

# ÉRUDIT : LE NUMÉRIQUE AU SERVICE DE L'ÉDITION EN SCIENCES HUMAINES ET SOCIALES

GUYLAINE BEAUDRY  
MARTIN BOUCHER  
TANJA NIEMANN  
GÉRARD BOISMENU

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/Studies in Book Culture*,  
vol. 1, n° 1, 2009, revue numérique.



## ÉRUDIT : LE NUMÉRIQUE AU SERVICE DE L'ÉDITION EN SCIENCES HUMAINES ET SOCIALES

### INTRODUCTION

Il est possible de retracer les principales composantes du parcours d'une institution lorsque celle-ci compte plus de dix années d'existence. Le recul devient possible et l'objet le mérite. Pourtant Érudit reste une jeune institution, même si dans le monde du numérique les institutions qui ont atteint cet « âge » font figure, non pas de pionnier, mais de précurseur d'une activité en matière d'édition et de diffusion de la recherche.

Que l'on retienne la trajectoire institutionnelle, le service mis en place ou la conception sous-jacente à cet « outil » majeur, certaines lignes de force se dégagent. Elles sous-tendent à la fois les ambitions et les réalisations. Le développement d'Érudit peut se concevoir comme un défi à la « loi de la gravitation commerciale ».

Cela est possible d'abord parce que certaines valeurs habitent cette entreprise. En ce sens, Érudit c'est un engagement ferme pour la valorisation et la diffusion du savoir, dans un modèle qui repose sur l'appropriation par la communauté des chercheurs de l'organisation et de l'infrastructure qui y œuvrent. Cela se veut une alternative professionnelle au modèle commercial ambiant. Au plan du corpus, Érudit nourrit un intérêt premier pour les revues scientifiques, puis transgresse les frontières en composant un système d'information des documents de recherche qui réunit, sans les confondre, les divers

genres éditoriaux. Très tôt, la mission de rayonnement de la recherche se double de la constitution d'un laboratoire de recherche mis à la disposition des chercheurs en fonction de leurs protocoles.

Érudit c'est une ambition mue par l'audace et une démarche innovante aussi bien dans le choix des façons de faire, des protocoles techniques retenus et du modèle organisationnel. Faire des choix en fonction de la qualité et de la durée en misant sur les normes qui ont cours au plan international est un leitmotiv depuis le tout début. Cela est possible grâce à une jeune équipe multidisciplinaire qui se projette vers l'avenir et qui veut faire œuvre durable. L'esprit d'entreprise est allé de pair avec la volonté de regrouper les compétences et de mettre à profit la collaboration interinstitutionnelle. Essentiellement, Érudit est fondée sur le mouvement et sur la projection vers le futur : c'est une idée, devenue institution dans notre société et portée par de nouvelles entreprises.

## HISTORIQUE

Depuis l'expérimentation des conditions techniques de production éditoriale pour deux revues savantes jusqu'à la constitution d'un fonds de plus de 50 000 articles, plusieurs phases ont marqué le développement d'Érudit. Cette section en retrace les différentes étapes.

### Expérimentation et projet d'un Centre de services d'édition numérique

En 1996-1997, les Presses de l'Université de Montréal (PUM) créent une « Direction des publications électroniques<sup>1</sup> » dont les activités portent principalement sur la transition des revues savantes vers le numérique. Les PUM ont alors mené un projet pilote d'édition numérique pour deux de leurs revues : *Géographie physique* et *Quaternaire et Surfaces*<sup>2</sup>. La valeur de ces publications allant bien au-delà des données, au sens informatique du terme, les choix technologiques ont été faits selon l'importance qu'accordent la communauté scientifique et la société en général aux contenus véhiculés par les revues savantes. La norme ISO 8879:1986, le *Standard Generalized Markup Language* (SGML), a été

retenue pour cette première phase du projet, en utilisant la Définition de Type de Document (DTD) ISO : 12 083 qui a dû être modifiée pour s'adapter aux besoins spécifiques des deux revues.

À la suite de ce projet pilote, l'étape suivante a été de porter cette expérimentation au niveau de plusieurs revues. L'hypothèse posée était qu'une infrastructure et des services dont profiteraient plusieurs revues permettraient de bénéficier d'économies d'échelle et de constituer une équipe disposant d'un haut niveau d'expertise éditoriale et technologique. C'est ainsi que projet VCOSP (*Virtual Centre for Online Scholarly Publishing*) a été conçu, sous l'impulsion d'Industrie Canada.

Ayant eu vent des expérimentations en cours, des fonctionnaires d'Industrie Canada viennent rencontrer la direction des PUM pour discuter de la possibilité de constituer un service commun de publication numérique de revues. L'agence fédérale est intéressée par cette question en raison de ses liens avec le (CRSH) du Canada qui gère un programme de soutien aux revues savantes. Le projet du VCOSP avait comme objectif d'être le lieu de production et d'hébergement des revues. Les services de production éditoriale devaient être offerts aux revues, à partir des textes prêts à être publiés. Ces services devaient ne concerner que les nouvelles parutions des revues. La numérisation et le balisage de la production rétrospective feraient partie d'une étape ultérieure. Après une période de mise en place d'une norme de publications, d'autres centres devaient être créés au Canada.

C'est à ce stade que l'Université de Montréal, par ses plus hauts représentants<sup>3</sup>, s'engage avec enthousiasme et octroie un budget pour l'aménagement et l'ameublement de nouveaux espaces pour les PUM ainsi que pour des équipements informatiques. Par ailleurs, un noyau de l'équipe des « publications électroniques » se met en place<sup>4</sup>.

Une entente de principe est conclue entre Industrie Canada et l'Université de Montréal pour la réalisation du projet VCOSP, qui prend le nom de projet *Virtuoso*. Le projet *Virtuoso* est annoncé en mars 1998 lors d'un colloque à Vancouver. Au moment de la présentation du projet par les fonctionnaires d'Industrie Canada, un tollé de protestations se fait entendre. Plusieurs collègues du reste du Canada

soulignent la trop grande centralisation du modèle : même si d'autres centres de publication sont prévus, un seul à l'Université de Montréal devait être établi dans la première phase du projet. Devant cette résistance, le contrat n'est pas signé et le projet achoppe. Néanmoins, des contrats préliminaires (*bridging contracts*) sont octroyés pour commencer les travaux, notamment la réalisation d'études portant sur les modèles économiques possibles pour la diffusion numérique des revues et sur les normes et les processus d'édition à mettre en place. Ces moyens permettent de procéder à une expérimentation d'outils et d'un processus d'édition numérique pour quelques numéros.

Dans la foulée du colloque de Vancouver et de discussions avec des collègues du reste du Canada, *Virtuoso* devient un projet interuniversitaire composé de Wilfrid-Laurier University Press, University of Toronto Press, Electronic Text Centre des bibliothèques de University of New Brunswick, la Canadian Mathematical Society, Canadian Association of Learned Journals, les Presses du Centre national de recherche du Canada et les Presses de l'Université de Montréal. Plusieurs réunions ont eu lieu entre les membres de ce réseau. Toutefois, ce projet n'a pas de réalisations tangibles, mis à part un lieu d'échange d'information et la naissance d'un réseau d'experts canadiens en édition numérique qui peut être identifié comme l'origine de ce que deviendra plus tard le projet *Synergies*.

## VERS UNE POLITIQUE QUÉBÉCOISE

À la suite de la déconvenue du projet d'Industrie Canada, c'est grâce au soutien et à la vision d'une agence du gouvernement du Québec que des activités d'édition numérique prennent forme.

### Projet pilote du Fonds FCAR : cinq revues en SGML

L'expertise développée au cours du projet pilote des PUM ainsi que les contributions de l'Université de Montréal de même que celles d'Industrie Canada pour le faux départ du projet *Virtuoso* sont significatives. Toutefois, le rythme rapide des développements technologiques en matière d'édition numérique risque de faire perdre les efforts et les res-

sources consentis jusqu'alors. Devant l'importance que revêt le projet pour le développement de l'édition numérique dans la communauté universitaire québécoise, les promoteurs du projet de l'Université de Montréal souhaitent mettre sur pied un nouveau partenariat pour réunir les conditions de réalisation d'activités d'édition numérique. Une rencontre est organisée le 15 août 1997 avec la direction du Fonds pour la formation des chercheurs et l'aide à la recherche (Fonds FCAR)<sup>5</sup> pour discuter d'un projet d'édition numérique pour les revues. À la suite de cette rencontre, Gérard Boismenu et Guylaine Beaudry présentent au Comité de régie du Fonds FCAR du 5 septembre 1997 le « Projet IRIS : pour l'édition et la diffusion de revues savantes électroniques ». IRIS propose un nouveau modèle pour la revue savante, pour effectuer une transition vers le numérique qui soit économiquement viable. C'est l'institutionnalisation de la revue savante « électronique » qui est visée.

La proposition du projet IRIS tombe à point nommé. Deux ans plus tôt, en 1995, le Fonds FCAR décidait, après une étude, de maintenir le programme d'aide financière aux revues. Dans la foulée, un comité d'experts sur la question de « l'édition électronique », présidé par Jorge Niosi, remet son rapport en juin 1997 et propose à la direction du Fonds FCAR de réaliser un projet pilote d'édition numérique. L'objectif principal du projet pilote proposé est « d'obtenir des recommandations ou des réponses aux questions posées par le comité pour que nous puissions offrir un soutien efficace et cohérent aux revues électroniques dans le cadre du prochain concours destiné aux revues scientifiques<sup>6</sup> ». Les questions posées par le comité d'experts sont les suivantes :

1. Quels sont les modèles techniques de publication à suivre (par exemple HTML, SGML, diffusion réseau, par cédéroms, intégrité et identification des textes, l'exploitation du multimédia, etc.) ?
2. Comment faire la construction de la visibilité sur les réseaux et du prestige de la revue dans le contexte de la numérisation ?
3. Quel est le modèle économique à adopter considérant les spécificités du document numérique en particulier, la visibilité de l'abonnement, les types de paiement et les différentes unités allant du micro-paiement portant sur la simple consultation

de quantités minimales d'information jusqu'à la constitution d'ensemble de revues regroupées en paniers et susceptibles d'être reprises en lot par des bibliothèques?

4. Quels sont les moyens à prendre pour assurer la légitimité scientifique et la reconnaissance par les milieux de la qualité de la publication électronique?
5. Quelles sont les modalités à retenir pour permettre la transition du papier à l'électronique tout en maintenant les acquis de la revue en tant qu'institution (aussi bien à l'égard des lecteurs que des auteurs)?
6. Comment faire le maintien de l'intégrité des textes soumis par les auteurs?

Après analyse du projet proposé par l'Université de Montréal et discussion sur les conditions de réalisation, le Fonds FCAR demande à cet établissement de réaliser le projet pilote. Un cahier des charges préparé par le Fonds FCAR précise les questions auxquelles le projet doit apporter des réponses, notamment concernant les aspects technologiques, économiques et institutionnels liés à l'édition numérique de revues savantes. Les six questions présentées plus haut font partie du cahier des charges. Pour discuter des questions posées ainsi que des orientations de la nouvelle plateforme, un comité aviseur du projet pilote est créé. Le comité aviseur<sup>7</sup> se rencontre quatre fois entre le 23 mars 1998 et le 2 mars 1999.


Les biens livrables de cette subvention, non renouvelable et provenant de l'enveloppe des projets spéciaux du Fonds, sont la mise en ligne de cinq revues subventionnées par le Fonds FCAR et un rapport. C'est grâce à cette première subvention qu'Érudit<sup>8</sup> est créé et commence à offrir des services d'édition et de diffusion numériques aux revues. La figure 1 présente la première page d'accueil d'Érudit et la figure 2, les titres des cinq premières revues qui joignent Érudit, à la suite d'un concours lancé par le Fonds FCAR.

Cette première phase de développement d'Érudit a permis d'examiner les aspects de la transition et d'expérimenter une chaîne de traitement pour la publication et la diffusion électroniques des revues. Les résultats sont décrits dans le rapport du projet pilote, déposé en



mai 1999<sup>9</sup>. La transition vers la publication numérique des revues y est décrite en précisant les modalités techniques pour y arriver. Le rapport préconise le regroupement des infrastructures pour la publication et la diffusion des revues par la création d'un « laboratoire » d'édition numérique. À la suite de ces travaux, le Fonds FCAR décide de revoir son programme d'aide à l'édition savante afin de stimuler la diffusion numérique des revues savantes québécoises.

Figure 1



Édition de revues universitaires et diffusion sur inforoutes

Le projet Érudit propose un nouveau modèle de publication et de diffusion qui correspond aux besoins des revues universitaires et de leurs lectorats.

---

[Revues électroniques](#)      [Information sur le projet Érudit](#)

[Recherche](#)      [Services offerts aux revues](#)

---

©1998 [Presses de l'Université de Montréal](#)

Première page d'accueil d'Érudit, en 1998.

Figure 2



Édition de revues universitaires et diffusion sur inforoutes

**Revues électroniques**

Pour vous procurer et installer les modules d'extension (*plug-ins*) nécessaires à la visualisation des textes en format SGML et PDF (respectivement *Panorama Viewer*, de SoftQuad, et *Acrobat Reader*, de Adobe), consultez la rubrique [Logiciels à télécharger](#).

Géographie physique et Quaternaire	Sociologie et Sociétés	Meta	Relations Industrielles / Industrial Relations	Revue d'histoire de l'Amérique française
<a href="#">Vol. 52, no 1 (1998)</a>		<a href="#">Vol. 43, no 1 (1998)</a>	<a href="#">Vol. 53, no 1 (1998)</a>	<a href="#">Vol. 51, no 3 (1998)</a>
<a href="#">Vol. 51, no 1 (1997)</a>				

---

[Projet Érudit](#) | [Recherche dans la collection d'articles](#) | [Services offerts aux revues](#)

---

Dernière mise à jour: 20 septembre 1998  
©1998 [Presses de l'Université de Montréal](#)

Premières revues diffusées dans Érudit.

## *Sociologie et sociétés* et *Meta* : premières collections rétrospectives en ligne en 1999

Le projet pilote du Fonds FCAR porte essentiellement sur la production éditoriale courante. Assez rapidement, il est apparu primordial de procéder à la numérisation rétrospective des collections des revues. C'est en rendant disponible l'ensemble des numéros des revues que le fonds acquiert une profondeur et un intérêt plus grand pour l'utilisateur.

Les premières collections rétrospectives ont été numérisées en 1999 grâce au programme « Initiatives spéciales » du Conseil de recherche en sciences humaines (CRSH) et une subvention pour emplois d'été pour étudiants d'Industrie Canada. Les revues *Meta* et *Sociologie et sociétés* ont reçu chacune une subvention du programme du CRSH pour procéder à la numérisation rétrospective de leurs fonds. Ces deux projets ont été réalisés par l'équipe de la Direction des publications électroniques des Presses de l'Université de Montréal<sup>10</sup>.

## Création de la Division du traitement de l'information : stabilisation de l'équipe

Depuis 1997 jusqu'en 2000, les activités d'édition numérique sont réalisées au sein des Presses de l'Université de Montréal en partenariat avec le Service des bibliothèques de la même université. La quasi-totalité des salaires de l'équipe de la Direction des publications électroniques est financée par des subventions. En 2000, la direction de l'Université de Montréal prend la décision de créer la « Division du traitement de l'information » au sein de la Direction générale des technologies de l'information et de la communication (DGTIC) et d'y mener les activités d'édition numérique. Quatre postes permanents sont ainsi créés pour réaliser des projets tel qu'Érudit ainsi que pour exécuter des mandats plus spécifiques en lien avec l'organisation et le traitement de l'information. L'idée du nouveau vice-recteur adjoint – technologies de l'information, Pierre Bordeleau, est de faire bénéficier à d'autres secteurs de l'Université les compétences fines développées en édition numérique. Ce changement de structure administrative assure une stabilisation d'un noyau dur de l'équipe, confirme un engagement de l'Université de Montréal dans ce

domaine porteur et permet de planifier à moyen terme les activités de développement du secteur.

## LE REGROUPEMENT DES FORCES

La réalisation du projet pilote, outre l'expérimentation technique, a donné lieu à des discussions entre la direction d'Érudit et des directeurs de revue, des chercheurs et des éditeurs sur les enjeux de la mise en œuvre d'une infrastructure et de services professionnels d'édition numérique au Québec. Un des éléments récurrents de ces échanges est la nécessité de travailler en collaboration sur une base interinstitutionnelle. Ce

besoin de se regrouper pour réunir les forces et les expertises des acteurs de l'édition des revues amène la formation du Groupe interuniversitaire pour l'édition numérique (GIEN). Le GIEN est à l'origine du Consortium interuniversitaire qui sera créé quatre ans plus tard. Le GIEN est composé des organisations suivantes : la Direction des technologies de l'information et de la communication de l'Université de Montréal, le Réseau d'informations scientifiques du Québec (RISQ), Les Presses de l'Université de Montréal, Les Presses de l'Université Laval, la Direction de la diffusion, le Service de la recherche et de la création de l'Université du Québec à Montréal et la Bibliothèque de l'Université Laval. Les principaux animateurs du GIEN sont Christian Allègre, Guylaine Beaudry, Benoît Bernier, Gérard Boismenu, Claude Bonnelly, Chantal Bouthat et Guy Teasdale.

## Étude de faisabilité commandée par le Fonds FCAR

Dans le processus de révision de la politique du Fonds FCAR à l'égard de la diffusion des revues, un comité d'étude s'est vu confier le mandat de faire le point sur le maintien du programme d'aide aux revues et de conseiller le Fonds sur la position à adopter face à l'introduction des technologies de l'information dans la communication des connaissances. Ce comité d'étude, par ses recommandations, a renouvelé la reconnaissance du besoin d'un programme d'aide financière — mais en le transformant — et a préconisé la promotion de la publication des revues en version numérique, tout en maintenant la rigueur de la sélection des

articles de façon à accroître la qualité et la diffusion internationales de la recherche québécoise. Afin de donner suite aux diverses recommandations, qui ont été acceptées par son conseil d'administration, le Fonds FCAR a lancé un appel d'offres pour la préparation d'une étude de faisabilité sur la conception d'un portail de production, de diffusion et de gestion de publications électroniques.

À la suite d'un concours, le Fonds FCAR choisit le « Groupe interuniversitaire pour l'édition numérique » (GIEN) pour réaliser le mandat. Cette étude consiste à proposer les paramètres permettant la réalisation d'un portail de production, de diffusion et de gestion de publications électroniques, en tenant compte de toute une série d'aspects. Les résultats de cette étude de faisabilité<sup>11</sup> ont été présentés à la communauté des directeurs de revue et des chercheurs en octobre 2000 lors d'un colloque organisé à Montréal par le Fonds FCAR. C'est lors de ce colloque que l'idée de distinguer et de traiter différemment la production et la diffusion émerge de la part du Fonds. Ainsi, la production éditoriale numérique doit demeurer sous le contrôle des revues. Par contre, la diffusion numérique pourrait faire l'objet de la création d'une infrastructure partagée et financée distinctement des mesures à prendre pour assurer la production.

## La diffusion : réponse à l'appel d'offre par le GIEN

Après l'étude de faisabilité et de la consultation qui en a découlé, le Fonds FCAR confirme la distinction entre production et diffusion. Au printemps 2001, la Conférence des recteurs et des principaux des universités du Québec (CREPUQ), à titre d'organisme regroupant l'ensemble des universités du Québec, est désignée responsable de la diffusion dans le milieu universitaire. Le Comité de la recherche de la CREPUQ crée le Comité de pilotage du Portail de diffusion des publications électroniques (PDPE) en mai 2001. Le rôle de la CREPUQ n'est pas de réaliser ce portail mais d'assurer la coordination générale du processus d'attribution du contrat et d'assurer la conformité des orientations du portail proposé aux objectifs du projet de mise en place et de gestion d'un portail pour la diffusion des revues en format numérique. Le 10 juillet 2001, un appel

d'offres accompagné d'un cahier des charges pour la réalisation d'un portail de diffusion des publications électroniques est diffusé.

Le mandat du portail est d'assurer, pour le compte des revues scientifiques qui en font partie, l'ensemble des opérations nécessaires à leur hébergement, leur archivage et leur diffusion en format numérique. Le portail vise d'abord à rendre visibles et accessibles, aux chercheurs et au grand public du monde entier, par le biais de l'Internet, les textes et le contenu des revues scientifiques qu'il héberge par les moyens de diffusion électronique à la pointe des technologies de l'information. Le portail est un service comprenant une équipe de personnes responsables de son opération, une structure de gestion, un ou des serveurs, un site web, une architecture et un environnement logiciel permettant d'accueillir, héberger et diffuser des revues savantes. La conception du portail doit reposer sur des principes et des standards reconnus et une architecture ouverte et évolutive compatible avec des standards ouverts comme XML, Open Archive Initiative, Open URL, etc.<sup>12</sup>

Le GIEN dépose une proposition<sup>13</sup> le 20 août suivant. En septembre, le mandat est donné au GIEN avec la demande de se constituer en organisme sans but lucratif (OSBL). Toutefois, les promoteurs du projet ont la surprise d'apprendre que la CREPUQ ne dispose pas du budget pour la réalisation de la plateforme. Le mandat est donné au GIEN de réaliser un portail, mais auparavant de compléter le montage financier du projet.

### La production : deux centres de production éditoriale (UdeM et ULaval)

De façon à alimenter en contenu la plateforme de diffusion, les membres du GIEN conviennent de créer deux centres de production éditoriale. Un centre de services d'édition numérique est créé à l'Université de Montréal, au sein de la Division du traitement de l'information de la DGTIC<sup>14</sup>, et un deuxième est établi à la Bibliothèque de l'Université Laval. Les équipes de production comptent utiliser les mêmes logiciels et les mêmes processus. Jusqu'alors, les services d'édition sont réalisés avec des outils et des processus pour répondre aux besoins de projets pilotes. Avec l'expertise développée de 1997 à 2001, les équipes des deux

universités sont en mesure de créer une chaîne de traitement qui soit plus robuste et qui permette un plus haut degré d'automatisation pour que les opérations soient assurées par du personnel technique. Afin de réaliser cette chaîne de traitement, le GIEN obtient une subvention du Fonds de l'autoroute de l'information du gouvernement du Québec pour le projet « Production et diffusion électronique de documents universitaires québécois » qui devait être conçu entre le 30 mars 2001 et le 31 décembre 2002. Les services d'édition numérique sont offerts dès l'automne 2002 aux revues du Québec. La tarification établie par les deux centres d'édition correspondait (comme c'est toujours le cas) aux coûts directs de production. Les frais de gestion, les frais de développements informatiques (autant pour la production que pour la diffusion) et les frais d'hébergement ne sont pas facturés aux revues.

Dès leurs débuts, les services d'édition, tant à Québec qu'à Montréal, rencontrent une demande solvable en raison du programme de soutien aux revues savantes qui intègre un « volet additionnel » pour financer spécifiquement la production numérique des revues. Les revues choisissent le prestataire ou l'institution qui procède à la préparation de leurs versions numériques. Seules les normes techniques adoptées par le Fonds FCAR doivent être appliquées pour la production des fichiers de façon à ce qu'ils puissent être diffusés convenablement par la plateforme de diffusion Érudit.

### Des revues à un système d'information multi genres : la deuxième génération de la plateforme Érudit (2002)

Outre la mise en œuvre des services d'édition numérique, les promoteurs du GIEN ont entamé des démarches pour le financement du développement de la plateforme de diffusion. La première génération de la plateforme Érudit ne pouvait pas répondre aux exigences de la montée en régime occasionnée par le nouveau programme du Fonds FCAR et l'intérêt croissant des revues pour effectuer une transition vers le numérique. C'est de nouveau le Fonds de l'autoroute de l'information (FAI) qui a financé le développement d'Érudit, cette fois-ci pour la deuxième génération de la plateforme de diffusion. Le projet « Portail québécois de diffusion de revues universitaires » est financé par le FAI tout en

étant complété par les contributions institutionnelles des membres du GIEN. C'est cette partie des développements qui correspond au portail de l'appel d'offres lancé par la CREPUQ à l'été 2001.

La figure suivante présente la page d'accueil de la deuxième génération de la plateforme Érudit<sup>15</sup>.



Le lancement de cette deuxième génération de la plateforme Érudit a eu lieu le 24 octobre 2002, à l'Université de Montréal. Les figures 4 et 5 présentent des photographies prises lors du lancement. C'est à partir de ce moment qu'Érudit ne se limite plus qu'aux seules revues. Les genres scientifiques du livre, des prépublications et des thèses sont également diffusés. L'idée directrice est que le chercheur ou l'étudiant ne cherche pas nécessairement par genre, mais bien selon des mots clés, des noms d'auteurs ou des thématiques. Bien que le type de publication soit d'une importance capitale pour la validation et pour l'évaluation de la source d'information, a priori, la recherche d'information gagne à se faire à travers des corpus de genres différents. Néanmoins, les différents genres scientifiques doivent être bien identifiés dans les résultats de recherche pour que le lecteur soit informé au premier coup d'œil du type de documents qu'il consulte.



En octobre 2002, près de 6 000 documents sont disponibles par la plateforme Érudit. Tous les documents sont en accès libre.

**Figure 4**



Lancement de la deuxième génération de la plateforme Érudit le 24 octobre 2002 dans le Hall d'honneur de l'Université de Montréal.

**Figure 5**



Les invités d'honneur au lancement de la deuxième génération de la plateforme Érudit le 24 octobre 2002 dans le Hall d'honneur de l'Université de Montréal.

## Création du Consortium Érudit s.e.n.c.

Depuis 2000, les membres du GIEN réalisent en collaboration plusieurs projets et participent à la mise en œuvre d'une infrastructure commune et partagée pour l'édition numérique de publications scientifiques. Un statut juridique est devenu nécessaire pour faciliter la gestion des subventions reçues ainsi que pour répondre aux enjeux et à la nécessité de prendre certaines dispositions contractuelles quant à la propriété intellectuelle. En 2004, après près de deux ans de discussion entre les établissements partenaires, le Consortium Érudit est créé légalement sous la forme d'une société en nom collectif, sans but lucratif. Les associés sont l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Les presses universitaires, qui étaient partie prenante des activités du GIEN, sont présentes dans la nouvelle organisation par un représentant au sein de son conseil d'administration<sup>16</sup>.

Avec ce statut légal confirmé par une convention, il a été possible de préparer des contrats de diffusion liant chacune des revues et le Consortium Érudit. Bien que les services de diffusion ne soient pas facturés aux revues, le contrat est essentiel pour préciser les obligations et les responsabilités des deux parties, comme des dispositions à prendre à l'égard des questions touchant la propriété intellectuelle.

## CONSOLIDER UNE EXPERTISE AU QUÉBEC AYANT UN RAYONNEMENT INTERNATIONAL

Depuis 1998, les animateurs d'Érudit sont en relation avec des collègues européens pour proposer et réaliser des projets en collaboration. Ces relations sont de plusieurs ordres, partant de la participation à un réseau d'échange d'information jusqu'au transfert technologique, comme dans le cas du projet avec le CNRS réalisé en 2004.

## Partenariats internationaux

Le type de collaboration le plus significatif et ayant des retombées profitables pour les lecteurs et les revues sont toujours à privilégier. En ce sens, la mise en commun de contenus et la réalisation de projets d'interopérabilité ont été priorisées. Pour faciliter l'échange des données, la

direction d'Érudit a proposé aux partenaires français d'utiliser le modèle de données Érudit Article<sup>17</sup>.

Les plateformes Cairn ([www.cairn.info](http://www.cairn.info)) et Persée ([www.persee.fr](http://www.persee.fr)) ont toutes deux adopté le format Érudit Article pour la production de leurs revues numériques. La collaboration est plus développée avec Persée, plateforme de numérisation patrimoniale et de diffusion des collections rétrospectives des revues françaises, dont la responsabilité incombait à la Sous-direction des bibliothèques du ministère de l'Éducation nationale de France. Une réelle interopérabilité des contenus est réalisée. Érudit donnant accès aux revues diffusées par Persée, y compris pour la recherche en texte intégral, et Persée offrant les mêmes services pour les revues diffusées par Érudit. Ainsi, un utilisateur effectuant selon des critères particuliers une recherche en texte intégral recevra des résultats portant sur des articles qui se trouvent dans Érudit et dans Persée, et ce, par l'une ou l'autre des plateformes. Cette entente, depuis 2003, offre aux revues québécoises et françaises une plus grande visibilité.

Avec l'intérêt manifesté pour ce service par les lecteurs et les directeurs de revues, la poursuite de cette démarche de collaboration avec des plateformes de diffusion des revues européennes a donné lieu à la création du Réseau francophone de diffusion de revues scientifiques en sciences humaines et sociales. Ce groupe s'est rencontré pour la première fois à la Délégation générale du Québec à Paris en décembre 2004 et a poursuivi ses travaux jusqu'en 2006. Les directions de Persée et d'Érudit ont invité des représentants des plateformes Revues.org, Revel, PEPS (projet belge d'édition numérique), CNRS/Centre d'édition numérique scientifique et Cairn à créer le Réseau dont les objectifs étaient :

1. d'offrir des services répondant aux normes internationales de qualité pour la diffusion et l'accès aux revues scientifiques ;
2. de donner accès, par l'une ou l'autre des plateformes du Réseau, à un fonds distribué de revues scientifiques, dans leur texte intégral; et
3. d'adopter les mêmes standards techniques afin de permettre la mutualisation des collections, et de participer à leur évolution<sup>18</sup>.

Des cinq réunions tenues entre novembre 2004 et juin 2006, plusieurs documents ont permis de préciser ces objectifs : objectifs, principes et mode organisationnel du Réseau francophone, critères d'admissibilité des revues au sein du Réseau, protocole de développement en collaboration du schéma Érudit Article, principes et services du Réseau francophone, mutualisation des fonds : considérations techniques. Pourtant, mise à part la collaboration entre Érudit et Persée, il n'a pas été possible de développer davantage le Réseau.

### Service de gestion des abonnements

Depuis le début des activités d'Érudit, les revues sont diffusées en accès libre. En raison de la réalité budgétaire de la majorité des revues, les directeurs de revue réunis en assemblée ont souvent exprimé, et ce dès 2004, la nécessité d'avoir un service de gestion des abonnements et de filtrage des accès. Entre 2004 et 2006, plusieurs réunions, rencontres et échanges ont lieu avec les directions de revue et les éditeurs pour préciser le modèle économique. Les directions de revue expriment des demandes diverses et difficilement réconciliables selon un spectre allant de l'accès libre jusqu'à l'accès aux seuls abonnés pour l'entièreté de leurs collections.

Un consensus s'est dégagé sur le modèle binaire suivant qui est toujours en application. Une revue diffusée dans Érudit est disponible soit en accès libre, soit selon le principe de la barrière mobile de deux ans. Ainsi, les revues qui optent pour un modèle économique comprenant la souscription d'abonnements voient leurs articles des deux dernières années de publication accessibles seulement à leurs abonnés. Par ailleurs, tout le reste de leur collection est disponible en accès libre. Il s'agit d'un modèle hybride qui a fait ses preuves. Il assure une large diffusion tout en garantissant aux revues de bénéficier de revenus d'abonnement.

Aujourd'hui, des institutions et des individus de partout dans le monde sont abonnés à Érudit. De plus, depuis 2008, les revues diffusées par Érudit sont présentes dans toutes les bibliothèques universitaires de recherche du Canada grâce à une entente pluriannuelle d'achat groupé négociée entre Érudit et le consortium des bibliothèques de

recherche canadiennes, le Réseau canadien de documentation pour la recherche.

### Pour un accès intégré à la documentation scientifique : la troisième génération de la plateforme Érudit (2008)

Le 23 avril 2008, le Consortium Érudit procède au lancement de la troisième génération de sa plateforme web. Grâce à une importante contribution financière du ministère du Développement économique, de l'Innovation et de l'Exportation (MDEIE), Érudit offre une bibliothèque numérique innovatrice pour la promotion et la diffusion de la recherche universitaire. Cette démarche inclusive donne toute sa richesse dans un site complètement revu, non seulement dans son apparence<sup>19</sup>, mais aussi dans ses outils et ses fonctionnalités. À partir de ce moment, la plateforme Érudit offre les services à valeur ajoutée les plus avancés dans ce domaine : interface de navigation trilingue (français, anglais et espagnol), outil de recherche dans plusieurs genres scientifiques (revue, livre/actes, thèse, autres documents), filtres des résultats de recherche selon plusieurs critères, index auteurs-titres par collection de revue, exportation des notices, notamment pour l'utilisation de logiciels de gestion bibliographique, interrogation Z39.50, indexation pour recherche et navigation des fonds de plateformes partenaires. Ces développements ont également pu être réalisés et une partie des activités courantes des services de diffusion de la plateforme est assurée grâce du soutien constant du Fonds québécois de recherche sur la société et la culture.

Figure 6



Troisième génération de la plateforme Érudit lancée le 27 avril 2008.

Une partie significative du projet subventionné par le MDEIE a rendu possible la numérisation et la mise en ligne de centaines de numéros des revues savantes québécoises publiés seulement sous forme imprimée. Près de 300 000 pages d'articles et de comptes rendus de revues ont été ajoutées au fonds d'Érudit et participent à une plus grande diffusion des résultats de la recherche publiés au Québec.

**Figure 7**



Les membres du comité exécutif du Consortium Érudit lors du lancement du 27 avril 2008 (Gérard Boismenu, Guylaine Beaudry, Chantal Bouthat et Silvie Delorme).

Cette réalisation marquant le dixième anniversaire d'Érudit a également contribué à l'augmentation significative des consultations. Seulement pour l'année 2008, la plateforme Érudit a reçu plus de 3,6 millions de visiteurs de partout dans le monde. Avec cette nouvelle plateforme, plus que jamais, Érudit propose une structure professionnelle de diffusion de la connaissance scientifique dans le cadre du service public. Les services éditoriaux de cette nouvelle plateforme sont décrits dans la prochaine section.

**Figure 8**



L'équipe d'Érudit du Centre d'édition numérique de l'Université de Montréal lors du lancement du 27 avril 2008.

**Figure 9**



Des membres de l'équipe d'Érudit de la Bibliothèque de l'Université Laval lors du lancement du 27 avril 2008.



## SERVICES D'ÉDITION NUMÉRIQUE POUR REVUES SAVANTES

Le numérique est devenu un outil indispensable à la production des publications scientifiques. Les auteurs, les éditeurs, les graphistes et les imprimeurs travaillent tous avec le numérique. Pourtant, bien souvent, on préfère encore la lecture sur papier à la lecture sur écran. À l'heure actuelle, la production d'une publication numérique ne se distingue pas fondamentalement de celle d'une publication « traditionnelle ». C'est sa diffusion qui se trouve transformée dans le nouvel environnement numérique. C'est là que se situe la véritable originalité des publications numériques. Dans ce contexte en mouvement continu, la capacité d'un document à être bien référencé devient une préoccupation éditoriale centrale.

Les processus de publication d'Érudit répondent aux exigences de la production de publications universitaires, leur diffusion, leur archivage et leur accessibilité à long terme. Érudit offre des services professionnels d'édition numérique adaptés pour les différents types de publications. D'une part, le service d'édition numérique de revues, de livres et d'actes permet de diffuser ces documents dans leur version PDF et/ou XHTML, tout en les intégrant dans l'outil de recherche. D'autre part, Érudit présente une structure d'accueil pour la diffusion des thèses et mémoires. Les visiteurs peuvent consulter les thèses sous forme numérique des universités partenaires, par deux moyens : 1) en parcourant la liste des mémoires et des thèses grâce à un index par auteurs, et 2) par les résultats de recherche qu'ils obtiennent en utilisant l'outil de recherche d'Érudit. Toute université québécoise peut faire une demande d'indexation de ses mémoires et de ses thèses numériques. Ce service est gratuit. La zone « Autres documents et données » de la plateforme Érudit est un dépôt interinstitutionnel pour les documents et les données d'une trentaine d'unités de recherche des universités québécoises. Selon les cas, les unités déposent leurs documents au fur et à mesure en disposant d'un accès au dépôt qui leur est propre ; ou encore, les documents sont traités par une personne de l'équipe d'Érudit. Plus de 2600 documents sont présentement archivés et diffusés dans ce dépôt.

Au cours de prochaines sections, notre attention sera portée sur les conditions de publication numérique des revues savantes.

## Les revues diffusées dans Érudit

À l'heure actuelle, Érudit diffuse 86 collections de revues et produit des numéros courants de 69 revues des sciences humaines, sociales et des sciences naturelles. Plus de la moitié des revues ont fait l'objet d'une numérisation rétrospective de l'ensemble de leur collection. 61 revues publient une version imprimée et une version numérique en parallèle. 15 revues sont offertes en version numérique uniquement.

La plateforme Érudit donne, par ailleurs, accès aux fonds d'autres diffuseurs et plateformes, notamment de l'*Electronic Text Centre*<sup>20</sup> de l'Université du Nouveau-Brunswick et du portail *Persée*<sup>21</sup>. Érudit et Persée offrent un accès à leurs fonds mutualisés, par l'un ou l'autre des sites, tant en navigation que par leurs outils de recherche respectifs. Une recherche dans Érudit permet ainsi de repérer et de consulter des articles publiés dans Persée et ceux publiés par l'*Electronic Text Centre*.

Érudit et Les Presses scientifiques du CNRC ont également mutualisé leurs fonds, ce qui permet d'avoir accès à l'ensemble des collections par l'un ou l'autre des sites, aussi bien en navigation que (très prochainement) par leurs outils de recherche respectifs. Dans l'ensemble, 136 revues et des centaines de milliers de documents sont ainsi disponibles depuis la plateforme Érudit.

## Services offerts aux revues

Érudit publie et diffuse toute revue savante quel que soit l'éditeur ou le pays d'origine. Pour garantir un cadre scientifique et de qualité, le comité éditorial d'Érudit a développé des critères d'admissibilité pour l'adhésion à la plateforme. Ainsi, une revue est acceptée pour diffusion si elle satisfait les critères suivants :

1. avoir comme objectif principal de publier les résultats de travaux de recherche avancée et d'érudition rédigés par des spécialistes à l'intention de spécialistes ou de praticiens ;

2. être en mesure de démontrer la qualité de sa direction scientifique, notamment sur la foi des compétences du directeur et des autres membres du comité éditorial, et du processus d'évaluation des articles par les pairs ;
3. paraître au moins deux fois l'an et avoir paru au cours des deux années précédentes ;
4. être rédigée en français ou en anglais, ou dans une autre langue internationale, tel l'espagnol ;
5. appliquer des pratiques commerciales dites d'« édition responsable », à défaut de choisir le libre accès.

Le travail éditorial sur le contenu des revues reste entre les mains des équipes éditoriales de chaque revue. Toutes les activités en relation avec les auteurs et les évaluateurs pour la préparation des textes à publier relèvent entièrement de la responsabilité de la revue. Ce n'est qu'au moment où l'équipe de la revue transmet les textes prêts à publier à Érudit que la production numérique s'effectue.

Différents processus d'édition sont offerts pour la production de numéros courants et de numéros ayant été publiés en version papier ultérieurement. Le souci premier est la production d'un numéro de qualité en respectant de courts délais. Les revues accordent à Érudit une licence non exclusive pour la diffusion numérique des numéros issus de sa production courante ou rétrospective. Les éditeurs détiennent tous les droits, titres et intérêts sur les articles publiés. Érudit s'engage à faire figurer à chaque consultation des articles la mention légale. Érudit participe au programme de dépôt légal des publications numériques de Bibliothèque et Archives Canada (BAC). Tous les numéros diffusés sur la plateforme sont envoyés à BAC pour satisfaire aux exigences du dépôt légal pour la version numérique.

## Modèles économiques

Les revues disponibles dans Érudit sont pour la plupart éditées et produites par des éditeurs non professionnels et sans but lucratif comme les unités d'enseignement, les sociétés savantes et les presses universitaires. Il s'agit rarement d'éditeurs commerciaux.

Pour la production de la version numérique de numéros courants, les revues doivent trouver une façon de récupérer les coûts engagés. Règle générale, les structures éditoriales des revues savantes bénéficient d'un financement assuré par des subventions publiques, éventuellement, par l'institution universitaire ou par la société savante d'attache, et par le revenu des ventes de la revue.

En accord avec les éditeurs et les auteurs, Érudit a adopté un modèle économique flexible et simple. Les directions des revues diffusées dans Érudit font le choix de l'accès libre (*open access*) ou adoptent le principe de la barrière mobile de deux ans : pour les revues qui en font la demande, les numéros des deux dernières années de parution sont disponibles sous abonnement. Tout ce qui est antérieur aux deux dernières années de publication est en accès libre. Au total, plus de 80 % du contenu est en accès libre, tout en permettant aux revues de tirer les revenus nécessaires pour assurer la poursuite de leur projet éditorial.

Érudit propose deux types d'abonnements institutionnels à la version numérique de ses revues : par titre ou à l'ensemble de la collection (panier). Les tarifs d'abonnement sont fixés par les directions de revue. L'accès se fait par reconnaissance des adresses IP. La durée d'un abonnement est établie sur une année civile. Un service d'authentification donne accès aux abonnés individuels, si la revue décide d'offrir un tel service à ces lecteurs.

## LE XML, LE FORMAT DE PRODUCTION ET DE DIFFUSION

Depuis ses débuts, Érudit a développé une expertise dans la publication basée sur la norme XML, assurant les meilleures conditions d'utilisation et de préservation des documents numériques. Tous les processus d'édition sont basés sur cette norme qui a été recommandée par le W3C en 1998. Le balisage d'un texte scientifique à l'aide de la norme XML permet de rendre lisible par ordinateur sa structure sémantique et hiérarchique. Chaque élément d'un document ainsi structuré peut être stocké, recherché, réutilisé et matérialisé par plusieurs médias. Tous les articles des revues diffusées par Érudit sont encodés selon le schéma

Érudit Article spécialement conçu pour présenter les caractéristiques des articles de revues scientifiques.

## Les processus de production éditoriale

Érudit a mis en place une chaîne de traitement basée sur la conversion de fichiers sources vers des documents XML. Ainsi, la préparation d'une version numérique se fait dans les formats XML, PDF et XHTML. Différents processus sont appliqués pour la production de numéros courants et pour la numérisation rétrospective des collections.

### PRODUCTION DE NUMÉROS COURANTS

Les textes qui proviennent des équipes de revues où des éditeurs pour la publication de numéros courants sont en format de mise en page (QuarkXPress, PageMaker, InDesign) ou de traitement de texte (Microsoft Word). Dans le cas des revues publiées sous format papier, il s'agit des mêmes fichiers transmis à l'imprimeur pour la production de la version imprimée.

Après l'analyse éditoriale du numéro permettant d'identifier tous les éléments sémantiques des articles à traiter, la première étape de la chaîne de traitement consiste en l'extraction des textes pour les convertir en format RTF. Un outil spécialement conçu par l'équipe d'Érudit transforme le RTF vers le format XML. Des métadonnées<sup>22</sup> préalablement saisies dans une base de données sont également ajoutés au moment de la création du format XML.

Pour garantir l'affichage adéquat, en XHTML, de tous les caractères graphiques autres que des lettres et des chiffres par tous les navigateurs, il est nécessaire de faire une gestion des caractères spéciaux. Certains codes de caractères doivent être remplacés par leurs équivalents en jeux de caractères normalisés (Unicode – UTF-8<sup>23</sup>).

Lorsque le traitement d'un article est effectué, on procède à une analyse des images à produire. Il peut y avoir des images sources, créées par le graphiste qui a produit l'article et fournies avec les fichiers de mise en page, et des images qu'il faut saisir à l'écran, à partir du logiciel de mise en page ou à partir d'un fichier PDF. Le traitement de chaque image dépend du type d'image reçue<sup>24</sup>. Une fois les images

de diffusion produites et leurs propriétés intégrées dans la base de données de production, les données XML relatives à l'image traitée sont générées pour être intégrées dans les documents XML des articles. Des traitements semblables doivent être effectués pour le traitement de fichiers vidéo et audio.

Avant la mise à disposition des épreuves numériques à la direction de revue, des étapes de validation éditoriale et de contrôle de qualité rigoureux sont effectuées par les membres de l'équipe de production. La procédure pour la correction des épreuves numériques est la suivante :

1. mise en ligne des épreuves numériques, non accessibles au public ;
2. envoi par courriel à la direction de la revue d'une URL non diffusée ;
3. contrôle des épreuves par la direction de la revue ;
4. acceptation par courriel des épreuves numériques avec, le cas échéant, les corrections à effectuer ; et
5. mise en ligne publique des articles.

#### LA NUMÉRISATION RÉTROSPECTIVE

Les numéros dont les fichiers sources ne sont plus disponibles ou lisibles doivent être numérisés et traités par des outils de reconnaissance optique de caractères (ROC). Une fois la collection à numériser restituée, elle doit être numérisée à l'aide d'appareils munis d'un alimenteur automatique. Un fichier PDF de type image – texte, des fichiers RTF et TXT enregistrés en UTF-8 et un fichier TIFF avec une résolution de 600 dpi pour l'archivage sont ainsi générés<sup>25</sup>. La reconnaissance optique des caractères est réalisée à partir du fichier TIFF. Les articles sont ensuite convertis en format XML à l'aide de la chaîne de production éditoriale pour le rétrospectif.

Un tiré à part en format PDF pour chaque article est rendu disponible permettant ainsi aux lecteurs de les consulter aisément. On estime que plus de 90 % des mots d'un texte numérisé seront correctement interprétés par les logiciels de ROC. Évidemment, la qualité initiale de l'impression et les polices utilisées influencent grandement ces résultats. Une vérification est donc nécessaire afin de

corriger les erreurs de ROC les plus flagrantes, dans les métadonnées essentiellement. La transformation du contenu du document en format XML permet son exploitation par l'outil de recherche et par les autres services de navigation tels que la création des index des auteurs ou des numéros thématiques.

Le travail éditorial effectué vise à opérer la translation d'une forme à l'autre. Un contrôle de qualité rigoureux et une validation éditoriale sont indispensables. L'évolution des conventions éditoriales et les nombreux changements de directions qui ponctuent l'existence des publications universitaires complexifient considérablement ce processus. Les incohérences typographiques ou éditoriales qui en découlent et qui restent plutôt invisibles en consultant la version imprimée surgissent souvent lors de la numérisation<sup>26</sup>. L'équilibre entre la consolidation de la collection numérique et l'impératif de l'authenticité des publications font de cette étape bien souvent une entreprise complexe.

## Le schéma Érudit Article

Le balisage en XML et l'exploitation des revues diffusées par Érudit sont assurés par l'utilisation du schéma Érudit Article<sup>27</sup>. La norme XML est un langage permettant de définir des formats de document. Ces formats sont en fait des modèles de documents, généralement appelés des schémas. Un schéma pour sa part permet d'exprimer l'ensemble des contraintes que doit respecter un document XML pour qu'il soit valide dans un certain contexte. Ainsi, à l'aide du schéma défini pour des articles de revues savantes, il est possible de vérifier si un article en format XML est valide en fonction du schéma Érudit Article.

Les documents structurés contiennent de l'information sur leur structure logique et sémantique. Dans le cas des articles de revues savantes, chaque élément du texte se voit attribuer des balises qui l'identifient. Par exemple, on y identifie explicitement le nom de l'auteur et son affiliation, les résumés dans les différentes langues, les mots-clés, les paragraphes du texte ainsi que les citations, les figures et les tableaux et les références bibliographiques. Voici un exemple de balisage du nom d'auteur et de son affiliation :

```
<auteur> <nompers> <prenom>Louise</prenom> <nomfamille>Nadeau</nomfamille> </nompers> <affiliation>Professeure titulaire, Département de psychologie, Université de Montréal</affiliation> </auteur>
```

Toutefois, le balisage XML des articles n'offre pas uniquement des capacités supérieures de recherche, grâce à la richesse de ses modèles de données, il permet également de diffuser largement l'information traitée. L'utilisation du XML a permis à Érudit d'implanter facilement une stratégie de diffusion qui assure l'indexation des publications tant par les outils de recherche spécialisés (tels que *Google Scholar*) que par les bases de données bibliographiques disciplinaires (Francis, Repère, *ABC-Clio*, *PubMed/MedLine*, *Nines*, *Philosophical abstracts*, etc.). Érudit est également compatible avec les applications Open URL, telles que SFX et Article Linker.

## Les formats de présentation : XHTML et un tiré à part en format PDF

Le balisage fin et l'identification de chaque élément du texte permettent de multiples possibilités d'exploitation. Transformé en XHTML à l'aide de feuilles de style CSS et XSLT pour la présentation web, le document structuré offre une flexibilité exceptionnelle. Le XHTML permet d'ajuster la disposition et la présentation du texte afin qu'elles soient adaptées à la lecture à l'écran et aux préférences individuelles de chacun des lecteurs. Les résumés, bibliographies, figures et images peuvent être extraits de leur contexte pour faciliter leur consultation ou pour donner un aperçu du contenu d'un article depuis la liste des résultats. Les titres peuvent être récupérés pour créer une table des matières alors que les notes de bas de page peuvent être reliées à leur appel par des liens bidirectionnels afin d'améliorer la navigation.

Les tirés à part diffusés dans la plateforme en format PDF sont optimisés pour diffusion web conformément aux normes de production et de préservation établis par Érudit. Celles-ci demandent, entre autres, l'ajout de métadonnées et l'intégration des polices pour les PDF texte. En plus, une page de garde, ajoutée par Érudit, fournit aux



lecteurs le contexte du document consulté en plus de l'information bibliographique complémentaire pour chaque article.

Le PDF a pour avantage de préserver l'apparence originale des documents. Toutefois, le PDF reste davantage un format d'impression sur demande qu'un format de visualisation ou d'exploitation. Ce format ne fait que restituer à l'écran ce que l'on retrouve sur papier. Anciennement un format propriétaire, donnant aucune assurance quant à la pérennité de l'information stockée, le PDF Archive (PDF/A) est maintenant une norme ISO<sup>28</sup> et a pour objectif d'assurer la conservation à long terme de fichiers archivés sous forme numérique.

### Imprimé vs numérique, de la translation à la création d'un nouveau modèle pour la revue

L'édition numérique ouvre de nouvelles possibilités : elle augmente spectaculairement l'efficacité du transport de l'information et améliore bien souvent la présentation des publications scientifiques. Grâce à elle, de nouvelles latitudes, impensables auparavant, s'offrent maintenant aux éditeurs, aux chercheurs et aux lecteurs. Dorénavant, il est non seulement possible d'illustrer des documents textuels avec des images et des graphiques, mais aussi de créer des œuvres multimédias avec des séquences vidéo et audio intégrées. Comme le démontre le prototype de l'article scientifique du futur récemment présenté par *Elsevier* et *Cell Press*, le numérique permet de décomposer « la structure organisationnelle de l'article afin que les lecteurs puissent naviguer entre les couches de contenu selon leur niveau d'expertise et d'intérêt<sup>29</sup> ».

À l'heure actuelle, 50 revues sur 71 diffusées dans *Érudit* sont des versions parallèles, c'est-à-dire qu'elles sont des translations de la version imprimée. Le potentiel d'innovation de ces versions parallèles est limité. En fait, il s'agit plutôt de la continuation des revues conventionnelles avec une impression de modernité. Selon l'expérience *Érudit*, les équipes éditoriales expriment rarement le désir de profiter des moyens du numérique. Très peu veulent ajouter du contenu audio ou visuel à la version numérique d'un numéro afin de l'enrichir. Il semble que pour le moment ce n'est pas encore un besoin réel dans les sciences humaines et sociales. Toutefois, l'émergence de nouveaux

concepts éditoriaux peut être observée : la nouvelle revue *CuiZine The Journal of Canadian Food Cultures/Revue des cultures culinaires au Canada*, disponible en version numérique uniquement, par exemple, a intégré des contenus créatifs et dynamiques à son projet éditorial. Pour la première fois dans l'histoire d'Érudit, du contenu vidéo et audio fait partie intégrante de chaque numéro diffusé.

## INFRASTRUCTURE TECHNOLOGIQUE

Au plan technique, la troisième génération de la plateforme de diffusion constitue un point tournant dans l'évolution d'Érudit. Le passage du format de présentation HTML statique vers une génération dynamique des contenus a amené l'équipe d'Érudit à redéfinir ses orientations technologiques ainsi que ses méthodes de développements.

### Les choix technologiques

L'adoption du langage de programmation Java fut l'un des premiers choix effectués en raison de sa grande portabilité et de l'étendue de sa communauté de pratique, permettant ainsi de tirer un meilleur parti des logiciels en code source libre (open source). Dans cette optique, et de manière non restrictive, l'essentiel de l'infrastructure technologique répond à cette orientation. Seuls des logiciels éprouvés et fiables sont implantés, tels que le serveur web Apache et le conteneur d'application J2EE Tomcat, également de la communauté Apache.

À ce jour, le système d'exploitation installé sur la plupart des serveurs de diffusion est Microsoft Windows, d'une part en raison d'anciennes politiques institutionnelles et d'autre part en raison de certaines technologies propriétaires utilisées dans les versions précédentes de la plateforme. Une migration vers Linux est prévue dans les prochains mois afin d'accroître la performance et la stabilité de l'ensemble des technologies de la plateforme actuelle.

### LE MOTEUR DE TRANSFORMATION XSLT

La transformation des fichiers XML de diffusion vers le format HTML requiert un moteur de transformation XSLT (XML Stylesheet

Transformations). À cette fin, un autre produit de la communauté Apache, le cadre applicatif Cocoon (version 2.1), a été retenu en raison de ses multiples avantages dont :

- **Simplicité** : les principes de base du canevas sont simples à maîtriser, permettant de mieux se concentrer sur la logique d'affichage plutôt que sur l'apprentissage d'un modèle de traitement complexe.
- **Flexibilité** : le développement de comportements particuliers et spécifiques à la plateforme Érudit est simplifié par l'architecture « ouverte » de Cocoon.
- **Performance** : la génération dynamique des contenus est très rapide, ce qui est essentiel en raison du nombre élevé de requêtes simultanées faites sur le serveur d'Érudit.
- **Support multilingue intégré** : la possibilité de définir un dictionnaire de traduction pour les trois langues de l'interface, à savoir le français, l'anglais et l'espagnol, simplifie grandement l'écriture des feuilles de style XSLT.
- **Abstraction des URL** : la définition de patrons d'URL facilite la maintenance et la pérennité des liens même lorsque les documents sont déplacés ou détruits.

Enfin, l'accès aisé aux bases de données et l'écriture de pages Web dynamiques de type JSP (Java Server Pages) viennent compléter la liste des caractéristiques à l'avantage de Cocoon.

## LE MOTEUR DE RECHERCHE

Outre l'accès aux documents en mode navigation, le moteur de recherche est indéniablement l'outil le plus puissant permettant à l'utilisateur de repérer ce qui l'intéresse dans le fonds d'Érudit. Pour cette raison, l'équipe d'Érudit a porté un soin particulier au choix de ce moteur et c'est sur le couple *SOLR/Lucene* que s'est arrêté le choix. Il est intéressant de noter que ces deux produits gratuits sont eux aussi issus de la communauté Apache.

Lors de l'évaluation à l'automne 2007, c'est le service web *SOLR* qui a démontré des caractéristiques alors innovatrices. De plus, la puissance de l'engin d'analyse et d'indexation *Lucene* n'était déjà plus

à démontrer. D'abord, l'accès à *Lucene* via le service web *SOLR* facilite l'accès aux fonctionnalités d'indexation et de recherche à travers une architecture distribuée. Ensuite, la création automatique de facettes de recherches (selon une indexation appropriée) s'est aussi avérée un facteur déterminant. Enfin, les excellentes performances du service pour la récupération des résultats de recherche ont convaincu de son intérêt.

L'introduction de *SOLR/Lucene* comme moteur de recherche a nécessité le développement d'une expertise particulière afin de pouvoir en assurer une exploitation optimale. La création des index de recherche et des facettes demande en effet une connaissance approfondie de *Lucene*.

Il est important de noter qu'à la différence des versions précédentes de la plateforme de diffusion, l'ensemble des genres scientifiques (articles, livres, actes, thèses, documents et données) est indexé en texte intégral par le moteur de recherche. On y dénombre à l'heure actuelle près de 135 000 documents.

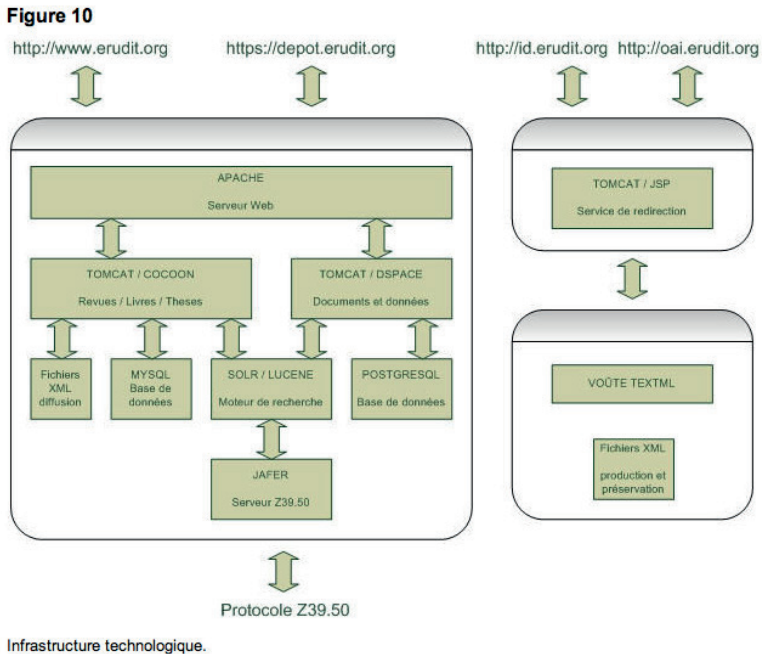
#### LE STOCKAGE DES DONNÉES

Les documents XML de diffusion résident sur le système de fichiers du serveur de diffusion dans une structure de répertoires appropriée au moteur de transformation XSLT. À ce système de fichiers s'ajoute également deux systèmes de gestion de base de données : *MySQL* et *Postgresql*. Le premier sert à accéder aux données du système d'abonnement et aux données de diffusion complémentaires (section « À propos de cette revue », filtrage des accès, etc.). Le deuxième est uniquement dédié à l'installation du dépôt *Dspace*.

Enfin, pour les besoins d'archivage et d'interopérabilité, tous les documents XML de diffusion sont stockés dans une voûte gérée par le logiciel propriétaire *TextML*. Toutefois, ce dernier sera remplacé par le dépôt d'objets numériques Fedora<sup>30</sup> d'ici la fin 2009.

## La plateforme de diffusion

La plateforme de diffusion est composée de plusieurs services auxquels on accède à l'aide d'un simple fureteur ou de logiciels spécialisés. Ces différents services sont répartis sur trois serveurs (voir figure 10).



### LA PLATEFORME WWW.ERUDIT.ORG

La composante centrale de la plateforme de diffusion est le site Web permettant d'accéder aux différents genres scientifiques (articles de revues, livres, actes de colloques, thèses). Parmi les principales caractéristiques de ce site, on retrouve :

- une interface entièrement trilingue ;
- le filtrage des accès pour les revues sous abonnement ;
- des formulaires de recherche simple et avancée ;
- un affichage par facettes des résultats de recherche ;

- la création d'une page de garde pour les tirés à part ;
- des services d'exportation de notices (texte, EndNote, RefWorks, courriel) ;
- un index des auteurs pour chaque revue ;
- un index des thèses des universités participantes.

L'outil de recherche intègre tous les types de documents qui sont repérables grâce à des icônes appropriés. L'utilisateur a donc le loisir de chercher parmi une très vaste collection de documents, tant ceux d'Érudit que ceux provenant des fonds des partenaires comme Persée ou le CNRC.

#### LA ZONE « DOCUMENTS ET DONNÉES »

Intégrant le visuel des autres zones du site, la zone « Documents et données » (<http://depot.erudit.org>) est en fait une application *DSpace* autonome, à l'exception du formulaire de recherche qui renvoie directement au moteur de recherche d'Érudit. On retrouve dans ce dépôt la littérature grise d'une trentaine de centres de recherche québécois. À la suite du dépôt manuel ou en lot des documents, ces derniers sont indexés (en texte intégral) dans le moteur de recherche *SOLR / Lucene* au même titre que les autres genres scientifiques.

#### LE SERVICE D'IDENTIFICATION PÉRENNE

À l'instar d'autres services d'identification pérenne tels Handle<sup>31</sup> ou DOI<sup>32</sup>, Érudit a créé en 2003 son propre service d'identification (<http://id.erudit.org>) basé sur une nomenclature répondant à ses besoins. Moyennant la connaissance d'un identifiant référant à un article particulier, par exemple 001345ar, le service redirige l'utilisateur vers le document demandé, de préférence vers sa version HTML si elle existe, sinon vers sa version PDF. Pour ce faire, une application web Java utilise les métadonnées nécessaires à la redirection en provenance de la voûte *TextML*.

#### LE SERVICE D'INTERROGATION Z39.50

Une adaptation du serveur *Jafer*<sup>33</sup> permet l'interrogation du moteur de recherche d'Érudit via le protocole Z39.50. Les requêtes à *Jafer* sont

alors transformées en requêtes à *SOLR* et la réponse de *SOLR* est ensuite reconvertie en format OAI MARC.

Tout bon logiciel de gestion bibliographique tel qu'EndNote ou ReferenceManager est en mesure d'interroger le moteur de recherche d'Érudit et d'importer les références bibliographiques ainsi obtenues.

### LE SERVICE D'INTEROPÉRABILITÉ

Le moissonnage des métadonnées par les partenaires d'Érudit passe par le service d'interopérabilité (<http://oai.erudit.org>) qui rend disponible une instance de l'application OAI Cat<sup>34</sup>, application conforme au protocole OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) version 2.0. Comme dans le cas du serveur *Jafer*, des connecteurs spécifiques ont été développés afin de pouvoir récupérer les données déposées dans la voûte TextML puis de les transformer dans les formats XML d'échange (Érudit-Article 1.0, Exodic, NLM).

### De la production à la diffusion

Trois environnements distincts hébergent l'ensemble des données et des applications des plateformes de production et de diffusion.

#### L'ENVIRONNEMENT DE PRODUCTION

Les étapes de la production sont effectuées dans un environnement distinct de la plateforme de diffusion. Un serveur dédié ainsi que les postes de travail des techniciens et techniciennes en édition numérique servent à exécuter les différentes applications impliquées dans le processus d'édition.

#### L'ENVIRONNEMENT DE PRÉ-DIFFUSION

À mi-chemin entre l'environnement de production et de diffusion, cet environnement permet d'assurer les dernières étapes du contrôle de la qualité, étapes préalables à la diffusion des documents. C'est dans cet environnement que les techniciens ainsi que la coordonnatrice à l'édition et à la direction de revues peuvent consulter les documents tels qu'ils apparaîtront en diffusion.

À cette fin, plusieurs opérations sont effectuées dans cet environnement, comme la copie des fichiers provenant de l'environnement

de production et l'ajout d'une multitude d'informations à la base de données de diffusion. Par la suite, un mécanisme de ticket (une clé secrète ajoutée aux URL) restreint le visionnement des documents aux seules personnes autorisées à faire le contrôle de la qualité.

### L'ENVIRONNEMENT DE DIFFUSION

Une fois le contrôle de la qualité terminé, le processus de diffusion rend les documents accessibles en supprimant le mécanisme de ticket, indexe les documents dans le moteur de recherche puis les copie dans la voute pour les besoins de préservation et d'interopérabilité.

### La nouvelle plateforme de production éditoriale

La plateforme de production actuelle est essentiellement centrée sur l'édition d'articles de revue scientifique. Or, à la lumière de l'expertise développée depuis 10 ans chez Érudit et de l'évolution des besoins en matière d'édition numérique, un exercice de révision des processus de traitement s'est avéré nécessaire et a conduit à la mise en place d'un projet de développement d'une nouvelle plateforme de production. Parmi les principaux facteurs ayant motivé cette décision, notons :

- la plateforme de production doit permettre l'édition de nouveaux genres scientifiques dont les livres, les actes de colloques et même les cours en ligne ;
- la plateforme actuelle ayant été conçue en 2002, de nouvelles technologies permettent aujourd'hui d'accélérer certains processus de la chaîne de production ;
- et, la plateforme de diffusion offre des possibilités qui ne peuvent être exploitées de manière optimale qu'avec une nouvelle chaîne de traitements conçue à cet effet.

Depuis 2008, l'équipe de développement d'Érudit a connu une forte croissance et atteint un bon niveau de maturité. L'utilisation de méthodes de développement et d'ingénierie logicielle systématique a permis d'améliorer de manière significative les temps de développement et la qualité des applications. Fort de ces acquis, l'ambitieux



projet de développement de la nouvelle plateforme de production peut donc prétendre au succès.

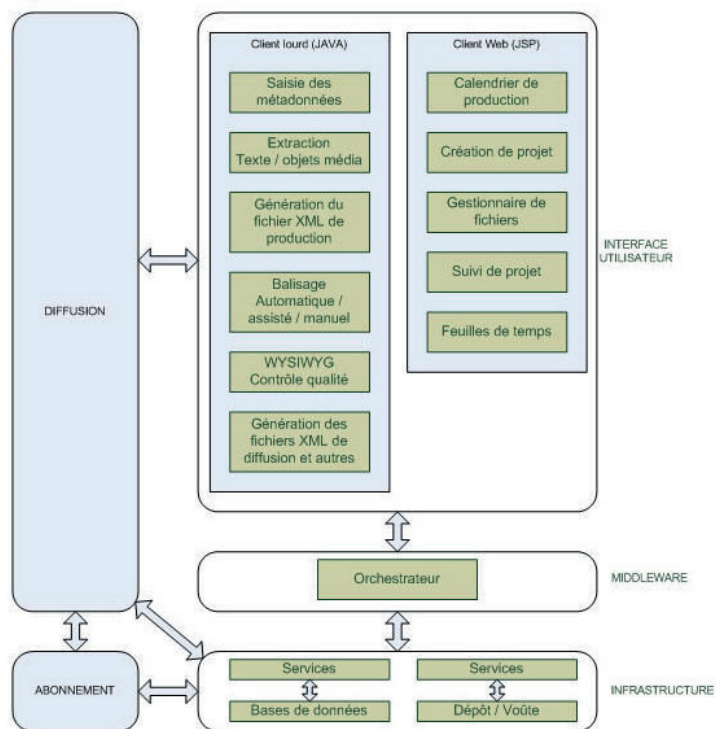
#### UNE PLATEFORME SOUS LE SIGNE DE L'INNOVATION

Afin de simplifier toutes les étapes du processus d'édition numérique, la nouvelle plateforme présentera les caractéristiques suivantes :

- prise en charge native du schéma Érudit Article version 3.0 ;
- édition de nouveaux genres scientifiques (livres, actes) ;
- interface intégrée pour toutes les étapes du processus d'édition ;
- interface web pour les opérations de gestion et de suivi de projet ;
- production facilitée de tous les formats de diffusion et d'interopérabilité ;
- applications hautement configurables.

La figure 11 présente une vue de l'architecture des composantes de la nouvelle plateforme de production.

Figure 11



Architecture de la nouvelle plateforme de production.

Les bénéfices d'une telle plateforme seront multiples. En plus de pouvoir offrir des services d'édition pour d'autres genres scientifiques, l'équipe de production d'Érudit sera aussi en mesure d'utiliser un processus simplifié, impliquant moins de logiciels, plus rapide, plus convivial et d'une qualité supérieure.

## CONCLUSION

Érudit est une institution en flèche qui se doit de poursuivre sa trajectoire. Sa force dès le début aura été de devoir compter sur l'intérêt et la volonté du milieu universitaire, pris comme communauté aussi bien qu'établissements. Cette volonté universitaire a rencontré celle

des pouvoirs publics, d'où la mise en place d'un cadre institutionnel satisfaisant pour laisser se manifester l'inventivité de ses principaux artisans. Le mouvement, qui est le mode de vie des créateurs, saura se frayer un chemin et se donner les outils pour s'exprimer au mieux. Les défis diffèrent : s'il fallait d'abord montrer la raison et la capacité d'être d'une telle infrastructure au service de la recherche, aujourd'hui il faut confirmer dans la durée le rôle de cette institution nationale qui a un rayonnement international, non seulement pour les contenus diffusés, mais aussi pour le savoir-faire de ses artisans. Nom commun, Érudit est aussi un nom propre qui est à l'honneur de la société qui la porte.

---

### **Annexe 1**

#### MEMBRES DU COMITÉ AVISEUR DU PROJET PILOTE DU FONDS FCAR

- Guylaine Beaudry, directrice, publications électroniques, Presses de l'Université de Montréal
- Gérard Boismenu, directeur scientifique, Presses de l'Université de Montréal
- Daniel Boivin, président, OCLC Canada
- Nicole Carette, secrétaire à l'édition, Géographie physique et Quaternaire
- André Clas, directeur, Meta Danielle Collignon, secrétaire à l'édition, Meta
- Jean-Pierre Côté, directeur des Bibliothèques, Université du Québec à Montréal
- Max D'amours, directeur Loisir et Société
- Anthony Giles, Relations industrielles/Industrial Relations
- Lucie Girard, Fonds FCAR
- Jean-Claude Guédon, professeur, Université de Montréal
- Arlette Joffe, directrice du Service des bibliothèques, Université de Montréal

- Claudine Leclerc, secrétaire à l'édition, Relations industrielles/Industrial Relations
- Lise McNicoll, secrétaire à l'édition, Revue d'histoire de l'Amérique française
- André Roy, directeur, Géographie physique et Quaternaire
- Fernande Roy, directrice, Revue d'histoire de l'Amérique française
- Jean Sexton, directeur, Relations industrielles/Industrial Relations
- Pierre Trudel, professeur, Faculté de droit, Université de Montréal

---

## Annexe 2

### MEMBRES DU PREMIER CONSEIL D'ADMINISTRATION DU CONSORTIUM ÉRUDIT

- Gérard Boismenu, président, Université de Montréal
- Claude Bonnelly, vice-président, Université Laval
- Chantal Bouthat, secrétaire-trésorière, Université du Québec à Montréal
- Michael Eberle Sinatra, représentant des revues, Romanticism on the Net
- Benoît Bernier, Les Presses de l'Université Laval
- Pierre Bordeleau, Université de Montréal
- Antoine Del Busso, Les Presses de l'Université de Montréal
- Lyne Sauvageau, observatrice, Fonds québécois de recherche sur la société et la culture
- Guylaine Beaudry, directrice générale, Université de Montréal

- 
- 1 Nous utilisons le terme « électronique » dans cet article pour respecter les appellations d'origine. Dans les autres cas, « numérique » est le terme privilégié.
  - 2 Ce projet pilote a été réalisé dans le cadre du travail dirigé de Guylaine Beaudry, partie prenante de son programme de formation à la maîtrise de l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal, sous la direction d'Yves Marcoux.

- 3 Particulièrement, René Durocher, alors directeur du Bureau de la recherche, et Maurice St-Jacques, vice-recteur à la recherche.
- 4 Martin Sévigny et Guylaine Beaudry d'abord. Se joignent ensuite Marc Audet, Marie-Hélène Vézina et Étienne Lavallée. Gérard Boismenu, alors directeur scientifique des Presses de l'Université de Montréal, est à l'origine des activités d'édition numérique dans cet établissement.
- 5 Gérard Boismenu, Guylaine Beaudry et Lucie Girard ont participé à cette rencontre. À cette occasion, le document « Le Projet IRIS : pour l'édition et la diffusion électroniques de revues savantes » est présenté.
- 6 Lucie Girard et Jacynthe Dolbec, *Projet pilote pour l'édition de revues électroniques : cahier des charges*, Fonds FCAR, 13 novembre 1997 (Doc. 97– CA-112-11).
- 7 La composition du comité avisé est présentée en annexe 1.
- 8 Au cours du projet, au printemps 1998, le nom d'IRIS est abandonné, et Érudit désigne depuis la plateforme de production et de diffusion éditoriale. Le nom de domaine « erudit.org » a été retenu le 15 avril 1998.
- 9 Gérard Boismenu, Martin Sévigny, Marie-Hélène Vézina et Guylaine Beaudry, *Le projet Érudit : un laboratoire québécois pour la publication et la diffusion électroniques des revues universitaires*. Rapport sur le projet pilote réalisé par les Presses de l'Université de Montréal, 1999, 260 p. <https://depot.erudit.org/id/001166dd>.
- 10 Les deux étudiants qui ont réalisé le projet sous la supervision des membres de l'équipe sont Fanny Chabot et Benoît Gervais-Laurendeau.
- 11 Guylaine Beaudry et Gérard Boismenu, en collaboration avec Guy Teasdale, Christian Allègre, Chantal Bouthat, Claude Bonnelly et Benoît Bernier, « Conception d'un portail de production, de diffusion et de gestion de publications électroniques. Étude de faisabilité », Québec : Fonds FCAR, 2000, 130 p.
- 12 CREPUQ, Appel d'offre et cahier des charges : portail de diffusion des publications électroniques, juillet 2001, p. 4.
- 13 Soumission du Groupe interuniversitaire pour l'édition numérique à l'appel d'offres de la CREPUQ portant sur la réalisation d'un portail de diffusion de publications électroniques, 20 septembre 2001, 81 pages + annexes.
- 14 Depuis 2006, le Centre d'édition numérique de l'Université de Montréal est rattaché au Bureau de la recherche, du développement et de la valorisation.
- 15 La conception graphique de cette nouvelle image a été réalisée par Gwenaél Bélanger.
- 16 L'annexe 2 présente la liste des membres du premier conseil d'administration du Consortium Érudit.

- 17 Ce modèle de données, d'abord sous forme de DTD (Définition de Type de Document) et par la suite de schéma XML, a été développé par Isabelle Spina de l'équipe d'Érudit de l'Université de Montréal. Érudit Article est disponible en code source ouvert. La documentation et les modèles se trouvent à l'adresse : [http://www.erudit.org/documentation/doc\\_erudit.htm](http://www.erudit.org/documentation/doc_erudit.htm).
- 18 Objectifs, principes et mode organisationnel du Réseau francophone, Paris 13 décembre 2005.
- 19 La conception de l'image graphique de la troisième génération de la plateforme Érudit a été réalisée par Jean – Christophe Charlier de la société Avion rouge.
- 20 <http://www.lib.unb.ca/Texts> et <http://www.erudit.org/revue/?mode=fond&idfond=3&lettre>
- 21 <http://www.persee.fr>
- 22 Une métadonnée est une donnée servant à définir ou décrire un document. Pour plus d'informations lire : <http://fr.wikipedia.org/wiki/M%C3%A9tadonn%C3%A9e>.
- 23 UTF8 : est un format de codage de caractères définis pour Unicode, voir : <http://unicode.org/>
- 24 Les formats de fichiers d'image reçus sont les suivants : EPS, GIF, JPEG, PNG, PSD, SVG ou TIFF.
- 25 Au sujet du processus de la chaîne de traitement du rétrospectif, consulter Émilie Paquin et Tanja Niemann, *Revue savantes québécoises : des milliers d'articles en accès libre*, Argus, vol. 38, n° 2, 2009 (à paraître).
- 26 Par exemple, le nom d'un auteur peut être écrit différemment d'un numéro à l'autre (Guy-Robert St-Arnaud ou bien Guy-Robert Saint-Arnaud) ce qui aura des impacts pour la recherche par nom ainsi que pour d'autres service tel que l'affichage de l'index des auteurs.
- 27 Pour la documentation technique sur le schéma Érudit Article, voir : <https://cen.umontreal.ca/pages/viewpage.action?pageId=5734590>.
- 28 Pour consulter la norme : [http://www.iso.org/iso/catalogue\\_detail?csnumber=38920](http://www.iso.org/iso/catalogue_detail?csnumber=38920).
- 29 Pour plus d'informations concernant ce prototype lancé par *Cell Press* et *Elsevier*, lire <http://lafeuille.homo-numericus.net/2009/08/larticle-scientifique-du-futur.html> et <http://beta.cell.com/>.
- 30 <http://www.fedora.info>.
- 31 <http://www.handle.net>.
- 32 <http://www.doi.org>.
- 33 <http://www.jafer.org>.
- 34 <http://www.oclc.org/research/software/oai/cat.htm>.