèle de laboratoire exploratoire de recherche (que nous verrons plus loin) puisqu'il permet des études exploratoires sur les caractéristiques des documents et leurs relations avec des événements historiques.

L'aspect instrumental est aussi considéré en termes de mesure dans les recherches scientifiques, c'est-à-dire lorsque le potentiel offert par les technologies peut servir au traitement de données dans une recherche, quoique cette notion soit plutôt connectée avec celle de *computing humanities* qui précède l'avènement des sciences humaines numériques.³¹

Les projets qui utilisent les plateformes numériques à des fins expérimentales peuvent s'inscrire dans cette perspective ; on pense aux projets qui visent l'étude de besoins ou la recherche de modèles de gestion et d'automatisation de tâches propres aux systèmes documentaires numériques. À cet égard, nous pouvons citer le projet *AHDS Projects and MethodDatabase (extension)*, le projet *LAIRAH* et le projet *Service-Oriented Architecture for Preservation and Ingest of digital objects*, les trois menés par les *Arts and Humanities Data Service*. Comme il a été mentionné, ces projets utilisent les plateformes comme sources de collecte d'information et pour des fins d'expérimentation scientifique, ce qui dénote l'aspect instrumental offert par les technologies dans le cadre de ces recherches.

Parmi les caractéristiques propres à cette imbrication, on compte les études d'analyse culturelle, qui cherchent à donner une

³¹ Svensson, Patrick. *The Landscape of Digital Humanities*. Digital Humanities Quarterly, HUMLab, Umeå University, 2010. <http://digitalhumanities.org/dhq/vol/4/1/000080>

interprétation qualitative et offrir une visualisation interactive de l'information à partir de l'analyse de grands ensembles de données. Les retombés de cette tendance sont décrits par Manovich :

« We believe that a systematic use of large scale computation analysis and interactive visualization of cultural patterns will become a major trend in cultural criticism and culture industries in the coming decades. What will happen when humanists start using interactive visualizations as a standard tool in their work, the way many scientists do already ? »³²

Le projet *Books as Expressions of Global Cultural Diversity*, mené par l'OCLC peut nous servir d'exemple de cette tendance, puisque son objectif est de fournir une vision globale de la littérature transnationale et des pratiques particulières de chaque pays sur la publication et la préservation de son héritage littéraire en utilisant les données de la base *World Cat* pour révéler les motifs de la production bibliographique dans une perspective mondiale.

Les projets qui abordent la question linguistique dans le but de faire évoluer les outils de classification automatique et le traitement textuel mettent également à profit l'utilisation des technologies comme instrument de recherche.

La technologie comme objet d'étude

Le modèle d'imbrication qui prend la technologie comme objet d'étude englobe les recherches qui sont centrées sur l'étude de l'émergence de nouvelles technologies. Cette

imbrication est à l'origine de la formation de champs ou de groupes de recherches interdisciplinaires qui se distinguent de ce qui a été convenu d'appeler les sciences humaines numériques, comme c'est le cas des études consacrées à l'Internet (*Internet studies*). Ces dernières, qui pratiquent souvent l'interdisciplinarité et analysent les impacts de nouvelles technologies, ne traitent cependant pas spécifiquement de l'utilisation d'un système documentaire numérique comme laboratoire de recherche ou lieu d'expérimentation.

La technologie comme moyen d'expression

Ce type d'imbrication se caractérise par l'entrecroisement de différents moyens d'expression dans un seul environnement (image en mouvement, texte, musique, 3D, base de données, etc.). Il est plus relié à l'exploitation du potentiel multimodal et interactif des technologies, dans des contextes éducatifs, ainsi que dans le contexte de la communication académique, lorsqu'on cherche à présenter les résultats d'une recherche ou utiliser les technologies dans le cadre d'une conférence, par exemple.

Le projet de recherche *Making the LEAP: linking electronic archives and publications* de l'ADHS peut nous servir d'exemple de ce type d'imbrication dans la mesure où il cherche à examiner comment les publications électroniques peuvent être interactives et comporter différentes couches de données et de moyens d'expression : base de données, GIS, VR, audio et vidéo, etc.

La technologie comme laboratoire exploratoire de recherche

La plupart des projets peuvent être catégorisés comme faisant partie de cette imbrication. Svensson le définit par « l'introduction d'une méthodologie

³² Manovich, Lev. *How to Follow Global Digital Cultures, or Cultural Analytics for Beginners*. Deep Search, ed. Felix Stalder and Konrad Becker. Transaction Publishers (English) and Studienverlag (German), 2009

exploratoire, où le chercheur ou l'étudiant est encouragé à exploiter les matériels, l'ensemble de données ou les questions dans une perspective expérimentale ». Les projets qui cherchent à offrir une interface dynamique et qui permettent la navigation sur les informations et différents modes de visualisation, s'inscrivent dans ce modèle. On pense aux projets *Armadillo: Historical Data Mining*, *Emancipating Digital Data - Scanning and Image Analysis of the Lincoln Papers*, *Digging into the Enlightenment*, *Watching the script* et *With Criminal Intent*.

Le projet TAPoR (TextAnalysis Portal for Research) utilise la technologie comme laboratoire de recherche tant en termes concrets, par l'établissement de différents laboratoires « physiques » de recherche, qu'en termes numériques, par la plateforme que le projet a développée pour servir d'espace exploratoire de recherche.

« Digital platforms such as TAPoR and Second Life can also function as exploratory spaces. When museology and culture analysis students at Umeå build exhibitions in Second Life and produce machinima films, the digital space is used as a laboratory that enables enactments and experimentation that would be quite difficult to facilitate in physical space. Similarly, when a researcher explores a text corpus through a collection of « interactive » tools provided in TAPoR, we may be concerned with a use of technology that extends beyond instrumental tools.³³

On pourrait aller dans le même sens pour le Laboratoire de fouille de textes en sciences humaines numériques (*Digital Humanities and TextMiningLab (DHTML)*) à l'Université de

Montréal. La dimension « physique » du laboratoire comprend l'établissement d'une infrastructure pour une recherche qui vise contribuer à l'avancement des connaissances théoriques et pratiques dans le domaine émergent des sciences humaines numériques. Le système documentaire sur lequel le programme s'appuie sera le corpus documentaire de la plateforme Érudit, qui donne accès à différents types de documents en sciences humaines et sociales : revues savantes, livres, actes, thèses et autres documents et données.

Conclusion

Plusieurs types de recherches en sciences humaines utilisent un système documentaire numérique comme laboratoire de recherche.

- Un accent particulier est mis sur l'aspect exploratoire en termes de méthodologie scientifique. La technologie comme laboratoire est perçue d'abord comme un environnement capable d'offrir un ensemble de variables, qui peuvent être manipulées par le chercheur en sciences humaines, en lui permettant une interaction de haute valeur ajoutée et la visualisation dynamique de leurs interventions.
- Les projets qui portent sur l'extraction de connaissances dans un grand corpus de documents numériques font partie aussi de l'aspect exploratoire, hautement stratégique du point de vue scientifique, dans la mesure où ils offrent la visualisation de motifs, permettant aux chercheurs de valider leurs hypothèses de recherche ou de découvrir les principaux concepts reliés à un domaine de recherche. Il est à noter que l'utilisation des technologies pour des fins méthodologiques a des

³³ Svensson, Patrick. *The Landscape of Digital Humanities*. Digital Humanities Quarterly, HUMlab, Umeå University, 2010. <http://digitalhumanities.org/dhq/vol/4/1/000080>

limites qui concernent chaque recherche en particulier³⁴, mais le potentiel qu'elles offrent demeure indéniable.

- Les projets, pour lesquels la technologie est vue comme un outil, font également appel à des plateformes comme laboratoire de recherche. L'utilisation de plateformes pour la collecte de données peut être considérée comme un type de laboratoire, surtout si le projet en question utilise la plateforme comme lieu d'expérimentation.
- L'analyse des principaux projets de recherche qui exploitent présentement des systèmes documentaires numériques permet d'identifier certaines caractéristiques qui favorisent le traitement des documents numériques. Ainsi, les deux principales dimensions des documents numériques qui ont un impact sur leur exploitation dans une perspective de recherche sont a) la flexibilité des formats des documents et b) la qualité des métadonnées qui leur sont associées.
- *La popularité auprès des éditeurs du format PDF s'avère un obstacle important à l'utilisation des documents numériques à des fins de recherche.* Les textes numérisés en format image et sauvegardés en format PDF sont très souvent inexploitable, même s'il peuvent être soumis à un processus de reconnaissance optique des caractères (OCR). Dans le même ordre d'idées, les documents

PDF contenant du textes sont plus facilement exploitables, mais les éléments de mise en page qu'il contiennent (du texte sur plusieurs colonnes, par exemple) causent des problèmes de traitement qui viennent biaiser l'exploitation que les chercheurs peuvent en faire.

- Par conséquent, le développement d'une infrastructure pour la publication numérique de livres universitaires qui voudrait avoir une valeur ajoutée considérable en permettant l'exploitation des documents à des fins de recherche devrait favoriser l'accès aux documents un format flexible et facilement traitable. Le format XML (TEI, DocBook, etc.) représente actuellement la norme en matière d'encodage des documents pouvant être exploités pour la recherche.
- Par ailleurs, l'absence de métadonnées ou de métadonnées de piètre qualité (données absentes ou incomplètes, par exemple) nuit à l'exploitation des documents numériques. Certains projets (donc les projets TAPoR ou MONK) ont mis en lumière la pertinence d'exploiter des métadonnées de qualités pour l'extraction de connaissances à partir de documents numériques. Il est donc primordial de s'assurer que des métadonnées de qualités sont associées aux ressources numériques dès les étapes de création et de diffusion. Ces métadonnées, en plus de permettre une analyse complexe des documents, favoriseront le moissonnage des documents dans les moteurs de recherche et dans les bases de données, permettant ainsi un meilleur accès aux documents numériques.

³⁴ Svensson, Patrick. *The Landscape of Digital Humanities*. Digital Humanities Quarterly, HUMlab, Umeå University, 2010.
<http://digitalhumanities.org/dhq/vol/4/1/000080>

8- Remarques finales et propositions

Il y a une dizaine d'années, il eut été possible de faire une synthèse semblable portant sur les revues savantes. Le paysage se déployait de façon comparable et posait des enjeux similaires. L'inaction en ce domaine eut été une erreur.

Il y a d'abord eu les revues

L'édition et la diffusion des revues savantes ont déblayé le terrain, ce qui n'est pas négligeable. Mais il faut remarquer d'emblée que les acteurs ne sont pas nécessairement les mêmes dans le monde du livre et l'économie politique du livre savant répond à des comportements et des réflexes particuliers.

Alors que le milieu des revues était caractérisé, à l'échelle nationale, par un morcellement des lieux d'édition (ce qui était déjà moins vrai pour les revues dominantes à l'échelle internationale), l'édition des livres est surtout l'affaire des éditeurs institutionnels, tels les presses universitaires.³⁵ Le paysage des éditeurs de livres est beaucoup moins éclaté.

De même, la structure des coûts d'acquisition chez les acheteurs institutionnels — on pense aux bibliothèques — est fortement déterminée par le prix de l'abonnement aux revues. On pouvait utilement opposer les périodiques numériques aux ouvrages et livres imprimés. Les budgets disponibles étaient objectivement associés aux supports. Or, désormais l'offre des livres numériques commence à se structurer et à s'amplifier et se dispute un espace budgétaire

des bibliothèques entre les revues numériques et les livres imprimés.

Il y a numérique et numérique

Ce n'est qu'une des manifestations d'un mouvement qui est plus large et qui investit tout le secteur de la communication scientifique. Évidemment, le mode d'accès aux résultats de la recherche, dont le support numérique est le plus courant, permet de considérer des méthodes de traitement et d'exploitation des données et des matériaux de recherche avec des protocoles renouvelés.

Si le support numérique présente certaines caractéristiques minimales, encore faut-il bien saisir que l'édition, comme processus de transformation d'un contenu pour une livraison dans les conditions les plus adaptées au genre éditorial, en l'occurrence scientifique, garde toutes ses vertus dans le monde numériques.³⁶

La question s'était posée pour le choix des processus d'édition, de conservation et de diffusion dans le secteur des revues pour la plateforme *Érudit*. Des choix porteurs, il y a plus de dix ans, permettent aujourd'hui d'assurer une diffusion sans autre forme de migration ou d'adaptation, tout en bénéficiant d'une qualité et d'une pérennité des données.³⁷ On voit d'ailleurs bien que la mise en phase de lieux de diffusion diversifiés avec une plateforme comme celle de *Synergies* (www.synergiescanada.org) impose l'observance de normes, la rigueur dans le traitement et l'enrichissement des documents

³⁵ G. Boismenu et G. Beaudry, *Le nouveau monde numérique. Le cas des revues universitaires*, PUM-La Découverte, Montréal, Paris, 2002.

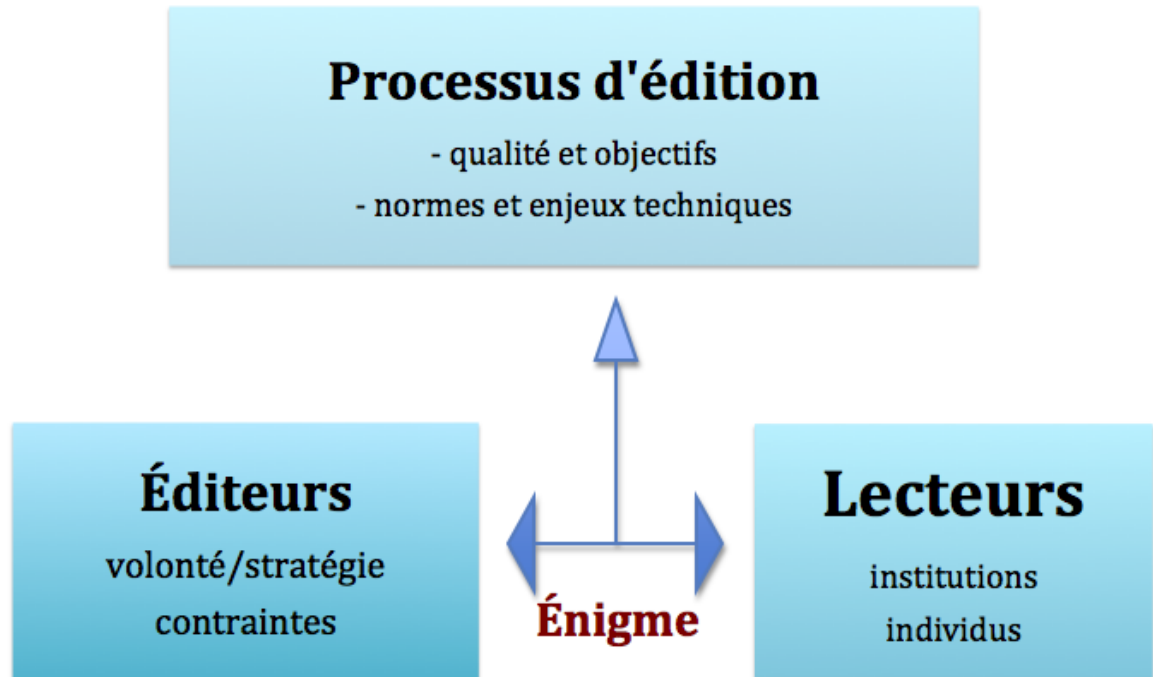
³⁶ G. Boismenu et G. Beaudry, « Du concept à l'outil », dans Claire Brossaud et Bernard Reber (dir.), *Les sciences humaines et sociales à l'heure des technologies de l'information et de la communication*, Éditions Hermès, 2007.

³⁷ G. Beaudry, M. Boucher, T. Niemann et G. Boismenu, « Érudit : le numérique au service de l'édition en sciences humaines et sociales », *Mémoires du livre*, vol. 1, n° 1, 2009.

afin d'assurer une stabilité et une opérationnalisation des services attendus. Les

attentes pour les services ne cessent de croître.

Livre savant numérique en SHS : options, choix, actions



Même si nous ne sommes pas là dans l'hypersophistication des protocoles, il est clair que l'esprit économe en cette matière n'est pas toujours le meilleur conseiller.

Or, nous sommes constamment dans un univers contraint. La décision de s'investir dans le monde numérique pose toujours la question des moyens.

Pour revenir aux revues, le fait de s'engager dans l'édition et la diffusion sous-entend que l'on réunisse les ressources nécessaires pour faire des choix qui s'inscrivent nécessairement dans la durée. À la rigueur, une recherche peut connaître un temps d'arrêt si elle est en panne de financement. C'est dommageable mais pas nécessairement irréparable. Or, une revue savante qui arrête de publier, une maison

d'édition qui met une pause à ses activités, cela ne pardonne pas : c'est un coup fatal à ces institutions dont la continuité des opérations est une caractéristique essentielle.

Les éléments d'une énigme

Une énigme (*puzzle*) se pose d'emblée. Trois grands pôles sont sous tension. L'arbitrage qui sera fait est porteur de sens et de conséquences.

L'édition universitaire n'échappe pas aux contraintes matérielles, car ses seuls revenus ne dégagent pas des marges pouvant permettre un investissement conséquent pour rendre compte des meilleurs standards dans l'édition numérique, une édition propre à rendre possible l'utilisation des plateformes de

diffusion sous la forme de laboratoire de recherche.

On voit que, parallèlement aux équipes de recherche, qui peuvent développer leurs outils de recherche et des fonds documentaires à l'aide de fonds de subvention, il est supposé que l'édition universitaire doit subvenir à ses besoins, y compris dans un contexte de changement technologique majeur. Certes les universités supportent bien souvent leurs maisons d'édition, mais le principe de financement de base est surtout celui des revenus commerciaux générés par leurs activités. Les bibliothèques ou les centres d'édition numérique peuvent puiser bien souvent dans des moyens hors de portée des éditeurs. Cet état de fait définit l'horizon et le champ des possibles de l'édition savante.

Cela contextualise les choix des éditeurs. La priorité est celle de diffuser une version numérique des livres savants, de rendre leurs titres disponibles éventuellement sur plus d'une plateforme, soit, en définitive, de tirer avantage du web en minimisant les investissements lourds.

Parallèlement, dans le monde de l'imprimé, les éditeurs mobilisent les compétences, en faisant affaire avec une constellation de prestataires de services, selon une finalité propre, tout en gardant près d'eux le savoir-faire essentiel, soit la sélection et la préparation du contenu du livre. Ils font, selon toute vraisemblance, de même dans le monde numérique : choisir les meilleurs fournisseurs à prix compétitif.

Cela n'engage pas à un processus innovant et encore moins à des initiatives pour constituer une infrastructure à forte densité technologique. La priorité est ailleurs : substituer les canaux de commercialisation de l'imprimé au numérique. Cette préoccupation tombe sous le sens. On peut difficilement expliquer le recours spontané et quasi-

systématique aux restrictions dans l'utilisation des livres numériques. Avec le recul, on peut voir dans les DRM l'expression d'un passage mal maîtrisé de l'imprimé vers le numérique. Tout cela est plus manifeste dans un environnement d'entreprise à but lucratif, mais la tentation est généralement forte dans d'autres logiques d'entreprise.

On peut estimer que les presses universitaires ont les ambitions de leurs moyens. Cela n'est que raisonnable. Mais la mission des presses universitaires se limite-t-elle à une gestion raisonnable de la contrainte commerciale, ou peut-on imaginer qu'elles puissent déborder sur la préparation, la valorisation, l'exploitation, le meilleur rayonnement, la préservation et la réutilisation des résultats de la recherche, dans un processus de développement de la connaissance.

Cette dernière opposition (gestion raisonnable et gisement de la connaissance) est un peu schématique, mais il découle, de la prise en compte de la marge de manœuvre définie principalement par les activités commerciales, un prisme qui réduit les ambitions, confine plusieurs choix, de telle manière que l'édition savante canadienne semble être consommatrice de services, afin de tirer son épingle du jeu sur la scène numérique en termes de présence sur un marché qui est en expansion. On peut s'arrêter là. Et d'une certaine manière est fait ce qui peut être fait.

Mais cet état de fait établit certaines limites. D'abord, l'éditeur retient les services d'un prestataire pour l'édition numérique et pour la diffusion. La conservation n'est pas précisément assurée, avec les migrations nécessaires le cas échéant. Le format retenu est généralement pauvre et le recours au XML ne semble porter que sur les métadonnées,

dont l'étendue et la richesse restent à démontrer.

Il y a là un paradoxe, dans la mesure où, comme le met en évidence *UBC Press* dans l'entretien que nous avons réalisé avec eux, on voudrait retenir le ePub ou davantage, mais il n'est pas certain que les plateformes de diffusion sont en mesure d'en tirer avantage. Pourtant cela reste assez limité comme ambition. D'où le fait que le format PDF, et bien souvent dans sa version « image », est utilisé à profusion. Avec l'avènement des nouveaux supports de lectures, partant du téléphone intelligent jusqu'à la tablette de lecture, il y a sans doute là une rigidité concernant la portabilité et les fonctionnalités offertes.

De la modestie en prime

L'édition numérique proposée aux éditeurs répond très souvent à des normes relativement modestes. Mise à part l'exploitation restreinte des textes édités par les supports de lecture, cette qualité modeste va de pair avec des limites concernant l'exploitation du fonds documentaire composé par la collection de livres : on pense, par exemple, aux limites pour la recherche qui ne peut pénétrer la structure sémantique des textes. Parallèlement, la pérennité des textes édités n'est pas assurée et suppose une migration future des documents numérisés pour les actualiser en fonction des versions de l'outil logiciel.

La capacité innovatrice des éditeurs universitaires canadiens est à sa plus simple expression, car, en tant qu'utilisateurs de services venant de prestataires externes, cela réduit les risques et la mobilisation de ressources, mais également limite à presque rien le savoir-faire en la matière.

Si cette position se prolonge, elle se perpétuera car la barrière à l'entrée pour ce secteur d'activités sera imposante et dissuasive. On peut penser que se limiter à consommer un savoir-faire et des services venant de tiers est une bonne posture. Si cela était le cas et se généralisait pour d'autres activités, cela placerait le Canada dans une situation particulièrement vulnérable sur l'échiquier numérique en raison d'une passivité et d'une capacité innovante à peu près neutralisée.

Si on ne veut pas généraliser cette position à l'ensemble des secteurs d'activités, la question à laquelle il faudrait répondre c'est pourquoi cette position serait-elle désirable pour le livre savant en sciences humaines et sociales.

L'introduction de limitations à l'utilisation des livres en version numérique, en raison du recours au DRM, heurte les usagers habituels, qu'ils soient institutionnels ou individuels.

Mais c'est également un double frein à l'usage de cette documentation dans les travaux de recherche sous la forme de forage dans des corpus ou d'investigation selon les protocoles des sciences humaines numériques. Pourquoi une double limitation ? C'est que la pauvreté relative des formats utilisés et leurs verrouillages partiels posent autant de freins à une recherche sérieuse dans les corpus. Il faudrait plutôt être en mesure d'utiliser une version des textes basée sur un balisage XML de bon niveau. On est souvent loin du compte.

Il s'agit, en fait, d'une condition assez communément acceptée pour procéder à une édition, diffusion, exploitation et conservation des textes savants, autant pour le livre que pour les revues. Mais une telle avenue exige plus de ressources que les bénéfices immédiats venant de la vente de titres ne sauraient

permettre et justifier. La valeur et l'impact de ce genre de choix se situent au-delà de l'entreprise individuelle, tant pour les ressources nécessaires que pour l'appréciation d'un tel besoin.

Soyons clair. Le coût d'opération pour une édition numérique de qualité, suivant des normes visant la meilleure exploitation du fonds documentaire, n'est pas exorbitant. Le surcoût sur la seule production des fichiers numériques reste dans l'ensemble de l'opération d'édition tout à fait raisonnable.

C'est le coût lié au développement d'un foyer d'expertise et d'une infrastructure de haute valeur ajoutée qui peut être dissuasif. Il s'agit à la fois d'un coût initial et d'un coût de développement pour la suite des choses. Ce n'est pas un hasard si, bien souvent, le financement public ou venant de fondations privées est mis à contribution pour l'implantation d'une telle infrastructure.

Une entreprise concertée

L'un des modèles les plus pertinents pour se doter d'une telle infrastructure est celle de l'entrepôt numérique.

Rappelons qu'il s'agit d'un lieu où l'on produit et héberge des fichiers numériques et où on gère les droits numériques, distribue les livrets et offre éventuellement de l'impression à la demande, ainsi qu'une structure de distribution aux libraires. Le modèle de l'entrepôt permet aux éditeurs d'avoir accès à une infrastructure technologique de pointe et à une main d'œuvre hautement qualifiée. Cela leur permet d'intégrer leurs contenus éditoriaux dans les principaux canaux de diffusion selon les normes en vigueur, afin de respecter une interopérabilité maximale. C'est la définition qui était présentée plus haut dans cette étude.

Il y a peut-être lieu d'y revenir, car on peut imaginer très bien que le contrôle d'un tel entrepôt soit entre les mains d'un consortium d'éditeurs universitaires canadiens. En même temps, chaque maison garde son label et son *branding*, et est totalement autonome concernant ses engagements financiers et commerciaux. La mise à disposition des livres peut se faire par diverses plateformes de diffusion, dont la plateforme maison de l'éditeur s'il le désire. L'entrepôt pourrait disposer d'une plateforme de diffusion non-exclusive, mais cela ne s'impose pas.

L'exemple de *BiblioVault*, cité dans cette étude, illustre bien cette architecture organisationnelle. Cet entrepôt agit pour plusieurs éditeurs. *Chicago University Press* y est lié. Ce lien n'est pas caché, mais l'accès aux livres de *Chicago University Press* se fait facilement par leur site et de façon transparente.

Un entrepôt ne fait pas écran. Il permet plutôt de proposer des services aux éditeurs qui répondent à leurs attentes, de faire évoluer ces services et, fondamentalement, de développer une expertise (canadienne en l'occurrence) en matière d'édition, de traitement et de conservation numériques du livre savant.

Arrimer une dimension recherche

Si ce type d'institution peut permettre une collaboration dans le milieu d'édition universitaire, ce peut également être l'occasion d'arrimer une dimension recherche de premier ordre.

Pour être exploitable par les chercheurs en sciences humaines numériques, il est d'un double avantage d'avoir accès à un fonds documentaire multigenre, qui réunisse les articles de revues et les livres (et sans doute

d'autres genres éditoriaux),³⁸ mais également un fonds édité dans des formats compatibles à la recherche, soit un XML de bonne tenue.

Nous avons fait au Canada des avancées non négligeables en matière d'édition de revues savantes. Imaginons que le travail se poursuive en développant une large collection de livres alimentée par les grands éditeurs universitaires et que, par le biais d'un entrepôt notamment, l'on puisse mettre à disposition des chercheurs cet ensemble avec des passerelles ou par un accès à une vaste collection. Nous aurions là une mine très riche et des perspectives de travail de recherche stimulantes.

Avec une infrastructure semblable et une expertise conséquente, c'est davantage qu'un entrepôt conventionnel dont nous disposons : cet entrepôt prend en soi la forme d'un laboratoire de recherche adaptable et adapté à une variété de protocoles de recherche. Encore faut-il que le corpus et la qualité de la collection correspondent également aux besoins des chercheurs. C'est là que l'on dépasse la mission conventionnelle de l'entrepôt pour les éditeurs et là, dans notre étude, l'exemple qui peut nous inspirer c'est celui de *Humanities E-Book*.

Dans ce projet, la collection, qui est constituée, est l'objet d'une sélection à l'initiative de spécialistes dans les différents champs de la connaissance. On pourrait tout autant, avec un financement de recherche adéquat, identifier les titres et les éditer, selon des normes de qualité, titres qu'il serait bon de réunir selon les besoins de l'investigation. Ce fonds devient un acquis patrimonial et une source d'investigation dans le cadre de

protocoles qui participent à l'avancement de la connaissance.

Le partenariat comme démarche stratégique

Ce genre de collaborations croisées (consortium d'éditeurs et collectifs de chercheurs) pourrait très bien s'inscrire dans un projet en partenariat, par lequel différents objectifs sont atteints dans le cadre d'une démarche intégrée de développement d'une économie du numérique du document universitaire. Parmi les partenaires, il faut pouvoir compter sur les autorités publiques sous une forme ou sous une autre. La volonté politique doit pouvoir infléchir, par des dispositifs adaptés et appropriés, ce mouvement vers la nécessaire inscription des acteurs canadiens dans ce secteur d'activité.

L'espace du document universitaire numérique (les revues comme les livres) est stratégique. Le grand défi pour le Canada, c'est de ne pas laisser le champ libre sans avoir livré bataille. Il est temps de se mettre en « ordre de combat », car le temps joue contre nous. Les acquis se confirment ailleurs. Pour un pays qui contribue autant à la connaissance dans le monde, trouvera-t-on les moyens au Canada d'être des acteurs significatifs dans la diffusion du savoir ?

³⁸ G. Beaudry, « La communication scientifique directe : un nouveau champ éditorial », *Hermès*, n° 57, 2010.

BIBLIOGRAPHIE

- Adema, Janneke, *OAPEN. Overview of Open Access Models for eBooks in the Humanities and Social Sciences*, <<http://project.oapen.org/images/OpenAccessModels.pdf>>, (consulté le 25 septembre 2010).
- Adema, Janneke et Paul Rutten, *OAPEN. Digital Monographs in the Humanities and Social Sciences: Report on User Needs*. <<http://project.oapen.org/images/D315%20User%20Needs%20Report.pdf>>, (consulté le 15 septembre 2010).
- Beaudry, G. « La communication scientifique directe : un nouveau champ éditorial », *Hermès*, n° 57, 2010.
- Beaudry G., M. Boucher, T. Niemann et G.Boismenu, « Érudit : le numérique au service de l'édition en sciences humaines et sociales », *Mémoires du livre*, vol. 1, n° 1, 2009.
- Benhamou, Françoise et Olivia Guillon, « Modèles économiques d'un marché naissant: le livre numérique », *Culture prospective*, <<http://www.culture.gouv.fr/deps>>, (consulté le 5 septembre 2010).
- Berg, Selinda Adelle, et al., « Not on the Same Page: Undergraduates' Information Retrieval in Electronic and Print Books », *The Journal of Academic Librarianship*, vol. 36, no 6, pp. 518-525.
- Bienvault, Hervé, *Le coût D'un livre numérique. Étude réalisée pour le MOTiF (Observatoire pour le livre et l'écrit en Ile-de-France)*, avril 2010, <http://www.lemotif.fr/fichier/motif_fichier/153/fichier_fichier_etude.sur.le.coa.t.d.un.livre.numerique.pdf>, (consulté le 1 octobre 2010).
- Boismenu G. et G. Beaudry, *Le nouveau monde numérique. Le cas des revues universitaires*, PUM-La Découverte, Montréal, Paris, 2002.
- Boismenu G. et G. Beaudry, « Du concept à l'outil », dans Claire Brossaud et Bernard Reber (dir.), *Les sciences humaines et sociales à l'heure des technologies de l'information et de la communication*, Éditions Hermès, 2007.
- Bui, Anh, *HighWire Press 2009 Librarian eBook Survey*, <<http://highwire.stanford.edu/PR/HighWireEBookSurvey2010.pdf>>, (consulté le 10 octobre 2010).
- Consortium universitaire de publications numériques Couperin, « Les 10 commandements de l'éditeur d'e-books: pour une offre idéale », <http://www.couperin.org/index.php?option=com_flexicontent&view=items&id=599:les-10-commandements-de-lediteur-de-books-pour-une-offre-ideale&Itemid=227>, (consulté le 10 octobre 2010).
- Gielen, Nina, *ACLS Humanities E-Book XML Conversion Experiment: Report on Workflow, Costs, and User Preferences*, <<http://www.humanitiesebook.org/heb-whitepaper-2.html>>, (consulté le 16 novembre 2010).
- Greco, Albert N. et Robert M. Wharton. « The Market Demand for University Press Book 2008-15 », *Journal of Scholarly Publishing*, Vol. 42, no 1, octobre 2010, pp. 1-15.
- Grigson, Anna, « Evaluating Business Models for E-Books Through Usage Data Analysis : A Case Study from the University of Westminster », *Journal of Electronic Resources Librarianship*, vol. 21, no 1, pp. 62-63.
- Guillaud, Hubert, « Qu'est-ce qu'un livre numérique? », in *Read/Write Book. Le livre inscriptible*, sous la dir. de Marin Dacos. <<http://cleo.revues.org/147>>, (consulté le 11 octobre 2010).
- Inger, Simon et Tracy Gardner, *How Readers Navigate to Scholarly Content*. <<http://www.sic.ox14.com/howreadersnavigatetoscholarlycontent.pdf>>, consulté le 10 novembre 2010.
- JISC, *JISC National E-Books Observatory Project. Key Findings and Recommendations*, <<http://www.jiscebooksproject.org/reports/finalreport>>, (consulté le 5 novembre 2010).
- Jonchère, Laurent, « Quel avenir pour le livre électronique dans les bibliothèques universitaires françaises? », *Documentaliste-Sciences de l'information*, vol. 42, pp. 26-30.
- Look, Hugh et Frances Pinter, « Open Access and Humanities and Social Science Monograph Publishing », *New Review of Academic Librarianship*, vol. 16, no 1 pp. 90-97. <<http://dx.doi.org/10.1080/13614533.2010.512244>>
- Le Ray, Éric et Jean-Paul Lafrance. *La bataille de l'imprimé à l'ère du papier électronique*. Montréal, Presses de l'Université de Montréal, 2008, 257 p.
- Manovich, Lev. *How to Follow Global Digital Cultures, or Cultural Analytics for Beginners*. Deep Search, ed. Felix Stalder and Konrad Becker. Transaction Publishers (English) and Studienverlag (German), 2009
- Maron, Nancy L. et al., *Sustaining Digital Resources: An On-the-Ground View of Projects Today. Ithaka Case Studies in Sustainability*. <<http://www.jisc.ac.uk/media/documents/publications/general/2009/scaithakaprojectstoday.pdf>>, (consulté le 20 septembre 2010).
- Nawrocki, François et al. « Contexte et offre. Le livre numérique en France: état des lieux et perspectives

- de développement du marché. », *Documentaliste - Sciences de l'information* vol. 47, no 2, 2010, pp. 30-42.
- Presner, Todd, et al.. The Digital Humanities Manifesto 2.0. UCLA Mellon Seminar in Digital Humanities. http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf.
- Steele, Colin, « Scholarly Monograph in the 21st Century : The Future More Than Ever Should Be An Open Book ». *The Journal of Electronic Publishing*, vol. 11, no 2, Spring 2008.
- Svensson, Patrick. *The Landscape of Digital Humanities*. Digital Humanities Quaterly, HUMlab, Umeå University, 2010. <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>
<http://dx.doi.org/10.3998/3336451.0011.201>
- Thompson, John B., *Books in the Digital Age. The Transformation of Academic and Higher Education Publishing in Britain and the United States*. Cambridge, Polity Press, 2005, 468 p.
- Williams, Peter et al., « The Role and Future of the Monograph in Arts and Humanities Research », *Aslib Proceedings : New Information Perspectives*, vol. 61, no 1, 2009, pp. 67-82.
- Center for Digital Scholarship
<<http://library.brown.edu/cds/projects/scholarly-grants>> (consulté le 19 octobre 2010)
- Center for History and New Media – CHNM
<<http://chnm.gmu.edu/>> (consulté le 19 octobre 2010)
- Centre for Open Electronic Publishing - Centre pour l'édition électronique ouverte (Cléo)
<<http://www.revues.org/>> (consulté le 19 octobre 2010)
- Centre Virtuel de la Connaissance sur l'Europe (CVCE)
<<http://www.cvce.lu/mce.cfm>> (consulté le 19 octobre 2010)
- CIRCAh
<<http://www.ucl.ac.uk/infostudies/research/circa/h/>> (consulté le 19 octobre 2010)
- Database machine drawings
<<http://dmd.mpiwg-berlin.mpg.de/home>> (consulté le 19 octobre 2010)
- Electronic Text Research at the University of Saskatchewan - ETRUS
<<http://etrus.usask.ca/?q=content/about-etrus>> (consulté le 19 octobre 2010)
- eScholarship Research Centre
<<http://www.esrc.unimelb.edu.au/>> (consulté le 19 octobre 2010)
- Historisch-Kulturwissenschaftliche Informationsverarbeitung
<<http://www.hki.uni-koeln.de/node/10>> (consulté le 19 octobre 2010)
- Humanities Advanced Technology and Information Institute
<<http://www.gla.ac.uk/departments/hatii/research/digitalhumanities/>> (consulté le 19 octobre 2010)
- Humanities E-Book
<<http://www.humanitiesebook.org/>> (consulté le 19 octobre 2010)
- Institute for Computing in Humanities, Arts, and Social Science - I-CHASS
<<http://www.chass.uiuc.edu/Index/Index.html>> (consulté le 19 octobre 2010)
- Istituto di Linguistica Computazionale
<<http://www.ilc.cnr.it/viewpage.php/sez=organizzazione/id=10/vers=ita>> (consulté le 19 octobre 2010)
- Le Laboratoire d'Analyse Statistique des Langues Anciennes
<<http://www.cipl.ulg.ac.be/Lasla/index.html>> (consulté le 19 octobre 2010)
- Loyola University Chicago's Center for Textual Studies and Digital Humanities
<<http://www.ctsdh.luc.edu/>> (consulté le 19 octobre 2010)

WEBOGRAPHIE

- Arbeitsstelle Computerphilologie
<<http://www.c-phil.uni-hamburg.de/Main/WebHome>> (consulté le 19 octobre 2010)
- Arts and Humanities Data Service UK - AHDS.
Enabling Digital Resources for the Arts and Humanities.
<<http://www.ahds.ac.uk>> (consulté le 19 octobre 2010)
- arts-humanities.net
<<http://www.arts-humanities.net/node/about>> (consulté le 19 octobre 2010)
- Australia New Zealand Digital Encyclopedia Group.
<<http://jod.id.au/anzdeg/collections.html>> (consulté le 19 octobre 2010)
- Australian E-Humanities Gateway
<<http://www.ehum.edu.au/>> (consulté le 19 octobre 2010)
- Australian Scholarly Editions Centre – ASEC
<<http://hass.unsw.adfa.edu.au/ASEC/>> (consulté le 19 octobre 2010)
- Bibliothèques Virtuelles Humanistes
<<http://www.bvh.univ-tours.fr/>> (consulté le 19 octobre 2010)

- Maryland Institute for Technology in the Humanities - MITH
<<http://mith.umd.edu/research/>> (consulté le 19 octobre 2010)
- Matrix: The Center for Humane Arts, Letters, and Social Sciences Online
<<http://www2.matrix.msu.edu/about/>> (consulté le 19 octobre 2010)
- Model Editions Partnership - MEP
<<http://xml.coverpages.org/mep-ceth.html>> (consulté le 19 octobre 2010)
- Networked Infrastructure for Nineteenth-Century Electronic Scholarship – NINES
<http://www.nines.org/about/what_is.html> (consulté le 19 octobre 2010)
- New Zealand Eletronic Text Centre
<<http://www.nzetc.org/tm/scholarly/tei-NZETC-About.html>> (consulté le 19 octobre 2010)
- ODIN <<http://odin.linguistlist.org/>> (consulté le 19 octobre 2010)
- Oxford Text Archive
<<http://ota.ahds.ac.uk/about/>> (consulté le 19 octobre 2010)
- Perseus Digital Library
<<http://www.perseus.tufts.edu/hopper/>> (consulté le 19 octobre 2010)
- Research in Computing for the Humanities
<<http://www.rch.uky.edu/index.html>> (consulté le 19 octobre 2010)
- Research Libraries Group
<<http://www.oclc.org/research/about/default.htm>> (consulté le 19 octobre 2010)
- Scholars' Lab, University of Virginia Library
<<http://www.scholarslab.org/>> (consulté le 19 octobre 2010)
- TAPoR
<<http://tapor.humanities.mcmaster.ca/home.html>> (consulté le 19 octobre 2010)
- Textgrid: community grid for collaborative editing, annotation, analysis, and publication of special
<<http://www.textgrid.de/en/ueber-textgrid.html>> (consulté le 19 octobre 2010)
- The National Science Digital Library
<<http://www.nines.org>> (consulté le 19 octobre 2010)
- Uni Digital
<<http://digital.uni.no/>> (consulté le 19 octobre 2010)
- Wired Humanities Project
<<http://whp.uoregon.edu/>> (consulté le 19 octobre 2010)
- Zenon Dai
<<http://opac.dainst.org/F/AISAEMCC7DHQPS>>
- KTGMJUG4VVVXNIK56HR56AIC5MLXSQPF9
HVM-70678?func=option-update-
lng&p_con_lng=eng> (consulté le 19 octobre 2010)
- Digital Humanities Observatory, Royal Irish Academy - DHO
<<http://dho.ie/resources>> (consulté le 19 octobre 2010)
- Nouvelles technologies, Nouvelles textualités - NT2
<<http://nt2.uqam.ca/>> (consulté le 19 octobre 2010)
- Humanities Research Institute - HRI
<<http://www.shef.ac.uk/hri/projects/completepps.html>> (consulté le 19 octobre 2010)
- centerNet <<http://digitalhumanities.org/centernet/>> (consulté le 19 octobre 2010)

Annexe 1 : Agrégateurs et entrepôt

ACLS Humanities E-Book	
Modèle organisationnel et unité institutionnelle	<ul style="list-style-type: none"> • American Council of Learned Societies (ACLS) http://www.acls.org/ • Personnel de l'ACLS composé de deux directeurs, trois rédacteurs en chef http://www.humanitiesebook.org/hebstaff.html • L'ACLS est composé de 71 sociétés membres : http://www.acls.org/societies/learnedsocieties.aspx?id=136&linkidentifiant=id&itemid=136
Projet éditorial	<ul style="list-style-type: none"> • Depuis 2002 • Environ 2,800 titres en texte intégral • Collaboration de 20 sociétés savantes de l'ACLS (http://www.humanitiesebook.org/societies.html), près de 100 éditeurs (http://www.humanitiesebook.org/publishers.html), et les bibliothécaires du University of Michigan's Scholarly Publishing Office (http://www.humanitiesebook.org/spo.html) • Comité éditorial du HEB (http://www.humanitiesebook.org/hebadvisoryboard.html) • Environ 500 titres sont ajoutés chaque année • La collection comprend à la fois des livres disponibles et épuisés (87% en librairie, avec droit d'auteur), datant de 1820 jusqu'à nos jours (nouveauautés, rééditions, domaine public) http://www.humanitiesebook.org/titlelist.html • HEB se voit accorder une sous-licence; les éditeurs originaux de l'imprimé ou les auteurs, dans le cas d'ouvrages épuisés, demeurent les détenteurs principaux du droit d'auteur • Sciences humaines et sociales (SHS), comprenant les domaines du savoir suivants (non traduits) : <ul style="list-style-type: none"> ○ Area studies: Australasian/ Oceanian, Byzantine, Canadian, Caribbean, Jewish Studies, Native Peoples of the Americas, Women's Studies ○ History: African, American, Asian, Comparative/World, Eastern European/Russian, Economic, European, Latin American, Law, Medicine, Methods/Theory, Middle East, and Science/Technology ○ Other fields: Archaeology, Art and Architectural History, Biblical Studies, Bibliographic Studies, Film and Media Studies, Folklore, History of the Book,

ACLS Humanities E-Book	
	<p>Linguistics, Literature, Literary Criticism, Musicology, Performance Studies (theater, music, dance, performance), Philosophy, Political Science, Religion, and Sociology)</p> <ul style="list-style-type: none"> • Leurs critères de sélection : <ul style="list-style-type: none"> ○ Qualité et utilité pour l'enseignement et la recherche ○ Travaux qui demeurent essentiels aux chercheurs et étudiants universitaires ○ Travaux qui sont fréquemment cités dans la littérature ○ Basée sur la qualité et la longévité des travaux ○ Fait partie d'un réseau scientifique de titres
Modèle d'affaires et modalités d'accès	<ul style="list-style-type: none"> • Les éditeurs ne sont tenus que de fournir une copie du livre (imprimé ou PDF), aucun frais ne leur étant demandé; HEB couvre les frais de la numérisation • 50% des revenus d'abonnements sont divisés parmi les détenteurs des droits d'auteurs en fonction du nombre de requête que chaque titre reçoit • Abonnements annuels (individuels ou institutionnels) ou achat d'une licence (10 ans d'accès aux titres actuels et à venir) • Les prix sont basés sur la taille de l'institution • La collection n'est pas divisée en panier/bouquet. Une seule collection est offerte à la clientèle • Les abonnements individuels sont offerts uniquement aux personnes qui sont membres d'une des 70 sociétés de l'ACLS • Prêt (consultation à distance) • Aucune restriction de type DRM • Il est possible d'enregistrer et d'imprimer certains ouvrages (390 sur 2,800) localement (http://www.humanitiesebook.org/pod-titlelist.html), trois pages à la fois (PDF) • Copier/coller disponible sur l'ensemble de la collection • Aucune fonction d'annotation ou de surlignage • Ajustement à la vue possible selon certains pourcentage (50% à 400%), pour les pages « image » • Recherche dans le texte intégral d'un ouvrage • Lien vers les critiques du livre sur d'autres portails • "Scholarly Fair Use rights & restrictions" pour l'impression et le téléchargement

ACLS Humanities E-Book	
Processus d'édition numérique	<ul style="list-style-type: none"> • Formats de restitution (PDF, XML [TEI avec sortie HTML], Image [GIF, PNG], Texte [OCR]) • Utilisation d'un modèle de données interne pour le balisage en XML, soit la DTD <code>acls-hebook.dtd</code>, version 1.5 (http://www.humanitiesebook.org/xml/doc/acls-hebook-doc.html) • Diversité des médias pour ce qui est des livres en format XML [78 titres] (http://www.humanitiesebook.org/xml-features.html) • Titres liés à leurs critiques dans JSTOR, Project MUSE, History Cooperative et ailleurs en ligne • Le flot du texte est fixé à un miroir d'écran qui ne peut être modifié • La taille du texte peut être modifiée ce qui modifiera légèrement le flot du texte, comme le miroir d'écran est fixé • N'offrent pas de support pour les appareils portatifs (iPod, iPad, Kindle, etc.) mais ils ont effectués un rapport et un sondage sur la question (http://www.humanitiesebook.org/heb-whitepaper-3.html) • Les titres XML sont liés à des titres connexes de la collection, d'archives externes, de sites Web, et d'autres ressources • Métadonnées (page couverture, titre de l'ouvrage, titre des chapitres, auteur, année de parution, éditeur, site Web de l'éditeur, pagination, mention de droit d'auteur)
Services de la plateforme	<ul style="list-style-type: none"> • Fonctionnalités de recherche (simple, booléenne, proximité, bibliographique) • Champs de recherche (titre, auteur, sujet, éditeur, texte intégral) • Filtre par publication en série (series) • Historique de recherche • Statistiques d'utilisation téléchargeables gratuitement • Notices MARC téléchargeables gratuitement, pour l'ensemble de la collection • Il est possible d'enregistrer et d'imprimer certains ouvrages (390 sur 2,800) localement (http://www.humanitiesebook.org/pod-titlelist.html), trois pages à la fois (PDF) • HEB offre des services de conversion de fichiers vers les formats de diffusion, d'hébergement de fichiers, et de commercialisation
Commentaires	<ul style="list-style-type: none"> • Interface très minimaliste, qui ne semble pas avoir fait le passage au Web 2.0, avec un graphisme désuet, non optimisé, et aucune fonctionnalité

ACLS Humanities E-Book

récente ou particulièrement avant-gardiste

- Au lieu d'avoir une navigation par discipline, ils proposent l'ensemble de la classification par sujet (vedette-matière) ce qui est très lourd, très spécifique (sans aucun regroupement de haut niveau) et qui est en fait une recherche qui est effectuée lorsqu'on clique sur un sujet et qui ramène à des résultats de recherche
- Apparition d'icônes à côté des titres sans qu'ils soient présents dans la légende, pour qu'on sache à quoi ils réfèrent (HEB Special Series, voir HELP)
- Numérotation des paragraphes pour les ouvrages qui sont diffusés sous forme textuelle (XML), des pages pour les ouvrages diffusés à l'aide d'images

Cairn	
Modèle organisationnel et unité institutionnelle	<ul style="list-style-type: none"> • Né de la volonté de quatre maisons d'édition (Belin, De Boeck, La Découverte et Erès) • En février 2006, la Bibliothèque nationale de France s'est associée à ce projet • Différents investisseurs institutionnels, notamment Gesval, la société ayant en charge la gestion des participations de l'Université de Liège • En France, le Centre national du livre apporte son soutien à cette initiative
Projet éditorial	<ul style="list-style-type: none"> • Depuis 2010 (pour les autres genres que les revues) • 647 ouvrages collectifs provenant de 87 collections, 745 ouvrages de référence de la collection « Que sais-je ? » en plus des dossiers de « L'état du monde » et des recensions de « La bibliothèque idéale des sciences humaines » • La collection ne semble contenir que des livres disponibles à la vente, datant de 1992 jusqu'à nos jours (nouveauautés, rééditions) • Disciplines couvertes : <ul style="list-style-type: none"> ○ Droit ○ Économie, Gestion ○ Géographie ○ Histoire ○ Lettres et linguistique ○ Médecine ○ Philosophie ○ Psychologie ○ Sc. de l'information ○ Sciences et techniques ○ Sciences de l'éducation ○ Sciences politiques ○ Sociologie et société ○ Sport et société • Il n'est pas précisé si un comité éditorial est en place mais il est mentionné que les « services sont proposés non seulement aux publications des quatre maisons d'édition à l'initiative du projet mais également à tous les acteurs, quel qu'en soit le statut, ayant en charge l'édition et/ou la diffusion de publications de sciences humaines et/ou sociales <u>de qualité</u> ». On présume donc qu'un comité éditorial est en place pour discriminer les publications dites « de qualité » des autres
Modèle d'affaires et modalités d'accès	<ul style="list-style-type: none"> • Les ouvrages collectifs et les ouvrages de référence sont proposés aux institutions dans le cadre de « bouquets » (bouquets thématiques ou

Cairn	
	<p>bouquets interdisciplinaires) auxquels les institutions peuvent s'abonner sous forme de licences d'accès forfaitaire (le nombre de consultations étant alors illimité)</p> <ul style="list-style-type: none"> • Le prix des bouquets n'est pas mentionné dans le site. Par contre, il est possible d'acheter les chapitres, à la pièce, ce qui semble coûter environ 3-5 € le chapitre • Les utilisateurs ont la possibilité de consulter ou d'imprimer les chapitres, et éventuellement de les reproduire, pour autant évidemment que cela s'effectue dans les limites de la réglementation en vigueur, concernant le droit d'auteur notamment • Abonnement (individuel, institutionnel) • Ne semble avoir aucune protection des données (DRM) • Copier/coller [PDF texte/image, HTML], pas d'annotation possible, pas de surlignage, à l'exception des termes recherchés, dans les résultats de recherche (possibilité de le désactiver), ajustement à la vue, recherche en plein texte, etc.
Processus d'édition numérique	<ul style="list-style-type: none"> • Formats de restitution (PDF, HTML) • Ce n'est pas précisé mais la DTD Erudit Article 3.0.0 (XML) semble être utilisée également pour les chapitres d'ouvrages • Un autre modèle de données a été utilisé dans le cas de L'état du monde dont la structure est unique par rapport aux autres ouvrages, mais qui ne cadre pas vraiment dans ce qui risque d'être publié, à court terme du moins, dans Érudit • Comme il s'agit essentiellement de rendre en format numérique, des livres publiés en format papier, le seul média géré est le format image • Ne précise pas les appareils portatifs (iPod, iPad, Kindle, etc.) qu'ils supportent • Métadonnées (titre de l'ouvrage, rédacteurs en chef, titre du chapitre, auteur, affiliation, pagination, résumé [ou premières lignes], intertitres du chapitre [plan de l'article], collection, éditeur, lieu de publication, année de publication, ISBN)
Services de la plateforme	<ul style="list-style-type: none"> • Fonctionnalités de recherche (simple, avancée) • Champs de recherche (titre, résumé, bibliographie, auteur, texte intégral, éditeur, type de publication, discipline, collection, année de publication [étendue]) • Filtres (termes associés, types de publication, disciplines, collection, date de parution) • Feuilletage en ligne (outil de type Acrobat)

Cairn	
	<ul style="list-style-type: none"> • Aucun service d'impression à la demande • Sauvegarde des notices (Ma bibliographie) • Leur principale mission étant la diffusion des œuvres SHS de leurs éditeurs partenaires, leur offre de services tourne autour de ça, soit l'hébergement, la promotion et la commercialisation des collections qu'ils ont ainsi constituées • Accessoirement, ils prennent en charge le travail de numérisation/conversion numérique, quand l'éditeur concerné le souhaite mais ce n'est pas obligatoire • À ce stade, ils ne recourent pas aux DRM; par contre, ils peuvent proposer aux éditeurs qui le désirent un service de tatouage numérique (watermarking) de leurs œuvres
Commentaires	<ul style="list-style-type: none"> • Ambiguïté au niveau des étiquettes de certains services (plan de l'article --> plan du chapitre) • Parfois le lien vers un service est disponible alors qu'il renvoie à une page sans contenu, par exemple, « Premières lignes » ce qui peut être frustrant • La navigation dans les chapitres offre la possibilité de voir directement les autres ouvrages d'un même auteur ou qui ont un sujet proche, qu'il s'agisse d'un ouvrage ou d'un article de revue • La barre de service qui se promène à gauche lors du défilement agace plus qu'elle n'aide. Par contre, c'est pratique d'avoir certains services à portée de clicks constamment (impression, taille de caractères, haut de page) • Certains services (PDF, ajouter à ma bibliographie) ne sont pas systématiquement disponibles pour les ouvrages collectifs et de référence • L'outil de feuilletage en ligne offert ressemble plus à un lecteur de fichiers PDF/image conventionnel. Il ne s'ajuste pas automatiquement à l'écran, ne donne pas l'illusion de « tourner les pages » et la consultation n'est pas chose facile (cadrage se perd d'une page à l'autre, piètre lisibilité, lecture par chapitre plutôt que pour l'ouvrage entier) • Fonctionnalité « Lire la suite... » qui offre d'ajouter au panier d'achat, soit le chapitre en cours, soit l'ouvrage en entier, en précisant le prix pour chacun

NetLibrary	
Modèle organisationnel et unité institutionnelle	<ul style="list-style-type: none"> • Une division d'EBSCO Publishing
Projet éditorial	<ul style="list-style-type: none"> • Depuis 2002 (achat par OCLC) • Plus 220 000 livres numériques, plus de 16 000 livres numériques sonores, et plus de 6000 livres compatibles avec les iPods • Le corpus est constitué de livres du domaine public, livres disponibles à la vente, livres épuisés, nouveautés, rééditions • Disciplines couvertes (non traduites) : <ul style="list-style-type: none"> ○ Business, economics and management ○ Computer science & information technology ○ Criminal justice & criminology ○ Education ○ Engineering/technology ○ Literature ○ Management & leadership ○ Medical (Allied Health, Nursing, Pharmacology) ○ Personal development ○ Reference ○ Social sciences ○ Teacher's resources • Politique de développement de collection (http://library.netlibrary.com/CollectionDevelopmentPolicy.aspx?ekmense1=18_submenu_38_link_3)
Modèle d'affaires et modalités d'accès	<ul style="list-style-type: none"> • Les fonctionnalités disponibles dépendent des conditions d'achat de chaque bibliothèque inscrite qui possèdent du contenu numérique ainsi que des restrictions propres à chaque éditeur • Portion de la collection en accès libre (3461 livres) et le reste par achat • Achat par la bibliothèque mais consultation à distance et téléchargement pour lecture sur Adobe Digital Editions pour les utilisateurs de la bibliothèque • Possibilité de protection des données (DRM) avec l'application Adobe Digital Editions • Impression (en partie) afin de ne pas violer le droit d'auteur (15 pages/heure si utilisateur connecté) • Copier/coller (en partie) afin de ne pas violer le droit d'auteur. Si le système constate un abus possible, un message d'avertissement apparaîtra

NetLibrary	
	<ul style="list-style-type: none"> • Au maximum 4 livres peuvent être consultés en même temps par un même utilisateur • Possibilité de check out/check in afin de télécharger pour consultation du document hors ligne • Annotation possible • Dictionnaire intégré • Lecture page par page ou chapitre par chapitre • Recherche en plein texte • Ajustement à la vue à l'aide du zoom du navigateur ou de l'application de lecture PDF
Processus d'édition numérique	<ul style="list-style-type: none"> • Formats de restitution (PDF, Djvu) • Ils fournissent une documentation aux éditeurs pour qu'ils puissent fournir les livres en format OeB (ancien format ePub) mais ça date de 2002 et il n'en est pas fait mention en nulle part ailleurs • Livres sonores disponibles (WMA protégé, MP3) • Adaptation aux différents types d'écrans et supports matériels (nook [B&N], Sony Digital Readers, COOL-ER [Interead]) et tous les appareils qui supportent l'Adobe Content Server (http://blogs.adobe.com/digitalpublishing/supported-devices) • Ils fournissent une documentation détaillée (qui ne semble pas à jour par contre) aux éditeurs sur la procédure à suivre pour fournir les livres (http://extranet.netlibrary.com/extranet/publisher/specifications/) ce qui nous permet d'en déduire qu'ils n'en font pas la production eux-mêmes mais offrent plutôt un service de soumission de livres pour publication dans NetLibrary • Métadonnées (titre de l'ouvrage, titre des chapitres, auteur, éditeur, lieu de publication, année de publication, langue, ISBNs, sujets, ID du produit)
Services de la plateforme	<ul style="list-style-type: none"> • Fonctionnalités de recherche (simple, avancée) • Champs de recherche (Texte intégral, Titre, Auteur, Mot-clé, Sujet, Éditeur, ISBN, Langue [fr, en]) • Limitations de recherche (Année de publication, Genre [livres, livres sonores, revues]) • Statistiques d'utilisation disponible (http://library.netlibrary.com/LibraryResourceCenter.aspx) • Notices Marc disponibles pour l'ensemble des livres numériques • Aucun service d'impression à la demande • La plateforme est disponible en 10 langues
Commentaires	<ul style="list-style-type: none"> • Interface offerte en 10 langues, ce qui est peu commun dans les plateformes explorées. Est-ce

NetLibrary

qu'ils ont vraiment une clientèle qui le justifie, dans la mesure où le contenu offert n'est qu'en français et anglais

- Lorsqu'on paramètre la recherche pour retourner tous les ouvrages, et non pas uniquement ceux de la collection de l'UdeM, ça nous renvoie un message ambigu lorsqu'on tente de le consulter, soit qu'il est déjà utilisé présentement par un autre utilisateur, et l'interface offre la fonctionnalité d'être averti quand il sera de nouveau disponible. Si on accepte, il s'ajoute même dans notre profil. Cependant, plus bas, il est indiqué que l'ouvrage ne fait pas parti de la collection de notre bibliothèque, ce qui signifie qu'il ne sera jamais disponible pour nous, mais ce n'est pas clair du tout
- C'est bien d'avoir la table des matières de l'ouvrage à côté du texte intégral (elle se réduit sur le côté, si non nécessaire), avec les différents niveaux qui s'ouvrent et se ferment au besoin, pour pouvoir naviguer plus facilement
- Intéressant de pouvoir ajouter des notes à un ouvrage mais je ne sais pas à quel point c'est utile comme ce sont des notes qui ne s'appliquent pas à une portion du texte mais à l'ensemble de l'ouvrage
- Ça manque qu'il n'y ait pas de possibilité de naviguer par discipline et d'avoir une bonne vision des orientations de la collection achetée par l'Université de Montréal
- Ça serait bien de pouvoir suggérer des ajouts aux bibliothécaires de l'Université par l'intermédiaire de NetLibrary
- Pratique de pouvoir limiter la recherche aux documents dans une langue en particulier même si dans les faits, les ouvrages ne peuvent être qu'en anglais ou en français
- Le fait d'avoir à « fermer » un livre pour qu'il redevienne disponible aux autres usagers n'est pas très intuitif, surtout que l'option n'est pas présentée clairement dans l'interface

MyiLibrary	
Modèle organisationnel et unité institutionnelle	<ul style="list-style-type: none"> • Partie intégrante des solutions proposées par Ingram Digital, pour l'industrie du livre : http://www.ingramdigital.com/ • Une compagnie d'Ingram Content Group : http://www.ingramcontent.com/
Projet éditorial	<ul style="list-style-type: none"> • 250 000 titres de livres (nouveau titres, rééditions) • Ajout d'environ 5 000 nouveaux titres par mois • En relation avec plus de 18 000 éditeurs • MyiLibrary Audio donne accès à des milliers de livres sonores http://www.ingramlibrary.com/myilibrary/default.aspx → n'a pas réussi à trouver comment y accéder. À croire que le service n'est pas encore disponible • Disciplines : <ul style="list-style-type: none"> ○ General Works ○ Philosophy. Psychology. Religion ○ Auxiliary Sciences of History ○ World History and History of Europe, Asia, Africa, Australia, New Zealand, etc. ○ History: America ○ Geography. Anthropology. Recreation ○ Social Sciences ○ Political Science ○ Law ○ Education ○ Music and Books on Music ○ Fine Arts ○ Language and Literature ○ Science ○ Medicine ○ Agriculture ○ Technology ○ Military Science ○ Naval Science ○ Bibliography. Library Science. Information Resources (General)
Modèle d'affaires et modalités d'accès	<ul style="list-style-type: none"> • Options d'achat flexibles : par titre ou par collection, à perpétuité ou par abonnement. Comprend des offres particulières telles que des prêts interbibliothèques et des options de paiement multiples telles que par éditeurs ou par disciplines. • Automatisation des achats : outils simples et efficaces pour automatiser l'acquisition de nouveau contenu • Plusieurs niveaux d'accès : une grande variété de méthodes d'authentification/d'accès contrôlés et

MyiLibrary	
	<p>sécurisés par l'utilisation de DRM appliqués à l'ensemble du contenu</p> <ul style="list-style-type: none"> • Téléchargement (en partie) : 10 pages à la fois • Impression (en partie) : 80 pages à la fois • Fonctionnalités : Copier/coller, annotation, ajustement à la vue, recherche en plein texte, etc.
Processus d'édition numérique	<ul style="list-style-type: none"> • Ils semblent offrir uniquement un service de diffusion et non pas de production • Formats de restitution (PDF, OeB) • Pas vraiment d'adaptation aux différents types d'écrans (<i>reflowable text</i>) et supports matériels (iPod, iPad, Kindle, etc.) • Comme il s'agit essentiellement de rendre en format numérique des livres publiés en format papier, le seul média géré est le format image • Métadonnées (titre de l'ouvrage, titre des chapitres, auteur, éditeur, EAN/ISBN [hardcover, paper, electronic, MIL], LCC, BIC, Dewey, lectorat visé, résumé)
Services de la plateforme	<ul style="list-style-type: none"> • Fonctionnalités de recherche (simple, avancée) • Champs de recherche (texte intégral, Titre, Auteur, mot-clé, éditeur, EAN/ISBN, LCC, sujet MESH, année de publication, langue) • Filtres (éditeur, sujet, première lettre du titre) • Résultats (surlignage des termes recherchés) • Notices MARC de qualité • Conforme à la norme COUNTER pour les rapports de statistiques de consultation des livres numériques • Aucune impression sur demande • Exportation des notices dans RefWorks et EndNote • Référence bibliographique d'un ouvrage fournie dans différents formats bibliographiques (APA, Chicago, Harvard, MLA, Vancouver) • Offrent une liste importante d'ouvrages de références (dictionnaires, encyclopédies, etc.)
Commentaires	<ul style="list-style-type: none"> • Lorsqu'on est dans la page d'accueil, plusieurs options de connexion nous sont offertes. Il n'est pas évident du tout qu'il suffit de cliquer sur « Search & View Login » pour être connecté automatiquement par l'intermédiaire de l'Université, sans avoir à saisir de nom d'utilisateur ou de mot de passe • La navigation n'est constituée, dans les faits, que de requêtes de recherche qui renvoient une liste de résultats (filtres par sujet LC, par éditeur, par titre) • Je trouve peu pertinent le filtre par le premier

MyiLibrary

caractère du titre

- L'affichage épuré des informations d'un ouvrage, en la séparant par des onglets est plus intéressante car on cible plus facilement ce qu'on cherche, sans être assailli d'informations inutiles
- Les notes que l'on ajoute à un ouvrage réfèrent à une page précise ce qui est plus pertinent que de ne référer qu'à l'ouvrage en entier
- La fonctionnalité « Go to page » est intéressante car en plus de permettre d'entrer un chiffre manuellement, elle offre une liste déroulante des pages possibles, à partir du premier chiffre saisi
- Il est possible de n'afficher qu'un style de référence bibliographique, lorsqu'on possède un compte usager
- Ouvrages en italien, espagnol, allemand, portugais, néerlandais
- Possibilité d'effectuer une recherche uniquement sur le critère de la langue, ce qui peut être pratique quand on veut connaître l'étendue de la collection dans une langue précise

OAPEN	
Modèle organisationnel et unité institutionnelle	<ul style="list-style-type: none"> • Regroupement d'universités et de presses universitaires européennes • Partenaires : <ul style="list-style-type: none"> ○ 9 principaux partenaires (http://project.oapen.org/partners.asp) qui sont tous actifs dans le mouvement de l'accès libre ○ Comité du consortium (9 membres) ○ Comité éditorial composé de 13 chercheurs provenant de différents pays d'Europe (et un du Canada) ○ Groupe de consultants externes (http://project.oapen.org/external.asp)
Projet éditorial	<ul style="list-style-type: none"> • Lancement officiel de la bibliothèque numérique en octobre 2010 (le projet existe depuis 1996, créé par le National Academies Press) • Collection évaluée par les pairs (OAPEN s'assure que les évaluations effectuées par les éditeurs partenaires correspondent à des critères minimaux déterminés par les membres du projet) • 724 livres disponibles • Réseau de 35 partenaires (http://project.oapen.org/partner_network.asp) • La fréquence et la quantité de nouveaux titres ne sont précisées nulle part dans le site ce qui est peu surprenant comme le projet vient d'être lancé • Titres de 2010 disponibles • Plus vieux livre date de 1969 • Disciplines (essentiellement en Sciences humaines et sociales) : <ul style="list-style-type: none"> ○ Biography and True Stories ○ Computing and information technology ○ Earth sciences, geography, environment, planning ○ Economics, finance, business and management ○ English language teaching (ELT) ○ Fiction and related items ○ Humanities ○ Language ○ Law ○ Lifestyle, sport and leisure ○ Literature and literary studies ○ Mathematics and science ○ Medicine ○ Reference, information and interdisciplinary subjects

OAPEN	
	<ul style="list-style-type: none"> ○ Society and social sciences ○ Technology, engineering, agriculture ○ The arts
Modèle d'affaires et modalités d'accès	<ul style="list-style-type: none"> • Projet en accès libre • Tout dépendant du livre, les droits diffèrent, référant à l'occasion à une licence Creative Commons • Modèle de publication provenant de la collaboration des institutions et éditeurs universitaires qui fournissent des fonds pour la publication • Les PDF peuvent être imprimés en entier sans problème tant que les droits annoncés pour le livre sont respectés • Copier/coller et autres fonctionnalités propres au format PDF sont disponibles
Processus d'édition numérique	<ul style="list-style-type: none"> • PDF (texte avec polices incorporées pour les plus récents, texte/image pour les plus anciens) ou TEI (XML) aussi offert (en naviguant au hasard, je n'en ai pas trouvé) • Métadonnées basées sur le Dublin Core, avec 19 champs disponibles, dont 10 sont obligatoires (http://project.oapen.org/specifications.asp)
Services de la plateforme	<ul style="list-style-type: none"> • Essentiellement, ce service s'adresse aux éditeurs associés aux universités européennes afin d'augmenter la visibilité de leurs livres et ainsi augmenter les ventes de la version imprimée • Offre à la fois une plateforme de publication et une bibliothèque numérique • Navigation par titre ou par auteur • Champs de recherche (titre, auteur, année, sujet) • Filtres de recherche (éditeur, sujet, date)
Commentaires	<ul style="list-style-type: none"> • Il est étrange qu'un des rares sites qui offrent des ouvrages en plusieurs langues (italien, allemand, néerlandais, anglais, français, espagnol, danois, etc.) n'offre aucun filtre par langue et qu'un projet européen de cette envergure n'offre pas d'interface dans toutes les langues représentées.

Entrepôt numérique

BiblioVault	
Modèle organisationnel et unité institutionnelle	<ul style="list-style-type: none"> • The University of Chicago Press http://www.press.uchicago.edu/ • Chicago Distribution Center http://www.chicagodistributioncenter.org/ • Chicago Digital Distribution Center http://cddc.uchicago.edu/
Projet éditorial	<ul style="list-style-type: none"> • Depuis 2001 • Le dépôt répond aux besoins de plus de 70 presses universitaires et comprend les fichiers numériques de plus de 20,000 livres • Livres disponibles pour la vente uniquement (comprenant des nouveautés), dont il n'est pas possible de discriminer ceux qui sont disponibles en version numérique de ceux qui ne le sont qu'en version imprimée • Aucune navigation par discipline ce qui rend plus ardu de trouver celles couvertes par la plateforme, mais à priori, la couverture semble privilégier les SHS. Voici la liste des disciplines dont il existe un extrait de livre (non traduit) : Anthropology, Art and architecture, Biography, Cartography and geography, Computer and information studies, Economics and business, Education, Fiction, History, Law, Literature and criticism, Music, Philosophy, Poetry, Politics and current events, Reference and writing, Religion, Science, Sociology • Toutes les presses universitaires peuvent devenir membre de la plateforme et voir une partie de leur collection intégrée • Sans comité éditorial
Modèle d'affaires et modalités d'accès	<ul style="list-style-type: none"> • Accès sous condition d'achat uniquement • Modalités de fixation des prix variables. Souvent le même prix ou un peu moins que la version « paper » du livre, avec parfois, une possibilité d'emprunter le livre pour une durée déterminée, à prix moindre • Achat (téléchargement) sur le site des presses concernées avec consultation partielle de certaines pages du livre (PDF, HTML, Google Preview), dans la majorité des cas (la quantité de pages disponibles varie beaucoup d'une presse à l'autre, d'un livre à l'autre) • Possibilité de protection des données (DRM) avec les applications Adobe Digital Editions (pdf, epub)

BiblioVault	
	<p>et MobiPocket Reader (mobi, prc)</p> <ul style="list-style-type: none"> • Impression (interdite, ou permise, généralement en partie, selon les droits mis en place, selon le type d'accès choisi [achat, emprunt]) • Les fonctionnalités (surlignage, annotation, etc.) sont tributaires des formats de restitution, et du lecteur, variables selon les choix que la presse a effectués
Processus d'édition numérique	<ul style="list-style-type: none"> • Formats de restitution (PDF, ePub, HTML, PRC, MOBI) • Comme il s'agit essentiellement de rendre en format numérique, des livres publiés en format papier, le seul média géré est le format image • L'un des services offerts par BiblioVault permet de préparer les livres afin de les transmettre aux vendeurs de livres (Amazon, B&N, netLibrary, etc.) • Fournit un dépôt à long terme des fichiers numériques des livres pour les presses qui sont membres • Adaptation aux différents types d'écrans et aux appareils mobiles (voir description des services http://www.bibliovault.org/about.epl) • Métadonnées (page couverture, titre de l'ouvrage, éditeur, année de parution, ISBN-10, ISBN-13, extraits, critiques, titres des chapitres, notes biographiques)
Services de la plateforme	<ul style="list-style-type: none"> • Fonctionnalités de recherche (par titre, auteur, éditeur, ISBN, sujet) • Fournit l'image de la couverture, la table des matières, des extraits, des critiques, la note biographique des auteurs, lorsque disponible • Services aux étudiants ayant une incapacité pour éviter qu'ils aient à utiliser la version imprimée des livres • Possibilités de prévisualisation d'une portion du livre (Google Preview, HTML, PDF, ePub) • Lien vers le site d'achat des presses • Format Marc ne semble pas fourni
Commentaires	<ul style="list-style-type: none"> • Du point de vue de l'utilisateur final, le lecteur, cet entrepôt ne peut se comparer avantageusement aux agrégateurs mais il offre tout de même un point d'accès supplémentaire aux éditeurs (des presses universitaires, essentiellement) qui souhaitent augmenter la visibilité de leurs ouvrages • Dommage qu'il n'y ait aucune possibilité de naviguer par discipline/sujet. Il est possible d'effectuer une recherche par sujet (vedette-matière) cependant

BiblioVault

- La fonctionnalité « Buy this book » renvoie parfois directement au panier avec le livre sélectionné, alors que dans d'autres cas, ça ne renvoie qu'à la page d'accueil de l'éditeur ou autre site, ce qui est moins intéressant, en plus de rendre la fonctionnalité non uniforme. Il faut alors refaire la recherche dans le site pour finalement en arriver à la conclusion qu'il n'est pas possible d'acheter le livre en ligne, ou encore qu'il n'est disponible qu'en format papier
- Il n'est pas spécifié quels livres sont disponibles en format numérique et quels autres ne sont disponibles qu'en format papier. Certains éditeurs semblent utiliser ce service uniquement pour avoir les fichiers numériques nécessaires à l'impression ou pour un meilleur référencement dans les outils d'achat, pour leurs exemplaires papier
- Pas de recherche en texte intégral possible

PLATEFORMES EXPLORÉES

Agrégateurs

- ACLS Humanities E-Book
<http://www.humanitiesebook.org/>
- Cairn
<http://www.cairn.info/>
- NetLibrary
<http://library.netlibrary.com/>
- MyiLibrary
<http://www.myilibrary.com/>
- OAPEN
<http://www.oapen.org>
<http://project.oapen.org/>

Entrepôt numérique

- BiblioVault
<http://www.bibliovault.org/>

Annexe 2. : Canevas d'entretien avec les dirigeants des presses universitaires canadiennes

1. Avez-vous déjà mené des expériences dans le domaine du livre numérique ?

Si oui :

- Quel pourcentage de votre catalogue est présentement accessible en version numérique (nouveau(s), fonds) ?
- Quels formats de diffusion employez-vous (PDF, AZW (Kindle), MOBI, iPhone Apps, LIT, DAISY, XML, Epub, HTML, etc) ?
- Avez-vous recours à une aide extérieure pour la production de vos fichiers de diffusion ?
- Quels sont les canaux que vous utilisez pour la commercialisation de vos livres numériques (Google Search Books, Amazon Search Inside, votre site internet, agrégateurs tels que Ebrary, MyiLibrary, etc.) ?
- Vos ventes de livres numériques sont-elles significatives ?

Si non :

- Quelles seraient les conditions économiques ou techniques à réunir pour permettre votre engagement dans le domaine du livre numérique (plus de personnel, une subvention, l'accès à une infrastructure de diffusion qui respecte votre mission, etc.) ?

2. La production et la diffusion de livres numériques font-elles parties de votre plan d'activités ?

- Avez-vous intégré la production et la diffusion du livre numérique à votre modèle d'affaires ? Comment ?

3. Comment décrieriez-vous votre expertise en matière de livre numérique ?

- Avez-vous intégré la production et la diffusion des livres numériques aux différentes étapes de votre production éditoriale ?
- Êtes-vous familier avec les meilleures pratiques en vigueur dans le domaine de l'édition numérique (standards de métadonnées, formats de fichiers, canaux de diffusion, etc.) ?

4. Seriez-vous intéressés à participer à un projet de plateforme de production et de diffusion de livres numériques multi-éditeurs ?

Si oui :

- À quelles conditions (contrôle des fichiers, préservation de l'identité de votre publication, etc.)? Si non :
- Quelles sont les raisons qui motivent votre refus?

5. Si vous aviez à vous engager dans l'édition de livres numériques, préféreriez-vous réaliser vous-même toutes les étapes allant de la production à la commercialisation ou voudriez-vous déléguer certaines parties du travail ?

- Quelles sont les étapes que vous préféreriez réaliser vous-mêmes et pour quelles raisons?
- Quelles sont les étapes que vous préféreriez déléguer et pour quelles raisons?

Annexe 3 : Projets de recherche issus de différentes institutions

Nom du projet	AHDS Projects and Method Database (extension)
Date d'activité	2005 – 2006
Description	Il s'agit d'une extension du projet AHDS Projects and Method Database (prototype) qui a été achevée en janvier 2005. L'objectif du projet était d'établir une base de données d'informations d'enregistrement pilote sur la recherche fondée sur les TIC dans le domaine des arts et des sciences humaines à travers le Royaume-Uni. Le but de cette extension est de créer une ressource consultable en ligne qui enregistre et relie des informations essentielles pour les chercheurs en sciences humaines.
Objectifs spécifiques	<p>Fournir une vue d'ensemble des projets utilisant les TIC à des fins de création de ressources numériques. Permettre aux chercheurs en sciences humaines d'identifier les méthodes en TIC basées sur des expériences passées et de partager l'expertise avec d'autres chercheurs qui utilisent les mêmes méthodes, indépendamment de la discipline. Permettre au AHDS des autres organisations, de bien informé les organismes subventionnaires sur les pratiques et les développements des recherches dans le domaine des TIC. Permettre aux services en arts et sciences humaines numériques, comme le AHDS, de développer leurs services selon les besoins et les habitudes des usagers. Cette ressource contient des renseignements sur:</p> <ul style="list-style-type: none">• Des projets de recherche en sciences humaines à l'aide des TIC au Royaume-Uni ;• Des méthodes de calcul employées par ces projets (rassemblés et organisés d'une façon contrôlée selon un point taxonomique des méthodes de calcul)• Des logiciels, des outils et des normes de métadonnées utilisés par de tels projets• Des publications, documents d'information et études de cas qui traitent de différentes méthodes de calcul en détail.
Méthodologie	Le projet identifie un besoin marqué pour cartographier une taxonomie de méthodes en TIC utilisées pour capturer, modeler, analyser et présenter les extraits numériques de recherches. Afin d'atteindre une source exhaustive d'information. La collecte de données s'étendra pour inclure tous les projets au Royaume-Uni qui utilisent extensivement les TIC, par exemple, ceux subventionnés par la British Academy et Leverhulme et les projets dans les bases Humbul ou Artifact. Les projets qui utilisent les TIC, mais ne produisent pas des ressources numériques seront aussi inclus.
Plateforme	ADHS : Le Arts and Humanities Data Service est un service national du Royaume-Uni subventionné par le JISC (Joint Information Systems Committee) et AHRC dont la mission est de préserver et promouvoir les ressources électroniques issues de la recherche et de l'enseignement en art et sciences humaines. Il comporte plusieurs collections numériques, comme des bases de données historiques, textes littéraires, corpus linguistique, collection multimédia en art et plusieurs d'autres.
URL	http://www.ahds.ac.uk/about/projects/pmdb-extension/index.htm

Nom du projet :	LAIRAH
Date :	2005-2006
Description :	Le projet LAIRAH a examiné l'utilisation de ressources numériques dans les sciences humaines pour déterminer s'ils sont durables du point de vue de leur utilisation et comment et pourquoi ils sont utilisés. L'objectif général est de créer un guide de bonnes pratiques lors de la création d'un projet en sciences humaines numériques.
Objectifs spécifiques :	<p>Déterminer le degré de l'utilisation de ressources en sciences humaines dans le portail Humbul, Artifact et ADHS.</p> <p>Déterminer si les ressources qui sont fréquemment utilisées partagent des caractéristiques communes.</p> <p>Mettre en évidence les bonnes pratiques et les aspects graphiques des projets plus utilisés afin d'améliorer leur utilisation et leur durabilité.</p> <p>Savoir si certaines caractéristiques (association entre disciplines, par exemple) ou médias (texte, image) ont un impact sur l'usage</p> <p>analyser l'impact des aspects institutionnels comme l'avis des spécialistes départementaux, la culture interne e l'organisation, les subventions et la gestion.</p>
Méthodes :	La recherche utilise la méthode de log analyses pour évaluer le degré d'utilisation des ressources en arts et sciences humaines. Cette technique permet d'identifier les motifs (patterns) d'usage et la sélection de ressources utilisées et non utilisées.
Plateforme :	Humbul, Artifact and AHDS portal sites.
URL :	http://www.ucl.ac.uk/infostudies/research/circah/lairah/

Nom du projet :	Service-Oriented Architecture for Preservation and Ingest of digital objects
Date :	2007-2009
Description :	<p>Le projet développera une architecture et une boîte à outils pour automatiser partiellement la préservation et ingérer les flux de travail dans les dépôts numériques, basés sur un ensemble d' « atomic web services », chacun mettant en capsule une unité fonctionnelle de préservation, combinée avec d'autres formes basées sur le web où une décision humaine est requise. Un outil d'orchestration de flux de travail sera utilisé pour combiner ces services à des fonctions de saisie des utilisateurs, créant les flux de travail visant à réduire le besoin d'intervention humaine. Une attention sera portée au processus de préservation requis lorsqu'un contenu est introduit dans un dépôt, incluant la génération technique et d'autres métadonnées liées à la préservation, ainsi que la normalisation du contenu à l'égard des formats de préservation.</p>
Objectifs spécifiques :	<p>Développer une boîte à outils de production qualifiée, basée sur une architecture flexible et extensible qui permettra aux gestionnaires de dépôts numériques de construire des flux de travail pour la préservation d'objets numériques.</p> <p>Étudier l'utilisation des annotations sémantiques des web services pour permettre la découverte dynamique de services de préservation et les exécuter au fur et à mesure, en opposition à la configuration figée de la boîte à outils basique.</p>
Méthodes :	<p>Le projet produira un ensemble détaillé de cas d'utilisation pour les services de préservation, et en particulier identifiera les exigences pour la préservation des métadonnées lors de l'alimentation du dépôt.</p> <p>Création de schémas XML pour représenter les entités de métadonnées PREMIS. La spécification du conteneur de métadonnées sera utilisée pour rassembler la variété de composantes et d'entités de métadonnées associées à un objet numérique.</p> <p>Développement d'une architecture pour la boîte à outils, incluant le design, le codage et le module de tests pour les composantes du logiciel.</p> <p>Tests et évaluation du logiciel. Le dépôt de l'AHDS sera utilisé pour l'étape de tests. La variété et la complexité des collections en font un défi particulier, en fournissant un haut degré de fiabilité et applicabilité de la boîte à outils.</p> <p>Recherche pour étendre l'utilisation de la boîte pour incorporer de l'annotation sémantique et les services de découvertes des besoins de préservation</p>
Plateforme :	AHDS
URL :	http://www.ahds.ac.uk/about/projects/soapi/index.htm

Nom du projet :	MetaTools - Investigating Metadata Generation Tools
Date :	2007-2008
Description :	Le but du projet est de développer une méthodologie pour évaluer les outils de génération de métadonnées, comparer la qualité des outils déjà existants et développer un prototype web qui intègre les meilleurs outils de génération de métadonnées.
Objectifs spécifiques :	Développer une méthodologie pour évaluer les outils de génération de métadonnées ; Comparer la qualité des outils de génération de métadonnées existants ; Développer, tester et disséminer un prototype de services web qui intègrent les outils de génération de métadonnées.
Méthodes :	L'approche adoptée par le projet sera de transformer les meilleurs outils en services dont l'interface est bien définie, basée sur standards XML comme SOAP et WSDL. Une ontologie sera développée pour permettre que les services web générant les métadonnées reçoivent des annotations et descriptions des machines capables d'interpréter la sémantique des objets, représentant explicitement la connaissance sur les services de façon flexible et extensible. Ces annotations permettront la découverte dynamique des services appropriés par les agents logiciels, qui pourront invoquer ces services, les combiner à l'intérieur de flux de travail, et surveiller les résultats avec une intervention minimale de l'utilisateur. L'efficacité du prototype de génération de métadonnées sera testée dans un large ensemble de ressources numériques de l'AHDS en coopération avec d'autres dépôts et portails.
Plateforme :	ADHS
URL :	http://www.ahds.ac.uk/about/projects/metatools/index.htm

Nom du projet :	Armadillo : Historical Data Mining
Date :	2008
Description :	<p>Le projet Armadillo: Historical Data Mining utilise un ensemble de ressources en ligne portant sur l'Histoire sociale britannique du XVIIe siècle pour évaluer les apports bénéfiques potentiels du Web sémantique en Arts et Sciences humaines. Il vise à examiner les nouvelles façons d'extraire de l'information issue de ressources électroniques non liées. C'est une tentative de répondre à la question sur la façon de placer et d'interpréter efficacement les informations contenues dans des ensembles de données de recherche distributives en ligne, utilisant des critères acceptables pour la recherche en arts et sciences humaines.</p>
Objectifs spécifiques :	<p>Évaluer l'efficacité d'utiliser le logiciel Armadillo pour faire de la recherche dans des bases de données distributives</p> <p>évaluer quels sont les avantages que le logiciel offre à l'égard d'autres technologies plus traditionnelles de repérage de l'information, comme la recherche par mots-clés et les bases de données d'origines XML;</p> <p>Évaluer le potentiel du logiciel d'offrir une réutilisation des ensembles de données en utilisant des techniques en fouille de texte;</p> <p>Évaluer si le logiciel peut être utilisé pour générer automatiquement des balises selon les standards émergents du Web sémantique.</p>
Méthodes :	<p>Le logiciel Armadillo fait le repérage de l'information en fonction d'une ontologie préétablie et en crée des instances avec les données extraites. Dans le cas particulier du projet, l'ontologie initiale sera axée sur les dates, les noms et les lieux centrés autour de Londres au XVIIIe siècle. Armadillo combine les données sur la base d'une approche qui met en valeur l'évidence. Il utilise les probabilités statistiques pour tenir compte des déviations orthographiques, des formatages typographiques et des informations contextuelles dans le but de déduire les éléments les plus probables à combiner avec d'autres qui réfèrent au même lieu, nom ou date.</p>
Plateforme :	The Old Bailey Proceedings Online, The Westminster Historical Database, Index to eighteenth-century fire insurance policies, St Martin's Settlement Exams Index
URL :	http://www.hrionline.ac.uk/armadillo/

Nom du projet :	Books as Expressions of Global Cultural Diversity
Date :	2009
Description :	Les données de la base de données bibliographique globale WorldCat révèlent des modèles (patterns) transnationaux dans la publication littéraire, la conservation de l'héritage littéraire des certains pays et la diversité culturelle présente dans les livres.
Objectifs spécifiques :	Le résultat attendu comporte un large ensemble de statistiques comparées par pays et accompagnées d'un rapport détaillant les découvertes de la recherche. Ces résultats fourniront une vision globale de la littérature transnationale et des études de cas féconds sur les pratiques particulières de chaque pays sur la publication et la préservation de leur héritage littéraire.
Méthodes :	Rassembler les détails de matériaux textuels non sériés (à l'exception de dissertations et documents gouvernementaux). La date de publication doit être valide et antérieure à 2010. Exclure les travaux dont la date de publication apparaît comme 19xx. Inclure les livres dont la date de publication, est plus récente que 1000 A.C, Un pré test dans six pays (délibérément mettant en lumière les travaux et les catalogues non anglophones) a été suivi par un raffinement des techniques d'extraction de données (fouille de données). La procédure a été appliquée dans le reste des données.
Plateforme :	World Cat : La base de données bibliographique de l'OCLC représente plus de 142 millions de notices, avec 1,43 milliard de copies tenues par des bibliothèques dans le monde entier (nombre toujours croissant). En outre, la base de données devient de plus en plus globale dans sa portée avec l'intégration des douzaines de données bibliographiques des bibliothèques nationales.
URL :	http://www.oclc.org/research/activities/globalbooks/default.htm

Nom du projet : **Emancipating Digital Data - Scanning and Image Analysis of the Lincoln Papers**

Date : 2008

Description : Le projet permet l'extraction de connaissances d'une grande collection numérisée des écrits d'Abraham Lincoln et fournit aussi un modèle de présentation multidimensionnel de l'information qui peut être appliqué à d'autres collections de documents historiques numériques.

Objectifs spécifiques : Numérisation de lettres, écrits et notes d'Abraham Lincoln
Création d'un dépôt de données distribuées et matériels liés au seizième Président;
Intérêt à optimiser la manipulation et le traitement d'un grand volume d'images composées de documents historiques et cartes géographiques;
Création d'un système web qui donne l'accès à ces documents à la communauté savante et au public en général. Ce système combine à la fois l'apprentissage traditionnel (par exemple, en l'utilisant le Google Maps) apprentissage interactif à travers de la formulation de requêtes sur le contenu des documents et l'extraction de données qui permettront des études exploratoires sur les caractéristiques de documents et leur relation avec les événements historiques.

Méthodes : Classification automatique et traitement d'une grande collection de documents numérisés;
Extraction de zones d'intérêt dans les documents numérisés (classification automatique d'images);
Intégration des informations spatiales, temporelles et documentaires;
Création d'une interface web qui permet d'interroger la base et d'offrir une représentation temporelle, spatiale et documentaire de l'information contenue dans les documents historiques.

Plateforme : The Lincoln Log: A Daily Chronology of the Life of Abraham Lincoln

URL : <http://isda.ncsa.uiuc.edu/lpapers/index.html>

Nom du projet :	TAPoR
Date :	Non précisé
Description :	<p>TAPoR a comme mandat la construction d'une infrastructure informationnelle et humaine unique d'analyse de textes à travers le Canada, en établissant six centres régionaux pour constituer un portail national de recherche en analyse de texte. Ce portail sera une passerelle à des outils d'analyse et repérage sophistiqués, accompagnés de textes représentatifs pour l'expérimentation. Les centres locaux incluront des laboratoires de recherche de texte avec les meilleurs logiciels et serveurs en texte intégral qui vont être coordonnés dans un portail vertical dédié à l'étude de textes électroniques.</p>
Objectifs spécifiques :	<p>Constituer une infrastructure avec 6 centres régionaux pour former un portail national d'analyse de texte.</p> <p>Constituer un portail vertical pour permettre l'accès aux ressources électroniques et aux outils d'analyse des textes.</p> <p>Le projet s'intéresse à trois dimensions de l'utilisation de ressources électroniques dans le cadre de la recherche en sciences humaines :</p> <p>la représentation du texte, c'est-à-dire, l'apprentissage à travers de la préparation méthodique des éditions électroniques, en choisissant quoi représenter et quels sont les objets multimédias (comme des images d'un manuscrit ou un audio d'une entrevue orale) devront être intégrés dans une édition.</p> <p>Les techniques d'analyse du texte et les outils, plus spécifiquement le développement et l'adaptation du logiciel pour implémenter des techniques de recherche avec lesquelles nous pouvons interroger les textes électroniques.</p> <p>L'accès au texte électronique par le moyen du développement d'un environnement académique basé sur des textes électroniques pour des équipes de chercheurs et la communauté en général dans le but d'étudier les enjeux à travers de l'agrégation d'outils pertinents, textes et médias.</p>
Méthodes :	Non précisé. Chaque laboratoire partenaire mène individuellement différentes recherches en utilisant le portail TAPoR
Plateforme :	Portal TAPoR
URL :	http://tapor.humanities.mcmaster.ca/home.html

Nom du projet :	PhiloMine / ARTFL
Date :	Non précisé
Description :	PhiloMine est une extension aux versions actuelles de PhiloLogic, pour assister à une variété de tâches liées à l'apprentissage des machines, fouille de texte et regroupement de documents. L'outil a été conçu originalement dans le cadre du projet ARTFL pour soutenir des bases de données de littérature française et il a été étendu pour être utilisé dans diverses bases de données textuelles et hypermédiatiques.
Objectifs spécifiques :	<p>Assister à l'extraction de connaissances des collections chargées dans la base du logiciel.</p> <p>Fournir un ensemble de fonctionnalités pour l'exploitation de textes électroniques (recherche et analyse) sans qu'il soit nécessaire de mettre en place un travail supplémentaire en termes de gestion et développement.</p>
Méthodes :	<p>Il s'agit d'un système modulaire dans le quelle une base textuelle est traitée comme un ensemble de bases inter reliées et coordonnées, généralement avec l'inclusion d'une base d'objets (unités de textes), d'une base de terme, d'un index de concordance (Word concordance) cartographiés comme objets textuels et d'un module de gestion d'objets « Mapping text objects to byte offsets in data file ». Chacune de ces bases est conservée et gérée en utilisant son propre sous-système.</p> <p>Création d'une spécification d'encodage (ATE : ARTFL Text Encoding) pour représenter les données dans le logiciel PhiloLogic qui permet la construction de base de données en différents schémas d'encodage, permettant l'utilisation du logiciel PhiloLogic pour indexer des collections arbitraires en HTML dans un espace World Wide Web avec un minimum d'encodage et aucune modification.</p>
Plateforme :	ARTFL : fondée en 1982 comme résultat d'une collaboration entre le gouvernement français et l'Université de Chicago, le projet ARTFL est un service basé sur un consortium qui fournit à ses membres l'accès à la plus large collection de ressources français en Amérique du Nord.
URL :	<ul style="list-style-type: none">- http://code.google.com/p/philomine/- http://perseus.uchicago.edu/- http://philologic.uchicago.edu/

Nom du projet :	Hermeneuti.ca /Voyer Tools
Date :	Non précisé
Description :	Hermeneuti.ca est un projet collaboratif élaboré par Stéfan Sinclair & Geoffrey Rockwell qui vise à penser certaines fondations de l'analyse de texte contemporaine, incluant les enjeux relatifs aux textes électroniques utilisés, les outils et méthodologies disponibles, et les formes diverses que peuvent prendre l'expression des résultats de l'analyse de textes.
Objectifs spécifiques :	Extraction de connaissances Fouille de texte Création du portail Voyer Tools, permettant faire de la fouille de texte dans le web, ou dans un corpus de document fourni par l'utilisateur
Méthodes :	non précisé
Plateforme :	Voyer (http://voyeurtools.org/) Voyer est un environnement d'analyse de textes basé sur le web. Il a été conçu en accord avec les principes de convivialité, pour permettre l'extraction de connaissances de façon flexible et puissante. Il fait partie du projet Hermeneuti.ca qui dont le but est de théoriser les outils d'analyse de textes et la rhétorique de l'analyse de textes.
URL :	http://voyeurtools.org/

Nom du projet :	Searching for the Victorians
Date :	Non précisé
Description :	Le projet vise à appliquer un processus de fouille de texte dans une large collection de livres de l'époque victorienne.
Objectifs spécifiques :	non précisé
Méthodes :	<p>Analyse de données textuelles de 1,681,161 qui ont été publiées en Angleterre et au Royaume-Uni tout au long du XIXe siècle, de 1789-1914</p> <p>Dans une première phase, limiter le processus de fouille aux mots présents dans les titres de livres</p> <p>Représenter graphiquement pour certains mots extraits la quantité de livres publiés chaque année pour permettre une comparaison visuelle et l'observation de l'évolution des publications sur une base temporelle et historique.</p> <p>L'impression multiple de livres dans une seule année a été condensée dans une seule « expression », selon la norme FRBR.</p> <p>Dans une deuxième étape, le processus de fouille sera appliqué au texte intégral</p>
Plateforme :	Google Books, Archive's Open Library et HathiTrust
URL :	http://www.dancohen.org/2010/10/04/searching-for-the-victorians/

Nom du projet :	Digging into the Enlightenment
Date :	Non précisé
Description :	« Digging into the Enlightenment » cherche à découvrir et cartographier l'échange de lettres, publications et rapports de voyages de la période connue sous le nom de « République des lettres », caractérisée par la constitution de communautés savantes et réseaux de connaissances qui ont émergé dans les premières décennies de l'avènement de la presse (entre 1500 et 1800). Ces données sont combinées avec l'expertise interprétative d'historiens et des savants littéraires. La cartographie de la République des Lettres est une occasion pour une collaboration unique entre les sciences humaines et les sciences pures pour produire un modèle d'un réseau mondial réel, utilisant des exemples riches et divers de ce matériel historique.
Objectifs spécifiques :	Démontrer comment les outils d'analyse visuels peuvent nous aider à extraire des nouvelles connaissances par des méthodes enracinées dans les sciences humaines, en utilisant les capacités d'annotations et l'habileté des chercheurs pour insérer de nouvelles données et dénouer les données déjà existantes.
Méthodes :	Combinaison entre les données géographiques, les événements historiques, relations interpersonnelles et données sociales pour générer une visualisation qui démontre comment la diffusion d'idées à l'échelle globale se rapporte à des processus dynamiques qui s'opèrent localement.
Plateforme :	http://enlightenment.humanitiesnetwork.org/
URL :	http://www.dancohen.org/2010/10/04/searching-for-the-victorians/

Nom du projet :	Watching the script
Date :	Non précisé
Description :	« Watching the Script » est un environnement de lecture, d'exploration des compositions théâtrales. Au lieu de présenter le texte d'une pièce dans une séquence, le système affiche le texte où les acteurs doivent dire les mots, en fournissant ainsi une représentation théâtrale du texte. L'environnement inclut aussi une perspective de mise en scène où certains traits (personnages ou lieux) peuvent être suggérés tout au long de la visualisation du manuscrit.
Objectifs spécifiques :	Non précisé
Méthodes :	La version actuelle du projet « Watching the Script » est dynamique et alimentée par plusieurs fichiers XML externes (les pièces de Shakespeare codée par le « Nameless Shakespeare projet »). Le premier fichier XML contient la composition, incluant toutes les paroles et directions des scènes. Le deuxième est le fichier en bloc de la composition qui détermine les coordonnées dans l'aire de la scène où chaque personnage se place selon ses paroles. Le troisième fichier XML est un menu qui contient tous les fichiers disponibles de la composition utilisée par le système.
Plateforme :	Nameless Shakespeare est une édition numérique construite à partir de la transcription du Projet Perseus du Globe Shakespeare, le texte standard du dernier dix-neuvième siècle et une étape majeure dans une tradition textuelle qui survit toujours dans des textes modernes comme la Rive de Shakespeare, l'édition de Bevington, ou Shakespeare Arden.
URL :	http://www.dancohen.org/2010/10/04/searching-for-the-victorians/

Nom du projet :	Digital Humanities and Text Mining Lab (DHTML)
Date :	Non précisé
Description :	Dominic Forest a élaboré un programme de recherche dont l'objectif est de concevoir, de développer et de valider des démarches méthodologiques de fouille de textes (text mining) sur de gros corpus de documents théoriques en sciences humaines. Le laboratoire de fouille de textes en sciences humaines numériques (Digital Humanities and Text Mining Lab (DHTML) permettra la réalisation de ce programme de recherche novateur, qui concevra une plate-forme modulaire pour l'analyse et l'interprétation des textes en sciences humaines.
Objectifs spécifiques :	- concevoir de nouvelles stratégies pour l'extraction, l'organisation et la visualisation automatique d'informations à partir de très grands corpus de documents textuels
Méthodes :	Le programme de recherche s'appuiera sur le corpus documentaire du consortium Érudit. Situé à l'Université de Montréal, Érudit donne accès à différents types de documents en sciences humaines et sociales : revues savantes, livres, actes, thèses et autres documents et données. Érudit a mis en place un traitement éditorial spécifique et appliqué à chacun des types de document qu'il propose. La qualité des documents fournis par Érudit (tant en termes de contenu que d'encodage) permettra d'exploiter tout le potentiel des démarches de fouille de textes qui seront développées dans le cadre de ce programme de recherche (métadonnées servant à la validation des approches de fouille de textes, balisage en format XML permettant de cibler des segments de textes spécifiques à l'intérieur des corpus, etc.).
Plateforme :	Érudit (http://www.erudit.org) - Érudit est un consortium interuniversitaire composé de l'Université de Montréal, de l'Université Laval et de l'Université du Québec à Montréal. Société sans but lucratif, Érudit propose un modèle innovateur de promotion et de diffusion des résultats de la recherche et constitue le « pôle Québec » du projet Synergies. La plateforme Érudit donne accès à plusieurs types de documents des disciplines des sciences humaines et sociales, ainsi que des sciences naturelles : revues savantes, livres, actes, thèses et autres documents et données. Un traitement éditorial spécifique est appliqué à chacun de ces types. L'outil de recherche permet d'interroger l'ensemble du fonds, c'est-à-dire, tous les types de documents, tout en distinguant les différentes catégories dans les résultats.
URL :	Non fourni

Nom du projet :	With Criminal Intent
Date :	Non précisé
Description :	L'« Old Bailey Proceedings » contient les preuves criminelles de presque deux cent mille procès à la Cour centrale de Londres à l'intérieur d'une période de 240 ans. Cette base de données comprend 120 millions de mots de texte structuré, le plus grand corpus cohérent des descriptions imprimées du comportement criminel jamais publié. Malgré cette richesse des données, les tentatives des historiens de se rapprocher de cette collection se limitaient à tamiser par les documents un par un, utilisant les histoires inhabituelles ou attrayantes pour illustrer les histoires culturelles, sociales ou intellectuelles. L'objectif du projet « Criminal Intent » est de démontrer des rôles potentiels de la fouille de texte dans la pratique de la recherche historique, en montrant que la rigueur historique peut être accomplie et que de nouvelles perspectives peuvent être gagnées, en passant de l'analyse d'un seul procès ou d'un ensemble étroit d'exemples appropriés à une analyse de modèles textuels statistiquement significatifs trouvés dans cette source.
Objectifs spécifiques :	En permettant l'analyse et la représentation statistiques des types de langages utilisés dans la cour et comment elles ont changé à travers le temps et en comparant les patrons des données extraites à ceux trouvés dans les données étiquetées, le projet réalisera deux choses. D'abord, une nouvelle façon de représenter graphiquement les changements du rapport de crime et de l'accusation sera créée; et la seconde une nouvelle méthodologie pour la découverte cohérente de descriptions liées sera évaluée. Il permettra aux chercheurs de passer de la méthodologie de ramassage et comparaison textuelle à une analyse structurée des couches de signification à motifs trouvés en formes spécifiques et distinctives de représentation historique.
Méthodes :	Non précisé
Plateforme :	Zotero, the Old Bailey Online project, and the TAPoR project.
URL :	http://criminalintent.org/about/

Nom du projet :	Fragmentary authors
Date :	2008 -2009
Description :	Le projet « Fragmentary authors » vise la création de liens dynamiques entre les œuvres classiques grecques et latins et les éditions dites « fragmentaire ». On fait ici référence aux œuvres qui sont des fragments de documents ou des débris de papyrus ayant survécu dans les sables de l'Égypte. Dans la plupart des cas, ces fragments ne sont que des citations à d'autres œuvres de l'antiquité qui elles sont restées intactes. Dans un monde numérique, ces fragments ne seraient que des liens pointant vers les œuvres originales où elles se trouvent.
Objectifs spécifiques :	Construction d'éditions dynamiques et hypertextuelles d'auteurs fragmentaires, en utilisant comme base la collection complète de textes sources grecs et latins de la base Perseus, avec des images de pages en texte consultable OCR-PORDUIT, et XML TEI-CONFORME transcrit pour certains.
Méthodes :	Dans 2008-09 nous commencerons le travail sur les historiens fragmentaires grecs, utilisant Müller's Fragmenta Graecorum Historicorum comme un point de départ. La production de ce travail sera tant édition initiale d'historiens fragmentaires grecs que les méthodes par lesquelles nous représentons des indicateurs dans des œuvres source et l'annotation savante associée.
Plateforme :	Perseus
URL :	http://www.perseus.tufts.edu/hopper/research/current#scalable

Nom du projet :	Making the LEAP: linking electronic archives and publications
Date :	Non précisé
Description :	Le but du projet est d'examiner des solutions novatrices pour les publications électroniques en ligne pour qu'elles fournissent un accès large aux découvertes du domaine des sciences humaines et de l'art ainsi que rendre les données sous-jacentes disponibles d'une telle façon qui permettra aux lecteurs d'aller plus profondément dans les archives disponibles en ligne afin de tester leurs interprétations et d'en faire leurs propres conclusions sur la recherche.
Objectifs spécifiques :	<p>Fournir une forme de dissémination additionnelle aussi novatrice qu'imaginative à quatre projets de recherche de haute qualité académique ;</p> <p>Examiner de quelle façon les publications électroniques peuvent être interactives, comporter différentes couches de données qui lui servent de support, comprenant une gamme de formats - base de données, GIS, VR, audio et vidéo, etc.</p> <p>Examiner comment les multiples formes de dissémination peuvent être utilisées par différents publics</p> <p>Exploiter les questions de référencement entre les archives distribuées dans le système documentaire AHDS et les publications électroniques.</p> <p>Examiner les questions soulevées par ces nouvelles formes de disséminations afin d'en fournir des directives et promouvoir de bonnes pratiques qui peuvent être appliquées à travers le domaine des arts et sciences humaines : questions éditoriales, citations, gestion de droits et changement de culture dans les institutions savantes.</p>
Méthodes :	Le projet utilisera l'infrastructure du journal Internet Archeology et de la AHDS pour fournir quatre exemplaires de publications électroniques à plusieurs couches et des archives électroniques, capables d'être implémentées dans le domaine des arts et sciences humaines.
Plateforme :	ADHS
URL :	http://ads.ahds.ac.uk/project/leap/

Nom du projet :	Grammaticométrie et classification des textes historiques latins (LASLA)
Date :	2008
Description :	Le but est de fonder la classification des textes sur des critères strictement linguistiques, évalués et analysés au moyen des outils de la statistique linguistique la plus récente et de la topologie textuelle ; d'établir ainsi une classification endogène des textes historiques latins, de la fonder objectivement et de la confronter aux classifications établies a priori (Annales vs commentaires vs biographies)
Objectifs spécifiques :	<p>Fonder la classification des textes sur des critères strictement linguistiques, évalués et analysés au moyen des outils les plus récents de la statistique linguistique et de la topologie textuelle.</p> <p>Établir une classification endogène des textes historiques latins, fondée objectivement sur la confrontation des classifications établies (Annales vs commentaires vs biographies).</p> <p>Faire progresser les outils d'analyse par le développement de méthodologies toujours mieux adaptées à la nature des données traitées et à la spécificité de leur organisation textuelle.</p> <p>Appliquer ces outils au classement des textes latins médiévaux ou néolatins, principalement dans le cadre des recherches menées par C. Philippart sur les hagiographies latines d'origine grecque et par Laurent Graillet sur les Lettres de Turquie de G.A. de Busbecq, et de valider par ceci une nouvelle méthodologie pour la modélisation des typologies textuelles.</p>
Méthodes :	non précisé
Plateforme :	Opera Latina
URL :	http://www.cipl.ulg.ac.be/Lasla/recherche.html

Nom du projet : **Lemmatisation assistée par ordinateur et analyse syntaxique automatisée (LASLA)**

Date : 2009

Description : La finalité de ce projet de recherche est de fournir de nouveaux outils pour l'étude de l'ordre des mots, pour la modélisation des structures énonciatives et pour la classification des textes latins et néo-latins.

Objectifs spécifiques : La recherche permettra de compléter les informations contenues dans la base Opera Latina. Les nouvelles données recueillies pourront être utilisées immédiatement dans les autres programmes de recherche du L.A.S.L.A. À cette fin, et également dans l'optique de pouvoir constituer un corpus homogène d'un point de vue générique, sur lequel tester les programmes d'analyse syntaxique que l'on aura développés, on a retenu comme corpus de travail l'ensemble des textes historiques latins de l'époque classique. Les mêmes méthodes pourront ensuite être appliquées à d'autres textes classiques, voire à des textes plus tardifs, notamment néo-latins, pour déterminer les similitudes et les écarts par rapport aux modèles classiques.

Méthodes : Sur le plan méthodologique et linguistique, il s'agit de mener une recherche originale et novatrice sur l'ordre des mots et sur les structures de l'énoncé latin : en faisant appel aux données déjà encodées et à la linéarité du texte, on implémente diverses règles d'ordre de mots proposées par des descriptions linguistiques récentes. En soumettant le corpus du L.A.S.L.A. à de véritables procédures d'analyse syntaxique automatisées, on cherche non seulement à pourvoir le texte de nouvelles annotations syntaxiques, mais aussi à évaluer le bien-fondé des descriptions linguistiques sur lesquelles ces procédures s'appuient. Une fois délimitées toutes les propositions subordonnées, ainsi que leur niveau de subordination, on pourra avoir une plus juste vision de l'ampleur des propositions principales, distinguer les divers types de propositions introduites par un même subordonnant, déterminer les zones de style indirect et les séries de phrases nominales, ou encore étudier avec plus de précision les mécanismes de coordination et de juxtaposition entre propositions.

Plateforme : Opera Latina

URL : <http://www.cipl.ulg.ac.be/Lasla/recherche.html>

Nom du projet :	Motifs syntaxiques et topologie textuelle (LASLA)
Date :	2009
Description :	Développer une méthodologie et un outil logiciel de reconnaissance automatique de certains motifs syntaxiques dans des textes déjà lemmatisés et étiquetés, français et latins ; étudier la distribution de ces motifs au fil d'un texte pour caractériser la dynamique interne de celui-ci, la comparer à d'autres structures textuelles et améliorer par là même les outils de classification automatique des textes.
Objectifs spécifiques :	<p>Développer une méthodologie et un outil logiciel de reconnaissance automatique de certains motifs syntaxiques dans des textes déjà lemmatisés et étiquetés, français et latins</p> <p>Étudier la distribution de ces motifs au fil d'un texte pour caractériser la dynamique interne de celui-ci, la comparer à d'autres structures textuelles et améliorer par là même les outils de classification automatique des textes</p> <p>Prendre appui sur la distribution des motifs au fil d'un texte pour produire des partitionnements automatiques de ce texte, repérer des zones sensibles d'apport d'information ou des zones transitionnelles par exemple</p> <p>Utiliser les résultats de ces diverses applications pour un retour réflexif sur la notion et en valider la pertinence conceptuelle et méthodologique</p>
Méthodes :	non précisé
Plateforme :	Opera Latina
URL :	http://www.cipl.ulg.ac.be/Lasla/recherche.html

Annexe 4 : An Annotated Review of Centers and Large Projects listed in the CenterNet website

(completed by Jordan Goldman and Matthew Halse,
Research Assistants for Michael Eberle-Sinatra, 13 November 2010)

- Advanced Computing in the Humanities - ACO*HUM <http://www.hit.uib.no/AcoHum/>
Bergen, Norway - A SOCRATES thematic network project aimed at developing an international dimension for investigating the educational impact of new technologies in all humanities disciplines.

ACOHUM is a project based out of The Universitetet i Bergen (University of Bergen, Norway) that began in 1996 and ended in 2000. The website contains many dead links, but there are still documents containing the minutes from various working groups and an overall project summary. ACOHUM's area of interest is best summarized as as : "The theme of our network has been the increasing use of advanced computing in teaching and learning in the humanities, with its problems and possibilities. " The group completed a book project in 1999, but the link no longer works.

- Ancient World Mapping Center <http://www.unc.edu/awmc/>
Chapel Hill, NC, US - Exists to promote cartography and geographic information science as essential disciplines within the field of ancient studies.

The Ancient World Mapping Center is an interdisciplinary research centre out of the University of North Carolina at Chapel Hill. The website promotes various "publications of interest" which predominantly include books and articles that fall within the centre's mandate. The Centre's website predominantly features (under 'table of contents' located midway through the page) a large collection of maps and geographical images. Many of the maps on the site are used for various scholarly publications, and are free for student use. Some maps that the centre has created have been submitted for publication to Routledge. The centre also maintains a relationship with the Pleiades project, located at <http://pleiades.stoa.org/>, which may be of interest as it provides a digital forum for academics to share geographic information about ancient Greece and Rome.

- Arbeitsstelle Computerphilologie <http://www.c-phil.uni-hamburg.de/Main/WebHome>
Hamburg, Germany - Conducts research in "computer philology" with strong emphasis on narratology, holds seminars and courses in collaboration with other departments of the university and offers a certificate in the subject.

The ACP is a research centre located at the University of Hamburg. Although most of this page is only available in German, the site contains links to various conferences in the digital humanities that are available in English. The site contains a lengthy list of publications by ACP members (in German). The centre also offers a variety of courses for students at the university. The ACP focuses on narratology, the digital humanities, and the intersections of technology and the study of language.

-
- Archaeological Computing Laboratory, USyd - ACL <http://pacl.arts.usyd.edu.au/>
Sydney, Australia - The somewhat misnamed Archaeological Computing Laboratory provides support to staff and students at the University of Sydney across a broad range of Humanities disciplines (and beyond), particularly Archaeology, History, Anthropology and Linguistics.

The Archaeological Computing Laboratory is part of the Spatial Science Innovation Unit, a joint project of the Faculty of Arts and School of GeoSciences. The ACL describes itself as focusing “on information management and delivery for the Humanities, with particular emphasis on the spatial and temporal dimensions (event modelling, spatial searching, mapping and map animation) and on new opportunities presented by the Internet (academic knowledge management, online databases and web mapping).” Their flagship project is the Heurist Database (www.heuristscholar.org) which is a bibliographic database for the Humanities. The ACL has three other large projects: an online encyclopedic history of Sydney, the “Greater Angkor Project” which investigates the “relationship between the vast extent of Angkor from the twelfth to the sixteenth centuries CE, land clearance for rice production and regional ecological impacts that may have played a part in its decline” and the Macquarie Atlas of Indigenous Australia in which the ACL coordinated/created 250 publication ready maps.

- Archaeological Computing Research Group
<http://www.southampton.ac.uk/archaeology/acrg/index.html>
Southampton, Hampshire, UK - Located at the University of Southampton, multidisciplinary research center focused on applications of computation, engineering and other disciplines to archaeology. In particular the ACRG focus on research around computer graphics, geographic information systems, data management, and linked data. The center is based in the School of Humanities at the University of Southampton.

The ACRG’s is both an academic centre (they award two different Master’s degrees as well as performing PhD thesis supervision) and a research centre. Their research interests include “work on Semantic Web data integration for archaeological datasets from the Mediterranean, online GIS systems, archaeological data management systems, object data capture via laser scanning, polynomial texture mapping, and digital photogrammetry, animation and digital representation theory, structural simulation of built environments, virtual repainting of Roman objects, simulation and physically accurate rendering of conservation procedures, spatial and network analysis of a range of landscape and urban datasets, lighting analysis, and the three-dimensional analysis of space and spatial experience.” They conduct research on (and create) CGI images and “GIS/CAD synergies.” The site contains a lengthy list of publications (by their affiliated members) as well as links to a variety of ACRG projects, mostly having to do with the intersection of archaeological computing. The centre also has a division that does multimedia research, with a particular focus on creating teaching/learning packages.

- Australia New Zealand Digital Encyclopedia Group <http://jod.id.au/anzdeg/index.html>
Australia - The Australia and New Zealand Digital Encyclopaedia Group (ANZDEG) is a loose affiliation of people working on, or associated with, online reference collections.

ANZDEG describes itself as “a loose affiliation of people working on, or associated with, online reference collections. We include people working on small, individual projects and members of large institutions - the definition of “digital encyclopaedia” is deliberately vague so as to encompass a broad range of projects and interests. The group includes people working in eResearch and digital

humanities, libraries and archives, museums, web publishing and computing science.” ANZDEG does not independently publish, but contains links to various encyclopedia’s and research databases. They hold a “small, relaxed” annual conference, which constitutes the majority of their mandate.

- Australian Scholarly Editions Centre - ASEC <http://hass.unsw.adfa.edu.au/ASEC/> Canberra, Australia - A research centre with involvement in the development of electronic scholarly editions and associated technologies.

The ASEC is a scholarly publishing house at the University of New South Wales at the Australian Defense Force Academy. Their two main projects are “The Academy Editions of Australian Literature,” which published ten volumes (in hard copy) of extensively researched Australian historical texts and the parallel “Electronic Editions” division which archived the Academy Editions electronically. The Electronic Editions attempts to digitize the original texts while using specific technology (which ASEC is involved in developing) that maintains “the integrity of the core text while it is being proliferated, translated across platforms, manipulated, supplemented and analysed.” Multiple pages on this site have not been updated since 2002.

- Bereich Historische Fachinformatik <http://hfi.uni-graz.at/> Berlin, Germany - A department with several scholars, focuses on informatics in history, lists courses that may substitute for regular history department offerings.

This page is only available in German. The Historische Fachinformatik (HFI) is an academic research centre located in the department of Computer Science and Historical Documentation at the Karl-Franzen University Graz. The website listed above is the departmental website. HFI does not publish, but lists a variety of research initiatives conducted by the faculty and students of the department. The site also contains a variety of information for students wishing to pursue degrees in Computer Science, History and Documentation.

- Bibliothèques Virtuelles Humanistes <http://www.bvh.univ-tours.fr/> France - Situé à l'Université de Tours dans le Centre d'études de la Renaissance, le programme des "Bibliothèques Virtuelles Humanistes" a été conçu en 2002 avec la collaboration de l'IRHT (section de l'humanisme). Il diffuse des documents patrimoniaux et poursuit des recherches associant des compétences en sciences humaines et en informatique.

The BVH is a large database that collects and shares French historical documents: the database contains roughly 2000 facsimiles from the fifteenth to the seventeenth century. The BVH is actively attempting to digitize a large number of texts of national historical importance. The site contains a variety of search techniques available to the general public and researchers alike. There is a list of upcoming and current projects, all of which consist of the digitization and scanning of texts.

- C3 MIT Convergence Culture Consortium <http://www.convergenceculture.org/aboutc3/> Cambridge, MA, US - The Convergence Culture Consortium at MIT (C3) is a partnership between thinkers and researchers from/affiliated with the Comparative Media Studies program at MIT, and companies with a keen interest in deciphering convergence culture and the implications it can have for their business.

C3, beyond its teaching duties in partnership with the Comparative Media Studies department (which offers a Master’s degree), is a consortium of researchers who maintain “a flexible and dynamic project of research activities. Working within a humanities framework, C3 applies qualitative research

strategies to explore media change, the cultural factors influencing consumer behavior, and changing sites for audience experiences with media properties.” Two book projects have emerged as a result of C3: *Convergence Culture* and *You Tube: Online Video and Participatory Culture*. C3 also engages in a variety of partnerships with private sector media companies, and maintains direct relationships with the Education Arcade, New Media Literacies, the Singapore-MIT GAMBIT Game Lab and Metamedia. (The links have been attached to each name as they may be of interest.)

- Canadian Institute for Research in Computing and the Arts <http://circa.ualberta.ca/> Edmonton, Alberta, Canada - CIRCA, the Canadian Institute for Research Computing in Arts, fosters the development of new computing methodologies and new computer-based resources for research and teaching in the humanities and advances the state of humanities research computing at the University of Alberta. CIRCA’s mission is to:* Conduct research in the applications of computers to research and teaching in arts disciplines.* Promote and facilitate the development of innovative computer-based projects in arts research.* Develop community-university research opportunities in this area.* Develop and maintain relationships with leading similar research institutes around the world in order to fully participate in international research in this area.

Beyond what is listed above, CIRCA has collaborations with the Experimental Reading Workshop, the GRAND Games Lab, Histories and Archive Collaboratories, Interactives Lab and New Media and the Visual Culture Studio. Each of these associated groups conducts a variety of research, including the development of ‘teaching’ video games. CIRCA promotes the various conference presentations and publications of its members. CIRCA also maintains a blog roll and an RSS feed with a variety of news relating to “Computing in Arts.”

- Center for Applied Technologies in the Humanities - CATH <http://wiz2.cath.vt.edu/> Blacksburg, VA, US - Provides instructional services, technical support, and web hosting for digital projects in the humanities.

CATH is associated with the English Department at Virginia Tech. CATH describes itself as bringing “together individuals engaged in humanities-based research, teaching, and/or outreach projects requiring intensive application of cutting-edge technology and robust infrastructural support. CATH’s goal is to be at the forefront in applying the technologies that are revolutionizing research, teaching, and outreach in the humanities.” CATH publishes a variety of its projects online, including a database of Poetry 1579-1830, John Cleland’s “The Dictionary of Love,” an index of Middle English verse and a database of watermarks from the collections of Thomas L. Gravell and C. M. Briquet among others. CATH’s newest projects (including the index of Middle English verse) use a technology called “TEI markup.” CATH also acts as a teaching centre.

- Center for Digital Humanities <http://humanistika.org/> Beograd, Serbia - Explores the theoretical basis and practical uses of digital technologies in humanities research and teaching, with a particular focus on digital libraries and electronic lexicography.

This website is completely in Serbian. The CDH has a variety of web-based information (including articles, news feeds and other information vehicles) on a variety of topics relating to the Digital Humanities. The site seems to be updated very regularly. They have three projects listed: the digitization of the Serbian dictionary on Facebook and Twitter, Trankspoetika (“a digital platform for the Serbian language and literature”) and a joint venture with the Association of Humanities Research

Institutes to create the Internet Library of Serbia for the Humanities and Arts. Each of these three initiatives are published on the Center's website.

- Center for Digital Research in the Humanities - <http://cdrh.unl.edu/>
Lincoln, NE, US - A joint initiative of the University Libraries and the College of Arts and Sciences at the University of Nebraska-Lincoln, is advancing collaborative, interdisciplinary research by creating unique digital content, developing text analysis and visualization tools, and encouraging the use and refinement of international standards through research, demonstration, and teaching.

The Center for Digital Research in the Humanities mandates itself to advance “interdisciplinary research in the humanities by creating unique digital content, developing tools to assist scholars in text analysis and visualization, and encouraging the use (and refinement) of international standards for humanities computing.” The CDRH's website provides a variety of forums for scholars and students to locate information pertaining to digital scholarship. The CDRH's website publishes a wide variety of archived material (too much to list here), mostly pertaining to American History and American Literature. Notable literature archives include images, texts and manuscripts by Walt Whitman and Willa Cather. Further, the CDRH has two ‘metadata demonstration projects’: one which helps users navigate 35 collections of material pertaining to Walt Whitman, and a second that demonstrates how to “create a model Metadata Encoding and Transmission Standard (METS) Profile for thematic research collection” based on their work with the Whitman archives. The CDRH is also a teaching centre, and provides student learning and post-doctoral research opportunities.

- Center for Digital Scholarship <http://library.brown.edu/cds/>
Providence, RI, USA - The Center for Digital Scholarship in the Brown University Library supports digital scholarship for the Brown community and beyond by supporting scholarly and academic activities that are conducted or enhanced through the use of digital technology or that engage with its effects.

The CDS at Brown both develops digital technology to be used for academic and archival purposes, and supports Brown academic project through a grant program to support collection building within the university: “Among the major activities of CDS is the development and ongoing support of digital scholarly projects and systems at Brown.” The CDS at Brown has also digitized a variety of the library's ‘signature collections,’ which are large archival stores of historical documents such as the “African American Sheet Music collection” among others. The CDS website publishes a variety of technical documents pertaining to digital scholarship, as well as a list of articles written by members of the unit.

- Center for History and New Media <http://chnm.gmu.edu/>
Fairfax, VA, US - Since 1994, the Center for History and New Media at George Mason University has used digital media and computer technology to democratize history—to incorporate multiple voices, reach diverse audiences, and encourage popular participation in presenting and preserving the past.

CHNM describes itself as using “digital media and technology to preserve and present history online, transform scholarship across the humanities, and advance historical education and understanding.” Their projects fall under the categories of ‘teaching/learning,’ ‘research,’ and ‘collecting/exhibiting.’ CHNM, under the auspices of each category, has a large quantity of projects

underway, each of which are published on their site for public use. Under the category of ‘research,’ the CHNM is engaged in “developing publications and applications to define digital humanities for a new generation of scholars,” and in tandem to publishing public educational texts, develops online tools to do such things as cite research sources, publish online collections, and “find and compare syllabi from thousands of universities and colleges on any topic.” The CHNM archives a variety of collections from 9/11 images to a biographical history of Martha Washington.

- Centre de recherche Hubert de Phal <http://www.cavi.univ-paris3.fr/phalese/centre/centre.htm>
Paris, France - L’équipe se donne comme objectif de développer les études littéraires assistées par ordinateur par l’information, la formation, la veille sur les nouveaux outils de recherche.

The CRHP, a collective of academics, combines technology with literary studies with specific archival material and scholarly essays pertaining to Claude Simon, Proust and Lautréamont. They also collectively author and publish a series entitled *Cap’Agreg*, which covers a variety of literary topics and is printed in hard copy by Nizet in Paris. There are currently seventeen volumes of *Cap’Agreg*. They are currently working on an ‘Edition hypertextuelle des Complaintes de Laforgue’ and a Bibliographie sélective of transgressive writing in *Un prêtre marié* by Barbey d’Aurevilly. The last activity on the website was in 2005.

- Centre de traitement automatique du langage <http://www.uclouvain.be/cental>
Louvain-la-Neuve, Belgium - Le CENTAL est un centre d’informatique rattaché à la Faculté de philosophie et lettres et spécialisé dans l’étude du traitement informatique des langues, une discipline qui s’est fait récemment connaître du grand public au travers d’applications commerciales à succès telles que la correction orthographique automatique, la reconnaissance de la parole, la traduction automatique, la recherche d’information, etc.

The CENTAL is a research and teaching centre located at the Université Catholique de Louvain that studies ‘natural language processing.’ They maintain a large variety of research projects on a variety of language processing topics such as speech-to-speech synthesis and building an “SMS corpus for research.” The CENTAL’s website maintains a list of publications written by the center’s researchers as well as a list of doctoral theses being produced. CENTAL also researches / produces electronic projects collaboratively with private sector companies as well as other departments at the university. CENTAL also acts as teaching centre, and offers regular seminars as well as a Master’s degree in Linguistics.

- Centre for Computing in the Humanities – CCH <http://www.kcl.ac.uk/schools/humanities/depts/cch/>
London, England - An academic department and research centre in the School of Humanities that offers undergraduate minor and MA programmes, promotes the appropriate application of computing in humanities research and provides collegial support to its sister department.

The CCH operates out of King’s College, London. They characterize their research method as collaborative:

“research scholars and research technologists come together as equals and grapple with some of the knottiest problems in arts and humanities research, creating innovative solutions and new knowledge. We try to find ways of answering old questions in new

ways, and of finding new questions. Our research is very often project-based, but we also carry out research into new technical methods that have application in the arts and humanities.”

The site maintains a large list of current and complete projects (listed [here](#)) of which most are archived digital material, including manuscripts and images. All projects are published on the internet. CCH also operates as a teaching centre and includes a series of seminars.

- Centre for Educational Technology <http://www.cet.uct.ac.za/>
Cape Town, South Africa - Enables and investigates the use of ICTs in higher education at UCT and beyond.

CET “enables and promotes and investigates the integration of learning technologies in teaching and learning at the University of Cape Town and in higher education.” Their focuses are curriculum development, the development and support of learning technologies and research from which they produce workshop and conference papers. Many of their projects focus on issues pertaining to South Africa, including the ‘Mobile Learning Project’ which “explores the use of prevalent mobile technologies in [a] resource-poor context to support teaching and learning in higher education curricula.” CET publishes reports on the status of higher education in a variety of African countries. Other CET projects pertain directly to education, including training for online facilitators. and projects that explore the educational value of information and communication technologies.

- Centre for eResearch - CeRch <http://www.kcl.ac.uk/iss/cerch>
London, England -

The CeRch is a research unit based out of King’s College, London. The centre lists its five key focuses as:

1. researching and developing sustainable e-infrastructures for research;
2. the theory and practice of digital libraries and digital archives;
3. knowledge organisation, digital information and records management, including digital content creation and representation, access and use, curation and preservation;
4. researcher and citizen practices in the digital domain;
5. ICT-Methods with particular expertise in arts and humanities e-Science, geo-spatial and geo-temporal methods, text mining, textual analysis, and the use of grid and cloud infrastructures.

CeRch predominantly researches the impact of ICT (information and communication technologies) on scholarship. They publish a series called the Digital Research in the Arts and Humanities, which are still being produced. Each book focuses on a different subject and how it is impacted by technology. CeRch also offers an MA in Digital Asset Management.

- Centre for Literary and Linguistic Computing
<http://www.newcastle.edu.au/school/hss/research/groups/cllc/>
Newcastle, Australia - Established to continue the development and application of statistical and computing tools for the analysis of (literary) texts.

The CLLC operates within the Faculty of Education and Arts at the University of Newcastle. They currently list two research collaborations:

-
1. Priority Research Centre in Bioinformatics “Biomarker discovery and Information-based Medicine – on the classification and description of literary texts using bioinformatics and information theory quantifiers.”
 2. Centre for the Study of Research Training and its Impact “the Stylistics of PhD examination reports.”

The CLLC also conducts research in which it uses statistical analysis on literary texts, which it publishes on its website: “Our main and most usual procedure is first to establish frequency tables for the thirty, fifty, or hundred most common words—whatever they may be—of a given set of texts; to extract a correlation matrix; and to subject the matrix to principal components analysis.” They have developed an ‘Intelligent Archive,’ which is a Java based piece of software used for text analysis” housed at the university The website maintains a list of authors they are currently working on, dating from the sixteenth century to the modernists. Further, members of the centre have published extensively, and the CLLC website lists a large number of academic books and journal articles.

- Centre for Open Electronic Publishing - Centre pour l'édition électronique ouverte (Cléo) <http://cleo.cnrs.fr/>
Marseille, FRANCE - The Centre for Open Electronic Publishing (Cléo), is a center funded by CNRS, EHESS, Université de Provence and Université d'Avignon. It offers a full range of solutions for electronic publishing and academic communication, producing three portals : Revues.org, Calenda and Hypotheses. Revues.org is the platform for journals in the humanities and social sciences, open to quality periodicals looking to publish full-text articles online. More than 250 journals have already joined the platform. The platform's role is enhanced with unequalled academic communication tools: Calenda, the social sciences calendar, and Hypotheses.org, a blog platform.

Revue.org allows academic journals and research notebooks to publish their texts on the web by both publicizing the journals on the Revues website and by providing them with internet hosting (each journal seems to end in '.revues.org'). It also allows journals to promote recent articles in a separate news feed. The website also features a service entitled ‘Calenda,’ which labels itself as “the social science’s calendar.” Calenda is a platform for research centres, universities and academics working in the social sciences and humanities to promote conferences, presentations, seminars, public lectures and to publish calls for papers. Hypothese.org (called the “platform for humanities notebook”) shares information relating to education in the humanities, literature, literary prizes, and allows various organizations to publish their research notebooks. Hypotese.org describes itself as “une plateforme de publication de carnets de recherches. Les carnets de recherche offrent une solution rapide et légère pour rendre compte régulièrement de recherches. Peuvent ainsi être créés des carnets de fouilles archéologiques, des chroniques scientifiques sur un thème précis, des carnets de bord d’une recherche collective en cours, des carnets de terrain, des blogs de revues ou de livres, des *newsletters* scientifiques, etc. *Hypothèses* accompagne le travail de mise en forme et de valorisation de projets éditoriaux en mettant à disposition des équipes les fonctionnalités souples d’un blog.”

- Centre for Research in Arts, Social Sciences and Humanities <http://www.crassh.cam.ac.uk/>
Cambridge, UK - CRASSH promotes interdisciplinary research and innovation and hosts the University of Cambridge's Digital Humanities Network

CRASSH describes their mission as being “to promote collaborations across the Arts, Social Sciences and Humanities, and beyond their edges, in order to stimulate innovative and interdisciplinary thinking and dialogue and to reach out to new networks of interest and new publics.” CRASSH is

predominantly a teaching centre, and caters to faculty, post doctoral students and graduate students alike. They fund teachers and students researching within CRASSH's mandate in the form of fellowships and to attend conferences. Among their many focuses are the digital humanities, in which they have attempted to map the digital work being done in the humanities at Cambridge. CRASSH also has a series of working groups that bring together faculty and students to discuss their research with each other.

- Centre for Textual Scholarship - CTS <http://www.cts.dmu.ac.uk/>
Leicester, England - The CTS is devoted to scholarly research in the fields of textual studies and history of the book, and to the emerging technologies that support them.

CTS is based out of the Humanities faculty at de Montfort University, Leicester. Their three focuses of study are bibliography, textual studies and book history. There are several on going projects at CTS, mostly involving the creation of digitally annotated literary texts. Examples include the correspondence between Elizabeth Barrett Browning and Robert Browning and an electronic edition of Woolf's "Time Passes." CTS publishes a series of master lectures from The Annual Symposium of Textual Scholarship. CTS is also a teaching centre and provides supervision for MA and PhD students.

- Centre Informatique de Philosophie et Lettres <http://www.cipl.ulg.ac.be/>
Liege, Belgium - Son objectif general est de promouvoir et coordonner l'utilisation de l'informatique au sein de la Faculte de Philosophie et Lettres.

The CIPL operates out of the Université de Liège, and describes its mandate in the following points:

1. Développer des recherches propres en informatique appliquée à des domaines de recherches de la Faculté de Philosophie et Lettres.
2. Encadrer les étudiants de la Faculté pour l'apprentissage ou l'utilisation de l'informatique, dans le cadre du D.E.S. en Documentation et Sciences de l'Information notamment.
3. Coordonner l'informatisation des sections, des départements et des services administratifs attachés au Décanat de la Faculté.
4. Gérer les équipements informatiques mis à la disposition de la Faculté de Philosophie et Lettres par le SEGI, ainsi que les équipements acquis à l'intention du Décanat et des secrétariats des sections de la Faculté.
5. Offrir aux utilisateurs du Centre un service d'information et d'aide technique.
6. Servir d'intermédiaire entre les services utilisateurs et le SEGI en ce qui concerne les demandes de crédit, de personnel ou d'équipement.

From a publishing perspective, the CIPL have written and edited ten series of critical texts on a variety of topics ranging from the German language to archeology. The center also has computer rooms for students and researchers to use.

- Centre national pour la numérisation de sources visuelles <http://www.cn2sv.cnrs.fr/>
Paris, France - CN2SV is a digital humanities center and network for images, iconographics collections and scientific photos. Is founded by National Center for Scientific Research (CNRS) and the large A&H equipment "Adonis".

The CN2SV is a national technological resources specializing in digital images used for research. They are also engaged in building technological infrastructure to "preserve and promote digital data"

by research organizations. They predominantly develop and share a variety of technologies, including 'MediHal' which is an open archive that shares scientific graphic images. They have a second project entitled 'Mon Carnet 2.0' which is a web software used for data collection. They have several reports published on their website that outline their yearly activities. CN2SV also liaises with a variety of other digital organizations, government branches and funding bodies.

- Centre of Culture, Language and Information Technology - AKSIS <http://digital.uni.no/> Allegaten, Norway - A research umbrella organization containing the Research Group for Text Technology (text encoding and editorial philology, of which the Wittgenstein Archive and participation in the Text Encoding Initiative are most significant), directed by Daniel Apollon; the Research Group for Language Technology (computational linguistics, corpus linguistics, terminology and lexicography and participation in ICAME), directed by Gisle Andersen; and InterMedia (research within ICT-based information and knowledge dissemination and learning), directed by Barbara Wasson.

This organization is now called Uni Digital. It is owned by the University of Bergen and describes itself as “a research and development centre for language, language testing, electronic publishing, digital media and electronic education.” They produce online dictionaries, conduct research on how language levels are assessed, perform language testing, conduct research “in semantic technology in order to facilitate efficient search in complex text collections” and research “the pedagogical potential and limitations of digital media and eLearning systems.” They have a variety of projects completed/underway, one of which is to “prepare a new scholarly and text-critical edition of Henrik Ibsen’s collected works” that will be published in electronic and printed formats. UniDigital has an e-publishing division that develops “IT solutions for text coding, digital publications and philology, including classification and preparation of large text collections.” Their flagship work is the Wittgenstein Archive Project.

- Centre Virtuel de la Connaissance sur l'Europe (CVCE) <http://www.cvce.lu/> Sanem, Luxembourg - The Centre Virtuel sur la Connaissance sur l'Europe uses information and communication technologies to provide a unique point of reference for creating and disseminating valuable content on the history of European integration. It is based in Luxembourg.

The CVCE mandates itself within three areas of focus:

1. the research, gathering, selection, processing and dissemination of information on the European integration process;
2. the promotion of cutting-edge technology and the establishment of a point of reference in the field of Digital Libraries;
3. the exchange of information, experiences and resources for the benefit of European citizens and the world of education.

The CVCE is the creator of the ENA (the European NAVigator) which is “a multilingual, multisource and multimedia digital library that contains more than 16,000 documents on the historical and institutional development of a united Europe from 1945 to the present day.” The library features a large quantities of original content in the form of photos, interactive maps, press clippings, audio and video files, interactive maps and explanatory synopses. It is updated regularly. The CVCE also hosts annual conferences pertaining to European History, including various workshops on how to use the ENA.

- Centro di ricerche informatiche per le discipline umanistiche

<http://www.signum.sns.it/index.php?id=54>

Pisa, Italy - Signum is an independent center which carries out research in different areas, (particularly innovating is their work on image digitization). It is based at Scuola Normale Superiore, but has nothing to do officially with other Pisa renown institutions, like ILC or the University.

SIGNUM develops solutions for “storing, cataloging, research and analysis of collections of structured sequences (relational store and catalog), semi-structured (XML) and unstructured (text or biological sequences)” by combining “technological experimentation and the development of applications and software libraries for use by the scientific community” and various partner agencies. The publishing component to SIGNUM’s mandate comprises of four different publications:

1. between 1981 and 2002 they published a tri-annual bulletin;
2. “Collana di Studi” which catalogues texts relating to Art History;
3. “Collana strumenti” which considers “lexicographic analysis of texts of art, literary, and technical and computer tools available in print or on computer media in the form of textual database”;
4. a division that has created a variety of CD ROM and multimedia projects.

Their website features a long list of projects that SIGNUM has completed which range from online Italian dictionaries, a “digital library of historical texts of the Italian Renaissance and Humanism” and a project called “text and semantics” which creates “topical maps for the analysis of philosophical texts.” Website is entirely in Italian.

- CIRCAh <http://www.ucl.ac.uk/infostudies/research/circah/>

London, England - London, UK. Studies the application of computing and digital technologies to the arts and humanities. CIRCAh brings together expertise in digital humanities, user studies, digital libraries, human computer interaction and e-Science. Its location within SLAIS also allows for collaboration with colleagues in library and archive studies.

CIRCAh is a project based research division within the UCL Department of Information Studies. Within the mandate listed above, CIRCAh participates in a variety of research projects (of which the majority are collaborative). They currently have four projects underway:

1. LinkSphere with the University of Reading;
2. INKE (Implementing New Knowledge Environments) with the University of Victoria;
3. Image, Text and Interpretation (with Oxford);
4. TEI By Example, which “provide example-based introductions to eight different aspects of electronic text markup for the humanities.”

Each of CIRCAh’s projects are published and publicized on the web. CIRCAh’s website features an extensive list of publications authored by members of their research team.

-
-
- Cologne Center for eHumanities - CCeH <http://www.cceh.uni-koeln.de/>
Köln, Germany - The Cologne Center for eHumanities (CCeH) is a "Teaching and Research Center" at the Faculty of Humanities at the University of Cologne. Funded in 2009 (but in a rather long tradition of digital humanities) it develops and offers dedicated digital humanities competences and services for research and teaching in all departments of the faculty.

The CCeH supports and promotes the use of technology for research in the Digital Humanities. They mandate themselves to promote cross-disciplinary collaboration under the understanding that the future of research in the Humanities in general is dependant on the use of digital analysis and the creation of digital publishing networks. CCeH promotes projects in the Humanities that use technology as well as researches the methodology of digital productions. The CCeH promotes and publishes a wide variety of research projects under way as well as providing links to information on the field of Digital Humanities. CCeH also facilitates a variety of training opportunities for students and professionals wanting to increase their ability to use technology in their own research initiatives. This website is available only in German.

- Cultures Anglophones et Technologies de l'Information
<http://www.cati.paris4.sorbonne.fr/>
Paris, France - Conducts computer-assisted research in the humanities, specifically the history of representation with a focus on urban life; conducts courses and colloquia; publishes proceedings, research results, multimedia products (CD-ROM Georgian Cities, Presses de l'Université de Paris-Sorbonne, 2000); puts online research and learning resources on its website; is project leader of a national consortium for training students in humanities computing skills and search strategies (Campus Numérique 2001, "Réseaux.doc" 2001).

CATI is a research centre at the Université de Paris-Sorbonne Paris IV which also houses a doctoral program in "Civilizations, Cultures, Littératures et Sociétés." CATI explores how technology can be used while conducting research in the field of cultural history, including the use of hypermedia, hyperdocuments and virtual space. They have a particular focus on the impact of hypermedia on interdisciplinary initiatives. CATI has several publications to their credit, including a CD ROM project entitled Georgian Cities, which is a multidisciplinary course on British cities during the Enlightenment. CATI has also published a series called "Villes en visite virtuelle" as well as a book called "Du CD-ROM à internet." Their current project, which will also be published, is a virtual tour of Montmartre during the age of Jazz. CATI also publishes various articles and is represented at international conferences relating to the digital humanities.

- Digital Humanities Center for Japanese Arts and Cultures
<http://www.arc.ritsumeikai.ac.jp/lib/GCOE/>
Kyoto, Japan - Located at Ritsumeikan University

The Digital Humanities Center for Japanese Arts and Cultures is a research body located at Ritsumeikan University, Japan. The DHCJAC aims "to promote talented researchers who can carry the responsibilities of Digital Humanities education and research in while maintaining close ties with similar programs in various foreign countries. Such researchers will be aware of trends in Japanese

studies overseas, while possessing a good command of the information in the digital archives. They will be able to take a leading role on the world stage to advance the progress of research. In terms of research activities, we seek a deeper study of the humanities based on a Digital Humanities program whose subjects are Japanese art and culture, centered on the historic city of Kyoto.” The DHCJAC has five research groups, respectively focusing on Kyoto Culture, Japanese Culture, digital archiving technology, historical GIS and web technology. Each research group has their own individual set of mandates, and is led by a team of university professors. The research is presented widely at conferences and university seminars, and documented extensively on the DHCJAC website, which includes videos, photos and text from each of the center’s ventures. Although the DHCJAC website does not seem to list individual publications, the body of work they produce seems to be prodigious: they have large research teams who seem to be very active within their academic communities. They list their current focus as “the 21st-Century COE, "Kyoto Art and Entertainment Innovation Research," went into constructing digital archives and assembling databases that focus on the tangible and intangible cultural properties of Kyoto and Japan. We are proud to have achieved some of the best results in the nation, with more than a million items of data already compiled. Our multi-media format digital archives and information technology have greatly enhanced the methods and range of research in the humanities. They are also serving as portals for information exchange and personal exchange, becoming a hub for Japanese studies institutions around the world.” NB: Although there are some links in English, the majority of the website is uniquely available in Japanese.

- Digital Humanities Hub <http://dhh.anu.edu.au/>
Australia - The Digital Humanities Hub at the Australian National University brings together researchers working in digital media across disciplines such as: anthropology, visual arts, social and environmental history and literary analysis.

The DHH was founded in January 2010 as part of the Research School of the Humanities and the Arts at the Australian National University, and is a continuation of the work completed by the now defunct CRIO (the Consortium for Research and Information Outreach). Although the DHH has several projects underway, they have four ‘showcase’ projects:

1. AUSTLANG: Australian Indigenous Languages database;
2. a database of ‘migration memories’;
3. Living Knowledge: Indigenous Knowledge in Science Education;
4. Fighting carbon with fire: a video brief on Indigenous people and climate change for the United Nations University.

They are also affiliated with Project Bamboo (at UC Berkeley/U Chicago) which aims “to establish an international consortium in the field.” Most (not all) of the DHH’s projects are published online and have websites that make their research publicly available. The DHH is also an academic centre and offers courses in the ‘Visual Culture Research stream’ at the Research School of the Humanities and the Arts.

- Digital Humanities Observatory, Royal Irish Academy - DHO <http://dho.ie/>
Dublin, Ireland - Funded under cycle four of the Programme of Research in Third Level Institutions (PRTLII), the Humanities Serving Irish Society (HSIS) initiative is a partnership of equals committed to developing an inter-institutional research infrastructure for the humanities.

The mandate of the DHO is best summarized as “a central component within the Humanities Serving Irish Society (HSIS) initiative.... It will enable research and researchers in Ireland to keep abreast of international developments in the creation, use, and preservation of digital resources. It will fulfill these objectives by:

1. serving as a knowledge base in Ireland via consultations with project partners;
2. setting national standards to ensure the interoperability, preservation, and long-term accessibility of digital resources;
3. establishing a central repository which will provide access to a wide variety of interdisciplinary, multilingual, and multimodal digital resources created on the island of Ireland.

The DHO co-ordinates and consolidates initiatives in humanities digitisation so that new lines of enquiry developing from the HSIS/DHO infrastructure and research programmes will benefit all researchers in Ireland.” The DHO both disseminates information pertaining to the digital humanities (they have an extensive news and events section on their website) and provides training and research opportunities (including collaborative opportunities) to scholars, students and researchers, including an annual DHO Summer School. Although the DHO itself does not publish, its members and directors are well published researchers. The DHO news feed promotes several of their publications. The DHO website also provides links to other organizations, and contains an online forum for academic discussion.

- Electronic Text Research at the University of Saskatchewan
<http://etrus.usask.ca/>Saskatoon, Canada - Saskatoon, Canada A forum for text-oriented humanities computing researchers, mostly at the University of Saskatchewan.

ETRUS describes itself as “a research group at the University of Saskatchewan. Our members are faculty, students, and others, at the University of Saskatchewan and elsewhere, who are engaged in or interested in research on electronic text in any form.” Their website lists a broad variety of research projects, mostly having to do with the creation of digital editions of literary works, including a hypertext edition of *The Sound and the Fury* and a digital edition of Percy Shelley's *Posthumous Poems*. Most, but not all, of their projects are published on the internet and linked to on the ETRUS website. Further, a large variety of presentations and seminars are given at ETRUS, which are listed. The ETRUS website also shares news pertaining to digital scholarship, as well as members of the research group.

- eScholarship Research Centre <http://www.esrc.unimelb.edu.au/>
Melbourne, Australia - The University of Melbourne eScholarship Research Centre is committed to contributing to an open and equitable world of useful knowledge, a world in which the University of Melbourne will be a major contributor. The Centre was created in 2007 to bring together expertise from across the University and to continue the work of the Australian Science and Technology Heritage Centre (1999-2006) and its predecessor the Australian Science Archives Project (1985-1999).

The eSRC is a research unit at the University of Melbourne. They describe their research interests as relating to “cultural informatics, archives, data preservation and sustainable information systems for research.” The eScholarship Research Centre has a (very!) long list of projects currently underway, each with links and information. The majority of their research is published online. Further, they are

the publishers of the Encyclopedia Australian Science, available as a digital publication. The centre also publishes (digitally) a variety of databases serving as research resources, including a database of Colonial Australian Popular Fiction, World History of Science Online and many others. eSRC publishes/shares online resources to help students and academics working in digital scholarship including an 'online heritage resources manager' and WARP, which is a "database tool which enables the scholarly web publication of reference texts." Their online software and systems are all aimed at allowing scholars to publish archival material with efficiency and at a low cost. The eScholarship Research Centre also participates in and hosts conferences on the subject of digital scholarship.

- European Language Resources Association - ELRA <http://www.elra.info/>
Paris, France - The driving force to make available the language resources for [language engineering and](#) to evaluate language engineering technologies; active in identification, distribution, collection, validation, standardisation, improvement, in promoting the production of language resources, in supporting the infrastructure to perform evaluation campaigns and in developing a scientific field of language resources and evaluation.

ELRA describes its mission as the promotion of "language resources for the Human Language Technology (HLT) sector, and to evaluate language engineering technologies." They research and promote the identification, promotion, production, validation, evaluation and standardization of language resources. (I had to look up their definition of 'language resource': "a set of speech or language data and descriptions in machine readable form, used e.g. for building, improving or evaluating natural language and speech algorithms or systems, or, as core resources for the software localisation and language services industries, for language studies, electronic publishing, international transactions, subject-area specialists and end users.") They publish a catalogue of language resources which enables researchers to access a variety of databases with different types of language resources (verbal/written, etc...). Access to the catalogue is not free, and part of the ELRA site is dedicated to promoting it to research groups. ELRA also publishes a quarterly newsletter which provides update both on ELRA itself and on "the latest news in the field of language engineering."

- Göttingen Centre for Digital Humanities - GCDH <http://www.uni-goettingen.de/en/136017.html>
Göttingen, Germany - The Göttingen Centre for Digital Humanities (GCDH) is an digital humanities collaboratory providing research, teaching, academic studies and continuing education in the field of Digital Humanities.

This organization is currently being redeveloped and the website states that a new constitution will be adopted soon. The GCDH, located at the University of Gottingen, is a collaborative research unit specializing in the digital humanities. They describe themselves as the following: "The GCDH handles fund-raising activities, organization, coordination, implementation and support for interdisciplinary research projects in the field of Digital Humanities and its applications, and the Center will also coordinate and maintain cooperative activities with national and international institutions as well as contacts with private-sector partners and research institutions." (All information is written in the future tense as to reflect what they will become under the new constitution.) Although the GCDH will conduct research/publication, they do not have anything listed as of yet. Their research has not yet begun. The GCDH will also be a teaching centre and will house a double major program in Computer Science and the Humanities/Social Sciences.

-
- Loyola University Chicago's Center for Textual Studies and Digital Humanities - CTSDH <http://www.ctsdh.luc.edu/>

Chicago, IL, US - The center supports research across the humanities, as well as in computer science, the School of Communication, the social sciences, law, and University Libraries.

CTSDH developed out of the English department's area of excellence in textual studies and digital humanities, spearheaded by Peter Shillingsburg, Martin J. Svaglic Chair in Textual Studies. In addition to supporting research projects, the CTSDH sponsors conferences, symposia, and lectures, and offers undergraduate and graduate students the chance to work with faculty on advanced research, and to take courses in and pursue research of their own in the interdisciplinary areas of textual studies and digital humanities. Beginning in fall 2011, the CTSDH will sponsor a new interdisciplinary MA program in Digital Humanities.

- Maryland Institute for Technology in the Humanities - <http://www.mith2.umd.edu/>
College Park, MD, US - Aims to generate and foster the development of projects in applied computing, to communicate new technological approaches within and outside the University and to provide resources for technologically-based scholarship and instruction; appoints resident fellows.

On a day to day basis, MITH functions as an applied think tank for the digital humanities, both in furthering the excellence of its Fellows' research and in cultivating its own innovative research agendas—currently clustering around digital tools, text mining and visualization, and the creation and *preservation* of electronic literature, digital games, virtual worlds. Our work unfolds in a generous physical space, complemented by programs and events that include team-consultations for faculty digital projects, weekly Digital Dialogues (brown bags), frequent visiting speakers, courses taught in our seminar room, and ongoing interaction among fellows, students, and staff. Under the Our Work tab, there are numerous in-house projects listed. CenterNet is listed as one of MITH's community projects.

- MIT Center for Future Civic Media - <http://civic.mit.edu/>
Cambridge, MA, US - Bridging two established programs at MIT—one known for inventing alternate technical futures, the other for identifying the cultural and social potential of media change—the Center for Future Civic Media is a joint effort between the MIT Media Lab and the MIT Comparative Media Studies Program.

The Center for Future Civic Media is working to create technical and social systems for sharing, prioritizing, organizing, and acting on information. These include developing new technologies that support and foster civic media and political action; serving as an international resource for the study and analysis of civic media; and coordinating community-based test beds both in the United States and internationally. The website hosts a blog which is upheld by members. These same members write for the blog, and it appears to have had at least one new entry everyday since August 17, 2006. The blogs are also in-depth, graphic, lengthy and fascinating.

- National Centre of Biography, Australian National University
Canberra, ACT, Australia - The National Centre of Biography at the Australian National University was established in 2008. It aims to be recognized as a world leader in the development of online biographical dictionaries and related resources, supporting and promoting innovation in life writing.

Perhaps one of the NCB's most cherished accomplishments, and on-going projects, is the Australian Dictionary of Biography (ADB). The ADB is Australia's pre-eminent dictionary of national biography. It provides concise, informative and fascinating descriptions of the lives of over 11,500 significant and representative persons in Australian history. The ADB is available both as a print publication and [online](#). So far 18 volumes (including a supplementary volume of 'missing persons') of the ADB have been published, with entries ranging from Dutch explorer, Dirk Hartog (1580-1621), through to people who died in 1990. It is anticipated that volume 18, which will complete the period 1981-1990, will be published in 2012. Entries will then be prepared for those who died in 1991-2000.

- Networked Infrastructure for Nineteenth-Century Electronic Scholarship – NINES

<http://www.nines.org/>

US - Aims to protect, sustain, and enhance digital scholarship [and criticism](#) in "the long 19th century" by federating peer-reviewed digital scholarship for research and analysis in an innovative discovery, collaboration, and publication environment.

NINES (*Networked Infrastructure for Nineteenth-Century Electronic Scholarship*) is a scholarly organization devoted to forging links between the material archive of the nineteenth century and the digital research environment of the twenty-first. Our activities are driven by three primary goals:

1. to serve as a peer-reviewing body for digital work in the long 19th-century (1770-1920), British and American;
2. to support scholars' priorities and best practices in the creation of digital research materials;
3. to develop software tools for new and traditional forms of research and critical analysis.

- New Media Literacies <http://newmedialiteracies.org/>

Cambridge, MA, US - New Media Literacies (NML), a research initiative based within USC's Annenberg School for Communication, explores how we might best equip young people with the social skills and cultural competencies required to become full participants in an emergent media landscape and raise public understanding about what it means to be literate in a globally interconnected, multicultural world.

The publications on the website are, for the most part, a track record of the publications by the two Research/Staff Directors of NML: Henry Jenkins III & Erin B. Reilly. The publications include articles, books and press, but they also include online videos. These videos are instructional, but some are also youth-made and serve as examples of clear instruction (e.g. How to dance to the popular "Soulja Boy" song).

- New Zealand Electronic Text Centre <http://www.nzetc.org/>

Wellington, New Zealand

The New Zealand Electronic Text Centre has four aims:

1. To create a digital library providing open access to significant New Zealand and Pacific Island texts and materials. This encompasses both digitised heritage material and born-digital resources;
2. To effectively partner with other organisations, as a collaborator and service provider, on a

-
- variety of digitisation and digital content projects;
3. To build a wider community skilled in the use and creation of digital materials through teaching and training activities and by publishing and presenting the results of research;
 4. To work at the intersection of computing tools with textual material and investigate how these tools may be used to make new knowledge from our cultural inheritance.

- Office for Humanities Communication - OHC <http://www.ohc.kcl.ac.uk/>

London, England - An umbrella organisation that fosters communication among scholars and others involved in computer-related projects and activities; organises major conferences and workshops and publishes monographs and collections of essays in humanities computing and related areas.

Beginning in 1992, the OHC has published a series of monographs and collected papers concerned with the impact of computers in humanities scholarship and higher education. This programme of formal, refereed publications is an active one, producing at least one or two publications each year. The publications page on the website has four links under “Electronic Publications” one of which (the third) is a dead link. Of the Paper Publications, the most recent is one dated August 2004. However, there is a “forthcoming” publication which can be ordered by consumers (by which they receive an advanced copy). This publication is entitled “A Guide to Good Practice in collaborative working methods and media tools creation”.

- Oxford Text Archive <http://ota.ahds.ac.uk/>

Oxford, England - The Oxford Text Archive develops, collects, catalogues and preserves electronic literary and linguistic resources for use in Higher Education, in research, teaching and learning.

The Oxford Text Archive hosts Arts and Humanities Data Service (AHDS) Literature, Languages and Linguistics, which is one of five centres offering archival services for a range of disciplines. The AHDS also provides specialist advice for applicants to UK Higher Education funding bodies who are planning to create electronic resources. The website has an extensive catalogue that includes titles such as “Speech, Thought and Writing Presentation Corpus (STWP)”, “British Academic Spoken English corpus” and “Cognition, Biology and Idealist Philosophy.” The topics range from Linguistics, Philosophy, Greek Drama, Discourse analysis and History and criticism. Most all of the titles listed in said catalogue are freely accessible.

- Research in Computing for Humanities <http://www.rch.uky.edu/index.html>

Lexington, KY, US - A "collaboratory" of computer scientists and humanities scholars investigating humanities research problems that provide intriguing challenges for computer science, and developing new applications of computer technology to research in the humanities; sponsors a lecture series in conjunction with the UK Center for Computational Sciences.

The most current event is in fact not current, but past. In May–August 2009 there was a “Complete hardware and software upgrade to the RCH network and the workstations in the RCH Project Office and ‘Digital Research Incubator’ project lab.”

- Singapore-MIT GAMBIT Game Lab <http://gambit.mit.edu/>

Cambridge, MA, US - The Singapore-MIT. GAMBIT Game Lab is a five-year research initiative that addresses important challenges faced by the global digital game research community and industry, with a core focus on identifying and solving research problems using a multi-disciplinary approach that can be applied by Singapore's digital game industry.

“Our name partly represents our existence as an opening gambit into a new stage of development for the games industry as a whole, for Singapore as a new digital arts and entertainment center, and for the examination and consideration of games in an academic environment.

1. Gamers - Player cultures and communities. Behavior, habits, and values.
2. Aesthetics - Concepts of expression and representation to engage the senses.
3. Mechanics - Rule design, constraints, system dynamics, and challenge.
4. Business - Best practices for risk and project management, industry trends.
5. Innovation - Novel and hybrid genres of gameplay and demographic appeal.
6. Technology - Software and hardware interfaces, computation, and rendering.”

GAMBIT also aims to bridge the strenuous gap between industry and academia through the Game Lab. “GAMBIT's game lab provides a place for students, academics and industry professionals to work together to develop games that both expand the boundaries of what is done in games while still keeping a close eye on whether the games are financially feasible and, perhaps more importantly, are fun to play.

- The Center for Digital Humanities at the University of South Carolina <http://cdh.sc.edu/>
Columbia, SC, USA - The Center for Digital Humanities (CDH) at the University of South Carolina explores innovative and experimental approaches to research, education, preservation, and public programs for the interdisciplinary field of digital humanities. Reaching across institutions, colleges, disciplines, and departments, CDH initiates, nurtures, and assists in the development and deployment of computational resources, collaborative tools, and educational programming.

Founded in 2008, the CDH at USC began a “pilot effort” entitled the Digital Humanities Initiative. Through this pilot phase, the Center “promoted individual projects featuring diverse technical challenges, fostered innovative uses of technology in the classroom, and served as an incubator for campus-wide efforts to promote digital technologies.” Seen as success, the center began a new initiative, beginning since June 2010, to offer “new opportunities in the research and teaching of digital humanities even as we continue our earlier successes.” Some of the Centers on-going projects include: American Heritage Vegetables, Braille SC, E-Collation, Handheld Art, Handheld Education for Public History and the Humanities Gaming Institute.