

CONCEPTION D'UN PORTAIL DE PRODUCTION, DE DIFFUSION ET DE GESTION DE PUBLICATIONS ÉLECTRONIQUES

ÉTUDE DE FAISABILITÉ

Étude préparée et rédigée
par Guylaine Beaudry et Gérard Boismenu,

avec la collaboration
de Guy Teasdale, de Christian Allègre, de Chantal Bouthat,
de Claude Bonnelly et de Benoît Bernier

au nom du Groupe interinstitutionnel sur la publication électronique,
composé de :

- 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, Service de la recherche et de la création de l'Université du Québec à Montréal
- La Bibliothèque de l'Université Laval

Déposée le 25 septembre 2000
au Fonds FCAR
Québec

RÉSUMÉ EXÉCUTIF

Cette étude de faisabilité s'inscrit dans le processus d'élaboration de la politique du Fonds FCAR qui manifeste la volonté de modifier son intervention en matière de soutien à la diffusion des résultats de la recherche en mettant à profit les technologies de l'information et de la communication. Un parcours des projets et centres de publication et de diffusion électroniques de revues savantes connus internationalement, ainsi que les récents développements des technologies du texte ont permis d'étayer la sélection des standards internationaux et de valider les choix technologiques et les procédures privilégiées.

Définir la vocation d'un portail de publication et de diffusion électroniques de revues ne doit entamer en rien l'autonomie éditoriale, ni l'indépendance administrative et la personnalité intellectuelle et physique dont s'est dotée une revue. En raison de la dispersion organisationnelle et institutionnelle des revues, le défi de l'électronique peut conduire à confirmer ou pousser encore plus loin le morcellement du mode de fonctionnement des revues et, de ce fait, à élaborer des solutions pour l'électronique sur une base individuelle et tenant compte des ressources limitées dont chacune dispose. Les solutions que l'on peut préconiser dans ce cas peuvent difficilement contribuer à un traitement de publication, de conservation et de diffusion en adéquation avec la vocation des revues savantes et avec la nature des documents publiés.

Le défi de la publication électronique peut être vu comme une occasion importante à exploiter, non seulement pour améliorer ses conditions de diffusion, mais aussi pour repenser la revue et le mode de formalisation et de communication de recherche. Le défi consiste à s'approprier les technologies de l'information afin de les mettre au service de la publication scientifique. En même temps, le chercheur et la revue savante doivent participer pleinement au renouvellement des conditions d'élaboration et de transmission des contenus, par lesquels ils communiquent les résultats de recherche.

L'architecture proposée pour le portail tient compte d'une diversité de services liés aux différentes formes d'expression de la communication savante. Premièrement, le portail s'appuie sur les revues qui devraient constituer une première zone. À ce premier noyau, une deuxième zone, comportant une collection de prépublications, pourrait se greffer. Les autres types de documents relevant de la littérature grise constituerait la troisième zone de ce portail. Sachant que ces différentes formes d'expression des résultats de la recherche répondent à des contraintes qui ne sont pas du même ordre, il importe de bien tracer les contours et les traitements qu'ils appellent. Ces distinctions permettent de définir des zones nettement identifiables avec leur cohérence propre, mais composant dans l'ensemble une collection riche et diversifiée.

Ajoutés à ces trois zones, des outils de diffusion, tel un outil de recherche performant, devront permettre d'exploiter de façon efficace les collections, en cherchant par différents critères de recherche à travers les documents des trois zones du portail. De plus, l'objectif de la diffusion sur le Web étant de porter les résultats de la recherche québécoise dans les créneaux de diffusion internationaux, des mécanismes et des protocoles assurant l'interopérabilité entre les différents portails de diffusion de revues savantes et de prépublications devront relier les

collections du portail québécois à d'autres collections d'importance.

La technologie proposée pour le portail en ce qui a trait à la production des revues savantes a déjà fait ses preuves et est utilisée par plusieurs éditeurs sans but lucratif et commerciaux. Nous proposons une chaîne de traitement basée sur la norme XML, ce qui permet la fabrication des versions papier et électronique à partir d'un même processus. Des métadonnées (description de l'article : auteur, titre, résumé, volume, numéro...) devraient être générées automatiquement à partir de la chaîne de traitement. Un fichier Postscript, un fichier PDF ou un prêt-à-photographier seront préparés pour l'impression de la revue.

Le portail devrait offrir un service d'intermédiaire entre la revue et les entreprises de numérisation en plus de créer les métadonnées des articles et ainsi permettre aux utilisateurs de chercher par un même outil de recherche à travers les collections courante et rétrospective d'une revue.

En plus des services de production, le portail devra créer et maintenir une infrastructure permettant la diffusion des résultats de la recherche. Ces services doivent toucher la diffusion par le portail même tout autant que l'accessibilité et le repérage du contenu informationnel du portail dans les bases de données importantes, qui dépouillent le contenu des périodiques dans le domaine des sciences humaines et sociales. Finalement, le portail permettra de mettre en lien ses propres collections avec celles d'autres portails de revues savantes et d'autres types de documents universitaires.

La conservation et l'archivage du patrimoine culturel des revues sont au cœur du projet de portail. Cette tâche devra être menée à bien en utilisant les meilleures pratiques en la matière. En premier lieu, le site d'hébergement devra offrir des assurances de sécurité à la fine pointe des technologies disponibles. Le portail devra s'assurer de la redondance des données en collaboration avec des partenaires choisis et répartis géographiquement (autres portails, bibliothèques nationales du Canada et du Québec). Le portail devra poursuivre une veille technologique en matière d'intégrité et d'authenticité des textes électroniques. Notamment, les travaux des groupes de travail du W3C sur la signature numérique et les métadonnées devront être suivis.

Pour assurer une viabilité économique, le portail doit proposer une vision d'affaires réaliste et applicable. En raison de sa responsabilité à l'égard de tiers — revues et institutions auxquelles le service est offert, ainsi qu'usagers et clients du portail —, il doit pouvoir se constituer en entité institutionnelle et légale. Cet organisme devrait être une institution sans but lucratif qui repose, autant que possible, sur un partenariat entre diverses composantes du milieu universitaire ; il serait souhaitable que ces partenaires aient déjà un engagement dans la publication des revues. Le cadre institutionnel devrait retenir la formule coopérative et témoigner de la propriété du milieu universitaire.

L'organisme serait dirigé par un conseil d'administration regroupant les partenaires qui se sont engagés dans le projet. Il faudra considérer la mise en place d'un comité « des clientèles » qui réunirait les revues desservies par le portail et des représentants des principaux utilisateurs du portail, comme les bibliothèques ou centres de recherche, et, le cas échéant, des bailleurs de fonds.

Les partenaires auraient une responsabilité professionnelle (qualité de service, stabilité et atteinte des objectifs) et financière dans la réalisation du plan d'affaires. Les contributions financières pour soutenir le portail pourraient être de plusieurs ordres. Les coûts d'implantation du portail devraient faire l'objet d'une subvention spéciale. Les services de production technique aux revues devraient être facturés sur une base d'autofinancement et soutenus par la politique de financement public des revues. Les coûts d'exploitation de la structure de diffusion du portail seraient assurés par les pouvoirs publics en association avec les partenaires.

Dans le but d'utiliser au mieux les compétences existantes et d'offrir un service proche des aires géographiques des directions de revue, l'architecture organisationnelle du portail devrait prévoir, d'une part, une structure de production distribuée avec un lieu de production à Québec et un à Montréal et, d'autre part, un portail de diffusion unique qui voit à l'ensemble des responsabilités de rayonnement et de visibilité des revues sur le Web.

L'organisme devrait être reconnu par le ministère de l'Éducation, par le ministère de la Recherche, de la Science et de la Technologie, et par le ministère de l'Industrie et du Commerce.

Un aperçu des coûts de démarrage du portail chiffre l'investissement selon deux scénarios, le premier représentant un investissement de 347 000\$ et le deuxième de 233 000\$. La différence entre les deux scénarios se justifie entre autres, par des questions de choix d'équipement. Les choix techniques sont invariablement faits en fonction d'un cadre institutionnel, des ressources dont on dispose pour la réalisation du projet et de l'expertise de l'équipe chargée de la mise en œuvre. Toutefois, nous soulignons l'importance de faire le choix d'équipement et de technologies de qualité pour le développement à long terme du portail. Il ne faut pas ici sous-estimer les possibilités d'acquisition en collaboration entre les différentes institutions et agences concernées par la diffusion des résultats de la recherche.

Le total des dépenses de fonctionnement du portail avant la gestion des abonnements d'élève à 344 000\$ par année. De cette somme, 182 000\$ représentent les coûts de production techniques qui seraient assumés par les revues à même leurs subventions, éventuellement ajustées, qu'elles obtiennent. La différence, soit 162 000\$, serait ainsi dédiée à la diffusion.

Deux scénarios sont proposés pour le financement du portail. Le premier propose de faire des revues savantes québécoises un bien public, donc diffusées gratuitement en version électronique sur le Web. C'est certainement la voie qui offre les meilleures conditions pour valoriser la publication électronique et pour échapper aux règles de la commercialisation, auxquelles on note peu d'attachement de la part des éditeurs de revues. Le financement de la gratuité serait assuré d'une part, par le paiement des coûts de production technique par les revues à même leurs subventions, et d'autre part, sur une subvention à la diffusion qui serait accordée au portail. Ce scénario ne nécessite pas la mise en place d'un système de gestion des accès au serveur, pas plus d'un service de gestion des abonnements électroniques.

Le deuxième scénario est celui de l'accès payant. Ici encore, la production technique des différentes versions doit être assumée par les revues. Toutefois, il nous faudrait augmenter les revenus pour permettre de financer le système de diffusion. Compte tenu de la situation financière des revues québécoises, on peut difficilement penser à une solution qui ferait

en sorte que leurs revenus en seraient diminués. S'appuyant sur des expériences diverses, les pratiques se situent autour d'une augmentation de 30% des tarifs d'abonnement reliée à l'accès à la version électronique des revues. Notons que la moyenne des tarifs d'abonnements aux revues québécoises est de 60\$, l'augmentation serait donc négligeable. En estimant que 90% des bibliothèques abonnées présentement aux revues s'abonneraient à la version électronique, le portail devrait tout de même s'appuyer sur d'autres sources de financement pour combler le manque à gagner.

On ne peut minimiser l'importance et la signification des transformations que suppose la transition vers la publication électronique pour les revues savantes. Fort heureusement, cette transition est supportée par des changements qui s'opèrent dans l'environnement général de la recherche et du milieu universitaire. Il s'agit d'une tendance lourde favorable aux actions à entreprendre dans les diverses sociétés. Pour une institution publique qui veut susciter ce mouvement dans une société particulière, il importe d'impulser cette transition sur un mode pro-actif, sans pour autant céder à la tentation volontariste. La politique de la table rase est mal avisée. Il convient plutôt, partant d'une bonne connaissance du milieu visé, de fixer des objectifs définis et spécifiques, qui pointent résolument dans la direction retenue. Parce qu'on ne peut engager, sur la base du plus petit dénominateur commun, une transition majeure mettant en cause les façons de faire, l'action publique exercera nécessairement une tension auprès du milieu. Il reste à savoir jusqu'où cette ligne de tension peut être poussée sans être rompue. La connaissance du milieu et une communication directe et franche avec les acteurs sont des guides précieux, mais une vision stratégique, prenant en compte la direction générale, les objectifs particuliers, les ressources disponibles, l'étalement des actions particulières et l'évaluation des résultats, possède de grands avantages pour clarifier les intentions, élaborer la politique d'intervention et définir les actions à entreprendre.

SOMMAIRE

INTRODUCTION

CHAPITRE 1 : LA CARACTÉRISATION DE LA RÉALITÉ DES REVUES

CHAPITRE 2 : L'INFRASTRUCTURE INTELLECTUELLE, ORGANISATIONNELLE ET MATÉRIELLE (AVEC LA COLLABORATION DE CHRISTIAN ALLÈGRE)

CHAPITRE 3 : UNE CHAÎNE DE PRODUCTION TECHNIQUE (AVEC LA COLLABORATION DE GUY TEASDALE)

CHAPITRE 4 : UN SYSTÈME DE DIFFUSION (AVEC LA COLLABORATION DE CHANTAL BOUTHAT ET DE CLAUDE BONNELLY)

CHAPITRE 5 : L'ARCHIVAGE (DE GUY TEASDALE, AVEC LA COLLABORATION DE GUYLAINE BEAUDRY)

CHAPITRE 6 : PRÉVISIONS BUDGÉTAIRES

CONCLUSION

TABLE DES MATIÈRES

RÉSUMÉ EXÉCUTIF	I
SOMMAIRE	V
TABLE DES MATIÈRES	VI
INTRODUCTION.....	1
ANTÉCÉDENTS.....	2
MANDAT	2
DÉMARCHE.....	3
QUELQUES RÉFLEXIONS.....	5
CHAPITE 1 LA CARACTÉRISATION DE LA RÉALITÉ DES REVUES.....	7
PRÉSENCE ET PRATIQUES DU SANS BUT LUCRATIF À L'ÉCHELLE INTERNATIONALE.....	8
EN BREF, SUR LES REVUES AU CANADA ET AU QUÉBEC	9
UN RÔLE ESSENTIEL DANS LA COMMUNICATION SCIENTIFIQUE.....	9
DISPERSION DES LIEUX DE PUBLICATION ET PRÉCARITÉ FINANCIÈRE	11
LE DÉFI DE L'ÉLECTRONIQUE : UNE OCCASION À EXPLOITER	12
CE QU'EN PENSENT LES DIRECTEURS DE REVUE.....	13
CHAPITRE 2 L'INFRASTRUCTURE INTELLECTUELLE, ORGANISATIONNELLE ET MATÉRIELLE.....	16
L'ARCHITECTURE GLOBALE ET LES SERVICES DU PORTAIL	17
<i>Architecture</i>	<i>17</i>
<i>La relation entre le portail et les revues</i>	<i>18</i>
<i>Offre de services aux revues et à la communauté des chercheurs</i>	<i>19</i>
UNE ENTITÉ INSTITUTIONNELLE	21
DES RESSOURCES INTELLECTUELLES	22
UN ENVIRONNEMENT TECHNIQUE ET MATÉRIEL	23
CHAPITRE 3 UNE CHAÎNE DE PRODUCTION TECHNIQUE.....	25
LES FORMATS	26
<i>De quel format parlons-nous?.....</i>	<i>26</i>
<i>Les formats structurés</i>	<i>27</i>
<i>Quelles utilisations pour le PDF et le Postscript?.....</i>	<i>28</i>
<i>Les formats images.....</i>	<i>29</i>
LA CHAÎNE DE TRAITEMENT.....	31
<i>Une chaîne de traitement documentaire.....</i>	<i>31</i>
<i>Les principes qui sous-tendent le modèle</i>	<i>32</i>
<i>La production de l'information</i>	<i>34</i>

<i>Le processus de travail dans son ensemble</i>	37
<i>La numérisation rétrospective</i>	39
LES PRÉPUBLICATIONS	40
<i>Open Archives initiative</i>	41
<i>Des prépublications et des revues</i>	41
<i>Un aperçu des serveurs de prépublications</i>	42
<i>Les prépublications et le portail québécois</i>	42
<i>Les documents de la zone prépublications</i>	43
<i>Implications techniques</i>	43
<i>Besoins et intérêt des chercheurs québécois pour un serveur de prépublications</i>	44
LA PRODUCTION DES FORMATS DE DIFFUSION DE L'INFORMATION	45
<i>La consultation à l'écran</i>	47
<i>L'impression à distance</i>	48
<i>L'impression professionnelle</i>	48
LES MÉTADONNÉES	49
<i>Les métadonnées : raison d'être et usages</i>	50
<i>Les métadonnées : les créer, les stocker</i>	52
<i>Les métadonnées : les superposer</i>	53
<i>Quelques modèles de métadonnées</i>	54
<i>Implications pour les revues</i>	55
CHAPITRE 4 UN SYSTÈME DE DIFFUSION	58
LA DIFFUSION : UN MANDAT ESSENTIEL	59
PRINCIPES DE BASE POUR LA CRÉATION D'UNE INTERFACE WEB POUR DIFFUSION DES REVUES SAVANTES	59
LE SYSTÈME DE DIFFUSION	60
<i>Services de base</i>	60
<i>Services complémentaires</i>	63
UNE OFFENSIVE DE DIFFUSION SUR LA TOILE	64
<i>Le référencement et le positionnement stratégique</i>	65
<i>La visibilité dans les bases de données</i>	66
<i>Les outils de promotion pour accroître la visibilité du site-portail</i>	67
DIFFUSION : ENJEUX ÉCONOMIQUES	69
<i>L'origine des revenus dans la situation actuelle</i>	69
<i>La gratuité</i>	70
<i>L'accès filtré et payant</i>	77
CHAPITRE 5 L'ARCHIVAGE	84
<i>Distinction entre différents termes</i>	85
OBSOLESCENCE TECHNOLOGIQUE	87
LES FORMATS D'ARCHIVAGE ET DE CONSERVATION LES PLUS UTILISÉS	88
LES SUPPORTS D'ARCHIVAGE DE CONSERVATION	88
LES STRATÉGIES EMPLOYÉES DANS LES DIVERS PROJETS	89
<i>L'émulation comme stratégie de conservation à long terme</i>	89
GARANTIR L'INTÉGRITÉ ET L'AUTHENTICITÉ DES TEXTES	90

LA RESPONSABILITÉ DE LA CONSERVATION ET DE L'ARCHIVAGE	91
EN RÉSUMÉ.....	91
CHAPITRE 6 PRÉVISIONS BUDGÉTAIRES	93
COÛTS DE PRODUCTION D'UNE REVUE.....	94
DÉMARRAGE DU PORTAIL	95
BUDGET DE FONCTIONNEMENT DU PORTAIL – SCÉNARIO POUR 30 REVUES	97
COÛTS DE TRANSITION POUR LES REVUES.....	100
CHAPITRE 7 CAHIER DES CHARGES	101
INFRASTRUCTURE INTELLECTUELLE, ORGANISATIONNELLE ET MATÉRIELLE	102
SYSTÈME DE PRODUCTION	103
SYSTÈME DE DIFFUSION	104
GESTION DE L'INFORMATION.....	105
PLAN D'AFFAIRES	106
CONCLUSION.....	107
ANNEXE 1.1 COMPTE RENDU DES GROUPES TÉMOINS DE QUÉBEC ET DE MONTRÉAL.....	111
REVUES PARTICIPANTES	111
<i>Québec</i>	111
<i>Montréal</i>	111
MISE EN SITUATION	112
LE PASSAGE À L'ÉLECTRONIQUE.....	112
MAINTIEN DU PAPIER.....	112
LE PORTAIL COMME INFRASTRUCTURE COMMUNE.....	113
LE FINANCEMENT	114
ANNEXE 1.2 SERVICE D'HÉBERGEMENT DE SERVEURS	116
ANNEXE 2.2 CAHIER DES CHARGES TECHNIQUE.....	118
DESCRIPTION TECHNIQUE DU PORTAIL.....	118
<i>Fonctionnalités (système de diffusion)</i>	118
<i>Architecture</i>	119
DESCRIPTION DES CONTENUS.....	121
<i>Estimation de l'espace disque nécessaire pour les contenus</i>	122
DESCRIPTION DES SYSTÈMES	123
INFRASTRUCTURES MATÉRIELLES	124
ANNEXE 3.1 SITES VISITÉS DANS LE CADRE DE CETTE ÉTUDE	126

INTRODUCTION

L'étude de faisabilité s'inscrit dans le processus d'élaboration de la politique du Fonds FCAR qui manifeste la volonté de modifier son intervention en matière de soutien à la publication de revues universitaires et de diffusion des résultats de recherche, en mettant à profit les technologies de l'information et de la communication.

ANTÉCÉDENTS

En 1995, le Fonds décidait, à la suite d'une étude, de maintenir le programme d'aide financière aux revues. Deux ans plus tard, un comité d'experts, qui se penchait sur l'intérêt de la publication électronique des revues, supportait cette hypothèse et recommandait de réaliser un projet pilote en ce sens. Le projet Érudite, qui devait examiner les aspects de la transition et expérimenter une chaîne de traitement pour la publication et la diffusion électroniques des revues, supportait dans son rapport, déposé en mai 1999, la transition vers la publication électronique des revues, proposait des modes techniques pour le faire, et préconisait le regroupement des infrastructures pour la publication et la diffusion des revues.

Dans les mois qui ont suivi, 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 politique 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 électronique, 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'offre pour la préparation d'une étude de faisabilité pour la conception d'un portail de production, de diffusion et de gestion de publications électroniques.

MANDAT

Le mandat premier de 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. La présente étude s'est attaquée à l'ensemble de ces aspects. Pour circonscrire l'objet du mandat, il est utile de revenir sur certaines options et préférences contenues dans le rapport du dernier comité d'étude et dans l'annonce de l'appel d'offres. Elles ont en effet balisé le travail que nous avons entrepris et que nous exposons dans ces pages.

La version électronique est appelée à être la version première de la revue. Cela sous-entend que l'introduction de la publication électronique doit être vue comme l'occasion d'exploiter des fonctionnalités nouvelles dans la communication des résultats de recherche, de revoir les conditions de production et de diffusion des revues et de repenser le mode de fonctionnement de la revue comme véhicule dans le transfert de connaissances. En parallèle

à cette idée, l'accent est mis dans ces documents sur l'intérêt de la mise en place d'une infrastructure commune qui aurait pour rôle de diffuser des normes de production et de fournir un lieu de diffusion et de rayonnement des publications. D'où la volonté de mettre en place un portail dont les responsabilités sont étendues. On parle ici d'un portail de production, de diffusion et de gestion des publications électroniques.

Pour concevoir et dégager les conditions de réalisation de ce portail, nous sommes invités à prendre avantage des potentialités des technologies de l'information et de la communication et à capitaliser sur les meilleurs standards de publication et de diffusion. C'est sans doute dans cette perspective que l'on privilégie, par exemple, le langage structuré dans la production des documents, la mise en place de métadonnées, la prise en compte de procédures d'archivage. De même, dans la mesure où les formats de diffusion sont appelés à varier dans le temps, il importe de prévoir une flexibilité pour le portail permettant de considérer la pratique de plusieurs formats. Cette dernière notion fait référence certainement aux formats dans lesquels les documents sont diffusés, mais aussi aux caractéristiques des types de documents qui évoluent ; mais, dès à présent, cette dernière question souligne l'influence que la publication électronique peut avoir pour la compréhension de la revue savante en elle-même, mais aussi met en relief le renouvellement des vecteurs de la communication savante avec, notamment, les collections de prépublications.

Enfin, on attend de cette étude qu'elle présente une conception et une évaluation des composantes d'un portail, compris comme un système collectif au service et au profit du milieu universitaire. C'est en ce sens que nous avons envisagé la conception et l'architecture du portail, les normes et les formats, les services pour la production, la diffusion, la gestion et la conservation des documents. C'est aussi dans cette perspective que les remarques sur les caractéristiques de l'infrastructure à mettre en place ont été formulées.

L'étude a été menée en s'appuyant sur ces principes de référence et sur ces balises.

DÉMARCHE

L'étude de faisabilité comprend évidemment des développements d'ordre conceptuel ; elle insiste aussi sur la dimension opérationnelle des choix qui sont proposés ou des hypothèses qui sont retenues.

Nous avons porté attention aux pratiques des principaux centres de publication et de diffusion électroniques de revues connus internationalement. Ce parcours des projets, ainsi que les récents développement des technologies du texte permettent d'étayer la sélection des standards internationaux et de valider les choix technologiques et les procédures privilégiées.

Par ailleurs, l'étude de faisabilité, qui s'alimente d'une pluralité d'expériences, tout autant pour la production et ses standards, pour la diffusion et ses services, que pour la gestion des publications et ses normes, débouche sur une estimation des coûts et sur l'esquisse d'un modèle organisationnel. En ce sens, l'étude cherchera à présenter la conception d'un portail, dans laquelle s'articulent la production, la diffusion et les considérations organisationnelles.

Cela dit, l'étude de faisabilité doit pouvoir rendre compte du contexte social général, mais surtout immédiat, dans lequel les politiques à mettre en œuvre devront s'insérer.

C'est l'une des dimensions dont on ne peut faire l'économie, à moins de tenir pour négligeables les résultats pratiques de la mise en œuvre de politiques. Pour ce faire, les paramètres et les principales hypothèses ont été soumis aux représentants de revues afin de valider les orientations générales de l'étude et d'apprécier leur accueil. Deux rencontres — qui ont fait office de « groupes témoins » — ont eu lieu (l'une pour la région de Québec, l'autre pour la région de Montréal) au moins quinze jours avant le dépôt de l'étude.

L'étude est divisée en chapitres pour découper les diverses étapes de la discussion, mais ceux-ci ont été conçus et élaborés en fonction d'une logique d'ensemble.

- Dans un premier temps (chapitre 1), la caractérisation du milieu des revues permet une reconnaissance du terrain visé par le changement de politique du Fonds FCAR ; cette caractérisation nous informe tout autant sur le rôle et la place occupés par les revues au Québec que sur les réactions que suscite la perspective de la mise en place d'un portail de production et de diffusion.
- De là, les dimensions de l'infrastructure que suppose le portail sont esquissées (chapitre 2), tant sur les plans intellectuel et organisationnel, que matériel. De même, ce chapitre permet de présenter le champ de compétence du portail à l'égard des revues et les services qui seraient offerts.
- Le chapitre 3 présente les composantes et leur articulation d'ensemble de la chaîne de production technique qui assurerait la publication électronique. Au-delà du propos technique, il faut y voir le socle sur lequel peuvent se mettre en place les divers services de diffusion et de gestion des documents numérisés.
- Cette présentation se poursuit (chapitre 4) par la mise en relief des éléments qui participent à la diffusion des documents universitaires, sachant qu'être disponible dans le Web, ce n'est pas être diffusé. Quelle est la démarche et quels sont les outils qui doivent être mis au service de la diffusion électronique ? telles sont les questions qui guident ce chapitre et elles conduisent à s'intéresser aux conditions d'accès au site et à la dimension pécuniaire qui lui est rattachée.
- Puis, la gestion des documents électroniques dans le cadre d'un portail fait ressortir la question de l'archivage et de la conservation (chapitre 5). De produit qu'elle est en version imprimée, la publication en version électronique devient un service, dont il faut assurer l'accès sur une période plutôt longue. La pérennité est une dimension qui prend un tout autre visage ici, tant pour la conservation des documents par le portail, que pour leur archivage en tant que patrimoine social.
- Le chapitre 6 présente de façon synthétique les prévisions budgétaires qui peuvent être dressées à ce jour. Les prévisions de revenus ne peuvent qu'être basées sur des hypothèses élémentaires qui témoignent du degré d'incertitude qui plane sur ces questions aujourd'hui.

- Enfin, dans le dernier chapitre, comme il était demandé dans l'appel d'offres auquel nous avons répondu, un cahier des charges est réuni de façon schématique afin de mettre en relief les principaux traits techniques et organisationnels qui devraient faire l'objet du prochain appel d'offres visant à implanter le portail.

La conception et l'élaboration de cette étude de faisabilité ont tiré profit de la qualité et de la diversité à la fois de l'expertise et de l'expérience des partenaires du Groupe interinstitutionnel. Chaque institution partenaire a délégué un représentant pour former un comité de travail, qui s'est réuni à toutes les semaines physiquement ou par conférence téléphonique. Un site Web de travail a été créé et notre liste de distribution a été très active. Les diverses parties de cette étude ont fait l'objet de discussions, avant leur rédaction ainsi qu'après. Les membres du comité ont aussi apporté tour à tour une contribution spécifique à l'une ou l'autre partie de l'étude et cette contribution a été soulignée dans la table des matières. Cette étude a été réalisée autour d'une période de sept semaines.

Le comité de travail a consulté deux experts : M. Martin Sévigny, de la compagnie AJLSM de Bordeaux, pour la chapitre sur la chaîne de production, et M. Robert Gérin-Lajoie, du Centre CIRANO de Montréal, pour la section sur le commerce électronique. Le comité a apprécié la compétence et le professionnalisme de leurs avis. Le comité a formé une petite équipe de recherche composée de : Nadège Arsa, Alexandre Delage, Anne-Marie Grenier, Marc Lemaire et Sébastien Tremblay. Qu'ils soient remerciés pour la qualité de leur travail. Le comité exprime sa reconnaissance à MM. Luc Simon, du vice-rectorat à la recherche de l'Université Laval, et à Gilles Deschatelets, directeur de l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal, qui ont assuré l'animation des « groupes témoins ». Enfin, la collaboration de M. Yves Mougeot, directeur des Programmes de subventions de recherche et de diffusion de la recherche au Conseil de la recherche en sciences humaines du Canada a été précieuse.

QUELQUES RÉFLEXIONS

Sur une base quotidienne, les chercheurs font l'expérience des nouvelles conditions dans lesquelles la communication des résultats de recherche se développe. Aucune communauté scientifique ne peut tourner le dos à cette réalité. Les véhicules, par lesquels les résultats de recherche sont diffusés, sont directement touchés par les innovations que suscite la pénétration des technologies de l'information et de la communication dans la diffusion du savoir. Les revues savantes, acteurs majeurs dans la communication scientifique, doivent faire face au défi de la publication et de la diffusion électroniques en capitalisant sur leurs fonctionnalités. Si la translation de la publication imprimée vers la version électronique apparaît une voie de passage naturelle, bientôt le support électronique conditionnera la façon même de concevoir et rédiger des articles. C'est un processus qui s'engage et qui est loin d'avoir donné tous ses fruits.

Le milieu universitaire d'une petite société comme le Québec ne peut que participer à ce mouvement plus vaste, non comme une fatalité, à laquelle on ne peut se soustraire, mais comme un défi stimulant, que l'on se doit de relever. En même temps, cette société québécoise, de dimension réduite, ne peut que se confronter à de grandes sociétés, par les institutions universitaires, par la population et par les ressources. Pour relever ce défi, dans

des conditions qui le rendent possible et prometteur, les institutions publiques doivent encourager les acteurs qui sont en première ligne (les revues et le milieu universitaire, plus largement), mais surtout les supporter et les accompagner. Cette question occupe une position centrale dans une politique de développement scientifique et suppose des ressources. Concevoir la transition de la communication scientifique vers l'électronique dans une perspective d'économie de ressources ferait fausse route. L'affectation des ressources change certes, mais leur volume loin de fondre se maintient, voire s'accroît.

Dans le contexte actuel, il est assez difficile de prendre la mesure exacte de ces ressources car les besoins réels sont subordonnés à plusieurs éléments non contrôlés, notamment les comportements et les anticipations des acteurs et des institutions. La part d'incertitude est grande. Dès lors, il est d'autant plus important d'avoir une très bonne maîtrise des données de base ; cela n'est possible que si l'on combine quatre dimensions fondamentales qui participent aux processus engagés et qui assurent une interaction étroite entre les objectifs poursuivis et les conditions de réalisation. Il importe, en effet, 1° de posséder une vision claire concernant les processus de transformation de la communication scientifique de même qu'une conception des modes institutionnels qui participent à ces processus, 2° de posséder une très bonne connaissance technique des moyens de réalisation des projets de publication électronique, 3° de compter sur une réelle expérience dans l'opérationnalisation de la publication électronique et, plus largement, de l'édition universitaire, 4° d'entretenir une interaction suivie avec le milieu des revues et de la recherche en général.

CHAPITE 1
LA CARACTÉRISATION DE LA
RÉALITÉ DES REVUES

L'élaboration de politiques et l'intervention publique auprès des revues savantes doivent prendre en compte les caractéristiques de ce milieu afin de retenir les avenues appropriées. Cette caractérisation doit comprendre des données essentielles sur l'environnement international, qui conditionne le milieu québécois et canadien, sur la présence relative dans les champs disciplinaires, sur le cadre organisationnel et la structure éditoriale, sur l'importance relative des revues dans la communication scientifique et sur leur contexte financier. Partant de là, il est plus facile de saisir la capacité d'agir sur certaines variables et sur l'impact prévisible de certaines actions.

PRÉSENCE ET PRATIQUES DU SANS BUT LUCRATIF À L'ÉCHELLE INTERNATIONALE

Les conditions d'existence des revues savantes sont bouleversées par une double dimension. D'abord, l'environnement pour la communication scientifique change avec à la fois l'application des technologies de l'information dans le fonctionnement des revues elles-mêmes et l'avènement et la diffusion sans contraintes de modèles de communication scientifique, qui étaient déjà présents dans le monde universitaire mais qui prennent beaucoup d'envergure avec ces technologies (pensons aux collections de prépublications ou aux forums de discussion). Nous parlerons longuement de cette dimension dans les sections qui suivent. Ensuite, concurrentement, la spirale inflationniste des prix d'abonnement des revues a provoqué des débats, voire des mises en accusation à propos des prix administrés et des rentes de situation que pratiquent quelques éditeurs oligopolistiques de revues savantes de rayonnement international. Les revues québécoises et canadiennes ne participent pas à ces grands groupes, sauf exceptions, mais elles ne sont évidemment pas à l'abri de l'évolution des pratiques dans ce secteur.

Cette dernière dimension conduit souvent à une vision très globale des acteurs en présence et de leurs comportements. Tout se passe comme si, dans cette perspective macroscopique, il n'y avait que des grands groupes commerciaux avec leurs pratiques oligopolistiques. Or, une étude le moins attentivement permet de souligner que, même dans l'univers restreint des grandes revues à l'échelle internationale¹, les grands éditeurs commerciaux publient un peu moins de la moitié du nombre de revues, car les sociétés savantes et les presses universitaires (surtout en sciences sociales pour ces dernières) éditent une petite majorité des revues. Cependant, les pratiques de prix — à quelques exceptions près pour les sociétés savantes (exceptions souvent citées pour supporter une vue monolithique des choses) — sont carrément différentes selon que l'on prenne les grands éditeurs commerciaux, d'une part, et les institutions et éditeurs sans but lucratif, d'autre part. Bien que ce soit dans des ordres de grandeur différents, la revue en sciences et la revue en sciences sociales sont de 2,5 à plus de 3 fois plus coûteuses quand elles sont publiées par un grand éditeur commercial plutôt qu'un éditeur sans but lucratif. Les revues publiées par des éditeurs sans but lucratif, dans l'un ou l'autre secteur, coûtent nettement moins cher (environ 110 \$ en moyenne en sciences sociales contre près de 340 \$ pour les éditeurs commerciaux, et environ 740 \$ en moyenne

¹ L'étude porte sur les 25 premières revues (avec les indices d'impact) dans huit disciplines (tant en sciences pures qu'en sciences sociales). Voir : Gérard Boismenu et Guylaine Beaudry, « Publications électroniques et revues savantes : acteurs, rôles et réseaux », *Documentaliste-Sciences de l'information*, vol. 36, no 6, novembre/décembre 1999, p. 292-305.

contre 1800 \$ en sciences) mais bénéficient par ailleurs d'un impact relativement plus important que leurs concurrentes des grands éditeurs commerciaux, ce qui augmente encore davantage leur intérêt.

EN BREF, SUR LES REVUES AU CANADA ET AU QUÉBEC

Quand on se penche maintenant sur la situation des revues dans les sociétés nationales — revues qui, pour la plupart, sont moins présentes dans les sommets de la pyramide internationale de la diffusion de la recherche auxquels nous venons de faire référence —, on constate que les éditeurs sans but lucratif sont proportionnellement très présents et que la position relative des sciences sociales est très grande. Un total de 201 revues savantes d'importance ont été répertoriées au Canada, en comptabilisant toutes les revues financées par le CRSH, le CRSNG et par le Fonds FCAR, ainsi que celles qui sont répertoriées dans des bases de données secondaires telles SCI, SSCI et/ou AHCI. De celles-ci, près des trois quarts publient en sciences humaines et sociales². Au Québec, des 52 revues répertoriées par la Banque de données sur les revues savantes québécoises³, 65 % sont en sciences humaines et sociales et 25 % en arts et lettres ; il reste une place réduite pour les sciences biomédicales, naturelles et le génie. C'est donc dire que les interlocuteurs premiers pour une intervention auprès du milieu des revues s'avèrent être en majeure partie des directeurs de revues en sciences humaines et sociales, voire en création et en arts.

À l'échelle canadienne, on sait que les trois quarts des revues sont éditées et produites par de petits éditeurs non professionnels, comme les unités d'enseignement (36 %) et les sociétés savantes (38 %). Les presses universitaires publient 16 % des revues et les éditeurs privés seulement 5 % (les éditeurs des autres revues n'ont pu être identifiés clairement). Cela signifie que, globalement, au moins 90 % des éditeurs sont des organismes sans but lucratif. Pour ce qui est des éditeurs privés (commerciaux), on peut aussi bien trouver, à un bout du spectre, Elsevier Science Ltd et Kluwer Academic Publishers et, à l'autre bout, les Éditions Bellarmin ou les Éditions Saint-Martin. Au Québec, cet éclatement des lieux d'édition et la place marginale des grands éditeurs commerciaux sont frappants ; compte tenu du bassin total de revues, le plus important éditeur universitaire québécois en la matière (les Presses de l'Université de Montréal), avec sept revues, ne peut se comparer avec University of Toronto Press qui a 27 revues et Wilfrid Laurier University Press, qui en publie 13. De façon générale, en matière de coût d'abonnement, on pratique une tout autre échelle de prix que celle des grandes revues. Tous éditeurs confondus, l'abonnement à la revue canadienne en sciences coûte en moyenne 175 \$ américains et, en sciences sociales, 40 \$ américains.

UN RÔLE ESSENTIEL DANS LA COMMUNICATION SCIENTIFIQUE

Les revues québécoises et canadiennes jouent un rôle essentiel dans les réseaux de communication scientifique, même si elles ne se situent généralement pas à la pointe de la pyramide de la communication scientifique. Plusieurs revues ont une histoire déjà fort longue et les revues ont constitué des institutions majeures dans la structuration de la communauté scientifique au Québec. Lieux de validation, de légitimité et de reconnaissance des résultats de

² Recherche inédite réalisée par Boismenu et Beaudry.

³ B. Godin, F. Vallières et É. Archambault, *La production scientifique québécoise : mesure basée sur la BRJQ*, 13 p. texte inédit.

recherche, elles remplissent leur rôle de revues savantes. Une étude montre⁴ que, dans la période récente (1991 à 1996), les auteurs québécois qui publient dans l'une ou l'autre des 52 revues répertoriées sont minoritaires (46 %) et, par voie de conséquence, que les auteurs étrangers (37 %) et canadiens (13 %) y occupent une place importante. L'origine des auteurs varie beaucoup : les trois quarts des collaborations viennent de 10 pays. Outre la France, qui vient nettement en tête, on retrouve par ordre d'importance les États-Unis et la Belgique, puis la Grande-Bretagne et d'autres pays européens. Les collaborations canadiennes sont principalement ontariennes, avec l'Université d'Ottawa, University of Toronto, York University, Queen's University, McMaster University, notamment. Ajoutons que le quart des textes sont écrits en collaboration avec des auteurs de l'extérieur du Québec et davantage avec des auteurs étrangers. Ce phénomène est à la hausse. Certes, les chercheurs établis publient à l'étranger mais publient aussi dans les revues québécoises et canadiennes. Si ces dernières attirent les chercheurs étrangers, elles n'ont pas été désertées par les chercheurs en pleine ascension ou établis. Pour les chercheurs, il est question surtout de pondération dans le choix des véhicules de diffusion des résultats de recherche et dans la manière de rejoindre le plus efficacement le public visé.

Publiées surtout en français ou lui donnant une large place, les revues participent, comme acteurs majeurs, à la diffusion des connaissances principalement dans ce sous-ensemble linguistique des réseaux scientifiques, dont on sait qu'ils sont plus polymorphes en sciences humaines et sociales. On pourrait espérer une diffusion plus soutenue de ces publications dans le monde francophone. Indépendamment de la valeur intrinsèque des revues, cette diffusion se bute à des contraintes et obstacles liés aux circuits commerciaux conventionnels. Cela ne doit pas nous conduire à sous-estimer l'importance relative des abonnements étrangers, venant d'institutions surtout, ce qui témoigne de la pertinence des revues dans les milieux spécialisés. On sait, pour notre échantillon de revues québécoises, que plus de 30 % des abonnements viennent de pays étrangers ; au total, cela compte pour un peu moins de la moitié des revenus d'abonnements, ce qui est loin d'être négligeable⁵. Dans l'état actuel des choses, et compte tenu des embûches rencontrées, ce niveau d'abonnements étrangers, surtout institutionnels, est une reconnaissance de la qualité du travail éditorial et de la participation des revues québécoises aux principaux forums internationaux des disciplines ou secteurs. Ce faisant, une diffusion électronique des revues contournerait les contraintes et obstacles à la diffusion de l'imprimé (sans pour autant devoir le supprimer) et intensifierait leur présence dans les milieux scientifiques extérieurs au pays.

⁴ B. Godin, F. Vallières et É. Archambault, *La production scientifique québécoise : mesure basée sur la BRSQ*, 13 p. texte inédit.

⁵ Dans une étude encore inédite, (de Gérard Boismenu et Guylaine Beaudry, « Pénétration des revues savantes canadiennes dans les universités américaines ») nous avons tenté d'apprécier la présence des revues savantes canadiennes dans les réseaux de recherche aux États-Unis. Un échantillon de 53 revues a été retenu et 24 grandes bibliothèques de recherche des universités aux États-Unis ont été étudiées. On constate qu'en moyenne les bibliothèques sont abonnées à 30 % des revues canadiennes, qu'un peu plus de revues en sciences humaines et sociales qu'en sciences (12 contre 10) sont présentes dans plus de la moitié des bibliothèques étudiées. Par ailleurs, en prenant l'ensemble des revues canadiennes recensées (plus de 200 revues), la moitié est répertoriée dans *Current Content* et les revues sont en moyenne répertoriées dans 14 *abstracts*, sans compter *Current Content*. Nous n'avons pas désagrégé pour les revues québécoises, mais on peut estimer qu'elles ne seraient pas atypiques, surtout si on considérait des indices de diffusion rendant justice à la langue de communication utilisée (le français).

DISPERSION DES LIEUX DE PUBLICATION ET PRÉCARITÉ FINANCIÈRE

À la suite de ces constats (revues massivement présentes en sciences humaines et sociales et publiées par des organismes sans but lucratif), on observe que la grande dispersion des lieux de publication caractérise le milieu des revues. Le fait qu'un grand nombre de revues soient publiées par des organisations indépendantes a pour conséquence un éparpillement des ressources et des compétences. La volonté des responsables de revue de préserver l'autonomie éditoriale a peut-être joué en faveur de cette situation, mais il faut souligner que les politiques publiques et universitaires y ont concouru. Cela a pour conséquence que, dans de nombreux cas, chaque revue, sur une base individuelle, se transforme en petite entreprise s'occupant de l'ensemble du processus d'édition, de publication et de diffusion, en répétant dans divers lieux, et sans économie d'échelle ou transfert de savoir-faire, les mêmes opérations : la préparation de copie, la mise en page, la négociation avec les fournisseurs, la mise en place d'un service d'abonnement ou la négociation avec une agence, l'organisation de la distribution en librairie, etc.

Ce morcellement a un coût qu'il est difficile d'évaluer de façon précise, mais il compte tout au long de la chaîne de publication et de diffusion pour la version imprimée. Ce coût apparaît nettement lorsque vient le temps d'envisager la publication électronique : une publication de qualité, ainsi qu'une diffusion conséquente, étant souvent hors de portée en raison du manque de ressources. On ne peut, par ailleurs, passer sous silence que ce morcellement provoque des contraintes sociales à l'introduction de changements. Les politiques antérieures des institutions, dont le Fonds FCAR, ont pu inciter les directions de revues à investir dans les tâches plus techniques de la production de l'imprimé (référence à l'acquisition d'équipement pour la mise en page et la préparation des prêts-à-photocopier). L'effet structurant que l'on voudrait voir dans les futures politiques du Fonds s'est révélé dans les politiques antérieures, à telle enseigne que la routinisation de pratiques de fabrication ainsi induite est susceptible de créer des contraintes à la mise en place d'une forme de regroupement, telle une infrastructure de publication et de diffusion s'adressant à l'ensemble des revues.

Il faut ajouter que les équipes éditoriales demeurent assez légères et comptent, pour une large part, sur des contributions non rémunérées et sur un personnel à temps partiel ou en prêt de service (avec ou sans compensation monétaire) d'un centre, d'un institut, d'une association, d'une faculté. Ces structures éditoriales assez légères bénéficient d'un financement assuré par des subventions publiques, éventuellement, par l'institution universitaire ou par la société savante de rattachement, et par le revenu des ventes de la revue. Ces structures ont été précarisées ces dernières années par l'effet, souvent cumulatif, de politiques de restrictions budgétaires de la part des organismes subventionnaires et des universités, pour ne pas parler de la baisse des revenus d'abonnements dans plusieurs cas. Tous ces éléments ont résulté en une pression plus forte sur les équipes éditoriales, alors qu'elles n'étaient déjà pas bien pourvues.

L'étude globale de l'état des revenus et des dépenses de 27 revues subventionnées par le CRSH ayant leurs bureaux administratifs au Québec⁶ permet de constater que près de 41 % des revenus des revues proviennent de la vente d'abonnements et, pour une portion congrue,

⁶ Données fournies par le CRSH et compilées par nos soins. Les données se rapportent à l'année 1998.

de la vente au numéro. Le soutien financier des deux organismes subventionnaires majeurs représente 47 % des revenus (à parts quasiment égales entre le Fonds FCAR et le CRSH). Le reste est assez diversifié, allant de la contribution monétaire des universités ou associations à la publicité. En moyenne, les revues enregistrent un déficit de 3 700 \$. Cette représentation ne prend pas en compte la contribution en nature ou en services des institutions ou associations, qui n'est pas moins présente.⁷ Par ailleurs, sans considérer cette dernière contribution, environ les trois quarts des dépenses servent à supporter la structure éditoriale et organisationnelle de la revue, ainsi qu'à préparer la première copie. Les coûts afférents au support papier, en termes d'impression, de reliure, d'affranchissement et d'expédition, correspondent au dernier quart⁸. Ce montage fragile, qui est reconduit d'une année à l'autre par les directions de revue, souvent avec une grande dose d'imagination, illustre la précarité des organisations et leur manque endémique de ressources.

C'est dans ces conditions que se pose la mise à contribution des technologies de l'information dans le milieu des revues au Québec dans le cadre d'un portail de publication et de diffusion électronique des revues.

LE DÉFI DE L'ÉLECTRONIQUE : UNE OCCASION À EXPLOITER

En raison de la dispersion organisationnelle et institutionnelle des revues, le défi de l'électronique peut conduire à confirmer ou pousser encore plus loin le morcellement du mode de fonctionnement des revues et, de ce fait, à élaborer des solutions pour l'électronique sur une base individuelle et tenant compte des ressources limitées dont chacune dispose. Les solutions que l'on peut préconiser dans ce cas peuvent difficilement contribuer à un traitement de publication, de conservation et de diffusion en adéquation avec la vocation des revues savantes et avec la nature des documents publiés. Le défi de la publication électronique peut être vu comme une occasion importante à exploiter, non seulement pour améliorer ses conditions de diffusion, mais aussi pour repenser la revue et le mode de formalisation et de communication de la recherche. Le défi consiste à s'approprier les technologies de l'information afin de les mettre au service de la publication scientifique. En même temps, le chercheur et la revue savante doivent participer pleinement au renouvellement des conditions d'élaboration et de transmission des contenus, par lesquels ils communiquent les résultats de recherche.

Le défi posé incite à s'interroger sur la structure organisationnelle de publication caractérisée par la dispersion et le morcellement, alors que l'ère actuelle favorise des pratiques pensées en termes de réseaux. L'isolement et l'individualisation des pratiques dans l'édition de revues savantes peuvent-ils être considérés comme favorables à la participation active aux changements en cours dans le système de communication scientifique ? Il n'y a pas de réponse simple à cette question, mais on peut voir dans ce défi une occasion pour repenser la structuration des activités de publication et de diffusion.

⁷ L'Association canadienne des revues savantes évaluait récemment que cette contribution s'établissait entre 5 000 et 10 000 dollars selon les institutions.

⁸ Un traitement électronique change les conditions de mise en forme du texte pour la première copie, d'où un besoin en ressources qui doit être réévalué ; cette évaluation viendra plus loin dans cette étude, mais on ne peut considérer qu'il y a là une source d'économie massive.

Ces activités de publication et de diffusion doivent s'inscrire de façon plus large dans une vision des changements dans la communication scientifique et être pensées en fonction d'une connaissance très fine des fonctionnalités des technologies de l'information appliquées à la vocation des revues. Cela dit, ces activités de publication et de diffusion constituent un apport pour les revues en ce qui a trait aux phases plus techniques de la production et de la mise à disposition des textes de la revue. L'autonomie éditoriale, le processus d'arbitrage, la sélection et la correction des textes sont l'apanage des directions de revue. Définir la vocation d'un portail de publication et de diffusion électroniques de revues n'entame en rien cette autonomie éditoriale, ni l'indépendance administrative et la personnalité intellectuelle et physique dont s'est dotée une revue. La revue est appelée à se délester de tâches de production directe et de diffusion (pour l'électronique) au profit du portail qui, de son côté, doit travailler étroitement avec les revues pour valoriser les contenus et accroître leur rayonnement avec une démarche systématique d'exploitation des potentialités du Web.

CE QU'EN PENSENT LES DIRECTEURS DE REVUE

Les directions de revue qui ont discuté le projet du portail, suite à une présentation des principales données qui sous-tendent la présente étude lors des groupes témoins, représentent la moitié des revues savantes qui ont un financement public au Québec (annexe 1.1). Les points de vue exprimés ne sont pas unanimes, mais de grandes tendances se découpent et elles sont instructives. Il ne faut pas négliger le fait que ces directions ont une connaissance très fine des conditions d'existence de leur revue, du travail éditorial jusqu'à la vente par abonnements, en passant par la production technique et la situation financière de la revue. De plus, les revues sont dirigées par des professeurs-chercheurs généralement très actifs professionnellement et qui ont une expérience d'auteur sur les plans national et international. Enfin, ils manifestent un intérêt certain pour l'évolution du mode de communication scientifique et comprennent les principaux enjeux.

La publication électronique ne heurte pas les directions de revue, bien au contraire. Elles saisissent bien la tendance lourde vers l'électronique et sont prêtes à aller de l'avant avec une version électronique de leur revue ; d'ailleurs, pour un certain nombre, la revue est déjà offerte en version électronique et/ou des informations et extraits de la revue sont offerts sur un site Web. Certes, l'électronique est une voie de l'avenir et la transition est essentielle pour conserver et augmenter la visibilité des revues québécoises. Partant de là, certaines considérations semblent essentielles pour assurer cette transition vers l'électronique avec la collaboration des directions de revue. Certaines considérations doivent ainsi être relevées.

Pour que les choses puissent se faire, le réalisme est de rigueur. Il n'est sans doute pas approprié que le passage soit imposé par une action volontariste et d'application immédiate. Aussi fascinante que soit l'idée de considérer la version électronique comme la version première de la revue, elle implique un changement de culture, elle impose de modifier et de revoir les contenus, et elle nécessite de redéfinir les méthodes de travail. D'où l'idée d'une transformation sous plusieurs facettes, qui relève aussi bien des auteurs, des équipes (éditoriales et administratives) en place et/ou à recomposer, de la capacité financière de chacune et d'un nouveau rapport aux activités de production technique et de diffusion électronique de la revue. De là, une transition, qui ne sera pas facile et pour laquelle la cadence accélérée semble brutale, du moins pour plusieurs revues ; elle pourrait avoir un effet contre-productif, c'est-à-dire

contraire à celui recherché, en raison des réactions négatives qu'elle pourrait provoquer. Il convient de donner du contenu à l'idée d'une transition maîtrisée.

Le grand défi, c'est de concilier l'édition indépendante et la mise en place d'un portail de production et de diffusion des revues. Cette jonction peut se faire si les rôles sont bien définis, c'est-à-dire en faisant en sorte que le portail n'investisse pas le processus éditorial et que la responsabilité d'ensemble de la revue, y compris la responsabilité financière, reste dans les mains des directions actuelles. Il faut bien reconnaître que le milieu des revues est, d'une part, un milieu de spécialistes et de chercheurs établis, mais que c'est aussi, d'autre part, un milieu d'artisans et de bénévoles qui a mobilisé beaucoup de ressources dans le passé pour valoriser les revues et accroître leur rayonnement. L'attachement que ce milieu porte aux conditions de production et de diffusion de la revue témoigne de l'investissement de chacun et du tissu social qui soutient une telle organisation. Un bouleversement de cette deuxième dimension du milieu provoque des réactions sociales prévisibles et vraisemblablement légitimes. En reconnaissant cette réalité, le portail — qui intervient comme producteur technique de la revue et diffuseur électronique — est vu par les revues comme un projet intéressant, précisément parce que la version électronique peut difficilement se faire selon un mode artisanal, comme c'est le cas pour la version papier.

Les directions de revue ont été agacées — pour employer une litote — par certaines affirmations concernant la raison d'être de la version papier. Ce n'est ni par nostalgie ni par seul souci d'avoir une image qui leur soit propre que les directions de revue soulignent la place toujours majeure de la version imprimée de la revue. La substitution de la version imprimée par la version électronique n'est pas un mouvement signalé du côté des revues importantes dans chaque discipline. L'adoption d'une version électronique pour les revues établies n'a pas, de façon générale, sonné le glas de la version imprimée. Les raisons évoquées tiennent principalement au changement progressif des habitudes des lecteurs, à un souci d'accroître la diffusion là où la version papier reste un support de grande pertinence, et aux attentes des lecteurs qui sont hors des milieux universitaires et qui n'ont pas les mêmes infrastructures de communication électronique. Par ailleurs, le papier n'occasionne pas que des coûts mais aussi des revenus qui dépassent ces premiers. Ces considérations ne militent pas contre la version électronique, mais soulignent l'intérêt indubitable de maintenir une version imprimée des revues. Ces considérations mettent en relief l'intérêt d'adopter une chaîne de production électronique de laquelle il est possible de tirer le matériel pour une version imprimée.

La question du financement est lancinante. Les revues ressentent vivement la précarité financière de la situation actuelle : c'est la capacité de maintenir, avec des équipes réduites et sous-financées, une grande qualité de contenu et une reconnaissance dans le milieu scientifique au Québec et à l'étranger qui est compromise à l'heure actuelle. On se demande, d'un autre côté, comment il serait possible pour le Fonds FCAR — qui a un discours de restrictions budgétaires à l'égard des revues tout au moins et qui a diminué le nombre de revues supportées financièrement au dernier concours — d'ajouter de nouveaux services sans déplacement de fonds en puisant dans les ressources actuelles des revues. Les économies à court terme occasionnées par l'introduction du portail de production et de diffusion électroniques des revues ne sont certainement pas acquises. La capacité de générer des revenus autonomes (par la vente d'abonnements et au numéro) et de diversifier les revenus publics est apparue, avec le temps, comme une garantie contre une trop grande vulnérabilité à l'égard des changements de politique d'une seule institution publique. L'idée de la gratuité

permise par la présence d'un « commanditaire public » est intéressante en soi mais accroît la dépendance à l'égard des fonds publics et élimine presque complètement les marges de manœuvre existantes à l'égard de décisions plus ou moins éclairées d'organismes publics, sur lesquels les revues ont peu de prise. La mise en place d'un portail serait mieux reçue si elle devait aller de pair avec un « nouveau pacte financier » entre les revues et le Fonds FCAR, voire avec d'autres sources publiques ; ce pacte devrait être caractérisé par la recherche d'une stabilité pour les revues et par des garanties sérieuses d'engagements publics s'inscrivant dans la durée.

Les revues sont des institutions qui ont participé activement au développement des sciences au Québec et à la communication scientifique sur les plans national et international. Il y a une histoire intellectuelle des revues savantes au Québec et une participation très actuelle au développement et au rayonnement des travaux des chercheurs d'ici et d'ailleurs. Là-dessus, les revues remplissent un rôle qui est dans leur nature même, et qui n'est pas spécifique au Québec. Les directions de revue ressentent une responsabilité très grande de protéger et de développer ce patrimoine collectif. Elles sont donc soucieuses d'éviter que des décisions mal assurées se transforment en coûts intellectuels et organisationnels majeurs et provoquent des situations irrécupérables. C'est dans cet esprit qu'elles sont peu enclines à l'aventurisme, qui n'est pas associé à l'électronique, mais plutôt à un processus de prise de décisions dans lequel tous les éléments n'auraient pas été pris en compte.

La conception du portail de production et de diffusion, qui est exposée par la suite, ne répond pas à l'ensemble de ces interrogations ou préoccupations, il va sans dire. Ces dernières dépassent largement les compétences du portail et mettent en cause plusieurs intervenants et facteurs systémiques. Pourtant, la conception du portail prend en compte plusieurs de ces observations, si bien que le portail s'appuie, notamment, sur une définition des compétences qui reconnaît le rôle et la contribution des revues ; de même, bien qu'il soit tourné vers les supports et formats électroniques, le processus de production du portail permet d'articuler sa mission première à la production d'une version imprimée, dont la responsabilité, cependant, ne relève pas de lui.

CHAPITRE 2
L'INFRASTRUCTURE
INTELLECTUELLE,
ORGANISATIONNELLE ET
MATÉRIELLE

Le portail a pour mission d'être une infrastructure commune mise au service de la communication savante et des revues en particulier. Cette infrastructure réunit plusieurs dimensions qui demandent un certain développement.

Le portail permet la mise en application de normes de production de qualité pour la publication électronique des revues et définit un lieu de diffusion assurant le rayonnement et la visibilité des revues. Le portail travaille en jonction avec le milieu des revues et intervient dans la prestation de services techniques et de conseils pour la production et la diffusion. Cela circonscrit la responsabilité et les services du portail, en regard des éditeurs de revue. De plus, le portail est appelé à prendre la forme d'une entité et d'une architecture institutionnelles, dont les caractéristiques seraient conformes à sa mission, et qui s'inscriraient dans la durée. Ajoutons que, sous-jacent à ces services et à une organisation stable, le portail fournira une expertise qui doit se renouveler, ce qui implique la mobilisation de ressources aptes à suivre l'évolution de ce secteur d'activités et à revoir les procédures, dans la recherche d'une optimisation des services rendus. Enfin, le portail doit réunir les conditions matérielles et les équipements à la hauteur de sa mission, ce qui signifie un environnement technique de qualité. C'est cet environnement multiforme que l'on va esquisser d'entrée de jeu, avant de traiter plus spécifiquement des fonctions de production, de diffusion et d'archivage remplies par le portail.

L'ARCHITECTURE GLOBALE ET LES SERVICES DU PORTAIL

Cette section a pour objectif de donner une vue d'ensemble de l'architecture du portail et des services qui seraient offerts. Les éléments, brièvement exposés, seront repris dans une autre section de cette étude.

ARCHITECTURE

On peut imaginer à terme l'architecture du portail en fonction d'une diversité de services liés aux différentes formes d'expression de la communication savante. Nous nommons « zones » les espaces propres à ces diverses formes d'expression. Premièrement, le portail s'appuie sur les revues qui devraient constituer une première zone. À ce premier noyau, une deuxième zone, comportant une collection de prépublications, pourrait se greffer. Les autres types de documents relevant de la littérature grise constitueraient la troisième zone de ce portail. Sachant que ces différentes formes d'expression des résultats de la recherche répondent à des contraintes qui ne sont pas du même ordre, il importe de bien tracer les contours et les traitements qu'ils appellent. Ces distinctions permettront de définir des zones nettement identifiables avec leur cohérence propre, mais composant dans l'ensemble une collection riche et diversifiée.

Ajoutés à ces trois zones, des outils de diffusion, tel un outil de recherche performant, permettraient d'exploiter de façon efficace les collections, en cherchant par différents critères de recherche à travers les documents des trois zones du portail. De plus, l'objectif de la diffusion dans le Web étant de porter les résultats de la recherche québécoise dans les créneaux de diffusion internationaux, des mécanismes et des protocoles assurant l'interopérabilité entre les différents portails de diffusion de revues savantes et de prépublications permettront de relier les collections du portail québécois à d'autres collections d'importance.

On peut imaginer l'ensemble en partant du noyau que composent les revues, ce qui correspond d'ailleurs au mandat précis de cette étude, puis en envisageant l'ajout d'autres zones. Cela donnerait au portail québécois la physionomie suivante :

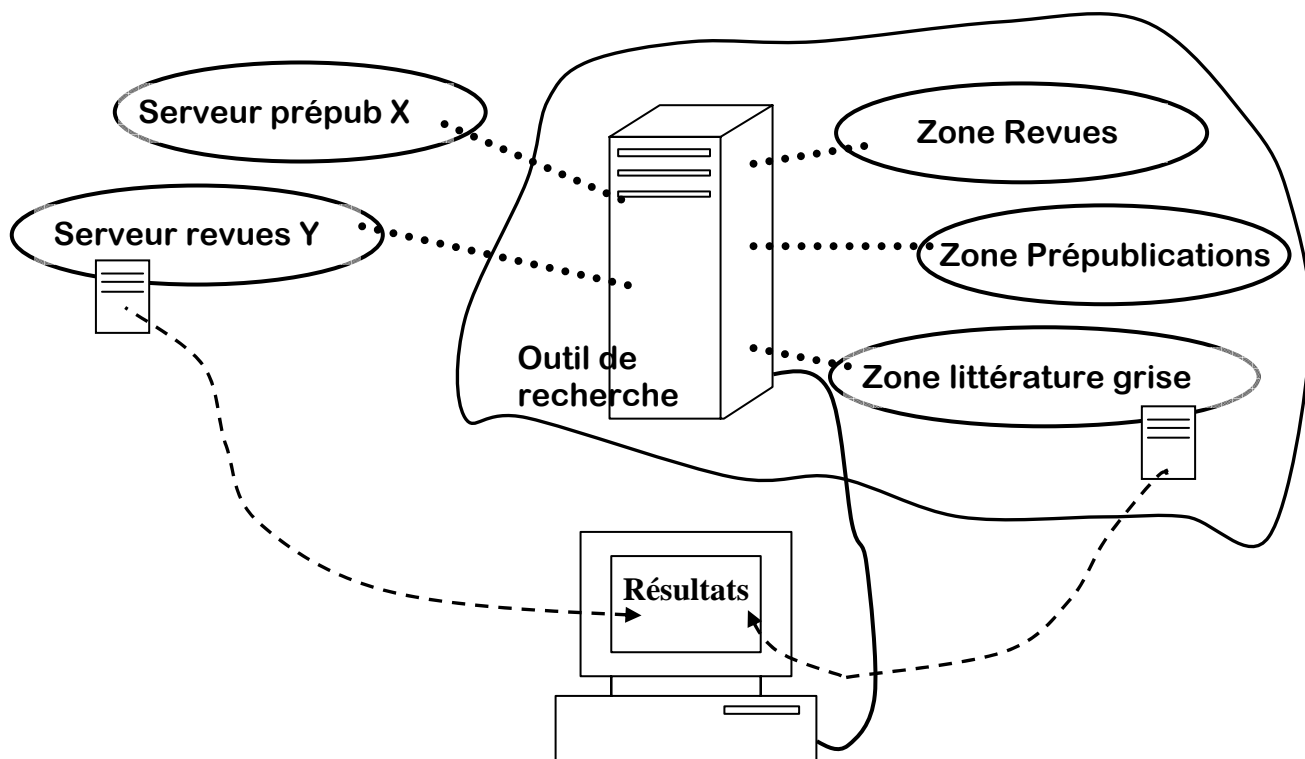


Figure 1 : Architecture globale du portail québécois

LA RELATION ENTRE LE PORTAIL ET LES REVUES

Comme l'indique la figure 2, la création du portail ne modifiera en rien les responsabilités éditoriales de l'équipe de la revue. La conception du projet éditorial et des numéros thématiques, l'appel de textes, l'arbitrage, le choix des articles à publier ainsi que toutes les autres activités que la revue mène en relation avec les auteurs et les évaluateurs pour la préparation des textes à publier continueront d'être la seule responsabilité de la revue. C'est au moment où l'équipe éditoriale de la revue reconnaît le texte prêt à publier que les équipes de production technique du portail interviendront.

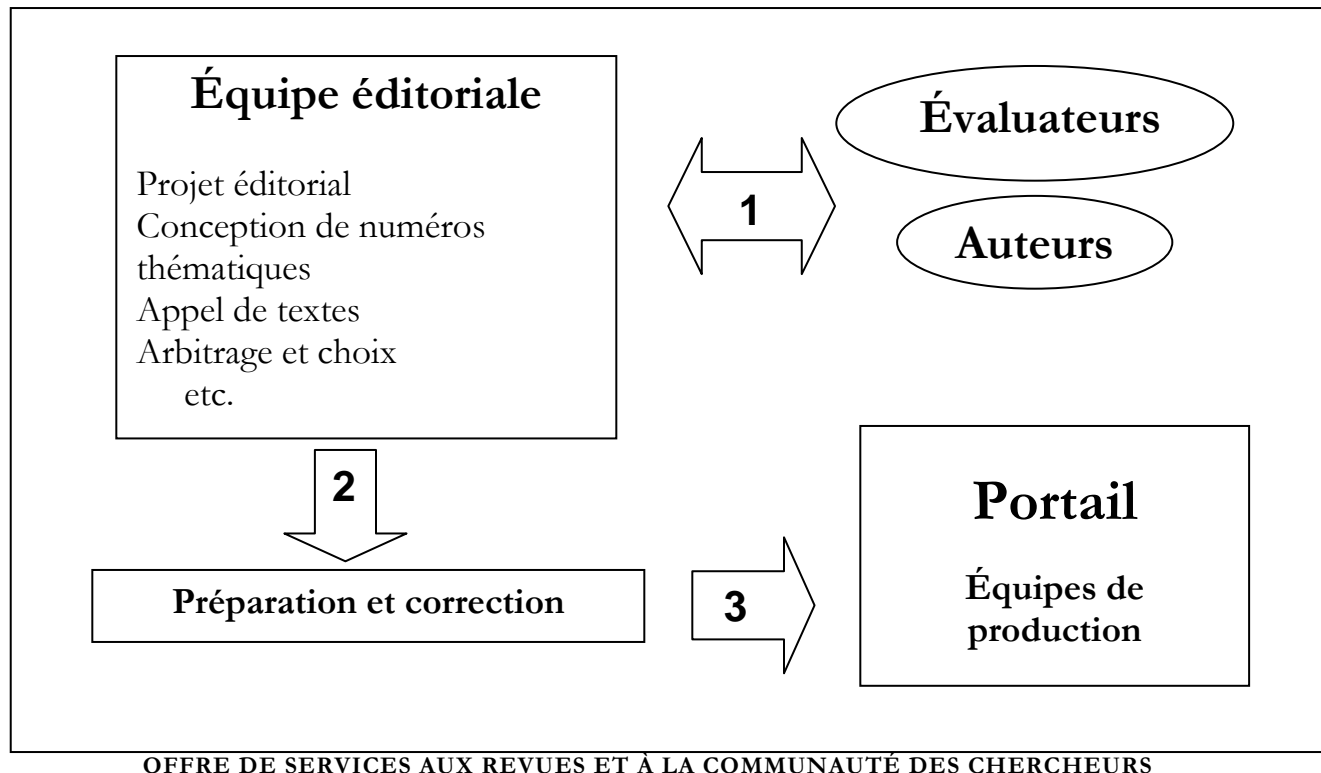


Figure 2 – Relation entre les revues et le portail

SERVICES DE PUBLICATION

Ces services, destinés principalement mais non exclusivement aux revues en sciences humaines et sociales, comprendraient la réception des textes, la préparation de la copie, la numérisation des images, la préparation des épreuves, la numérisation des figures ou photographies, la saisie des corrections, la préparation de la version électronique et la mise en page papier.

La technologie proposée pour la production des revues savantes a déjà fait ses preuves et est utilisée par plusieurs éditeurs commerciaux et sans but lucratif. Les revues savantes seraient produites par une chaîne de traitement basée sur la norme XML, ce qui permet la fabrication des versions papier et électronique à partir d'un même processus. Des métadonnées (description de l'article : auteur, titre, résumé, volume, numéro...) seraient générées automatiquement à partir de la chaîne de traitement. Un fichier Postscript, un fichier PDF ou un prêt-à-photographier seront préparés pour l'impression de la revue.

SERVICES DE NUMÉRISATION RÉTROSPECTIVE

L'intérêt d'une collection d'articles de revues savantes réside bien sûr dans la qualité des articles mais aussi dans le nombre de revues et d'articles sur un sujet donné. Une étude⁹ récente a permis de noter l'intérêt pour les articles parus jusqu'à dix ans avant le moment de la cueillette d'informations. Cet intérêt est sensiblement le même pour les chercheurs des sciences humaines et sociales et ceux des sciences exactes. La numérisation rétrospective des numéros antérieurs s'avère donc répondre à un véritable besoin, autant pour faciliter l'accès aux résultats de la recherche que pour offrir une collection d'articles plus riche.

Pour résumer, le portail offrirait un service d'intermédiaire entre la revue et les entreprises de numérisation en plus de créer les métadonnées des articles et ainsi permettre aux utilisateurs de chercher par un même outil de recherche à travers les collections courante et rétrospective d'une revue.

SERVEUR DE PRÉPUBLICATIONS

Le portail devrait, à terme, être composé d'une zone dédiée aux prépublications et ainsi permettre aux chercheurs de faire connaître rapidement à leur communauté disciplinaire l'état d'avancement de leurs travaux. Outre les documents destinés à une publication ultérieure sous un autre média, une zone du portail devrait accepter et diffuser d'autres types de documents relevant de la littérature grise des disciplines, comme par exemple, les thèses, les rapports et notes de recherche, les documents utilisés lors de conférence, etc.

SERVICES DE DIFFUSION ÉLECTRONIQUE

La mer d'informations accessible dans le Web fait en sorte que la seule présence sur Internet d'une revue n'équivaut en rien à sa réelle diffusion électronique. Le système de diffusion du portail devrait permettre de relier par un outil de recherche les revues portant sur les mêmes sujets et étant diffusées par d'autres portails, au Canada et ailleurs dans le monde. L'objectif est ici de rendre accessible une collection d'articles non pas par frontières géographiques, institutionnelles ou même disciplinaires mais bien par sujets. Outre les liens ainsi créés entre les différents portails actuels et à venir, le *portail* devrait porter une attention particulière aux performances de ses différentes pages dans les outils de recherche les plus consultés. De plus, des ententes devraient être conclues pour assurer la diffusion des articles par les différentes bases de données bibliographiques qui permettent les liens avec le plein texte.

Les outils proposés pour la mise en œuvre des services de diffusion sont la création et l'application dans la chaîne d'un modèle de métadonnées, l'identification de chacun des articles par un DOI (*Digital Object Identifier*), l'utilisation du service *CrossRef* permettant de relier les références bibliographiques aux articles cités et bien sûr, la création d'un portail où les utilisateurs auront accès à un outil de recherche performant. Cet outil de recherche permettrait d'offrir, par une même interface, l'accès aux collections des différentes zones du portail. Ces services devraient également permettre, à chaque parution, de diffuser automatiquement la

⁹ Gérard Boismenu et Guylaine Beaudry, « Publications électroniques et revues savantes : acteurs, rôles et réseaux. » *Documentation et Bibliothèques*, vol. 45, numéro 4, 1999.

table des matières accompagnée des résumés à une liste de personnes intéressées ou abonnées à la publication. Finalement, à la base, ce système d'information devrait être hébergé sur une partie du réseau assurant l'accès à une bande passante importante et dans un environnement sécurisé selon les normes reconnues.

SERVICES D'ARCHIVAGE DES DOCUMENTS ÉLECTRONIQUES

La responsabilité de l'archivage de la documentation produite par les citoyens d'un pays revient aux bibliothèques nationales. Il en va de même pour la documentation sur support électronique. Comme le modèle et les pratiques ne sont pas encore établis, le portail devrait offrir des services d'archivage qui correspondent aux normes fixées par les archivistes pour garantir l'accès aux revues savantes. Cependant, le portail devrait assurer la conservation des revues pour garantir l'accès pendant tout le cycle d'utilisation des revues.

GESTION DES ABONNEMENTS

Deux scénarios sont évoqués dans cette étude quant à l'accès gratuit ou payant aux revues sous forme électronique. Le cas échéant, selon la solution retenue, le portail devrait offrir des services de gestion des abonnements.

UNE ENTITÉ INSTITUTIONNELLE

Il n'est pas approprié, dans cette étude, de décrire les composantes institutionnelles spécifiques du portail. Contentons-nous de soumettre les caractéristiques générales qui devraient être associées au portail sur le plan organisationnel.

Le portail est une institution qui s'engage dans la prestation de services de production et de diffusion auprès des éditeurs de revue ; les services doivent être professionnels, fiables et s'inscrire dans la durée. Pour assurer une viabilité économique, le portail doit proposer une vision d'affaires réaliste et applicable. En raison de sa responsabilité à l'égard de tiers — revues et institutions auxquelles le service est offert, ainsi qu'usagers et clients du portail —, il doit pouvoir se constituer en entité institutionnelle et légale. Cet organisme devrait réunir certaines caractéristiques qui découlent de principes généraux.

Cet organisme devrait être une institution sans but lucratif qui repose, autant que possible, sur un partenariat entre diverses composantes du milieu universitaire ; il serait souhaitable que ces partenaires aient déjà un engagement dans la publication de revues ou dans les technologies du texte. Le cadre institutionnel devrait retenir la formule coopérative et témoigner de la propriété du milieu universitaire.

L'organisme serait dirigé par un conseil d'administration regroupant les partenaires qui se sont engagés dans le projet. Il faudrait considérer la mise en place d'un comité « des clientèles » qui réunirait des représentants des revues desservies par le portail et des principaux utilisateurs du portail, comme les bibliothèques ou centres de recherche et, le cas échéant, des bailleurs de fonds.

Les partenaires auraient une responsabilité professionnelle (qualité du service, stabilité et atteinte des objectifs) et financière dans la réalisation du plan d'affaires. Les contributions financières pour soutenir le portail pourraient être de plusieurs ordres. Les coûts d'implantation du portail devraient faire l'objet d'une subvention spéciale. Les services de production technique offerts aux revues devraient être facturés sur une base d'autofinancement et soutenus par la politique de financement public des revues. Les coûts d'exploitation de la structure de diffusion du portail seraient assurés par les pouvoirs publics en association avec les partenaires.

Dans le but d'utiliser au mieux les compétences existantes et d'offrir un service proche des aires géographiques des directions de revue, l'architecture organisationnelle du portail devrait prévoir, d'une part, une structure de production distribuée avec un lieu de production et de services conseils à Québec et un à Montréal et, d'autre part, un portail de diffusion unique qui voit à l'ensemble des responsabilités de rayonnement et de visibilité des revues dans le Web.

L'organisme devrait être reconnu par le ministère de l'Éducation, par le ministère de la Recherche, de la Science et de la Technologie, et par le ministère de l'Industrie et du Commerce.

DES RESSOURCES INTELLECTUELLES

Le portail doit pouvoir réunir des ressources intellectuelles de premier ordre. Il est très précieux de compter sur une connaissance assez fine du milieu des revues et des conditions de publication de celles-ci sur support électronique, mais aussi sur support papier. L'équipe chargée de créer et de consolider le portail doit réunir des compétences diverses afin de concevoir, d'implanter et d'assurer la prestation de services, en tenant compte de l'évolution des normes et pratiques dans un horizon large du milieu desservi, des ressources financières disponibles et du plan d'affaires. Cela implique une direction intellectuelle arrimée aux conditions de réalisation, une gestion professionnelle en phase avec un programme de « livraison » de services, soit des compétences professionnelles et techniques complémentaires.

Comme cela sera souligné dans le chapitre sur la diffusion, l'accroissement de la visibilité, l'augmentation du nombre d'utilisateurs et l'impact du portail dans et sur la communauté scientifique internationale varient en fonction de la capacité de faire du portail un lieu intellectuellement dynamique, qui nourrit, accueille et provoque des activités en mesure d'interpeller le milieu des chercheurs nationaux et internationaux. Partant d'une collection de revues diversifiée et de très bonne qualité intellectuelle et technique, la meilleure promotion pour le portail, c'est d'en faire un foyer intellectuel dynamique qui fait place à des activités et offre des services de premier niveau. Cela demande des ressources conséquentes en qualité et en importance.

On doit ajouter que la publication électronique est un domaine dominé par l'innovation : tant les technologies utilisées que les usages et les pratiques dans le Web évoluent très rapidement et interpellent régulièrement les attentes et les prévisions concernant la publication

et la diffusion électroniques. On est loin d'un univers où les conventions relatives aux processus de publication et à l'interaction avec les utilisateurs sont stables. On ne peut échapper à la nécessité d'investir dans la recherche appliquée, dans l'expérimentation et dans l'amélioration des fonctions de diffusion.

Cela demande des ressources consacrées à un personnel qualifié. Dans un secteur où l'évolution est aussi rapide et où les paramètres sont aussi mobiles, la nécessité de se situer à la frontière technologique a pour contrepartie de considérer les ressources dans la recherche et le développement d'applications comme le « nerf de la guerre », la condition indispensable à la fois à la stabilisation prochaine du processus de production et au développement de services qui révèlent tout l'intérêt de la publication électronique en exploitant les fonctionnalités.

De la sorte, le portail doit s'investir dans des *activités de recherche et de développement*. En étant très conscient de ce qu'est la recherche, l'expression « petit r grand D » correspond davantage à la mission du portail, car l'essentiel portera sur le développement d'applications tant pour la production que pour la diffusion.

Ajoutons que des activités consacrées à la *veille* sont fondamentales compte tenu de l'évolution des pratiques et des normes en matière de formats, d'outils de production et de diffusion, et de définition et transformation des modèles de diffusion tant technique qu'économique.

De plus, la grande mobilité des ressources humaines dans ce secteur rend nécessaire l'identification des conditions permettant d'assurer une redondance des expertises pour la gestion et le fonctionnement quotidiens des systèmes d'information. La *documentation et la diffusion des pratiques* et des décisions deviennent, dans ce contexte, très importantes. C'est en fait ce qui permet à tout organisme de mieux gérer les connaissances et les savoirs des forces vives de son personnel.

UN ENVIRONNEMENT TECHNIQUE ET MATÉRIEL

Le portail doit réunir les conditions nécessaires pour assurer la sécurité, la stabilité et la constance des services, aussi bien pour l'hébergement des documents produits que pour leur diffusion et leur archivage. Ces caractéristiques sont à la fois techniques et organisationnelles.

Le portail contient des données sensibles et stratégiques ; il est, dès lors, recommandé qu'il ne soit pas installé sur un serveur partagé, où d'autres sites Web et d'autres applications pourraient résider et compromettre son fonctionnement. Le portail devrait être installé sur sa propre machine.

En raison de son statut interinstitutionnel, il semble souhaitable que le portail soit hébergé dans un lieu qui reflète bien cette caractéristique. Le RISQ (Réseau d'informations scientifiques du Québec), comme lieu interinstitutionnel, serait approprié, d'autant plus qu'il dispose directement de l'accès le plus direct à la dorsale de l'enseignement supérieur et de la recherche au Québec. Le RISQ offre ce service d'hébergement et est relié à très haut débit à tous les réseaux de recherche nationaux et internationaux. (voir annexe 2.1 ; une offre d'hébergement)

Au-delà de l'hébergement sur lequel nous reviendrons, le portail suppose des composantes matérielles et logicielles en adéquation avec ses spécifications et ses fonctionnalités, qui seront exposées dans les prochains chapitres. Pour prendre la mesure de ces besoins, ainsi que de l'environnement nécessaire, une brève énumération suffit pour le moment. Cette énumération est complétée par le cahier technique qui apparaît en annexe :

- Estimation de l'espace disque nécessaire pour les contenus : pour les articles de revues savantes (publications courantes et numérisation rétrospective) et les textes en prépublication, les besoins d'espace disque pour les cinq premières années vont de 4 Giga-octets (Go) la première année à 30 Go la cinquième.
- Autres besoins d'espace disque : cette dernière estimation ne comprend pas l'espace requis pour le système d'exploitation, les divers ensembles logiciels nécessaires à la diffusion, dont l'outil de recherche ; on peut compter jusqu'à 2 Go pour le système d'exploitation.
- Le portail est de type vertical et doit pouvoir permettre, au moins progressivement, de proposer diverses fonctionnalités pour le système de diffusion ; on pense, entre autres, à un outil de recherche plein texte et exploitant des métadonnées, à la personnalisation des accès des visiteurs en fonction de leurs intérêts et de leurs droits d'accès, à l'enregistrement des habitudes d'usage et à l'adaptation de l'affichage à la machine visiteuse.
- La technologie sous-jacente à l'architecture du portail doit utiliser les technologies récentes les mieux éprouvées (Java/servlets, architecture ouverte, multi-plates-formes, CDF ou RDF, etc.).

L'hébergement (voir annexe 2.2) devrait comprendre, en particulier, l'accès du serveur à une très large bande passante et une position stratégique sur les circuits d'interconnexion Internet, la surveillance continue des équipements et la continuité du service. Il devrait comprendre également la surveillance 7/24 des appareils clients par des logiciels de surveillance, la climatisation et l'alimentation garantie (UPS et génératrice), la prise de copie de secours, la maintenance des DNS pour établir les noms logiques de serveurs.

Cet environnement technique vient compléter les ressources humaines qui seraient mobilisées et il serait à leur service. Il constituerait une condition nécessaire de la mise en place d'une structure de production, de diffusion et d'archivage des revues électroniques. L'ensemble de ces éléments — de la définition des responsabilités du portail à l'environnement technique, en passant par le cadre institutionnel et les ressources intellectuelles — compose la charpente sur laquelle s'élèverait le portail et dont on peut décomposer les principaux mandats dans les chapitres qui suivent.

CHAPITRE 3
UNE CHAÎNE DE
PRODUCTION TECHNIQUE

La revue des pratiques techniques et des services sur le plan international montre une variété dans les choix qui sont faits pour publier électroniquement les textes. On observe cependant une constante en ce qui a trait à la diversité des services offerts, la publication et la diffusion électroniques allant bien au-delà de la simple mise à disposition d'articles de revues en ligne¹⁰. Les portails de publication et de diffusion de revues savantes offrent des services qui répondent aux trois fonctions : production, diffusion et gestion. Pour chacune de ces fonctions, une pluralité de pratiques est observée, que ce soit pour les formats de production, les choix en ce qui a trait à l'archivage des documents électroniques ou les modèles économiques. Ce chapitre de l'étude s'attarde plus particulièrement à la première fonction, soit la production, dont les pratiques vont de la réception des textes qui parviennent des équipes des revues jusqu'à la production des formats de diffusion électronique et papier.

LES FORMATS

Avant de passer à la production proprement dite, il importe de définir les outils avec lesquels nous allons travailler. L'édition électronique amène plusieurs distinctions par rapport à l'édition papier. Une des distinctions fondamentales est que l'encodage numérique permet une multitude de transformations du document. Alors que le document papier est fixé au terme de sa production et qu'il demeurera tel pour toutes les autres étapes de ses cycles d'utilisation, le document électronique se voit prendre plusieurs formes, que ce soit au moment de la production, de la diffusion ou de la gestion et de l'archivage. Le document électronique est multiforme. Tantôt il est utilisé pour la diffusion dans le Web, tantôt il sert à créer les films et les plaques pour l'impression des exemplaires papier ou encore son contenu est exploité pour créer des métadonnées ou une base de données. Il importe donc, d'une part, de bien distinguer formats d'acquisition, de production, de diffusion et d'archivage et, d'autre part, de déterminer quels sont les formats à utiliser pour chacune de ces étapes du cycle de vie d'un article.

DE QUEL FORMAT PARLONS-NOUS?

Pour distinguer les différentes formes que le document électronique prend tout au long de ses cycles d'utilisation, retenons les formats d'acquisition, de production, de diffusion et d'archivage et situons-les par rapport au portail.

Les *formats d'acquisition* pour le portail sont, pour les documents textuels, les fichiers de traitements de texte commerciaux, le plus souvent Word. Il s'agit du format utilisé par le secrétariat de la revue pour l'étape d'uniformisation des textes (références bibliographiques, corrections linguistiques, etc.). Du format Word, on passe ensuite au format « Word++ », format sur lequel nous reviendrons plus loin. Disons pour l'instant que ce format est en fait l'application de la feuille de styles de la revue au fichier Word¹¹. Avec le format « Word++ », on passe à un format de production : c'est à partir de ce format que sera produit le *format XML de production*. Une fois toutes les corrections saisies et le contenu validé, on passe au *format XML*, qui servira également pour *l'archivage*. Les *formats de diffusion* seront produits automatiquement à partir du format XML. Dans un premier temps, les formats de diffusion

¹⁰ Voir la liste des sites visités dans le cadre de cette étude à l'annexe 3.1. On peut consulter aussi : Caroline Clément et Marc Bonvin, sous la direction de Françoise Khenoune. *Les périodiques électroniques en sciences humaines et sociales*, Mars 2000. http://www.unil.ch/BCU/recherche/l_art_bi.htm

¹¹ Idéalement, cette étape devrait être réalisée par le secrétariat de rédaction de la revue. Toutefois, pour des raisons particulières, l'équipe du portail pourrait offrir ce service aux revues.

seront HTML et PDF. On peut penser que dans un avenir assez proche, le XML pourra être utilisé directement pour la diffusion.

Pour ce qui est des formats images, les revues les reçoivent dans une foule de formats allant du fichier électronique en format vectoriel jusqu'à une mauvaise reproduction par photocopie d'une gravure en passant par des photographies. Les équipes de production doivent être très souples et avoir accès à plusieurs outils logiciels et matériels pour récupérer ou numériser les images.

Dans la mesure du possible, les fichiers comprenant les figures doivent être conservés dans un format reconnu pour l'archivage, le format TIFF par exemple. Les formats de diffusion pour le Web sont le plus souvent GIF ou JPEG.

LES FORMATS STRUCTURÉS

Les formats structurés permettent, en plus d'encoder le contenu d'un texte, de rendre lisible par l'ordinateur la structure sémantique et hiérarchique d'un document. Chaque élément d'un texte encodé à l'aide d'un langage structuré se voit attribuer des balises qui le délimitent et l'identifient. Par ce balisage structuré, le contenu d'un document est nettement distinct des diverses représentations qu'on peut en faire. Cette particularité fait en sorte que l'apparence ou la mise en page d'un document ne sont pas encodées à même le contenu, comme c'est le cas pour les formats de traitement de textes propriétaires. Une application et un fichier distincts déterminent les attributs de style, de mise en page ou d'affichage de l'information.

Les documents structurés permettent donc l'échange et la réutilisation de textes électroniques tout en préservant le contenu, les données et la structure sémantique d'un document des différentes utilisations qu'on fera de l'information dans le présent et le futur. Chaque élément d'un document structuré peut être stocké, recherché, réutilisé, extrait pour créer un autre document ou une base de données. Les deux formats de balisage structuré les plus utilisés sont le SGML et le XML.

LE SGML

Le SGML (Standard Generalized Markup Language) est un langage structuré normalisé par l'ISO en 1986. Le SGML est un métalangage qui permet de décrire la structure logique d'un document. Le cœur d'un système SGML, la DTD (définition de type de document), est en quelque sorte la « grammaire » d'un genre de texte (article, livre, dictionnaire, etc.) où on retrouve la description des éléments, de leurs contenus et des relations entre les éléments. À partir du moment où les éditeurs ont voulu diffuser des contenus à la fois sur support papier et électronique, l'implantation d'un système SGML est apparue très intéressante. Seulement, les investissements financiers ainsi que l'expertise requise pour la mise en place et la gestion d'un système SGML ont freiné l'implantation d'un tel système dans certains milieux.

POURQUOI PAS SEULEMENT LE HTML?

À première vue, le HTML (HyperText Markup Language) peut sembler être une solution intéressante. Ce format est une application simple du SGML. HTML est facile à apprendre et sa diffusion dans le Web ou sur d'autres supports tels que le cédérom se fait assez

aisément. Cependant, on rencontre rapidement les limites du HTML. Le jeu de balises limité du HTML ne permet tout simplement pas d'identifier et de représenter adéquatement les nombreux éléments souvent complexes d'une revue savante. De plus, le HTML s'est souvent développé selon un navigateur ou un autre, ce qui lui confère un comportement proche d'un format propriétaire et ne lui accorde pas le statut de format d'archivage. Le HTML est encore pour un certain temps le format de diffusion le plus utilisé dans le Web mais, compte tenu de ses limites, il n'est définitivement pas adéquat pour l'édition de revues savantes.

ENTRE LE SGML ET LE HTML... LE XML

Le XML (eXtended Markup Language) a été recommandé par le W3C (World Wide Web Consortium¹²) en 1998. XML est un langage de balisage de document structuré basé sur le SGML qui a été développé pour pallier les limitations du HTML sans pour autant posséder les difficultés d'application du SGML.

L'intérêt du XML apparaît à partir du moment où un document doit être publié, non seulement sur format papier mais sur d'autres supports. Dès qu'il y a une deuxième version d'un même document, XML offre des avantages significatifs. Le texte d'un article en XML est structuré de telle sorte qu'il peut être matérialisé par plusieurs médias. Tout comme le format SGML, le XML permet de distinguer le texte et les données qu'il contient des représentations visuelles qu'on lui donne, papier ou électronique, en tenant compte des besoins actuels et futurs.

Le XML, c'est aussi une famille de technologies. Le XLink (XML Linking Language), encore à l'étude par le W3C, est la norme qui décrira la façon d'ajouter des liens hypertextes à un fichier XML. Le Xlink permet notamment de qualifier la nature des liens. On pourra, par exemple, à partir d'une zone ou d'un mot sensible, accéder à la biographie d'un auteur, sa bibliographie, ses coordonnées ou son affiliation. Le XSL (eXtensible Stylesheet Language) est un langage qui permet de développer des feuilles de styles pour la représentation à l'écran des documents. Il est basé sur le XSLT (eXtensible Stylesheet Language Transformation), un langage de transformation qui permet de faire la conversion d'un document XML à un autre type de document XML. Dans un système XML, les *schémas* permettront aux utilisateurs de développer leurs propres applications XML. Finalement, le XHTML (eXtensible Hypertext Markup Language) est une reformulation du HTML 4.0 en XML, en quelque sorte une passerelle pour favoriser le passage du HTML au XML.

QUELLES UTILISATIONS POUR LE PDF ET LE POSTSCRIPT?

Les formats de description de page, bien que diffusés dans le Web, sont surtout destinés à l'impression. Les deux formats de loin les plus utilisés sont Postscript et PDF (Portable Document Format).

Postscript, introduit en 1985 par la compagnie Adobe, a été conçu pour faciliter l'impression de documents, peu importe les environnements logiciel et matériel. Un fichier Postscript fait la description des différents éléments de la page, graphiques ou typographiques. Il n'y a aucune organisation structurée de l'information dans un document Postscript et les

¹² <http://www.w3c.org/>

ajouts de liens hypertextes sont impossibles. Postscript est surtout utilisé comme format de transmission pour les ateliers d'impression.

Le PDF (Portable Document Format) est un autre format propriétaire introduit par Adobe. Grâce au logiciel distribué gratuitement par Adobe, *Acrobat Reader*, il peut être lu sur plusieurs plateformes. Ce format permet l'intégration d'hyperliens, de signets et de métadonnées. Le PDF a pour avantage de préserver au document son apparence originale ; il est, de plus, habituellement facile de produire un document PDF dans un environnement contrôlé. Toutefois, le PDF est davantage à un format d'impression sur demande qu'un format de visualisation ou d'exploitation. Format attrayant par sa simplicité, il est assez souvent très simple de créer un fichier PDF à partir de n'importe quelle application qui nous permet d'imprimer un fichier. Outre le fait que le PDF soit un format propriétaire — ce qui ne donne aucune assurance quant à la pérennité de l'information —, ce format ne fait que restituer à l'écran ce que l'on retrouve sur papier. Devant un document PDF à l'écran, on se retrouve devant un *papier de verre*. De plus, comme l'information d'un document PDF n'est pas structurée, les possibilités de recherche sont moindres. On peut dire en ce sens que le PDF facilite la transition entre les supports papier et électronique. C'est en quelque sorte comme si on appliquait une technologie nouvelle à des outils d'un autre âge. C'est peut-être le propre des périodes de transition, comme le laisse entendre Febvre et Martin dans *L'apparition du livre* :

Les premiers ouvrages imprimés n'ont pas modifié immédiatement l'apparence et la forme du manuscrit médiéval ; en effet, les premiers imprimeurs se sont appliqués à réaliser des imitations précises, à produire en fait de faux manuscrits. Leur travail était tellement remarquable qu'il est virtuellement impossible pour le non-initié de distinguer un manuscrit réel d'un de ces premiers imprimés. Par exemple, on rapporte que la Bible à 42 lignes a été imprimée avec des caractères qui reproduisaient fidèlement ceux des missels du Rhin.¹³

LES FORMATS IMAGES

Une discussion détaillée des contenus multimédias dépasse le cadre de ce document, mais quelques principes importants à respecter doivent être soulignés.

D'abord, la question des formats se pose tout aussi naturellement aux contenus multimédias qu'aux contenus textuels. Ces formats doivent être choisis en fonction de la pérennité de l'information et de la structuration.

Ensuite, il faut essayer, le plus possible, d'inclure des métadonnées dans les documents multimédias, afin de faciliter leur repérage et leur manipulation. Cette pratique est encore peu courante, mais différents standards sont en train de se développer à ce sujet¹⁴. Enfin, bien entendu, il faut relier le document maître (en format XML) avec les fichiers externes de façon à être indépendant du stockage physique des différentes ressources.

Un document encodé en format image est en quelque sorte une représentation assez près d'une photographie d'un texte, d'une image ou d'un objet. En ouvrant un tel type de fichier, ce que nous avons à l'écran est bel et bien l'*image* d'un texte, par exemple. Cela implique

¹³ Lucien Paul Victor Febvre et Henri-Jean Martin, *L'apparition du livre*, Paris, Michel, 1971.

¹⁴ Voir notamment RDFPic. Adresse URL : <http://jigsaw.w3.org/rdfpic/>

que pour tous les formats images, aucune manipulation de texte n'est possible, que ce soit « copier-coller » ou encore la recherche plein texte.

Il existe deux grandes familles de formats images : les formats vectoriels et les formats image en mode point¹⁵. Les formats vectoriels permettent de décrire l'image en une série d'objets géométriques (lignes, ellipses, polygones, etc.) dont les propriétés sont précisées. L'information contenue dans le fichier d'un format vectoriel décrira, par exemple, la position, l'épaisseur et la couleur d'une ligne à tracer. Il s'agit donc d'une série d'instructions qui sont réutilisées par l'outil de visualisation pour reconstruire l'image. Ce type de format permet d'éditer les différents objets d'une image indépendamment : les images peuvent être modifiées sans perte de résolution et la taille des fichiers est relativement petite.

Les fichiers image en mode point sont faits d'une mosaïque d'éléments d'image appelés pixels. Chaque pixel contient les informations concernant la couleur d'un point de l'image en particulier. La combinaison de tous les pixels forme l'image. Les fichiers image en mode point sont habituellement très lourds à cause de la quantité d'information que contient chaque pixel et du grand nombre de pixels nécessaires pour obtenir une image de qualité. Il faut toujours privilégier un format vectoriel à un format image en mode point, sauf lorsque le type d'information ne s'y prête pas.

Pour la production des revues, comme en témoignent le nombre de formats différents qui sont acheminés aux rédactions de revue, une multitude d'outils sont utilisés par les auteurs et leurs assistants. Pour ce qui est de la diffusion électronique, les formats les plus utilisés sont GIF et JPEG. Le GIF (Graphics Interchange Format), créé par CompuServe en 1987 et amélioré en 1989, est un format très utilisé dans le Web. La compression du format GIF ne permet d'utiliser que 256 couleurs ou tons de gris, ce qui offre comme avantage de générer des fichiers très légers. Le GIF est utilisé en général pour les figures.

Le JPEG (Joint Photographic Expert Group) permet l'utilisation de 16 millions de couleurs et est surtout employé pour représenter des photographies. C'est également un format très répandu dans le Web, particulièrement reconnu pour son excellent algorithme de compression.

Le format TIFF (Tagged Image File Format) a été développé par Aldus et Microsoft. Le principal avantage du TIFF est son algorithme de compression qui assure aucune perte d'information. Ce format de type image en mode point est reconnu par les archivistes et plusieurs guides des meilleures pratiques pour l'archivage des documents en formats images¹⁶. Les fichiers TIFF sont particulièrement lourds avant compression et ne peuvent être lus par les navigateurs Web. Ce format est surtout utilisé comme format de capture dans les projets de numérisation rétrospective. Le TIFF est encore pour le moment le format recommandé pour l'archivage des documents sous forme image.

L'équipe du portail devra suivre le développement et les utilisations de nouveaux formats images. Le SPIFF (*Still Pictures Interchange File Format*) a été commandé par l'ISO afin de garantir

¹⁵ Creating Digital Resources for the Visual Arts: Standards and Good Practice. Adresse URL: http://vads.ahds.ac.uk/guides/creating_guide/sect32.html

¹⁶ Creating Digital Resources for the Visual Arts: Standards and Good Practice. Adresse URL: http://vads.ahds.ac.uk/guides/creating_guide/sect32.html

l'interopérabilité des systèmes manipulant les images¹⁷. Ce format, créé par le *Joint Photographic Expert Group*, fait partie du domaine public. Il s'agit d'une amélioration de leur norme JPEG. Le PNG¹⁸ (*Portable Network Graphics*) est la nouvelle étoile filante des formats d'images. Il est destiné à remplacer le GIF et le TIFF. Il s'agit également d'un format non propriétaire.

Il est possible d'obtenir des fichiers en PDF Image lors de la numérisation d'une image¹⁹. Pour ce faire, il faut utiliser le logiciel *Adobe Capture*. Ce format peut être utile pour les projets de numérisation rétrospective de revues savantes, comme on le verra dans une prochaine section de ce chapitre.

LA CHAÎNE DE TRAITEMENT

La production et la diffusion électroniques de revues savantes impliquent des opérations humaines et informatiques effectuées avec rigueur et efficacité. La chaîne de traitement décrite dans cette section de l'étude permet la publication de revues savantes en vue de leur diffusion sur plusieurs supports (électronique et imprimé). Bien qu'une simple translation du contenu de la version papier vers l'électronique puisse déjà contribuer à la plus grande diffusion des revues, la chaîne de traitement proposée conduira éventuellement à créer de véritables publications électroniques. En effet, la version principale sera destinée à la consultation en ligne, ce qui amènera les auteurs à penser en termes de publication électronique : cette nouvelle tendance pourrait notamment avoir pour effet de modifier, à moyen terme, le type de représentation du contenu, en plus du texte, que l'on retrouve dans ces revues.

Cette section ne contient pas toutes les questions relatives à la chaîne de traitement. Nous avons plutôt voulu soulever les principales questions, et y apporter les premiers éléments de réponses. Lorsque c'est possible, nous avons proposé des pistes de solution, voire des systèmes, mais le but premier n'était pas de donner une liste précise et exhaustive des étapes et outils de la chaîne documentaire. À ce point-ci du projet, trop d'informations sont encore inconnues ou imprécises pour être en mesure d'effectuer des choix définitifs, pensons aux ressources financières disponibles, au nombre de revues à publier et à diffuser, au modèle économique retenu ainsi qu'à l'expertise disponible pour la réalisation du projet.

UNE CHAÎNE DE TRAITEMENT DOCUMENTAIRE

Une chaîne de traitement documentaire comprend l'ensemble des processus, humains et technologiques, nécessaires pour produire et transformer de l'information. Dans le projet qui nous intéresse, l'objectif de ces processus est bien entendu de publier des revues savantes sur différents supports. Il est possible de regrouper les processus en trois grands ensembles que nous présentons brièvement ici.

La *production* de l'information consiste à consigner sur support informatique des idées, concepts, relations, réflexions issus d'un travail intellectuel. Il s'agit, en quelque sorte, de prendre le résultat de la création humaine et de l'enregistrer dans un fichier. Très souvent, un tel processus se limite à démarrer un logiciel de traitement de texte, à saisir l'information, puis à

¹⁷ Daniel Lecomte, *et al.*, *Les normes et les standards du multimédia*, Paris, Dunod, 1999, p. 91.

¹⁸ A Basic Introduction to PNG Features. Adresse URL: <http://www.libpng.org/pub/png/pngintro.html>

¹⁹ Gary Cleveland, *Selecting Electronic Document Formats*, Juillet 1999. Adresse URL : <http://www.ifla.org/VI/5/op/udtop11/udtop11.htm>

enregistrer le fichier correspondant. Toutefois, la publication électronique de revues savantes exige un processus un peu plus complexe.

La *gestion* de l'information consiste à organiser un ensemble d'unités d'information logiquement reliées dans le but de pouvoir les manipuler facilement ou les conserver de façon sécuritaire. La gestion de l'information implique l'archivage, sujet dont il est question dans un autre chapitre de cette étude.

La *diffusion* de l'information consiste à transférer l'information consignée sur support informatique vers des personnes qui ont besoin de la consulter. Il s'agit d'un processus de transfert d'information depuis un ordinateur vers une personne. Il est très important de réaliser que la mise en ligne d'une revue ne signifie pas nécessairement qu'elle sera consultée. Là comme pour les autres aspects de la chaîne documentaire qui nous occupe, des moyens concrets et précis doivent être mis en œuvre pour répondre aux besoins des lecteurs en matière de diffusion.

La notion de *publication* est au cœur de ce projet car elle conditionne plusieurs aspects de la chaîne de traitement. En effet, l'objectif du portail est de produire des versions publiées des articles de revues, puis de les diffuser et de les archiver.

LES PRINCIPES QUI SOUS-TENDENT LE MODÈLE

La chaîne de traitement est basée sur certains principes fondamentaux permettant d'atteindre les objectifs reliés à la publication de revues savantes dans un contexte électronique. Nous allons présenter brièvement ces principes afin d'aider à comprendre les différents aspects de la chaîne.

LA NORMALISATION

La chaîne de traitement utilise des formats normalisés pour la représentation de l'information et tente d'utiliser le plus souvent possible des mécanismes de traitement normalisés, sans mettre en péril l'efficacité et la souplesse de l'ensemble. Toutefois, aucune normalisation des outils n'est proposée, au contraire, l'indépendance par rapport aux logiciels et systèmes doit toujours être gardée.

La normalisation des *formats* de représentation est nécessaire, car elle permet d'atteindre une certaine indépendance par rapport aux outils utilisés, mais aussi d'assurer la pérennité de l'information. La normalisation des mécanismes de *traitement* de l'information est intéressante et de plus en plus en vogue, même si, en général, ces normes sont moins pérennes que celles concernant les formats. En effet, les « langages » de manipulation de l'information évoluent sans cesse, poussés par l'expérimentation de nouvelles approches mais surtout par l'augmentation de la capacité des ordinateurs.

EXPLOITATION DE L'ÉLECTRONIQUE

Les documents numériques amènent des changements aux façons de faire tout en ouvrant des perspectives nouvelles pour la création, la diffusion et l'exploitation des résultats de la recherche. La chaîne de traitement proposée permet d'exploiter facilement

l'information électronique selon les besoins et les outils *actuels et futurs*, tout en réduisant au maximum les inconvénients qu'amène invariablement l'introduction de ce nouveau mode de production.

Deux aspects fondamentaux distinguent l'information imprimée de l'information électronique. D'abord, cette dernière permet d'inclure différents types d'information, du texte à l'animation 3-D en passant par l'image, la vidéo, le son et la réalité virtuelle. La chaîne de traitement proposée doit donc permettre d'exploiter le caractère multimédia et interactif de l'information électronique. Ensuite, la publication imprimée s'effectue dans un contexte où l'on connaît et contrôle le support de diffusion de l'information. Dans le monde électronique, il existe une variété de supports de diffusion, mais surtout une grande variété de qualités ou de caractéristiques techniques de ces supports. La chaîne de traitement adoptée par le portail doit être d'une grande souplesse dans la production de l'information, électronique ou imprimée, afin de satisfaire le plus grand nombre de supports de diffusion, actuels et futurs, ainsi que le plus grand nombre d'expériences de consultation de l'information.

UNE SOURCE, PLUSIEURS PRODUITS

Le principe précédent nécessite de fournir l'information aux lecteurs sous différentes formes, en différentes versions²⁰. Il est impensable de produire en parallèle ces différentes formes ou versions. La modification du processus, ou sa ré-ingénierie, vise à produire une source d'information unique qui sera déclinée en différentes versions.

Cette source d'information devra être suffisamment riche pour que toutes les versions de diffusion s'obtiennent par appauvrissement, c'est-à-dire par la suppression d'information plutôt que par l'ajout. C'est la seule façon d'automatiser la chaîne de traitement : un traitement automatique peut difficilement ajouter de l'information, alors qu'en enlever est une opération assez simple.

LES DOCUMENTS STRUCTURÉS

Les documents structurés contiennent de l'information sur leur structure logique et sémantique. Par exemple, on y identifiera explicitement les différentes sections du document et leur titre, ou encore certaines parties du texte comme des lieux géographiques ou des noms de personne. Cette structure leur confère une grande richesse, car il est aisé de traduire une information de structure en une information de restitution. Ainsi, convertir une information telle que « ceci est un titre de section premier niveau » en une instruction telle que « ceci doit être imprimé en caractères gras et un corps de 14 pts » peut se faire aisément et surtout automatiquement.

L'utilisation de documents structurés est la seule façon de respecter le principe « une source, plusieurs produits » dans un contexte électronique. La chaîne de traitement permettra donc d'obtenir, pour chaque article, un document structuré, suffisamment riche pour

²⁰ On entend ici par « versions » essentiellement les mêmes contenus diffusés sur différents supports. On aura ainsi, la version électronique d'un article et sa version papier. Il pourrait arriver que la version électronique soit enrichie par rapport à la version papier, par exemple, par l'ajout d'icongraphie couleur ou de données brutes.

permettre toutes les exploitations envisagées. Il existe deux normes permettant de représenter des documents structurés : le SGML et le XML.

L'AUTOMATISATION

L'intérêt *économique* de la chaîne de traitement repose sur l'automatisation des différents traitements effectués sur l'information. La création de l'information demeurera toujours un processus intellectuel. Une fois cette information créée, sa déclinaison en différents produits de gestion et de diffusion devrait toujours être réalisée automatiquement.

TRANSITION EN DOUCEUR

Les revues auxquelles s'adressera le portail existent déjà en tant que publications imprimées. Il est donc important de leur offrir une transition vers l'électronique qui s'effectue le plus simplement possible, tout en ne limitant pas les possibilités offertes par ce nouveau support. La chaîne de traitement proposée relève le défi de respecter les pratiques issues de l'imprimé, tout en permettant une exploitation intéressante de l'électronique.

LA PRODUCTION DE L'INFORMATION

La première étape de la chaîne de traitement consiste à obtenir un document électronique structuré, suffisamment riche pour servir de source unique aux multiples formes de publication et de diffusion. Ce sont les revues qui « approvisionnent » la chaîne de traitement. Il devient essentiel que les équipes éditoriales des revues acheminent des textes dont le contenu est prêt à publier, de sorte que les corrections soient saisies une seule fois dans la source unique, pour ensuite servir à la production des différents formats de diffusion.

Afin de tenir compte de ce contexte, la chaîne de traitement utilise une approche indirecte et à deux étapes pour obtenir le document structuré (en format XML²¹) qui est la source unique d'information. Dans une première étape, l'information est saisie (ou récupérée) dans un document de traitement de texte conventionnel (nous préconisons le traitement de texte Microsoft Word, le plus utilisé) en respectant des règles d'écriture assez strictes. Ensuite, ce document Word est converti *automatiquement* en format XML, sans aucun traitement manuel ou intellectuel.

Cette procédure permet de faire un compromis optimal entre la bonne connaissance du traitement de texte chez les différents contributeurs et la souplesse ainsi que la normalisation du format XML. Le facteur qui permet ce compromis, et sur lequel cette partie de la chaîne de traitement repose, est le respect des règles d'écriture du document Word. Puisque le document Word peut être converti automatiquement vers le format XML, il devient en quelque sorte la source unique d'information, sur laquelle les corrections seront effectuées, et à partir de laquelle (indirectement) les différents produits seront élaborés.

²¹ Nous préconisons l'utilisation du format XML plutôt que le SGML afin de profiter des nombreuses applications XML disponibles. La norme SGML permet d'aller encore plus loin dans la structure et la gestion de l'information, mais ses fonctions additionnelles ne présentent aucun intérêt pratique pour le portail.

LE DOCUMENT WORD

Le format Word (comme tous les formats de traitement de texte) ne se prête pas à la représentation de documents structurés, et ce, pour deux raisons principales :

- le format n'est pas hiérarchique, alors que les documents textuels le sont presque toujours ;
- les traitements de texte ne permettent pas de valider le contenu ou la structure des documents.

Ce dernier point est très important, car une bonne utilisation des documents structurés implique la notion de validation (de la structure), validation qui permet d'obtenir des documents à structure homogène ; cela permet leur exploitation de façon structurée par la suite. Il s'agit exactement du même principe que l'on applique aux bases de données, où l'on sépare en champs les différentes unités d'information, pour ensuite pouvoir les exploiter très facilement et efficacement. Les documents structurés appliquent ce principe aux documents textuels.

Même si le traitement de texte ne permet pas de valider la structure d'un document, il permet d'identifier (en partie) cette structure, et ce, à l'aide des feuilles de styles. On peut facilement indiquer qu'un paragraphe est le titre d'une section de niveau deux en lui appliquant le style « Titre 2 », par exemple. Les styles de type caractères permettent d'obtenir le même effet sur quelques lettres ou quelques mots dans le texte, à l'intérieur d'un paragraphe.

La validation de la structure devra donc être effectuée d'une façon détournée, soit par la formation des personnes qui appliqueront les styles, mais aussi par des programmes informatiques. L'application des styles doit nécessairement être effectuée par des personnes qui comprennent à quoi cette étape peut servir, qui connaissent les documents qu'elles s'approprient à « styler » et qui connaissent la feuille de style à appliquer. Leurs connaissances de la revue et de leur discipline constituent donc la première validation.

Une validation humaine n'est toutefois pas totalement fiable, et c'est pourquoi une validation informatique est nécessaire. Celle-ci sera obtenue lors de la conversion du document Word au format XML, ce dont nous parlerons dans la prochaine section.

LE DOCUMENT XML

La norme XML n'est pas à proprement parler un format de document. Il s'agit plutôt d'un langage permettant de définir des formats de document²². Ces formats sont en fait des modèles de documents, que l'on appellera ici des *schémas*²³. Ces schémas permettent d'exprimer

²² Le XML permet aussi de représenter des documents qui ne respectent pas un schéma particulier, seulement les règles syntaxiques d'XML. On parle ici de documents « bien formés » plutôt que « valides ». Toutefois, pour bâtir une chaîne de traitement XML, l'utilisation de schémas est essentielle, pour assurer une certaine cohérence à l'ensemble.

²³ Jusqu'à maintenant, les schémas ont toujours été représentés, autant en XML qu'en SGML, par des DTD, soit des *définitions de types de documents*. Toutefois, une norme en construction du W3C (XML Schema, voir <http://www.w3.org/XML/Schema.html>) permet également de définir un format XML. Cette future norme est appelée à remplacer les DTD, mais pour l'instant l'utilisation de celles-ci est recommandée, en particulier dans les applications documentaires où les XML Schemas apportent peu de nouvelles fonctionnalités intéressantes.

l'ensemble des contraintes que doit respecter un document XML pour qu'il soit valide dans un certain contexte. Par exemple, à l'aide d'un schéma défini pour des articles de revues savantes, il est possible de vérifier si un article en format XML est valide en fonction de ce schéma.

Les schémas, qui sont les véritables formats de documents, et le respect de leurs contraintes constituent le cœur d'un système de gestion documentaire basé sur la norme XML. C'est pourquoi il faut leur accorder une grande importance lors de la conception et lors du développement des différents outils. Une des premières activités du portail sera de définir un schéma spécifique aux revues savantes. Les parties génériques du schéma devraient être inspirées de schémas de référence tels que Docbook²⁴, TEI²⁵ ou ISO 12083²⁶. Normalement, un seul schéma devrait être suffisant pour l'ensemble des revues, car les caractéristiques communes des articles sont plus nombreuses que leurs différences. De plus, en utilisant une approche générique pour certaines caractéristiques des documents, il est possible d'en arriver à un schéma unique. Toutefois, il faudra évaluer l'intérêt et la nécessité d'utiliser un schéma de référence *modulaire* et de l'adapter pour obtenir un schéma spécifique pour chaque revue. En définissant correctement le schéma de référence et les schémas dérivés, il serait possible d'obtenir des outils uniques pour l'ensemble des revues et ainsi rendre les traitements efficaces. Le choix entre un seul schéma ou bien un schéma de référence plusieurs fois adapté devra être fait après une étude plus approfondie.

La production du document XML se fait à partir du document Word, automatiquement. Pour y arriver, il est beaucoup plus facile de passer par le format RTF²⁷. La conversion d'un document RTF au format XML (respectant un schéma spécifique) n'est pas une tâche générique pour laquelle il existe des outils déjà préparés. En effet, le XML n'étant pas un format, il est impossible de créer un outil de conversion du format RTF vers n'importe quel schéma XML.

Il existe deux types d'outil pour effectuer cette conversion. Les solutions basées sur la programmation tirent profit d'un langage de programmation pour manipuler les différentes structures du format RTF pour les convertir au XML. D'autres solutions basées sur des règles (par exemple « les paragraphes qui sont en style *Titre 2* sont des titres de section ») sont en général plus faciles à mettre en place, mais moins souples, moins puissantes.

L'utilisation d'un langage de programmation s'avère nécessaire si le contenu du document Word doit être manipulé de façon à changer l'ordre des éléments, à ajouter du contenu, à intégrer de l'information provenant de sources externes, etc. Une telle solution sera nécessaire dans le cas où, par exemple, les documents Word ne sont pas conçus en fonction du document XML à produire, mais plutôt en fonction d'autres impératifs, tels que les pratiques antérieures. Le langage de programmation Omnimark²⁸ peut s'avérer une solution intéressante pour ce genre de conversion, tout comme le langage de transformation XSLT²⁹. Dans ce

²⁴ Il s'agit d'une DTD développée pour l'industrie informatique, largement utilisée et bien adaptée aux documents textuels. Voir : <http://www.docbook.org/>.

²⁵ Le projet « Text Encoding Initiative » a notamment donné comme résultat une DTD permettant de représenter des textes littéraires (et, par extension, tout document textuel). Voir <http://www.uic.edu/orgs/tei/>.

²⁶ La norme ISO12083 contient quatre DTD, dont une permet de représenter des articles de revue. Pour en savoir plus : <http://www.xmlperts.com/12083.htm>.

²⁷ Rich Text Format, le format d'échange de Microsoft que l'on peut considérer comme strictement équivalent au format Word quant à l'information représentée, mais qui utilise une syntaxe différente plus facile à manipuler.

²⁸ Produit de la société Omnimark Technologies, voir <http://www.omnimark.com/>.

²⁹ Le XSL Transformations, norme du W3C, voir <http://www.w3.org/TR/xslt/>.

dernier cas, il faudra d'abord convertir le document Word vers un format XML quelconque³⁰, générique.

L'utilisation d'un outil basé sur des règles sera suffisant si les documents Word sont construits de façon très similaire aux documents XML. Dans un tel cas, la structure peut être induite directement des styles utilisés, peu de contenu devra être déplacé et l'écriture de règles de structuration devient très efficace. Les outils les plus connus pour effectuer ces transformations sont Majix³¹ et UpCast³².

Le choix final de l'approche et des outils dépasse le cadre de ce document et devra tenir compte de critères comme les connaissances et l'expérience des personnes qui programmeront les conversions ou établiront les règles.

LE CONTENU NON TEXTUEL

La discussion précédente a porté exclusivement sur le contenu textuel des documents. Le contenu textuel des documents hypermédias est habituellement celui qui sert de ciment pour l'ensemble des composantes, un peu comme dans une page HTML où l'on trouve le texte et des références aux images et autres objets multimédias stockés à l'extérieur du document.

Dans le contexte du XML, les solutions sont les mêmes : le document XML contient le texte et des références aux objets multimédias. Un document XML est souvent constitué de plusieurs fichiers informatiques : la chaîne de traitement propose l'utilisation de Word pour la création de l'information textuelle et toutes les informations non textuelles seront stockées non pas dans Word mais dans des fichiers externes. Toutefois, le document Word contient l'information nécessaire pour pouvoir faire référence aux objets externes.

LE PROCESSUS DE TRAVAIL DANS SON ENSEMBLE

La description qui précède a volontairement évité d'intégrer l'intégralité du processus de travail pour se concentrer sur les deux objets manipulés dans cette partie de la chaîne de traitement, le document Word et le document XML. Cette section fait un survol du processus de la soumission des projets d'articles aux revues jusqu'aux dernières corrections d'épreuves. Ce survol effectué, il ne restera qu'à intégrer à la chaîne de traitement la création des métadonnées dans les documents, ce dont nous parlerons plus loin.

De façon générale, un projet de publication s'enclenche par la soumission, normalement en format électronique, d'un article par un auteur au comité éditorial d'une revue. Ce comité, voire le ou la secrétaire de rédaction, peut demander des ajustements si le protocole de rédaction n'a pas été respecté. Dans ce cas, l'article retourne à l'auteur puis reviendra au comité.

Par la suite, l'article est acheminé à quelques personnes qui forment le comité de lecture, dans le but d'approuver ou de rejeter la publication de l'article. Ces évaluateurs font des

³⁰ Depuis la version 2000 des outils Office de Microsoft, il est possible de sauvegarder un document en format XML. Toutefois, ce format XML est difficile à manipuler, car il contient beaucoup d'information de présentation qui nous sont inutiles ainsi que de la logique. C'est pourquoi il est préférable de passer par d'autres outils.

³¹ Produit de la société TetraSys, voir <http://www.tetrasys.fr/majix.html/>.

³² Produit de la société Infinity-Loop, voir <http://www.infinity-loop.de/en/prodUpcast.html/>.

commentaires et suggèrent des modifications avant de retourner l'article au comité de la revue. Le comité éditorial et le rédacteur en chef de la revue déterminent, selon le projet éditorial de la revue et à la lumière des commentaires et évaluations des pairs, quels articles seront publiés. Une fois le choix fait, les auteurs reçoivent les commentaires, questions et demandes de modifications de leur article.

Une fois les modifications effectuées et le protocole de rédaction de la revue appliqué, le contenu de l'article est considéré comme final et l'article est prêt à entrer dans le circuit de fabrication. C'est précisément ici que le portail interviendrait. Des épreuves sont alors produites et acheminées au secrétariat de la revue ainsi qu'à l'auteur. S'il y a lieu, des corrections sont apportées aux épreuves et l'article est finalement publié.

Ce processus traditionnel a fait ses preuves et fait l'objet de conventions dans la communauté universitaire. Dans l'état actuel des choses, on doit l'adapter et le rendre plus efficace, notamment en utilisant l'informatique et les réseaux pour échanger les différents documents (ce que nous n'aborderons pas ici), mais aussi en accélérant la production des épreuves. De plus, ce processus doit être adapté pour permettre la correction non seulement des épreuves destinées à l'impression, mais également celles destinées à la consultation en ligne.

La chaîne de traitement proposée adaptera ce processus de la façon suivante. Tout d'abord, aucun changement n'est à prévoir entre le moment où l'auteur soumet l'article et le moment où le comité éditorial détermine que l'article est prêt à publier. Ainsi, tout ce travail est effectué à l'extérieur du portail, par les responsables de la revue, les auteurs et les évaluateurs. Ces acteurs travaillent tous sur le document Word.

Par la suite, une version Word du document est produite en respectant un modèle établi spécifiquement pour permettre une publication électronique sur plusieurs supports. La production de cette version « Word++ » devrait être effectuée par les secrétariats des revues ou éventuellement par le portail pour les revues où cette étape est difficile à intégrer. La création de la version « Word++ » est une tâche manuelle et intellectuelle qui consiste à baliser le document pour y ajouter de la valeur en identifiant la structure pour qu'elle soit manipulable par des outils informatiques.

Une fois cette version « Word++ » créée, une version XML est obtenue *automatiquement*. Le reste du processus consiste à créer des épreuves, imprimées et électroniques, des articles et à vérifier ces épreuves pour identifier et corriger les dernières erreurs. Les épreuves sont produites *automatiquement*, avec les différents outils de la chaîne de traitement qui seront présentés plus loin. Soulignons toutefois que dans le cas des épreuves imprimées, c'est un document PDF qui sera utilisé et transmis aux personnes qui effectueront la vérification. La saisie des dernières corrections dans la version « Word++ » terminée, le document XML de référence sera produit et, de là, toutes les versions de diffusion.

Le point le plus important de cette partie de la chaîne consiste en l'équivalence (à court terme) des versions « Word++ » et XML des articles, équivalence obtenue par la réalisation d'un puissant outil de conversion et qui permet de tirer partie des forces de chacun des formats, soit l'utilisation presque généralisée dans le cas de Word et la pérennité et la facilité de manipulation du format XML. La figure 3 illustre cette partie de la chaîne de traitement.

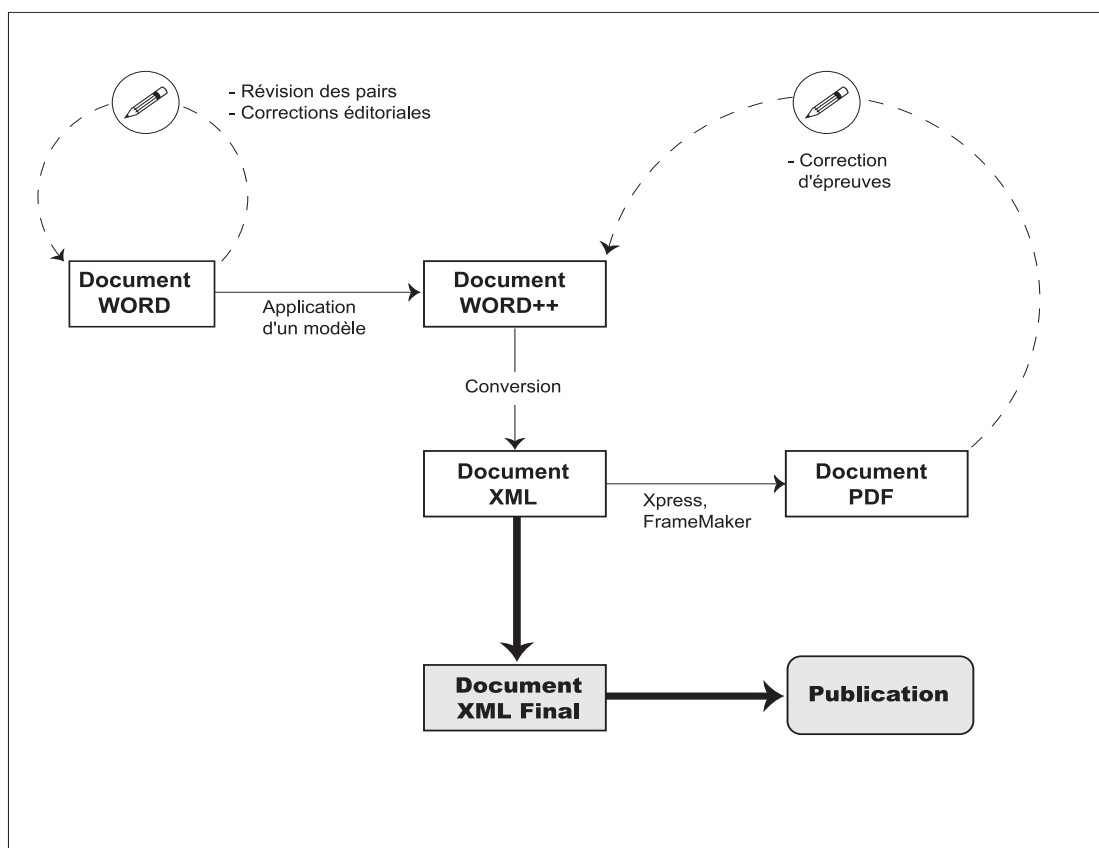


Figure 3 – Production du document XML

LA NUMÉRISATION RÉTROSPECTIVE

Les articles de revues déjà publiés pourraient sans problème être intégrés dans une telle chaîne de traitement. Les articles dont nous possédons la source électronique devraient être convertis au format « Word++ » et suivre le reste de la chaîne. Les articles dont nous ne possédons pas la source doivent être numérisés et traités par la reconnaissance optique de caractères, ou encore ressaisis, puis amenés dans le format « Word++ » et suivre le reste de la chaîne.

Les efforts à consentir sont très grands et les implications économiques de cette dernière solution ne permettent peut-être pas de faire de grands projets de numérisation rétrospective et d'application de la chaîne de traitement compte tenu des priorités. Par exemple, gardons à

l'esprit que la numérisation rétrospective en format image coûte entre 30 et 40¢ la page et que le coût de la numérisation + ROC se situe entre 3 et 4 \$ la page. Pour cette dernière formule, il faut ajouter les coûts de traitement pour la vérification des fichiers (coquilles ROC) et la conversion vers le XML. Pour cette seule considération, on peut être porté à utiliser une approche différente, qui consiste à diffuser en mode image les articles déjà publiés, dans un format PDF par exemple, permettant ainsi aux lecteurs de les consulter assez aisément. Cette solution temporaire doit être implantée en pensant à l'avenir, afin de rendre possibles de futurs traitements en reconnaissance optique de caractères, voire de conversion vers le XML. Il est donc important de s'assurer de la qualité de la numérisation et de la production d'un format image reconnu pour l'archivage à long terme, le format TIFF par exemple, qui rendra possible aisément la récupération des fichiers pour les traitements futurs.

Un format image, comme nous l'avons vu plus haut, ne permet toutefois pas la recherche plein texte. Des métadonnées doivent donc être associées aux articles ainsi numérisés pour permettre l'intégration des collections courante et rétrospective par l'outil de recherche et l'interface de navigation. L'ajout de ces métadonnées constitue toutefois une tâche manuelle qui peut être fastidieuse, même s'il est probablement possible de récupérer une partie des métadonnées depuis une base de données bibliographiques.

LES PRÉPUBLICATIONS

La notion de prépublication peut connaître plusieurs usages. Il s'agit, au départ, d'un document qui n'a pas été arbitré par des pairs, contrairement aux procédures suivies par les revues savantes. Certains auteurs parlent plutôt de solutions d'auto-archivage et de diffusion mises en place par les auteurs³³ ou de *Skypwriting*³⁴, faisant ici référence à la libre circulation des résultats de la recherche pour toutes les personnes ayant accès au Web.

Agissant *a priori* comme véhicules informels de communication de prépublications ou d'autres types de littératures grises non arbitrées, quelques serveurs de prépublications sont devenus, pour les communautés de chercheurs de certaines disciplines, des lieux incontournables et essentiels pour la diffusion des résultats de la recherche³⁵. On ne saurait parler de ce nouveau mode de communication des résultats de la recherche sans souligner le succès de *arXiv*³⁶, créé par Paul Ginsparg en 1991, au *Los Alamos National Laboratory*. Ce serveur de prépublications, principalement en physique, reçoit des millions d'accès par mois. Les succès en boucle de la mise en ligne de centaines de serveurs de prépublications depuis 1991, ainsi que la coordination entre ces différents sites qui s'organise avec notamment l'*Open Archives initiative*³⁷, font penser à plusieurs que la généralisation de ce mode de communication est la solution à la crise des revues savantes.

³³ *Author self-archiving solutions*, Herbert Van de Sompel, Carl Lagoze. 2000. « The Santa Fe Convention of the Open Archives Initiative. » *D-Lib Magazine*, vol. 6, no 2. <http://www.dlib.org/february00/vandesompel-oai/02vandesompel-oai.html>

³⁴ Stevan Harnad, 1999. « Free at Last: The Future of Peer-Reviewed Journals. » *D-Lib Magazine*, vol. 5, no 12. <http://www.dlib.org/dlib/december99/12harnad.html>

³⁵ Herbert Van de Sompel, Carl Lagoze, 2000. « The Santa Fe Convention of the Open Archives Initiative. » *D-Lib Magazine* vol. 6, no 2. <http://www.dlib.org/february00/vandesompel-oai/02vandesompel-oai.html>

³⁶ <http://arXiv.org/>

³⁷ <http://www.openarchives.org/>

OPEN ARCHIVES INITIATIVE

L'objectif de l'*Open Archives Initiative*³⁸ (OAi) est de contribuer de façon concrète à la transformation de la communication savante. Depuis la première rencontre des participants à l'OAi à Santa Fe en octobre 1999, plusieurs outils sont en développement pour arriver à réaliser l'interopérabilité entre les différents serveurs de prépublications. En effet, les serveurs de prépublications ayant des pratiques et des modes de fonctionnement très différents, les membres de l'OAi voient, dans l'application de solutions concrètes pour réaliser l'interopérabilité entre les différents serveurs, le moyen de promouvoir ce nouveau mode de communication des résultats de la recherche. La Convention de Santa Fe recommande trois mécanismes pour permettre l'interopérabilité entre les serveurs : 1° la création et l'adoption d'un modèle simple de métadonnées appelé *Open Archives Metadata Set* (OAMS) qui permettra de décrire les documents, 2° l'utilisation du XML pour l'encodage et l'échange de ces métadonnées, et 3° la définition d'un protocole commun pour réaliser l'extraction des métadonnées OAMS des serveurs participant au projet. L'objectif étant de permettre l'accès simultané à toutes les prépublications des serveurs reliés à l'OAi par une même interrogation.

DES PRÉPUBLICATIONS ET DES REVUES

Malgré la reconnaissance du grand intérêt des serveurs de prépublications, à part l'exemple de Los Alamos pour la communauté des chercheurs en physique, « les serveurs des autres disciplines ne contiennent qu'une minuscule portion du corpus annuel d'articles publiés³⁹ » (notre traduction). De plus, on ne constate aucun impact négatif depuis la création de Los Alamos sur la revue *Physical Review* dont les champs d'intérêts recoupent largement ceux des prépublications déposées sur le serveur⁴⁰. C'est dire que les auteurs valorisent les revues savantes comme mode de validation et de reconnaissance de leurs travaux qui va au-delà de la communication entre chercheurs⁴¹. Car, à ce jour, une prépublication n'est pas une publication, notamment en ce qui a trait à la reconnaissance des auteurs par les pairs et à l'importance des publications arbitrées dans le processus de promotion des professeurs et des chercheurs.

On peut estimer que nous assistons à la transition qui va du modèle actuel de communication scientifique, qui s'appuie sur la revue, vers un modèle ayant pour centre un serveur de prépublications avec ou sans arbitrage. On peut se demander aussi si la véritable question n'est pas plutôt, certes à court terme, mais peut-être aussi à plus long terme, celle d'une cohabitation des prépublications et des revues. Ce n'est souvent pas parce qu'une nouvelle technologie ou une nouvelle forme de communication émerge qu'elle doit nécessairement remplacer celle qui était utilisée auparavant. La communication de messages sonores ou textuels par satellite entre les capitaines des bateaux a eu raison du code Morse mais on utilise aujourd'hui le téléphone et la messagerie vocale sans penser que le courriel seul suffit.

³⁸ Pour davantage d'informations sur l'*Open Archives Initiative*, consulter l'article suivant : Herbert Van de Sompel, Carl Lagoze (2000). « The Santa Fe Convention of the Open Archives Initiative. » *D-Lib Magazine* 6 (2). <http://www.dlib.org/february00/vandesompel-oai/02vandesompel-oai.html>

³⁹ Stevan Harnad, 1999. « Free at Last : The Future of Peer-Reviewed Journals. » *D-Lib Magazine*, vol. 5, no 12. <http://www.dlib.org/dlib/december99/12harnad.html>

⁴⁰ Arthur P. Smith 2000. « The journal as an overlay on preprint databases » *Learned Publishing*, vol. 13, p.43-48.

⁴¹ Sur cette question, voir l'article de Rowland (<http://www.ariadne.ac.uk/issue7/fytton/intro.html>) où on identifie quatre fonctions fondamentales de la revue savante : communication de l'information, contrôle de la qualité, l'archivage canonique et la reconnaissance par les pairs.

UN APERÇU DES SERVEURS DE PRÉPUBLICATIONS

La revue des serveurs de prépublications permet d'observer certaines tendances qu'il faut considérer dans l'élaboration d'un portail de diffusion de l'information scientifique. Ces tendances touchent les services et types de documents diffusés sur ces sites, les domaines couverts et les normes utilisées pour la gestion de ce type de document. Bien que la très grande majorité de sites de prépublications soit dans le domaine des sciences, avec en tête la physique et les mathématiques, plusieurs sites de prépublications en sciences humaines et sociales existent depuis plus ou moins longtemps. La psychologie, l'économie et la philosophie sont les disciplines des sciences humaines et sociales qui semblent utiliser le plus les possibilités de ce mode de communication.

La majorité de ces serveurs ne distribuent que des prépublications. Toutefois, certains diffusent également des thèses et des articles.⁴² Il n'existe aucune norme quant aux formats d'encodage des prépublications. Les serveurs de prépublications donnent généralement des directives assez précises aux auteurs⁴³. Les formats de fichier les plus fréquemment demandés sont : PDF, Postscript, Tex/LaTeX et HTML. Tous les sites de prépublications sont à accès libre pour les lecteurs. Certains sites, tels *Los Alamos*⁴⁴, *RePEc*⁴⁵ et *Cogprints*⁴⁶, demandent aux auteurs de s'enregistrer pour obtenir un code d'utilisateur et un mot de passe afin de déposer des prépublications sur le serveur.

Dans la majorité des cas, les prépublications se trouvent sur le même serveur ou dans le même environnement que le site Web d'entrée. Toutefois, certains projets tels *GrayLit Network*⁴⁷ et *MPRESS*⁴⁸ font plutôt office de portail fédérateur pour un ensemble de serveurs de prépublications. Les services offerts permettent alors de chercher à travers plusieurs bases de données qui se trouvent sur plusieurs sites. En plus de l'utilisation d'un outil de recherche ou d'un plan de classement disciplinaire, plusieurs sites offrent la possibilité de s'inscrire à une liste de diffusion qui signale les dernières prépublications mises en ligne.

LES PRÉPUBLICATIONS ET LE PORTAIL QUÉBÉCOIS

Compte tenu de l'intérêt que suscitent les serveurs de prépublications pour plusieurs communautés de chercheurs dans diverses disciplines, il apparaît fondamental d'accorder une attention particulière et une place à ce mode de communication dans la conception et la réalisation du portail. Les dimensions que nous avons traitées portent sur le type de documents que l'on pourrait déposer et recueillir sur le serveur, les implications techniques pour l'intégration des prépublications au portail, les développements à suivre dans ce milieu, ainsi que la nécessité de connaître les besoins et l'intérêt des chercheurs québécois pour ce nouveau mode de communication.

⁴² Voir par exemple *l'Index national des prépublications de mathématique en France*. <http://www-mathdoc.ujf-grenoble.fr/prepub.html/>

⁴³ Voir par exemple *CH Working Papers : Submission of papers*. <http://www.kcl.ac.uk/humanities/cch/chwp/submit.html> et *Guidelines for Submission to the Education-line database*. <http://www.leeds.ac.uk/educol/guide.htm/>

⁴⁴ <http://xxx.lanl.gov/>

⁴⁵ <http://www.repec.org/>

⁴⁶ <http://cogprints.soton.ac.uk/>

⁴⁷ <http://graylit.osti.gov/>

⁴⁸ <http://mathnet.preprints.org/>

Le portail devrait devenir le point d'entrée à toutes les formes de communication savante. Malgré une nette démarcation des différentes zones qui composeront le portail, les lecteurs auront accès à des outils qui permettront de chercher par sujets à travers des collections comprenant tout autant des articles publiés sous forme papier qu'on retrouve en version électronique sur le portail que des thèses ou encore des notes de recherche. Le critère principal de recherche est ainsi le sujet, sans limite quant au type de documents.

LES DOCUMENTS DE LA ZONE PRÉPUBLICATIONS

Les prépublications rejoignant les mêmes auteurs et lecteurs que ceux des revues savantes, nous proposons qu'une section du portail soit construite pour recevoir, archiver et diffuser des prépublications, certes, mais également d'autres types de documents que l'on qualifie de littérature grise. Les prépublications, quant à elles, sont le plus souvent des articles destinés à une publication dans un autre média et pour lesquels la diffusion rapide permettra à la communauté des chercheurs d'une discipline de contribuer de façon plus active et efficace aux développements de travaux de recherche. Cette partie du portail pourrait également donner accès à des thèses, à des notes de recherches, à des rapports, à des textes et à du matériel de conférence, à des discours, à des articles non destinés à la publication sur un autre média, et autres documents qui pourraient intéresser la communauté de chercheurs d'une discipline visée⁴⁹.

Le portail pourrait se porter garant de la conservation⁵⁰ et de la diffusion des documents qui lui seraient acheminés. De plus, différents moyens techniques pourraient permettre d'intégrer des collections qui ne se trouvent pas nécessairement sur le serveur du portail. Par exemple, on peut penser que l'outil de recherche du portail permettra de chercher dans la collection des thèses québécoises accessibles dans le Web. De la même façon, d'autres collections de prépublications rattachées tantôt à des centres de recherche ou à d'autres serveurs pourraient être accessibles à partir du portail. Cette zone du portail serait, comme la zone « revues savantes », reliée à d'autres collections se trouvant sur d'autres portails.

Comme nous l'avons indiqué plus haut, les outils techniques permettant l'interopérabilité des différents serveurs de prépublications sont en voie de réalisation à la suite des travaux de la Convention de Santa Fe de l'OAi. Il sera donc d'une importance capitale pour le portail de suivre les développements de ces travaux de façon à développer des outils et des protocoles compatibles avec ceux de l'OAi, notamment le modèle de métadonnées à adopter.

IMPLICATIONS TECHNIQUES

Une fois que l'on dispose de l'espace disque requis pour le dépôt et la diffusion des prépublications, les implications techniques du développement d'une zone de prépublications pour le portail sont d'abord reliées à la sécurité. Le dépôt par qui le veut bien d'un fichier sur un serveur implique nécessairement des mesures de sécurité très strictes et le contrôle quotidien des répertoires ouverts au public. De plus, comme les fichiers déposés proviennent

⁴⁹ Mentionnons également au passage un autre type de document appelé « article expérimental » que l'on retrouve notamment sur le site des *CHWP*. Un article expérimental est une publication en ligne sans comité de lecture qui exploite les propriétés d'Internet et qui représente soit une recherche en cours, soit un instantané d'une idée en voie de développement. Pour plus de détails, voir <http://www.kcl.ac.uk/humanities/cch/chwp/titles.html>

⁵⁰ Voir dans le chapitre 5 la distinction entre *conservation* et *archivage*.

d'une foule de sources différentes et qu'habituellement plusieurs formats d'encodage sont acceptés, des mesures particulières pour la détection des virus doivent être mises en place pour des raisons évidentes.

Ensuite, compte tenu du coût de conversion des documents vers un format structuré, il est absolument impensable qu'un traitement vers XML soit appliqué à tous les documents déposés sur un serveur de prépublications. L'investissement que nécessite le traitement doit nécessairement être justifié par la reconnaissance du document dans un milieu donné.

Toutefois, de façon à permettre l'accès aux différentes zones du portail par un même outil de recherche, des métadonnées doivent être créées. Les métadonnées pourraient être créées par les auteurs eux-mêmes au moment du dépôt de leur document en remplissant un formulaire Web où les informations concernant entre autres le nom de l'auteur, son affiliation, le titre du document, le résumé et les mots-clés seraient saisis. L'équipe du portail aura donc à définir un modèle de métadonnées et à développer l'interface de saisie pour les auteurs.

Toutes les opérations, tels le dépôt et la mise en ligne des textes, le traitement et l'intégration à la base des métadonnées, devront s'effectuer automatiquement, sans intervention humaine. Il faut atteindre un taux de 100% d'automatisation pour l'ensemble des opérations techniques du système. À cet effet, les auteurs seront incités à utiliser les meilleures pratiques, non seulement pour la saisie de leurs métadonnées mais également pour la présentation de leurs textes. Des directives précises devront être transmises aux auteurs.

BESOINS ET INTÉRÊT DES CHERCHEURS QUÉBÉCOIS POUR UN SERVEUR DE PRÉPUBLICATIONS

Autant il apparaît essentiel d'intégrer les prépublications dans l'architecture générale du portail dès le moment de sa conception, autant il est capital de mieux connaître les besoins et l'intérêt des chercheurs québécois pour la création d'un tel serveur.

Contrairement aux comportements et activités des chercheurs québécois vis-à-vis leurs revues, que nous connaissons assez bien, nous avons très peu d'informations sur leurs habitudes en ce qui a trait aux serveurs de prépublications existants. On sait déjà que plusieurs d'entre eux utilisent ces moyens pour communiquer rapidement leurs résultats de recherche. Toutefois, nous avons peu d'informations sur la place que prend ce moyen par rapport aux autres modes de communication et, particulièrement, de publication des articles produits dans des revues. De plus, certains sites de prépublications⁵¹ créent un comité d'arbitrage des documents soumis. Est-ce le genre de sites qui serait privilégié? Cette possibilité doit également être soumise aux chercheurs pour évaluer l'intérêt de mettre en place une pareille structure d'évaluation des documents.

Sur ces aspects, cités à titre d'exemples, il convient d'ouvrir une discussion plus large pour connaître l'ordre de priorité que l'on doit accorder à ce service, sa nature et sa place dans l'infrastructure de la communication savante. Cet exercice permettrait de définir une action qui correspond au mieux aux attentes du milieu.

⁵¹ Voir par exemple les *CHWP* <http://www.kcl.ac.uk/humanities/cch/chwp/> .

LA PRODUCTION DES FORMATS DE DIFFUSION DE L'INFORMATION

La chaîne de production permet l'utilisation de plusieurs formats de diffusion, chacun ayant ses avantages et ses inconvénients. Les formats retenus pour l'avenir prévisible ne sont pas limitatifs : d'autres formats pourraient s'ajouter pour répondre à d'autres besoins ou pour s'adapter à de nouvelles technologies. La richesse des documents sources (documents de production) permettra de produire ces formats assez facilement. Après avoir défini les formats au début de cette section, nous verrons ici comment il sera possible de les convertir des documents XML.

Les formats de diffusion sont tous produits à partir d'une seule source, soit le document XML de référence. Les différentes étapes de conversion, de même que les outils préconisés pour les réaliser, sont identifiées dans la figure 4.

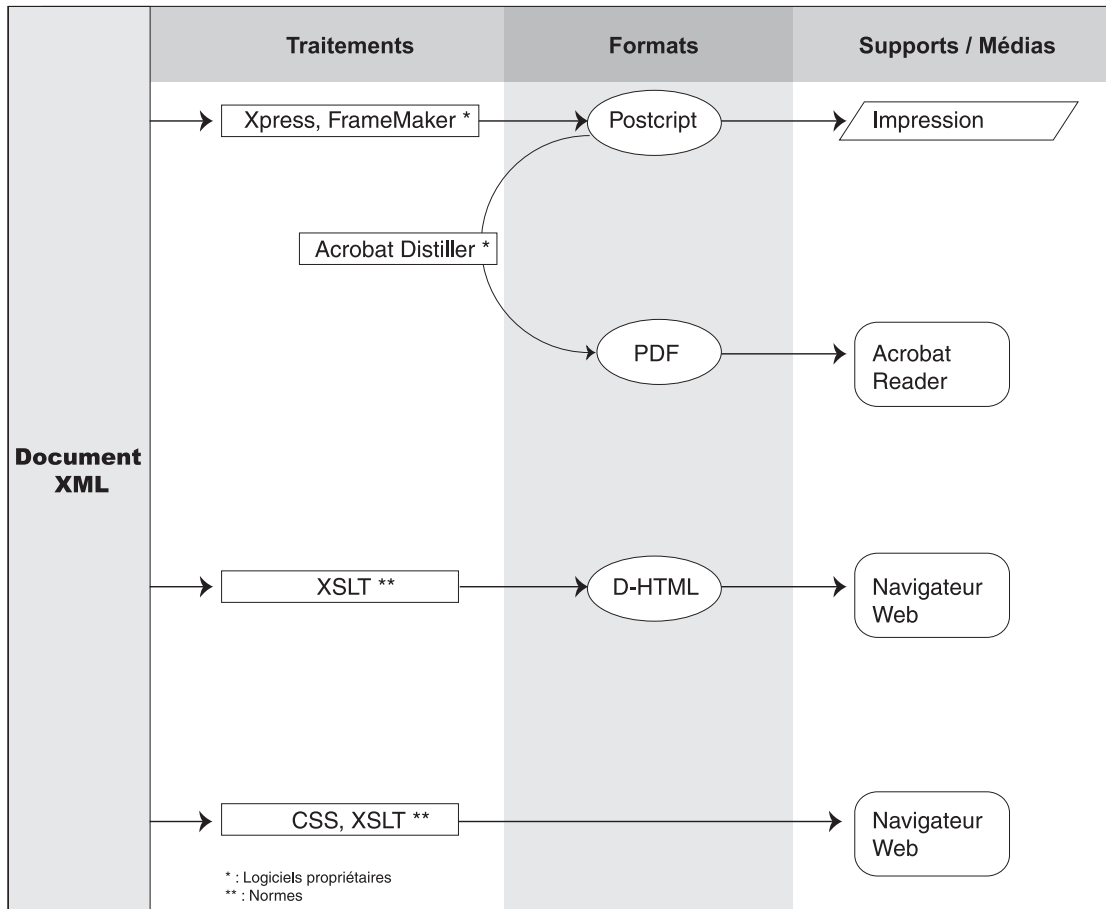


Figure 4 – Production des formats de diffusion

LA CONSULTATION À L'ÉCRAN

On peut affirmer que la consultation des articles à l'écran n'est pas toujours d'un grand confort et que, comparativement à l'imprimé, la vitesse de lecture est réduite et la capacité de rétention de l'information est plus faible⁵². Mais ce serait faire peu de cas du fait que l'ordinateur peut fournir des outils très utiles et performants pour l'exploitation des informations disponibles dans le Web. Cela est d'autant plus vrai lorsque l'environnement de consultation offre des aides à la lecture comme des annotations personnelles ou partagées ou encore des signets.

De plus, cette consultation à l'écran est en quelque sorte indépendante de l'encodage initial du document en format XML. En effet, à partir d'articles de revues encodés en XML selon le même schéma, il est possible d'obtenir des présentations visuelles très variées pour les articles des différentes revues. Ainsi, cette consultation à l'écran permet aussi de respecter les chartes graphiques des revues, et même de les mettre en valeur en utilisant les avantages de l'électronique.

Le format HTML — plus précisément le *Dynamic HTML* (DHTML) — constitue, encore pour quelques années au moins, le format de choix pour la diffusion d'information électronique. Judicieusement utilisé, il permet une mise en page intéressante et efficace, un dynamisme utile ou agréable, et une efficacité de transfert et d'affichage. La conversion des documents XML vers un format DHTML ne pose aucun problème : plusieurs outils permettent de l'effectuer automatiquement, moyennant un effort initial de programmation.

Depuis novembre 1999, la norme XSLT, issue du W3C, est un langage de choix pour de telles transformations, en particulier dans le contexte d'un portail de revues savantes. Plusieurs raisons militent en ce sens. D'abord, les processeurs XSLT sont particulièrement efficaces pour des documents de taille petite ou moyenne, ce qui est tout à fait approprié pour les articles de revues. Cette norme, de plus, facilite la création de feuilles de styles modulaires qui permettent de créer plusieurs variations d'une même transformation. Ce dernier aspect est important, car pour offrir un grand confort de navigation aux différents utilisateurs, il y a probablement lieu de préconiser la production de plus d'une version DHTML des articles. Ensuite, d'une revue à l'autre, la mise en page des articles pourra varier mais, fondamentalement, de nombreux aspects de la conversion seront identiques. Cela fait en sorte que la mise en place de cette partie de la chaîne de traitement demandera la création d'un grand nombre de conversions légèrement différentes, d'où l'importance d'une approche modulaire et par conséquent d'une technologie qui le permette. Ajoutons à ces facteurs favorisant le choix du XSLT le fait qu'il s'agit d'une norme.

Les conversions au format HTML peuvent être effectuées de façon dynamique, c'est-à-dire à chaque fois qu'un article est demandé sur le serveur de diffusion. Elles peuvent aussi être faites d'avance. La première approche minimise les besoins en espace disque, mais demande une plus grande puissance du serveur. La deuxième approche inverse ces besoins. Puisque les articles, une fois publiés, ne sont pas modifiés, et que le coût de l'espace disque est peu élevé, le prétraitement des articles semble l'avenue la plus appropriée. Toutefois, si on veut offrir aux utilisateurs une représentation HTML qui dépend de certains paramètres inconnus à l'avance,

⁵² Voir à ce sujet : Jeff Grabmeier, « Texts on Computer Screens Harder to Understand, Less Persuasive », *Research News*, août 2000. Adresse URL: <http://www.acs.ohio-state.edu/units/research/archive/comptext.htm>

une approche dynamique sera nécessaire. Par exemple, un utilisateur pourrait saisir un ou plusieurs mots, et on pourrait lui présenter un document où ces mots sont mis en évidence dans le texte, afin de faciliter la lecture. Ou encore, si le serveur offre des outils d'annotation publique et privée, la présentation des documents dépendra des annotations et des préférences d'affichage, ce qui exclut le prétraitement des articles.

Concurremment, la diffusion des documents XML directement dans le Web est possible, et ce, de deux façons. D'abord, on peut associer à ces documents une feuille de style CSS. Pour l'instant, les navigateurs Internet Explorer, Opera et Netscape — encore en version *beta* — permettent de les afficher. Cette approche est peu recommandée en raison de ses limites : les feuilles de styles CSS permettent de mettre en page correctement un document XML, mais offrent peu d'intérêt en matière de dynamisme des pages ou de mise en contexte de l'information dans l'ensemble d'un site. L'autre approche consiste à associer une feuille de style XSLT au document et, dans ce cas, la transformation s'effectue sur le poste client. Pour l'instant, seul le navigateur Internet Explorer offre cette fonctionnalité, mais il est probable que le navigateur Netscape 6 la permettra, même si cette dernière n'est pas présente dans la version *beta* qui est actuellement disponible. À moyen terme, il s'agit d'une stratégie intéressante, pour minimiser à la fois les traitements du côté du serveur et l'espace disque requis.

L'IMPRESSION À DISTANCE

L'impression à distance consiste à fournir aux lecteurs, habituellement via Internet, un document mis en page et destiné à être imprimé sur leur imprimante personnelle. Cette approche permet d'éviter la lourdeur de l'impression professionnelle et de répondre à une diversité de besoins. De plus, les lecteurs peuvent imprimer seulement les pages ou les articles dont ils ont besoin, ce qui évite une consommation démesurée de ressources.

Les logiciels de consultation de documents électroniques permettent normalement d'imprimer les documents consultés. Cette impression n'est pas de bonne qualité, car les documents n'ont pas nécessairement été préparés pour l'impression. C'est le cas, entre autres, des documents en format HTML. Le meilleur format d'impression à distance reste le format PDF d'Adobe. C'est pour cette raison que le portail de revues produira et diffusera des versions PDF des articles. La production de ces documents implique les mêmes opérations que la production des documents Postscript (que nous allons voir dans la rubrique suivante). Puisque l'impression professionnelle est encore d'actualité, il est tout à fait logique d'utiliser le produit intermédiaire de l'impression professionnelle, soit le document Postscript, pour obtenir un document PDF. Cette conversion est automatique et ne pose pas de problème particulier.

Par conséquent, la production des versions pour l'impression professionnelle et l'impression à distance est semblable ; seule une étape automatique s'ajoute pour obtenir des documents PDF à partir des documents Postscript. Notons au passage que lors de cette conversion, des ajustements peuvent être effectués pour, notamment, réduire la qualité des images, évitant ainsi de trop grands temps de téléchargement des articles.

L'IMPRESSION PROFESSIONNELLE

L'impression professionnelle permet d'obtenir des articles sur support papier, avec une très bonne qualité d'impression et habituellement reliés en numéros. Cette

impression demande une expertise et des équipements particuliers. C'est pourquoi cette tâche est habituellement effectuée par des sociétés spécialisées. Pour ce faire, on doit pouvoir fournir à l'imprimeur les documents ayant une mise en page adéquate. Le format Postscript⁵³ est bien souvent utilisé pour l'échange de documents destinés à l'impression.

La technique la plus intéressante pour produire des documents Postscript à partir des documents XML est d'utiliser un logiciel de mise en page. Il s'agit d'importer l'information (les articles) dans ce logiciel et de produire un document Postscript ; cela se fait sans problème. Mais en raison de la complexité de la mise en page pour l'impression, cette opération peut difficilement se réaliser de façon purement automatique, mais on peut s'en approcher. Par exemple, les logiciels Quark XPress⁵⁴ et FrameMaker⁵⁵ permettent d'importer de l'information dans un langage balisé (XPress Tags pour XPress, MIF pour FrameMaker) que l'on peut obtenir assez facilement par conversion d'un document XML. Il existe également des solutions entièrement automatisées, avec par exemple 3B2⁵⁶ ou Adept Publisher⁵⁷, mais les efforts de paramétrage et de programmation à faire sont très grands. Il existe une version de FrameMaker (FrameMaker+SGML) permettant d'importer directement des documents SGML, mais cette option n'est pas particulièrement intéressante pour notre chaîne de traitement basée sur XML, entre autres parce qu'il faut de toutes façons manipuler le XML pour le convertir en SGML et l'adapter aux exigences de l'impression.

Il est difficile de produire un document entièrement imprimé de façon automatique, sans vérification manuelle, mais la chaîne de traitement proposée permet de minimiser ces interventions humaines. Si bien que les premières épreuves pour l'impression se font presque automatiquement, avec seulement quelques petits ajustements manuels. Éventuellement, il sera possible de créer les épreuves papier avec la version XML et une feuille de styles CSS appropriée mais, pour l'instant, le support CSS est trop peu développé dans les logiciels pour utiliser cette approche.

LES MÉTADONNÉES

Nous venons d'identifier tous les éléments qui permettent de produire des articles. Or, comme on le soulignait plus haut, on ne peut prétendre *diffuser* une collection d'articles simplement en la rendant accessible dans le Web. C'est l'équivalent d'ouvrir les portes d'une immense bibliothèque sans fournir de catalogue : on se doute bien qu'elle contient des informations intéressantes mais aurions-nous le temps de les chercher dans ces conditions ? Le Web ressemble actuellement à cette « bibliothèque ». Le portail devra mettre en place plusieurs outils pour assurer la diffusion efficace et optimale des articles. L'un de ces outils s'appelle « métadonnées ».

Le processus traditionnel de la communication savante mobilise plusieurs acteurs : des auteurs produisant des textes, ces textes sont par la suite édités et publiés après que les éditeurs

⁵³ Ce format d'impression est fort populaire mais propriétaire (la compagnie Adobe le contrôle). De plus, le format PDF (propriété d'Adobe également) est de plus en plus utilisé pour les tâches d'impression. Ces deux formats sont relativement semblables.

⁵⁴ Produit de la société Quark, voir <http://www.quark.com/>.

⁵⁵ Produit d'Adobe, voir <http://www.adobe.com/>.

⁵⁶ Voir <http://www.3B2.com/>.

⁵⁷ Produit d'Arbortext, voir <http://www.arbortext.com/>.

leur aient conféré une importante valeur ajoutée — grâce au contrôle de la qualité, à la révision par les pairs et à leur diffusion —, les bibliothèques les achètent et apportent également une valeur ajoutée — qui permet l'identification, la sélection, le catalogage, l'organisation physique et la conservation du « contenant » qu'est ce périodique — et, enfin, des banques de données spécialisées indexent ces ressources pour permettre le repérage d'un contenu tout aussi spécialisé et offrent aux chercheurs des outils de consultation efficace des sources.

L'édition savante électronique transforme ce processus, notamment au niveau du repérage. L'exploitation des métadonnées appliquées à des articles de revue en version électronique permettra de tirer de grands avantages dans le repérage et la mise en réseau de sources documentaires diverses.

LES MÉTADONNÉES : RAISON D'ÊTRE ET USAGES

QUE SONT LES MÉTADONNÉES ?

L'analogie classique pour décrire les métadonnées, c'est la fiche de carton d'un catalogue de bibliothèque. Cette petite fiche servait de substitut au document original et permettait de faire un choix sans voir le document qu'elle décrivait. Les informations bibliographiques, les résumés, les termes d'indexation, les *abstracts*, tout ce qui peut être un substitut au document original et qui libère les usagers potentiels de la nécessité de connaître à l'avance l'existence et les caractéristiques de ce document sont des métadonnées.

Dans l'univers électronique, les métadonnées sont des renseignements, des informations, des ensembles d'énoncés à propos d'objets numériques qui peuvent être soit des articles, soit d'autres objets numériques insérés dans ces articles (par exemple des fichiers image ou son). Les métadonnées décrivent les attributs et le contenu de ces objets. Elles sont utiles au repérage mais également à l'administration, à la description, à l'accès et à la conservation de l'information. Elles identifient le contenu de l'objet ou son sujet, son contexte (le qui, le quoi, le pourquoi) et sa structure.

POURQUOI UTILISER DES MÉTADONNÉES ?

L'utilisation des métadonnées est rendue nécessaire par une simple raison de logique mathématique. En juin 1993, on comptait 130 sites Web⁵⁸, en août 2000, la société Netcraft en comptait 19,8 millions⁵⁹, soit une augmentation de 14 800 000 %!⁶⁰ Les chiffres sont encore plus impressionnants si on considère les documents électroniques présents sur ces sites : quelque 800 millions de documents publiquement accessibles par les robots de recherche dans le Web étaient recensés en 1999⁶¹. Michael Dahn, sans contester l'enquête de Lawrence et Giles, en vient à la conclusion qu'en novembre 1999 le Web publiquement *indexable* comprend 1,16 milliards de documents, mais que le Web publiquement *accessible* se chiffre plutôt entre

⁵⁸ Matthew Gray, MIT, *Internet Statistics; Growth and Usage of the Web and the Internet*. <http://www.mit.edu/people/mkgray/net/>.

⁵⁹ Netcraft, *August 2000 – Web Server Survey*. Adresse URL: <http://www.netcraft.com/survey/Reports/0008/>.

⁶⁰ De plus, les chiffres cités par Netcraft ne sont qu'indicatifs car les sites répertoriés ne sont que des adresses primaires (www.mon_site.com); ils ne comptent pas les hébergés à cette même enseigne (www.mon_site.com/mon_site) ni les pages dynamiques, de plus en plus présentes dans Internet.

⁶¹ Steve C. Lawrence et Giles Lee, «Accessibility of Information on the Web», *Nature*, vol. 400, 8 juil. 1999. <http://www.nature.com/>.

1,45 et 2,33 milliards de documents⁶². De leur côté, les producteurs du moteur de recherche Inktomi ont atteint leur milliardième document le 18 janvier 2000⁶³. Plusieurs motifs plaident pour le recours à des métadonnées dans cet univers à la fois riche et pléthorique.

Tout d'abord, l'accessibilité et l'utilisation accrues des documents électroniques, grâce notamment aux facilités de recherche, doivent être supportées de façon conséquente par les outils offerts aux utilisateurs, ce que permettent au premier chef les métadonnées. Elles améliorent aussi grandement la recherche d'information dans de multiples collections et permettent, par exemple, l'interopérabilité entre différents portails. De même, nous avons là la capacité de restituer le contexte de création d'une source électronique et de lier différentes pièces au même fonds d'archives, par exemple en indiquant, aujourd'hui et dans cent ans, que tel auteur est relié à tel article et à tel périodique. Il s'agit, de plus, d'un outil précieux pour diversifier les points d'accès à l'information, la présentation des résultats et les possibilités de manipulation de l'information, et ce, sans toucher à l'intégrité du document (il devient aisé de trier avec les métadonnées les articles par date de création, par thèmes, par critères linguistiques, intellectuels ou autres). Elles sont aussi précieuses pour garder la trace des différentes déclinaisons d'un même document (par exemple, en XML, HTML ou PDF) lorsque celui-ci a plusieurs versions. Ajoutons que, dans la problématique de la conservation et de la « migration de l'information », les métadonnées techniques et descriptives, qui décrivent le contexte de création du document, permettront de procéder avec toutes les informations nécessaires. Enfin, en considérant la question des droits et la dimension légale, on doit noter que, par les métadonnées, il est possible de garder la trace des diverses couches de droits attachés à l'information, à la protection des droits, aux restrictions de consultation, etc.

Même si l'utilisation des métadonnées peut être vue comme une procédure supplémentaire, avec des coûts inhérents, on doit reconnaître qu'il s'agit d'un instrument particulièrement précieux pour un portail de revue, tant pour les services offerts, pour les liens à entretenir avec d'autres portails que pour la gestion interne des documents produits et conservés. À cela, il faut ajouter une autre dimension, en s'appuyant sur la distinction entre le Web accessible et le Web indexable.

WEB ACCESSIBLE VS WEB INDEXABLE

En inventoriant le Web indexable, c'est-à-dire le Web librement accessible par les robots de recherche qui le sillonnent, on passe à côté des pages accessibles uniquement à l'aide d'un formulaire de recherche, les pages protégées par des mots de passe, les pages dans des répertoires d'où les robots sont exclus, les pages utilisant des cadres (*frames*) mal définis, les pages non HTML ou en texte ou en PDF, etc. Certains estiment que ces pages représentent la moitié du Web indexable. Cette « nuance » entre le Web accessible et indexable est capitale pour définir une caractéristique essentielle du portail.

Afin d'atteindre la plus grande diffusion possible, la « fiche descriptive » des articles (mentionnée dans le chapitre sur le système de diffusion) sera située dans une partie du portail accessible aux moteurs de recherche. Nous prévoyons mettre en place une structure de

⁶² Michael Dahn, «Counting Angels on a Pinhead : Critically Interpreting Web Size Estimates », Online, January/February 2000, p. 35-40. Voir plus loin le commentaire concernant Web accessible/Web indexable.

⁶³ Inktomi, News and Events. Adresse URL : <http://www.inktomi.com/new/press/billion.html/>.

répertoire permettant une navigation par « bouquinage » (*browsing*), comme on peut le faire sur les rayons d'une bibliothèque. Cette structure de répertoire faciliterait la sélection, dans un premier niveau, des titres de périodiques présents dans le portail. En cliquant sur un titre, on atteindrait les volumes (l'année), d'où l'on pourrait sélectionner un numéro, où serait placée la « fiche descriptive » de chaque article. Un hyperlien permettrait alors d'atteindre le texte intégral de l'article. En tout temps, un bouton permettrait d'accéder au moteur de recherche. Si on choisit l'accès par moteur de recherche, le système nous fournira, en réponse à nos requêtes, les fiches descriptives des articles correspondants.

Les métadonnées recueillies pour chaque article permettront de créer de la façon la plus automatisée possible la « fiche descriptive », du côté « Web indexable ». Elles détermineront également, du côté « Web accessible », les zones qui seront recherchables grâce à un puissant moteur de recherche qui indexera la structure XML et le texte intégral des articles.

LE TORRENT ET LE SAS

On estime que seulement 6 % du Web est à caractère scientifique ou éducatif ; la plus grande masse du contenu (83%) étant à caractère commercial⁶⁴. Dans ces conditions, le simple fait de rendre un texte savant disponible dans le Web équivaut aujourd'hui à verser un verre d'eau de plus dans l'océan.

De là, deux tendances lourdes se dégagent pour mettre un peu d'ordre dans ce chaos : la création de portails spécialisés et l'utilisation de métadonnées. On assiste de plus en plus à une spécialisation du Web. Bientôt, on retrouvera des Webs à l'intérieur du Web. Une de ces manifestations est la création de portails : portails de divertissement, portails de services financiers, portails de revues savantes, etc. Lorsque les portails scientifiques et l'infrastructure de navigation et d'interopérabilité entre ces portails seront implantés et consolidés, les chercheurs n'auront plus à naviguer sur tout l'océan Web ; on s'orientera vers la qualité plutôt que vers la quantité. À cet égard, les métadonnées sont le moyen pour assurer cette mise en ordre du Web.

LES MÉTADONNÉES : LES CRÉER, LES STOCKER

Dans un environnement de documents structurés, un certain nombre de métadonnées sont déjà présentes grâce à la structure des documents. Par exemple, le titre d'un article est déjà identifié comme tel par les balises XML. Il est possible d'extraire ces éléments d'information du contenu du texte pour générer des schémas normalisés de métadonnées, comme le Dublin Core. Il est également possible de récupérer des éléments de métadonnées se trouvant dans les « propriétés du document » ; dans ce cas une extraction est nécessaire. Tous les éléments de description bibliographique de l'article, de même que différents types d'indexation matière, devraient se retrouver dans les métadonnées.

Les métadonnées ne sont pas nécessairement toutes connues au même moment ; elles seront créées à divers moments dans la chaîne de production du document. Idéalement, il faut chercher à rapprocher le plus possible la création des métadonnées des créateurs de l'information ; le principe à suivre est d'associer les métadonnées le plus tôt possible, car elles

⁶⁴ Steve C. Lawrence et Giles Lee, «Accessibility of Information on the Web», *Nature*, vol. 400, 8 juil. 1999.
<http://www.nature.com>

peuvent être utiles dans la chaîne de traitement elle-même. Cela dit, en raison de la rigueur requise pour cette opération, nous privilégions, au stade actuel, de confier aux éditeurs de revues savantes plutôt qu'aux auteurs la création des métadonnées bibliographiques dans le document « Word++ » et dans les propriétés du document Word, et ce, au tout début du processus de mise en forme du document reçu des auteurs.

Le format XML nous permet d'associer des métadonnées très facilement, et de façon très structurée, aux documents qui utilisent ce format, et donc aux articles des revues du portail. Ainsi, toute métadonnée propre à l'article devrait s'y retrouver en premier lieu, par insertion, quitte à ce qu'elle soit reprise ailleurs dans le système, par souci d'efficacité.

Pour les articles rétrospectifs, les métadonnées devraient être associées au moment de la numérisation, à l'aide d'un traitement manuel ou encore par récupération semi-automatique des informations depuis des bases de données bibliographiques. Concernant les prépublications, les auteurs pourraient eux-mêmes fournir les métadonnées de base nécessaires à la gestion et la diffusion des articles, à l'aide de formulaires disponibles dans le Web par exemple.

LES MÉTADONNÉES : LES SUPERPOSER

Rien n'est simple en ce bas monde, fût-il virtuel. Il n'y a (malheureusement) pas un schéma universel de métadonnées. Toutes les fonctions des métadonnées, mentionnées plus haut, nécessitent un format spécialisé : par exemple, fonction bibliographique, Dublin Core ; fonction évaluation, PICS ; fonction contrôle de la privauté, P3P ; fonction sécurité, *digital signature*, etc. De plus, plusieurs disciplines spécialisées ont besoin de métadonnées disciplinaires, de thésaurus, de réseaux sémantiques, bref de systèmes de représentation de connaissances particuliers. Toutes ces couches de métadonnées s'ajouteront et se superposeront à l'objet numérique. Comment s'assurer que ces normes empilées seront compatibles entre elles et qu'il n'y aura pas de conflit entre les différents éléments de métadonnées ? Une des solutions mises de l'avant par le W3C est le RDF⁶⁵. De façon simplifiée, le cadre (*framework*) RDF est un réceptacle permettant de recevoir différents schémas de métadonnées. Le RDF repose sur une structure XML.

Une fois que toutes ces couches de métadonnées seront interopérables, que l'infrastructure de support logiciel sera mise en place et maîtrisée, ces métadonnées deviendront compréhensibles par machine (*machine understandable*). Alors, la découverte des ressources et la recherche d'information dans le Web pourront en grande partie être réalisés de façon automatique par des robots de recherche spécialisés dans chaque discipline et ainsi sera réalisé le « Web sémantique » annoncé par Tim Berners-Lee depuis quelques années.

Pour conclure, retenons que les métadonnées permettront d'offrir aux utilisateurs un outil de recherche performant. De plus, elles faciliteront les échanges d'information avec des partenaires extérieurs, tels que des services d'indexation ou d'autres portails similaires au nôtre. Enfin, elles seront essentielles pour la gestion des articles.

⁶⁵ Resource Description Framework, recommandation du W3C, voir <http://www.w3.org/RDF/>.

QUELQUES MODÈLES DE MÉTADONNÉES

On vient de le voir, différents modèles de métadonnées doivent être associés aux articles afin de rendre la collection cohérente et organisée, mais aussi pour faciliter les échanges et le repérage d'information. Un aperçu des principales métadonnées nécessaires et de leur origine est utile.

Les éditeurs utilisent, de plus en plus, un standard d'identification unique des articles, le DOI.⁶⁶ Cet identificateur permet entre autres de donner des adresses permanentes aux articles, ce qui facilite leur repérage et leur gestion. L'ajout d'un identificateur unique aux articles ouvre la porte à de nombreuses applications, y compris une navigation facilitée dans les références bibliographiques.

En effet, en apportant une réponse à la problématique des liens URL rompus (erreur 404), le DOI nous permet d'implanter électroniquement, et de façon pérenne, cette caractéristique fondamentale de l'édition savante, le recours aux citations. Le service CrossRef⁶⁷ a pour objectif de lier les références bibliographiques aux contenus des articles diffusés en ligne. On prévoit que d'ici la fin de l'an 2000, trois millions d'articles provenant de milliers de périodiques seront liés au moyen de CrossRef et que la croissance annuelle sera de 500 000 articles. Notons que les liens ne seront pas seulement entre des revues savantes mais pourront pointer vers des articles d'encyclopédie, des actes de colloque, des manuels scolaires, etc., ce qui permettra un enrichissement important de la lecture. CrossRef s'appuie fortement sur les DOI. L'utilisation des DOI et la participation à un service comme CrossRef s'imposent pour un portail de revues savantes ; nous proposons d'associer un DOI à chaque article.

Pour réaliser cette interopérabilité, le portail devra expédier un ensemble limité de métadonnées au serveur CrossRef, dans un format bien défini. Nos propres métadonnées seront suffisamment riches pour générer les éléments de métadonnées nécessités par CrossRef.

DUBLIN CORE⁶⁸

Dublin Core (DC) est le standard qui semble devoir s'imposer : il est le standard le plus avancé pour la description des ressources Internet. Élaboré depuis 1995 à Dublin en Ohio (siège de Online Computer Library Center), le développement du Dublin Core est assuré par le Dublin Core Directorate, supervisé par le OCLC Office of Research and Special Projects⁶⁹.

Le Dublin Core comprend 15 éléments de base pour décrire les ressources électroniques. Il est conçu autour de 5 principes fondamentaux, à savoir :

- Tous les éléments sont optionnels.
- Tous les éléments sont répétables.

⁶⁶ Digital Object Identifier, voir <http://www.doi.org/>

⁶⁷ Voir <http://www.crossref.org/faqs.htm>

⁶⁸ Pour plus de détails, voir : Susan Haigh, «Le projet de métadonnées Dublin Core», *Flash Réseau*, no 63, décembre 1999. Adresse URL : <http://www.nlc-bnc.ca/pubs/netnotes/fnotes63.htm>

⁶⁹ Cliff Morgan, «Journals metadata : information about content», *Learned Publishing*, Vol. 12, no. 3, July 1999, p. 192 ; Jean Hudgins et al., *Getting mileage out of Metadata, Application for the Library*, Chicago, American Library Association, 1999, p. 14.

- DC est extensible. DC est un plus petit dénominateur commun (DC Simple) mais permet aussi, si on le désire, d'avoir une description plus riche, au moyen de sous-éléments (DC Qualified)
- DC est multidisciplinaire.
- DC est international (plus de 20 langues actuellement).

Ce modèle de métadonnées est déjà utilisé dans le milieu de l'édition savante. L'éditeur John Wiley & Sons, qui publie chaque année plus de 35 000 articles dans plus de 400 revues, en a fait son standard. Ces revues sont codées en SGML. L'en-tête (*header material*) comprend des informations sur la revue, le volume et le numéro, mais on y retrouve aussi le titre de l'article, l'auteur et son affiliation, le résumé, des mots-clés, la date de réception de l'article, etc⁷⁰. Une base de données de métadonnées, comprenant tous les articles publiés, a été constituée. Pour les champs pertinents, tous les éléments SGML ont été traduits directement en métadonnées Dublin Core Simple, par exemple <ATL> = DC.Title.

Le portail de revues savantes devra adopter une pratique similaire ; les éléments *Dublin Core* seront obtenus en grande partie par extraction et conversion.

ET TOUTES LES AUTRES

La norme PICS⁷¹ (Platform for Internet Content Selection⁷²) a été conçue à l'origine pour bloquer l'accès des enfants aux contenus offensants dans Internet. Toutefois, PICS permet aussi d'annoter nos articles au moyen de métadonnées qui seront utilisés par les moteurs de recherche pour déterminer, entre autres, le caractère académique de nos ressources. Ces normes sont conçues pour faciliter la recherche automatique d'informations dans le Web et permettront de sélectionner plus facilement l'information savante.

Toujours dans un cadre RDF, il sera possible d'insérer des métadonnées qui informeront les lecteurs des pratiques en matière de protection des renseignements personnels grâce au projet « Platform for Privacy Preferences » (P3P⁷³). Par exemple, dans une transaction électronique, il sera possible d'informer les clients que leurs informations nominatives ne seront pas transmises à des tiers. Ces échanges d'information se feront automatiquement et de façon transparente entre le navigateur du client et le serveur Web. Cette norme est essentielle au développement de la confiance en matière de commerce électronique (« Web of trust »). Le portail devra suivre l'évolution de ce projet et éventuellement le mettre en pratique.

IMPLICATIONS POUR LES REVUES

La mise en place de la chaîne de production technique suppose que le portail et les équipes de revue travaillent en phase. L'introduction de la publication électronique a nécessairement des implications pour les revues qui sont de deux ordres : elles concernent le

⁷⁰ Cliff Morgan, «Journals metadata : information about content», *Learned Publishing*, Vol. 12, no. 3, July 1999, p. 194.

⁷¹ Pour plus de détails, voir : Terry Kuny. «Filtrage du contenu d'Internet : plate-forme PICS, étiquettes et filtres». Flash Réseau no53, 23 mars 1998. Adresse URL : <http://www.nlc-bnc.ca/pubs/netnotes/fnotes53.htm/>.

⁷² <http://www.w3.org/PICS/>.

⁷³ <http://www.w3.org/P3P/Overview.html/>.

travail éditorial et de préparation de la copie et elles ouvrent des perspectives nouvelles sur le contenu et les formes d'existence de la revue.

LA PRÉPARATION DE LA COPIE

Les équipes de la revue poursuivent les mêmes activités éditoriales et ont la responsabilité de livrer au portail une copie, dont le contenu est prêt à publier. Pour que l'automatisation des procédés produise ses bénéfices, il faut pouvoir compter sur des textes dont le protocole de présentation est appliqué de façon rigoureuse. De plus, l'utilisation du langage structuré demande à ce que les textes soit préparés avec l'application d'une feuille de style. Si ce n'est pas la revue qui fait cette dernière préparation, le portail devrait s'en charger, mais il est de beaucoup préférable que la préparation de la copie se situe le plus près possible des auteurs et de la direction éditoriale, et donc de l'équipe de la revue. Il s'agit de tâches assez simples qui demandent un savoir-faire facile à acquérir dans le cadre d'une formation que le portail devrait donner.

Sur ce plan, l'impact le plus significatif pour l'équipe se manifesterait par l'adaptation des procédures de travail et par l'acquisition d'une formation assez légère. Par ailleurs, il faut pouvoir s'assurer que les revues possèdent des équipements informatiques de bon niveau pour exécuter les tâches attendues. Ces deux volets, formation et acquisition d'équipement, ont une incidence financière dont il faut tenir compte. On retrouve cette préoccupation dans plusieurs projets de publication électronique ; on peut citer, à titre d'exemple, J-Stage, Electronic Journal Publication & Dissemination Center.⁷⁴ Ce sont plutôt des coûts d'implantation ou de démarrage, encore qu'un certain suivi dans la formation puisse s'avérer nécessaire avec le renouvellement des équipes de secrétariat.

LA CONCEPTION DE LA REVUE

La « fréquentation » des outils électroniques pour la production et la diffusion des revues va provoquer, sous l'influence notamment des auteurs mais aussi de l'usage des fonctionnalités nouvelles, un transfert par lequel la version électronique de la revue sera considérée comme la version première de référence. De là, une réflexion sur la façon de concevoir la revue, sa forme d'existence et les pratiques que cela peut induire devrait s'imposer. Cette réflexion et les innovations qu'elle peut susciter, dépendent de la sensibilité des directions de revue, mais certainement des secteurs disciplinaires et des types de document que cela suppose. C'est ainsi que, selon les secteurs, on jugera avec plus ou moins d'intérêt la possibilité d'intégrer dans le corps de la revue les fonctionnalités du multimédia, les liens hypertextes, des données dynamiques, etc.

Ajoutons que l'introduction des processus électroniques dans les activités de publication devrait participer à une ré-ingénierie des diverses étapes qui les ponctuent et de la façon de travailler, depuis la soumission des articles à leur diffusion en version électronique. Cette ré-ingénierie ne sera pas instantanée, mais elle devrait faire l'objet de collaborations entre les équipes du portail et des revues. Par exemple, on peut imaginer, comme c'est le cas pour le CNRC, que le portail puisse faire l'acquisition d'un système logiciel de monitoring informatisé

⁷⁴ <http://www.jstage.jst.go.jp>

pour l'ensemble du processus éditorial de la revue et pour sa production et que ce système soit diffusé auprès des revues. L'effet de l'introduction des processus électroniques ne sera certainement pas instantané et suivra des rythmes difficiles à prévoir pour le moment.

CHAPITRE 4
UN SYSTÈME DE DIFFUSION

LA DIFFUSION : UN MANDAT ESSENTIEL

Le nombre de documents html qui alimentent la Toile croît à une vitesse vertigineuse. De même, le nombre d'internautes double presque chaque année. Le Web est assurément un média privilégié pour la communauté scientifique qui cherche à maximiser la communication des connaissances. Pourtant, la multitude d'informations disponibles, le manque de catalogage et les déficiences dans leur organisation constituent des entraves à l'accès au savoir. Si bien qu'une simple présence dans Internet des revues québécoises en sciences sociales, serait-ce par un portail unique, ne garantit pas nécessairement une visibilité et un rayonnement adéquats à l'information scientifique véhiculée par ces médias. En somme, être disponible sur le Web n'est pas synonyme d'être diffusé.

La diffusion comme phase de l'édition électronique est souvent sous-estimée. Il s'agit pourtant d'une composante essentielle, voire stratégique, du mandat d'un portail. D'où la nécessité d'élaborer des stratégies de diffusion et de mise en place adaptées à cet environnement virtuel et d'y allouer des ressources humaines et financières suffisantes. Cette section permet de distinguer les différents aspects impliqués dans le processus de diffusion d'un portail, d'en montrer les enjeux et les retombées possibles. On partira de l'accessibilité et de l'ergonomie du site, pour s'intéresser aux formats de diffusion et aux services offerts ; de là, on traitera des actions stratégiques de positionnement dans le Web afin de maximiser la visibilité et le rayonnement des documents mis en ligne.

PRINCIPES DE BASE POUR LA CRÉATION D'UNE INTERFACE WEB POUR DIFFUSION DES REVUES SAVANTES⁷⁵

Une démarche visant à optimiser la diffusion de l'information réunie dans un portail se compose de plusieurs éléments mais ne peut faire l'économie, au point de départ, d'une attention particulière pour la conception du site. L'accessibilité est l'une des caractéristiques permettant à un portail de devenir un outil de communication et un foyer de référence pour la recherche. Cette accessibilité découle aussi bien de l'environnement donnant accès à l'information que de l'aisance avec laquelle on peut mener la consultation de l'information.

L'accès à l'information est façonné par trois grands paramètres. Il prend forme à travers l'*architecture* du site, soit la structuration de l'arborescence des fichiers et l'organisation de l'information. Plusieurs configurations sont possibles. La nature du contenu et sa diversité, les fonctionnalités et les outils de navigation qui sont offerts, et les habiletés de navigation des utilisateurs sont autant de repères qui doivent inspirer cette architecture qui, en l'occurrence, devrait notamment varier les modes d'accès à l'information, permettre l'exploitation de la puissance du lien hypertexte, tout en tirant profit des fonctions offertes par le site.

⁷⁵ Les informations de cette section proviennent de : Patrick J. Lynch et Sarah Horton, *Yale Style Manual - Site Design*, 1997.

Adresse URL : <http://info.med.yale.edu/caim/manual/contents.html> ;

Communications Jean Lalonde, *Les bâtisseurs de l'inforoute- Site indépendant d'assistance et de référence en conception de sites Web au Québec (section : 5 questions pour éviter l'échec d'un site Web d'entreprise)*, 1996-1998. Adresse URL :

<http://www.cjl.qc.ca/batisseurs/5questions.htm/>;

Jakob Nielsen, *Top Ten Mistakes in Web Design*, 1996. Adresse URL : <http://www.useit.com/alertbox/9605.html/>;

Jakob Nielsen, *Top Ten New Mistakes in Web Design*, <http://www.useit.com/alertbox/990530.html/>.

Précisément, cet accès est supporté par les *infrastructures de repérage* de l'information, d'où l'intérêt de diversifier les modes d'accès au contenu ; on pense à l'accès par catégories de répertoire, à l'accès par un module de recherche plein texte et par champs (titre, auteur, description, mots-clés, etc.), ainsi qu'à l'accès par une classification documentaire.

La consultation de l'information sera facilitée par le soin apporté à l'ergonomie de l'interface et par la puissance des outils d'aide à la navigation. L'ergonomie commence par des précautions toutes simples. Par exemple, le graphisme, qui confère au site son identité visuelle, a davantage à être sobre et mis au service de la valorisation du contenu informationnel. Un environnement stable et familier doit être offert à l'utilisateur, un environnement qui se caractérise par la simplicité, la rapidité et la clarté. De même, les outils d'aide à la navigation peuvent être puissants, mais encore faut-il fournir des outils facilitant leur utilisation et le positionnement sur le site. On peut souligner, à titre d'exemple, la présence d'un plan du site, d'indications de positionnement pour que l'utilisateur puisse se situer dans l'architecture du site, de fichiers d'aide à propos notamment des modes d'accès à l'information.

Les questions de l'accessibilité au contenu informationnel et des conditions de consultation doivent occuper une place privilégiée lors de la conception du site puisque ces aspects contribuent à une large diffusion. Il s'agit de conditions de départ, au sens où la richesse du contenu et sa mise en forme sont des éléments de pertinence tout à fait essentiels. Mais là encore le contenu ne se suffit pas à lui-même, il faut donc pouvoir s'assurer que le portail sera répertorié par les outils de recherche afin que sa visibilité soit maximale auprès des utilisateurs potentiels. Ce sont ces composantes que nous considérons dans les sections qui suivent.

LE SYSTÈME DE DIFFUSION

SERVICES DE BASE

NAVIGATION

Le premier type de navigation est celui qui suit *l'organisation logique des périodiques scientifiques*, à savoir : les relations entre une revue, ses volumes, ses numéros et ses articles. Le lecteur doit pouvoir naviguer aisément dans une telle organisation et toujours comprendre ce qu'il peut trouver en poursuivant sa route. Pour y arriver, outre les liens à créer entre les différentes unités d'information, un important mécanisme de fiches descriptives doit être mis en place. Ces fiches descriptives permettront de prendre rapidement connaissance du contenu d'une unité d'information sans nécessairement avoir à la consulter au complet. Une fiche descriptive d'une revue pourrait être son nom, sa périodicité, ses objectifs, ses personnes responsables, etc. Pour un numéro, la table des matières remplit cette fonction, alors que pour un article, les informations comme le titre, les auteurs, le résumé se retrouvent naturellement dans une telle fiche. Le défi du système de navigation consiste à modéliser correctement les différentes fiches descriptives et à les produire automatiquement depuis les unités d'information fondamentales, à savoir les articles des revues. Bien entendu, certaines informations ne pourront être extraites des articles et devront être produites à part, mais il faut minimiser ces interventions.

Une autre forme de navigation devra être offerte, navigation que l'on pourrait qualifier de *thématique*. Il s'agit d'exploiter les métadonnées associées aux articles pour offrir une navigation non plus basée sur l'organisation par revues ou numéros, mais sur des concepts comme des sujets ou des auteurs ou des relations thématiques. Une analyse des métadonnées sera nécessaire afin d'identifier celles qui pourront donner lieu à un système de navigation.

CONSULTATION

La consultation d'un article consiste à récupérer une version du document et à lire ce dernier, l'imprimer, l'annoter, etc. La consultation devrait débiter par la fiche descriptive d'un article. En effet, puisque la diffusion d'un article pourra se faire selon plusieurs formats et représentations, et qu'aucun de ces formats n'est, *a priori*, prioritaire, la fiche descriptive devient le point d'entrée de l'article⁷⁶. Cette fiche descriptive devra contenir un lien vers toutes les versions disponibles de l'article et, une fois ce lien activé, le lecteur pourra consulter l'article à sa guise.

RECHERCHE

La recherche d'information est essentielle dans un portail, qu'il soit constitué de revues savantes ou d'autres sources d'information. La navigation n'est en général pas suffisante, en particulier lorsque la collection de documents est assez volumineuse. L'outil de recherche offert devra permettre de distinguer *l'information primaire*, c'est-à-dire les articles scientifiques, *l'information secondaire*, par exemple la description d'un numéro ou d'une revue, et *l'information de navigation*. Cette dernière catégorie devrait normalement être ignorée en recherche.

Un bon outil de recherche devrait permettre, d'un côté, de composer des requêtes de recherche variées et puissantes et, d'un autre côté, de présenter les résultats de façon claire et utile pour les utilisateurs.

Pour répondre au premier objectif, les documents indexés devront être les versions XML de référence des articles. L'outil de recherche devra pouvoir tenir compte non seulement du texte des articles, mais aussi de leurs métadonnées ainsi que de la structure des documents. Les métadonnées et la structure devront permettre de restreindre les critères de recherche, par exemple de chercher des mots seulement dans les sujets ou les auteurs ou les titres de sections, mais aussi d'influencer le tri par ordre de pertinence des résultats.

Quant au deuxième objectif, relatif à la présentation des résultats, la situation peut être plus complexe. Dans un premier temps, la possibilité d'effectuer différents tris, y compris par ordre de pertinence, ne semble pas poser de problèmes fondamentaux. Toutefois, une fonction intéressante des outils de recherche consiste à mettre en évidence les mots cherchés dans les documents trouvés. Puisque la recherche s'effectue dans des documents XML, en utilisant notamment la structure, mais que la consultation pourra être effectuée dans des documents HTML ou PDF, il n'est pas facile d'établir une correspondance entre la requête exprimée et le contenu des articles dans ces formats. Des recherches et du développement sont encore à faire

⁷⁶ D'ailleurs, l'adresse d'un article devrait être reliée à celle de sa fiche descriptive. Par exemple, dans un système basé sur les DOI, la résolution d'un DOI devrait mener le lecteur sur la fiche descriptive de l'article en question.

de ce côté, c'est pourquoi il est préférable de suivre ces développements et d'intégrer de telles fonctions à plus long terme.

GESTION DES UTILISATEURS ET DROITS D'ACCÈS

La diffusion libre et gratuite des articles de revue dans Internet ne nécessite pas *a priori* une gestion des utilisateurs et des droits d'accès. Toutefois, dès que des restrictions sont imposées sur l'accès aux documents, mais aussi dès que des services personnalisés sont offerts aux utilisateurs, cette gestion devient nécessaire.

La conception d'un bon système de gestion des utilisateurs et des droits d'accès pour un portail de revues savantes n'est pas chose aisée, et il n'est pas possible de le faire dans le contexte de ce document. Contentons-nous de souligner les aspects les plus importants. Par ailleurs, seuls les aspects touchant aux restrictions d'accès à l'information non gratuite sont nécessaires à court terme, les autres aspects pourront être implantés lors de phases ultérieures.

Le premier problème à résoudre concerne la définition de ce qu'est un utilisateur. Pour un portail de revues savantes, cette définition peut être problématique ou à tout le moins diversifiée. Une personne qui a accès à un portail, parce qu'il fait partie d'une institution qui possède les droits d'accès, jouit des privilèges de l'institution mais, en même temps, possède son propre environnement personnel. Si on ajoute qu'il peut faire partie de plusieurs institutions, qui ont toutes des droits d'accès différents, la situation se complique. Une telle description nous amène à proposer une solution complexe mais souple, faisant appel à la notion d'utilisateurs et de groupes, sachant que les utilisateurs peuvent faire partie de groupes et que des groupes peuvent inclure d'autres groupes. Par défaut, les utilisateurs héritent des permissions et restrictions des groupes dont ils font partie.

Toutefois, il est nécessaire d'identifier les différents objets ou services dans le portail qui peuvent faire l'objet de restrictions. Cette identification peut aussi être complexe, car on doit distinguer au moins les différentes revues et probablement les volumes ou numéros, si on suit le modèle des publications imprimées. On peut ajouter un grand nombre de contraintes comme, par exemple, celle de donner accès à un ou quelques articles seulement. On pourrait même aller plus loin, en permettant de voir certaines parties des articles. Pour éviter une trop grande complexité du système, il serait préférable de donner un libre accès à la fiche descriptive et de réserver l'accès à l'article aux personnes autorisées.

L'accès aux différents services peut aussi être contrôlé. Un service de recherche d'information se doit d'être disponible pour tous, mais il n'en va pas nécessairement de même pour la diffusion sélective d'information ou l'annotation des articles.

La mise en place d'un système de gestion des utilisateurs passe donc par la rédaction d'une politique d'accès à l'information, par l'identification des unités d'information et des services, qui peuvent faire l'objet de restrictions, et par la définition de ce qu'est un utilisateur. Une fois ces éléments définis, il faut mettre en place un système qui permette de jongler avec tous ces concepts, toutes ces définitions, et ainsi permettre toutes les combinaisons possibles. La plupart des systèmes de gestion documentaire offrent une gestion des utilisateurs par groupes et personnes, de même qu'une gestion des droits d'accès par document ou par dossier.

En manipulant intelligemment ces outils, on peut arriver à offrir une gestion qui remplit une bonne partie des besoins exprimés.

SERVICES COMPLÉMENTAIRES

À ces services centraux, il est permis d'en ajouter d'autres qui viendront valoriser le portail pour ce qui est de ses attributs relevant de la consultation et de l'exploitation de la documentation offerte. Ces services, aussi intéressants soient-ils, se greffent au noyau de base, si bien qu'il n'est pas impératif (même si ce serait souhaitable) qu'ils soient mis en place dès la première phase d'implantation du portail.

DIFFUSION SÉLECTIVE DE L'INFORMATION

La diffusion sélective de l'information, c'est-à-dire l'envoi d'information directement aux personnes concernées, est importante dans un tel portail à cause du caractère périodique de la parution des revues. Puisque les numéros d'une revue particulière sont publiés selon une certaine périodicité, il est tout à fait naturel d'en avertir les personnes intéressées. La principale difficulté du système de diffusion sélective de l'information consiste à faire en sorte de minimiser les besoins en gestion. Ce système pourrait être autogéré par les utilisateurs, de leur identification à leurs préférences, en plus de leurs besoins en information.

Comme pour la navigation, la dissémination devrait être possible selon deux orientations, la première, basée sur la publication des articles, la deuxième, thématique. Dans le premier cas, les utilisateurs devraient pouvoir être avertis de la publication d'un article, et donc habituellement d'un numéro d'une revue, s'ils se sont déclarés intéressés par cette revue. Le deuxième cas permet aux utilisateurs de s'abonner à certains sujets ou certains auteurs, et d'être avertis lorsqu'un article est publié.

AIDE À LA LECTURE

L'aide à la lecture désigne toutes les petites fonctionnalités qu'il est possible d'ajouter afin de faciliter l'acquisition des connaissances par les lecteurs. L'implantation de ces fonctionnalités peut être progressive, en ce sens que rien n'oblige de le faire dès le début du projet. Toutefois, pour certains utilisateurs, il s'agira de raisons suffisantes pour apprécier le système mis en place.

Différents logiciels de navigation de contenu documentaire peuvent nous inspirer concernant les fonctionnalités d'aide à la navigation, en particulier les logiciels qui ont été conçus pour la consultation de cédéroms documentaires. Les navigateurs Web standard offrent des fonctionnalités intéressantes que l'on peut brièvement rappeler.

- Un système d'*annotation* est intéressant car les lecteurs ont souvent l'habitude de prendre des notes lors de la consultation d'articles scientifiques. Si les annotations peuvent être partagées entre différents utilisateurs, le système devient encore plus intéressant.

- Les *signets* permettent de revenir rapidement à certains éléments d'information. Les navigateurs Web permettent de mettre des signets sur des pages, mais pas sur des parties quelconques de pages Web. Pour des articles importants, la possibilité de marquer un endroit précis à l'intérieur du document peut s'avérer indispensable.
- Les *liens* créés par l'utilisateur peuvent aussi être un outil intéressant d'acquisition de connaissance dans un corpus documentaire, en particulier lorsque ces liens peuvent être partagés ou « publiés » afin de les faire connaître à d'autres personnes.

Ces trois systèmes peuvent tous être implantés à l'aide d'outils respectant la norme (en construction) XLink, qui permet de créer des relations entre des documents ou des parties de documents sans nécessairement toucher au contenu de ces documents. Afin de permettre le partage des annotations, signets et liens, ces outils devront être implantés « côté serveur ».

Une autre fonctionnalité intéressante est la *manipulation* des images et des autres contenus multimédias. Ainsi, il peut être très utile de pouvoir agrandir une image pour mieux apprécier les détails, de pouvoir démarrer et arrêter une séquence vidéo à sa guise, etc. Pour y arriver, des logiciels « côté client » doivent être fournis, car les navigateurs Web actuels ne vont pas assez loin dans ce sens.

PERSONNALISATION

La personnalisation d'un service Web consiste à offrir à l'utilisateur une expérience de navigation et de consultation qui soit adaptée à ses besoins, et donc potentiellement différente de celle d'un autre utilisateur. Les premiers services à offrir une telle personnalisation furent les portails comme *My Yahoo !*, où l'on peut définir sa propre page d'accueil.

La personnalisation la plus intéressante pour le portail des revues savantes serait de recueillir les préférences des utilisateurs en matière de thématiques, d'auteurs, de revues, etc. et d'offrir à ces utilisateurs, en première ligne, des articles qui respectent ces préférences.

Une telle personnalisation n'est pas essentielle, mais il pourrait s'agir d'une valeur ajoutée, justifiant ainsi son implantation éventuelle.

UNE OFFENSIVE DE DIFFUSION SUR LA TOILE

Le référencement et le positionnement d'un site portail dans les différents outils de recherche Internet (index et répertoires) s'avèrent d'une importance cruciale pour attirer une audience et maximiser le transfert du savoir.⁷⁷ Une esquisse des principes pour rendre

⁷⁷ Les informations de cette section proviennent : Tracey Stanley, «Moving Up The Rank», *Ariadne*, no 12, novembre 1997. Adresse URL : <http://www.ariadne.ac.uk/issues12/search-engine/>;
Tracey Stanley, « Keyword Spamming : Cheat Your Way To The Top », *Ariadne*, no 10, juillet 1997. Adresse URL : <http://www.ariadne.ac.uk/issues10/search/engines/>;
Tracey Stanley, «The relavant of underpants to seaching the Web», *Ariadne*, no 24, juillet 1997. Adresse URL : <http://www.ariadne.ac.uk/issues24/search/engines/intro.html>;

opérationnels ces deux aspects peut être brossée. Mais la diffusion du contenu de revues savantes nécessite également une visibilité dans les principales bases de données disponibles au chercheur d'information dans le domaine des sciences sociales. D'où l'intérêt d'identifier ces bases de données potentielles et leurs revendeurs, d'examiner la représentation des revues québécoises dans celles-ci, d'analyser les possibilités offertes par ces bases et d'établir les principes d'une stratégie pour accroître leur dépouillement au sein de ces outils. Finalement, pour attirer de nouveaux publics, fidéliser les utilisateurs actuels et adapter les services offerts par le portail aux nouvelles tendances, une campagne de *webmarketing* peut être envisagée.

LE RÉFÉRENCIEMENT ET LE POSITIONNEMENT STRATÉGIQUE

Les internautes utilisent massivement (85 %) les outils de recherche pour trouver de l'information dans le Web et, dans une proportion encore plus grande, ils ne prennent pas connaissance de plus de deux ou trois pages des résultats de leur requête de recherche⁷⁸. On constate ainsi toute l'importance, pour un portail de revues savantes en sciences humaines et sociales, à la fois d'apparaître dans les bases de données des principaux outils de recherche du Web — c'est ce que l'on appelle le référencement — et de sortir dans une position privilégiée dans les résultats des requêtes acheminées par les utilisateurs à ces moteurs — ce qui renvoie à la notion de positionnement stratégique. L'un et l'autre, référencement et positionnement, sont étroitement liés et font partie intégrante des objectifs de diffusion d'un portail.

RÉFÉRENCIEMENT : MOTEURS DE RECHERCHE ET RÉPERTOIRES

Le référencement, à savoir l'inscription d'un site Web dans les outils de recherche et dans les répertoires afin que le site soit intégré dans les bases de données, est un processus circulaire et évolutif : l'inscription n'est jamais définitive et doit être adaptée aux nouveaux besoins de visibilité. Il faut y consacrer des ressources, mais surtout considérer cet aspect au moment même de la conception du site Web. Une action fructueuse demande une connaissance à la fois des outils de recherche, des modalités d'indexation et des politiques éditoriales des répertoires. Ici on doit distinguer les outils de recherche des répertoires ou annuaires.

Pour les premiers, immenses bases de données, ce sont des robots qui de façon automatisée indexent un site ayant ou non soumis une demande pour être intégré à la base. La connaissance des modalités d'indexation peut permettre d'optimiser les pages Web pour les divers outils (pensons à : AltaVista, Google, Northern Light, America Online Search, FAST Search, Netscape Search, etc.).

Pour les seconds, qui classifient les ressources Internet à la suite de l'évaluation « d'éditeurs » humains et en référence à une politique de classement, la démarche est différente. Les « éditeurs » humains décident, en fonction de critères tels que la qualité du contenu, l'organisation de l'information, l'accessibilité au contenu, la fréquence de mise à jour, la

Karen Liberatore, «Getting to the Source», *Macworld*, 22 septembre 2000. Adresse URL :

<http://macworld.zetnet.com/features/pov.4.4.html/>;

Kathleen Murphy, «Cheaters Never Win», *Internetworld*, Mai 1996.

<http://www.internetworld.com/print/1996/05/20/undercon/cheaters.html/>.

⁷⁸ Georgia Institute of Technology, Graphic, Visualization and Usability Center, *Tenth User Survey*, 1998.

http://www.gvu.gatech.edu/user_surveys/survey-1998-10/.

pérennité de l'information, etc., d'intégrer ou non le site au répertoire (à titre d'exemple, citons : Yahoo!, La toile du Québec, Lycos, Looksmart, Francité, Nomade, Open Directory).

À tout prendre, il est impossible d'être partout, aussi est-il préférable de canaliser les ressources et l'attention pour le référencement aux outils de recherche majeurs, qui sont les plus susceptibles d'attirer beaucoup d'utilisateurs. Cela, évidemment, sans négliger l'intérêt d'apparaître dans des outils de recherche spécialisés, susceptibles de s'adresser à des populations plus ciblées mais qui correspondent à la mission du portail. À ce stade, seulement une fraction du chemin est parcourue, car il ne suffit pas d'apparaître ou d'être répertorié dans les principaux lieux de consultation et d'interrogation pour la recherche documentaire dans le Web : une bonne visibilité dépend de la capacité à être dans les premiers résultats de recherche pour certains mots-clés ou certaines requêtes susceptibles d'être formulés par les utilisateurs potentiels du portail, ce qui conduit à parler du positionnement stratégique.

LE POSITIONNEMENT STRATÉGIQUE : DÉVELOPPER UN MODÈLE D'OPTIMISATION

Par cette notion de positionnement, on fait référence à un ensemble de méthodes employées au cours de l'élaboration, de la publication et de l'entretien du site du portail pour assurer une visibilité optimale dans les outils de recherche et les répertoires. Le positionnement sera qualifié d'optimal si le site du portail apparaît dans les 20 ou 30 premiers résultats à la suite d'une requête dans un outil de recherche. Pour les répertoires, l'objectif, c'est d'être intégré dans la base de données. Pour ce faire, il convient de recenser, pour s'y adapter, les critères utilisés par les outils de recherche des répertoires.

L'étude des mots-clés et des catégories par lesquels le portail devrait être référencé, ainsi que des rubriques sous lesquelles le site devrait apparaître dans les répertoires ou annuaires, est capitale non seulement pour assurer le meilleur positionnement, mais pour avoir une bonne correspondance avec le vocabulaire que l'utilisateur potentiel est susceptible d'utiliser dans sa recherche. Aussi banale que puisse sembler cette démarche, elle devient très rapidement de nature technique.

Une surveillance régulière est de mise pour s'assurer du positionnement du site dans les outils de recherche, car les algorithmes de calcul de pertinence des outils changent, ce qui demande des adaptations en conséquence. Concurrément, de nouveaux sites apparaissent et désirent occuper les mêmes créneaux. Il est en ce sens parfois nécessaire d'ajuster la démarche en référence/positionnement, quitte à changer de stratégie. Les retombées du référencement peuvent parfois tarder à se manifester, mais, sur une trajectoire moyenne, les résultats sur le trafic sont généralement significatifs.

La préoccupation de bon aloi pour le traitement accordé au portail par les outils de recherche et les répertoires constitue une dimension qui reste partielle dans une démarche intégrée de rayonnement sur et par le Web, car la place et le rôle des bases de données ne doivent pas être négligés.

LA VISIBILITÉ DANS LES BASES DE DONNÉES

En plus de la visibilité atteinte par les outils de recherche et les répertoires, il faut se préoccuper de l'accessibilité et du repérage du contenu informationnel du portail dans les

bases de données importantes qui dépouillent le contenu des périodiques dans le domaine des sciences humaines et sociales. Ce n'est pas en soi une nouveauté, car les directions de revue prennent déjà plusieurs initiatives pour s'assurer que l'indexation de leurs publications imprimées soit faite par plusieurs bases de données, en se concentrant surtout sur les plus prestigieuses ou les mieux placées selon le domaine de spécialisation.

Ces initiatives sont parfois vaines, car des contraintes peuvent faire obstacle. Il est clair que ces bases de données se veulent sélectives au sens où elles retiennent les revues qui présentent des assurances de qualité et de rayonnement dans leur domaine de spécialisation. On ne peut ignorer que la langue joue un rôle. Mentionnons que très peu de bases de données indexent massivement des documents en français — très peu de bases proviennent d'ailleurs du marché francophone — et que les bases de données de langue anglaise sont assez sélectives, notamment sous l'angle de la langue de communication. L'objectif consiste à accroître la visibilité des revues savantes québécoises en sciences humaines et sociales, tant dans les bases de données (les index) francophones qu'anglophones. Dans le but d'améliorer l'accueil dans ces dernières, il faudrait s'assurer que des résumés « consistants » en anglais soient produits pour les articles de chaque revue diffusée par le portail. D'ailleurs, ces résumés, plus formalisés que ceux généralement proposés, auraient un triple avantage supplémentaire.

Pour, d'abord, améliorer les conditions de sélection dans les grandes bases de données, faciliter, ensuite, la consultation sur le portail et la prise en compte d'articles par des utilisateurs non francophones et, enfin, enrichir les métadonnées (qui puisent dans les résumés), l'adoption de résumés formalisés et plus explicites devrait s'imposer. Ces résumés⁷⁹ pourraient comprendre certaines rubriques, telles que 1° contextualisation, 2° objectif, 3° méthodologie, 4° résultats, et 5° conclusion. Ce genre de protocole, par lequel l'auteur de l'article communique ses motivations, son but, sa démarche, ce qu'il trouve et, finalement, en quoi cela fait avancer la recherche, s'est avéré approprié en sciences naturelles et en sciences médicales, mais aussi en sciences humaines et sociales. Il s'agit d'un outil de communication scientifique dont l'utilité serait certaine, non seulement, mais notamment, pour l'indexation dans les bases de données.

Cela souligne que la collaboration entre les directions de revue et le portail s'impose afin d'amplifier les actions à entreprendre sur ce terrain. Ce travail ne peut se faire en parallèle et dans l'ignorance mutuelle des actions, si l'on veut améliorer significativement la présence des revues du Québec dans ces grands instruments de recherche que sont les bases de données, instruments qui gagnent en importance et en impact avec l'utilisation courante de la communication électronique dans la recherche. Il serait judicieux, par ailleurs, d'offrir aux bases de données la possibilité d'accéder au plein texte à partir de leurs propres sites, ce qui s'inscrirait dans la tendance qui s'affirme de plus en plus. Cela pourrait être à l'avantage des bases de données, tout autant que des revues.

LES OUTILS DE PROMOTION POUR ACCROÎTRE LA VISIBILITÉ DU SITE-PORTAIL

Une offensive de diffusion sur la Toile doit aussi passer par une préoccupation touchant la promotion du site-portail lui-même. C'est l'intérêt de son contenu et sa capacité d'attraction en tant que lieu de diffusion d'information scientifique qui doivent plaider pour le portail. Si

⁷⁹ James Hartley, «Is it appropriate to use structured abstracts in social science journals ?», *Learned Publishing*, vol. 10, no 4, octobre 1997, p.313-317.

bien que l'accroissement de la visibilité, l'augmentation du nombre d'utilisateurs et l'impact du portail dans et sur la communauté scientifique internationale sont fonction de la capacité de faire du portail un lieu dynamique intellectuellement, qui nourrit, accueille et provoque des activités en mesure d'interpeller le milieu des chercheurs nationaux et internationaux. La meilleure promotion pour le portail, ce n'est pas de « vendre » un produit par des arguments plus ou moins superficiels, c'est plutôt d'informer que ce portail, partant d'une collection de revues riche, diversifiée et de très bonne qualité intellectuelle et technique, constitue un foyer intellectuel dynamique qui fait place à des activités et offre des services de premier niveau.

La *communication* et *l'interaction* du portail avec les chercheurs et les étudiants constituent un élément de première importance. Pensons aux actions suivantes :

- Mettre en place des infrastructures pour recueillir les commentaires des utilisateurs, ce qui permet de recevoir du feed-back sur les services offerts afin de les améliorer, donc d'être proactif.
- Activer une liste de discussion ou un forum favorisant l'échange d'idées scientifiques.
- Organiser des événements sur le site : débats, conférences virtuelles, diffusion en primeur de certains contenus, entrevues en direct, etc.
- Publier une lettre d'information à périodicité fixe par courriel pour informer les utilisateurs des nouveautés du site et de ses services.
- Alimenter régulièrement une section « actualité » à l'intérieur du site pour tenir au courant les utilisateurs des plus récents développements dans le domaine et des activités animant la communauté scientifique.
- Fournir l'accès libre au maximum d'information pour donner un aperçu de la collection de revues et d'articles, et accroître la portion du corpus qui est gratuitement mise à la disposition des chercheurs en réduisant les délais pour le basculement des articles courants dans le fonds rétrospectif.

La pertinence, la qualité et l'intérêt des *services* parlent d'eux-mêmes. Leur conception et leur mise en application doivent prendre en compte ces dimensions. Parmi ceux-ci, relevons :

- Un service de diffusion sélective de l'information.
- La personnalisation de l'interface avec le portail et un service d'aide à la lecture.
- La mise à disposition d'une liste de liens pertinents vers des sites d'intérêts connexes et le développement d'une bibliothèque virtuelle à même le site.

Le portail se doit d'être un acteur de premier plan dans la communication scientifique et s'inscrire dans une logique de réseau à l'échelle internationale. Les *partenariats* occupent, en ce sens, une position stratégique dans le développement du portail. Certaines pratiques simples y concourent :

- Établir un réseau de portails de revues savantes : intégrer des notices provenant de chaque membre du réseau dans les résultats de recherche de chacun de ceux-ci afin de générer du trafic sur l'ensemble de ces sites, partage de services, spécialisation dans certains domaines/aspects de la publication de revues savantes.
- Procéder à des échanges de liens réciproques avec des sites poursuivant les mêmes missions.
- Mettre en place un programme d'échange de bannières publicitaires (forme de publicité complètement gratuite et ciblée).
- Parrainer des sites de moindre envergure traitant des mêmes sujets afin de les aider à obtenir de la visibilité tout en favorisant l'échange de services.

Le recours à la publicité aurait d'abord pour objectif de faire connaître les services du portail, ce qui s'imposerait dans les premières phases du programme de réalisation. Il s'agit, pour l'essentiel, d'une démarche *d'information*. Pensons à :

- Des envois massifs aux groupes de discussions et aux listes de distribution pour « publiciser » le site-portail.
- Des envois massifs par courriel à des usagers potentiellement intéressés par les services offerts par le portail (prospection).
- Un communiqué de presse annonçant le site aux journalistes susceptibles de relayer l'information dans leurs publications respectives (médias traditionnels et nouveaux médias).
- Une conférence virtuelle avec des journalistes du domaine et des acteurs importants du milieu pour faire connaître les services offerts par le portail.

Le portail a pour mission de servir de catalyseur à l'activité scientifique en sciences humaines et sociales dans le Web. Une campagne de promotion dans le Web, ainsi qu'un référencement et un positionnement adéquats devraient donner une excellente visibilité au site-portail. La promotion devrait miser sur le dynamisme et la créativité du portail. Il n'en reste pas moins que le facteur promotionnel le plus crucial pour un portail de revues savantes, c'est sa richesse documentaire, ses services, la qualité et le renouvellement de son information.

DIFFUSION : ENJEUX ÉCONOMIQUES

La mise en place d'un portail de publication et de diffusion des revues savantes exige des modalités de financement. À ce jour, l'édition des revues est rendue possible grâce à la réunion de divers facteurs.

L'ORIGINE DES REVENUS DANS LA SITUATION ACTUELLE

On pourrait surtout évoquer les ressources publiques qui soutiennent la recherche et

celles qui contribuent à la conservation et à la diffusion de ces documents avec les bibliothèques, elles aussi financées publiquement. Pour cerner de plus près les enjeux concernant le modèle économique pouvant supporter de façon durable un portail de revues, il est préférable de se concentrer sur le montage financier qui est actuellement mis en place pour la production et la diffusion des revues elles-mêmes dans l'environnement actuel qui est centré sur la version imprimée. Ce montage est composé, pour l'essentiel, de deux sources principales. D'abord, par ordre d'importance, il y a les subventions publiques qui, au total, comptent pour près de 47 % des revenus ;⁸⁰ il est important de noter que deux organismes sont mis à contribution à l'occasion de concours et toutes les revues ne sont pas en position de cumuler les deux subventions (au total, on a 33 revues subventionnées par le Fonds FCAR et environ 55 par le CRSH). Ensuite, la deuxième composante majeure (près de 41 %), ce sont les revenus provenant de la vente des revues ; en cette matière, les ventes au numéro sont assez marginales, les abonnements étant responsables de 95 % des revenus de vente. Les autres revenus sont assez disparates et finalement peu nombreux : ce sera la publicité, une contribution de l'université, une autre subvention. Retenons que la moitié des revenus résiduels n'est pas spécifiée.

C'est avec ces données, auxquelles il faut ajouter, évidemment, la structure globale des dépenses, que se pose la question de la gratuité. Cette question se pose dans son principe et dans son application. Pour ce qui est du principe, on peut évoquer deux dimensions : l'accessibilité des résultats de recherche et la gratuité d'un service rendu auprès d'utilisateurs.

LA GRATUITÉ

Il est hors de doute que la diffusion gratuite d'information, et de documents universitaires en particulier, favorise l'accessibilité, si bien que les statistiques de consultation seront nettement supérieures à celles se référant à un accès filtré payant. C'est la gratuité du service pour l'utilisateur qui est certainement déterminante ; en ce sens, l'accès « gratuit » pour l'utilisateur, qui découle d'un abonnement institutionnel, peut avoir un résultat comparable, car la barrière à l'entrée pour le chercheur ou le lecteur, qui utilise le service en ligne, est levée en raison d'un abonnement de l'institution auquel il est rattaché — les services de *proxy* ont beaucoup élargi cet accès apparemment libre pour l'utilisateur en première ligne. Il reste que, dans ce cas, l'accès est lié au paiement d'un abonnement institutionnel. Si on revient à l'idée de la gratuité, on peut sans doute affirmer que c'est une pratique désirable, voire souhaitable. Abstraction faite du délicat montage financier qui permet aux revues d'exister comme institutions de communication scientifique, la gratuité serait d'emblée une hypothèse à retenir. Mais la gratuité ne coûte pas rien (*price free is not cost free*) ! Qui doit assumer la production et la mise en ligne d'un service gratuit ? C'est la question qui se pose et à laquelle trois hypothèses de réponse ont été proposées. On y revient tout de suite.

L'autre dimension liée à la gratuité repose sur les considérations « éthiques » que l'on associe à la commercialisation en tant que telle. On peut prétendre que la commercialisation et ses résultats donnent un indice à la fois de la valeur que les lecteurs attachent à la revue et de sa diffusion réelle, ainsi qu'elle permet de sanctionner l'adéquation entre la politique éditoriale de la revue et les attentes des lecteurs. Toutes ces considérations peuvent avoir une certaine

⁸⁰ On se base ici, comme pour les estimations qui suivent, sur un échantillon de 27 revues supportées financièrement par le CRSH. Cet échantillon a été choisi de façon aléatoire par l'organisme subventionnaire et représente la moitié des revues financées qui ont leurs bureaux administratifs (leur adresse de secrétariat) au Québec.

vraisemblance et justifier l'engagement dans le circuit marchand, mais fondamentalement les pratiques marchandes des revues sont beaucoup plus liées à la nécessité de tirer des revenus permettant de supporter les coûts engagés dans la production des revues — ce raisonnement est correct, dans ses grandes lignes, pour les éditeurs sans but lucratif et non, évidemment, pour les grands groupes commerciaux oligopolistiques. Pour l'essentiel, ce n'est pas pour des considérations « éthiques » que les revues sont inscrites dans des pratiques marchandes : il s'agit davantage d'une condition nécessaire pour assurer leur existence. Si tant est que cette existence n'est pas mise en péril, la gratuité constitue une option particulièrement intéressante.

Dans les conditions actuelles, 41 % des revenus proviennent des ventes du support papier. Le support papier en tant que tel correspond à environ 25 % des dépenses d'une revue. Les grandes données de l'équation permettent de constater que la gratuité mettrait à mal la situation financière des revues, même dans l'hypothèse d'une diminution substantielle des dépenses liées à la suppression du support papier. Or, cette hypothèse de la suppression du papier est difficilement envisageable, du moins dans un proche avenir. Cela exacerbe le problème. Partons de là pour en discuter.

Si nous considérons (comme cela est affirmé dans le rapport du comité du Fonds FCAR) que le support papier devrait trouver son mode de financement, on peut envisager qu'un système d'abonnement pour la version papier pourrait exister parallèlement à une diffusion gratuite de la revue en ligne. On aurait deux systèmes de diffusion : l'un semblable à celui que l'on connaît pour le papier, l'autre pratiquant l'accès libre au texte des revues en ligne. L'abonnement se devrait de couvrir au moins les coûts du papier, alors qu'actuellement il couvre ce coût plus environ 15 % du budget consacré à d'autres frais. Restons-en là, et creusons le mode de financement pour le reste.

La chaîne de production qui est retenue permet de procéder à la production technique de la revue en langage structuré pour la production et la conservation et, de là, de diffuser en divers formats pour une consultation Web et pour l'impression à distance⁸¹. Dans le sillage de la production technique, nous avons la production de métadonnées, l'archivage électronique, l'application de normes de référencement permanent, les services d'accès et de consultation, une stratégie de diffusion sur le Web, un service de diffusion sélective de l'information et des éléments d'infrastructure intellectuelle et matérielle (tels la veille technologique et l'hébergement et la sécurité du serveur). Ce rappel partiel permet simplement de souligner la qualité et l'étendue des services du portail. Maintenant, à la question des modalités permettant d'assurer la gratuité du service pour l'utilisateur, trois hypothèses peuvent être évoquées.

LA GRATUITÉ PAR LA RÉDUCTION RADICALE DES DÉPENSES

Dans les textes qui portent sur cette question, on trouve une affirmation allant dans le sens suivant : les coûts pour produire une revue sont beaucoup trop lourds ; il faut, à la faveur de l'introduction de l'électronique, revoir à la baisse ces coûts. La réduction de la structure de coûts peut viser les coûts liés à la production matérielle de la revue (ou production technique) et les coûts qui découlent de l'existence organisationnelle de la revue.

⁸¹ Avec le fichier Postscript, qui est l'un des « produits » de la chaîne de traitement, la revue dispose de l'équivalent d'un prêt-à-photographier pour l'imprimeur : c'est en fait le fichier Postscript qui est acheminé à l'imprimeur.

Pour ce qui est des coûts de la production technique, l'introduction de l'électronique est loin de faire jaillir un trésor caché. Certes, si on élimine le papier, on vient de dégager un espace budgétaire qui reste assez loin de certaines estimations, car, il faut être clair sur une chose : l'économie dégagée pour l'impression et l'expédition lorsqu'on supprime le papier, c'est 100 %, mais ces postes budgétaires ne représentent environ que 25 % du budget d'une revue. Par ailleurs, il n'est pas du tout certain que l'hypothèse radicale de la suppression de la version imprimée soit souhaitable pour l'immédiat. Abstraction faite de cette considération, la publication électronique ne peut être porteuse d'économies majeures que si, avec l'électronique, on supprime le travail d'édition et de normalisation. Si on laisse à l'auteur le soin de faire sa propre mise en page et d'utiliser le format qui lui convient ou un format propriétaire, on ramène la publication des revues à la pratique de plusieurs serveurs de prépublications et on met en touche le langage structuré, les formats multiples, la production des métadonnées, la conservation, le référencement permanent, etc. On fait, de plus, très peu de cas des différents services qui découlent d'une diffusion conséquente et efficace. En d'autres termes, il n'y a pas de mystères : ou, au nom d'économies dont l'ampleur reste à démontrer, on adopte une vision minimaliste de mise en ligne d'articles et, de ce fait, on tourne le dos aux standards internationaux pratiqués par les grands portails de référence pour les revues (et donc du mandat confié par le Fonds FCAR) ou, au contraire, on peut chercher à produire techniquement dans les meilleures conditions, avec rigueur et professionnalisme, des revues en respectant et en s'inscrivant dans les standards et les normes qui permettent au portail de remplir sa triple mission de production, de diffusion et de conservation, et il devient alors vain de faire miroiter l'existence d'un trésor caché. Les mêmes fonctions n'exigent pas nécessairement les mêmes ressources que pour les conditions actuelles liées à l'imprimé, mais toutes les expériences tendent à montrer que, pour des services professionnels de qualité et similaires à ceux pratiqués dans l'ancien environnement, aucune économie substantielle ne permet de supporter la gratuité de la diffusion de la revue.

Du côté de l'organisation de la revue, la capacité de réduire la structure de coûts est, pour employer un euphémisme, fiable. Les frais liés au travail de l'équipe éditoriale et à l'existence organisationnelle de la revue sont difficilement compressibles dans la mesure où, d'une part, ils sont comprimés à leur strict minimum et, d'autre part, les revues fonctionnant déjà avec un travail gratuit, soit par prêt de services soit par bénévolat, les frais actuels ne comptabilisent qu'une partie des ressources nécessaires à leur existence. Les hypothèses de réingénierie des processus éditoriaux, d'arbitrage, de correction, de préparation de la copie, etc. sont particulièrement stimulantes et devront, dans un avenir qui n'est pas éloigné, être introduites. Si la remise en cause des procédures d'arbitrage par les pairs et des propositions alternatives sont appelées à jouer le rôle de mécanismes de réduction des coûts administratifs, il faut commencer probablement par discuter cette question au mérite et voir si elle est reçue favorablement. Dans l'ensemble, on peut informatiser et numériser l'ensemble des processus allant de la soumission d'un manuscrit jusqu'à la parution de la revue. Cette opération, prise globalement, sera étalée dans le temps, elle demande un investissement initial non négligeable et sera la source d'économies durables dont l'estimation est pour le moins approximative et incertaine. Pour l'avenir prévisible, à moins de mettre en péril l'existence matérielle et organisationnelle des revues ou de demander aux institutions de rattachement, quelles qu'elles soient, d'augmenter significativement leur contribution en service et en espèces, la gratuité ne peut reposer sur la réduction des frais liés aux structures éditoriales.

Une autre option serait d'assurer l'accès gratuit aux revues en ligne par un financement qui viendrait de la perception, auprès des auteurs, d'une contribution monétaire. Comme si les auteurs devaient payer un droit pour être publié. Ici ce n'est pas la revue, ni l'utilisateur de la revue qui assure la gratuité, mais celui qui est diffusé, soit l'auteur.

Cette pratique, qui existe déjà dans l'environnement de l'édition papier, est davantage présente dans certaines disciplines, plutôt associées aux sciences naturelles et aux techniques⁸² ; à l'occasion, on y a recours dans les sciences sociales, mais la contribution reste modeste et, bien des fois, est associée à un coût particulier dans le traitement de l'article — par exemple, lorsque l'iconographie est riche.

Il n'est pas utile de reprendre certaines questions sur la gratuité traitées précédemment, contentons-nous plutôt d'examiner les conditions de réalisation et le réalisme de cette hypothèse.

D'abord, si cette pratique a cours dans certains milieux, elle est somme toute assez exceptionnelle en sciences humaines et sociales. La contribution financière des auteurs pour être publié ne fait pas partie des usages (surtout si on se situe dans les revues publiées par des organismes sans but lucratif).

Ensuite, l'introduction de cette pratique risquerait de rencontrer de grandes réticences. Elle supposerait que les auteurs possèdent la capacité financière pour répondre aux attentes des revues. À moins de modifier substantiellement le mode de financement de la recherche, cette supposition est gratuite. Les professeurs d'université n'ont pas d'allocation pour payer de tels droits, à moins de bénéficier de subventions de recherche. Or, la réussite aux grands concours de subventions en sciences humaines et sociales au Canada touche environ 35 % de ceux qui ont présenté un projet de recherche. Ces derniers ne sont qu'une fraction de tout le corps professoral. En cela, on induirait une ségrégation entre les auteurs en fonction non plus seulement de la qualité du travail à publier mais aussi de la capacité de payer pour la publication. Les réticences seraient sans doute aussi fort grandes du côté des auteurs hors-Québec et étrangers qui représentent plus de la moitié des auteurs des revues québécoises.

Enfin, la seule façon de rendre cette pratique quasiment indolore serait de ramener ces droits à un niveau tellement bas qu'ils auraient un caractère symbolique. Nous revenons par là à la discussion de la rubrique précédente, à la seule différence que l'on n'affirmerait pas que la structure de coûts pour l'existence, la publication, la diffusion et la conservation (archivage) d'une revue est quasiment nulle, mais qu'elle est somme toute assez faible. Car quel est le seuil du symbolique ? 50 ou 100 dollars de droits à payer par article ? Dans un cas comme dans l'autre, on voit bien que le coût de l'introduction d'une telle pratique, en termes de relations avec les auteurs, d'image, etc., n'en vaut pas la chandelle étant donné la modestie des sommes recueillies : on parle d'une somme totale de 900 \$ ou de 1 800 \$ pour une revue de 18 articles par année.

La possibilité de réaliser cette hypothèse semble très faible ; elle représente donc une solution douteuse.

⁸² Pensons à *New Journal of Physics*, par exemple : <http://njp.org/>.

LA GRATUITÉ DÉCOULANT DE LA RECONNAISSANCE DES REVUES ÉLECTRONIQUES COMME BIEN PUBLIC

Selon cette hypothèse, la gratuité n'est pas imputable à l'effondrement supposé des coûts d'une revue ni aux auteurs qui sont publiés, mais au statut qui est accordé à la revue comme document de diffusion des résultats de la recherche. L'identification des revues à titre de bien public change radicalement la façon de considérer ces documents et leurs conditions de production et de diffusion.

Toute une série de biens sont identifiés de cette façon, pour lesquels l'accessibilité est largement facilitée ou sans restrictions. Pour ce faire, des modes de financement appropriés sont trouvés, qui généralement passent par les canaux publics. Un bien n'est pas public par définition ; il n'est pas non plus exclu de cette désignation d'emblée. Les choses évoluent et se différencient selon les sociétés : on investira significativement dans le logement social ou on s'en retirera, selon les époques ; on facilitera l'accès au livre et à la documentation par une politique de développement du réseau des bibliothèques ou on relèguera cette préoccupation dans le fouillis des questions non prioritaires.

Alors que la plupart des processus et stades participant à la recherche échappent à la sanction commerciale — et sont supportés par un financement public, direct ou indirect —, la diffusion des résultats de recherche par les revues savantes y est encore soumise. Or, on pourrait tout aussi bien considérer qu'en raison de l'intérêt et de l'utilité du contenu qu'elles révèlent et qu'en raison de la visibilité qu'elles donnent au système de recherche national, les revues devraient bénéficier de ce statut de bien public et jouir d'un accès le plus large possible. Il est vrai que les subventions publiques supportent les revues pour environ la moitié de leurs budgets. Même à cela, elles sont obligées de recourir aux circuits marchands pour une part notable de leurs revenus, ce qui devient une condition *sine qua non* d'existence. On pourrait imaginer la reconnaissance des revues, et du contenu qu'elles diffusent, comme un bien public, car elles constituent le prolongement naturel d'une chaîne dont les pouvoirs publics supportent les principaux maillons. Dès lors, il deviendrait important de leur assurer le meilleur rayonnement.

Cette hypothèse n'est pas que pure spéculation. Tout près de nous, elle s'est matérialisée récemment, avec l'annonce par le Conseil national de recherches du Canada, de la diffusion gratuite de la version électronique de leurs quatorze revues pour les utilisateurs ayant une adresse IP se terminant par < .ca >. Cela a été rendu possible par l'engagement d'une agence gouvernementale (Travaux publics et Services gouvernementaux de l'État fédéral) à combler la perte de revenus découlant de l'application d'une telle mesure pour le Canada. Le mode de calcul de la contribution gouvernementale a été basé non sur le remplacement de la perte de revenus, mais sur le principe du recouvrement des coûts. On a donc pris l'ensemble des coûts directs, desquels ont été soustraits les coûts d'impression et de distribution, pour établir un coût de première copie, sans égard au format. De ce coût ont été déduits les frais rattachés à l'organisation de la revue, à l'arbitrage par les pairs, c'est-à-dire les frais liés à l'existence de la revue comme institution et processus de sélection des textes. Le résultat de ces opérations donne une estimation des coûts de production de la version électronique. C'est sur cette base que l'évaluation de la contribution publique a été calculée.⁸³

⁸³ Ces informations nous ont été transmises par Aldyth Holmes qui est directrice des Presses scientifiques du CNRC.

Certains points ressortent de la mise en œuvre de cette mesure. D'abord, la contribution financière du CNRC pour ces revues n'a pas fléchi et la façon de les produire n'est pas modifiée par cette mesure. Ensuite, un nouvel acteur (public) est intervenu pour les fins de la diffusion en assurant un accès libre aux utilisateurs canadiens et le prix de cette intervention a été établi en référence au coût de production de la revue, mis à part les coûts liés au support papier. Enfin, l'accès filtré et payant a été maintenu pour les utilisateurs non canadiens et autres organismes ayant une adresse IP sans < .ca > et ces revenus sont censés pouvoir équilibrer l'ensemble du budget des revues au CNRC.

On peut considérer une autre avenue qui mettrait à contribution les bibliothèques universitaires dans un mode de financement pouvant assurer la gratuité aux utilisateurs. D'une certaine manière, les bibliothèques « subventionnent » les revues par les abonnements institutionnels ou, selon une pratique maintenant courante pour les revues électroniques, par des licences d'exploitation négociées institutionnellement ou souvent en consortium. Or, les bibliothèques pourraient subventionner à la source la production des revues, dont la diffusion serait ensuite assurée gratuitement par le portail. Ce genre de partenariat, conclu dans des ententes pluri-annuelles, permettrait de stabiliser le financement des revues et les coûts d'abonnement pour les bibliothèques. La *Scholarly Publishing and Academic Resources Coalition* (SPARC) préconise des modèles similaires. Cela dit, compte tenu des sommes en jeu, les contributions des bibliothèques ne pourraient suffire à assurer un financement adéquat des revues. L'inventaire rapide qui a été dressé permet de constater que les plus grandes bibliothèques sont généralement déjà abonnées aux revues supportées par le Fonds FCAR (notre inventaire s'étant limité à ces revues) et que les autres le sont assez largement. Dans le cadre d'une entente collective, les rentrées nouvelles — en raison du petit nombre de bibliothèques, du taux de pénétration dans les bibliothèques québécoises plutôt bon et des tarifs d'abonnement relativement modestes — seraient assez faibles. Dans l'ensemble, il s'agirait d'un déplacement d'une somme assez semblable, mais assurée. On pourrait chercher à accroître ces revenus en impliquant les bibliothèques collégiales ou celles des grandes municipalités. Cependant, dans le cadre du portail, si cette somme lui était entièrement créditée, elle serait « soutirée » en quelque sorte des revenus actuels des revues. Ce qui pose un problème d'un autre ordre. En tout état de cause, s'il y a un apport net de ce côté, il ne pourrait être que complémentaire à des subventions gouvernementales ou à d'autres sources de financement ; c'est ce montage qui pourrait offrir un modèle économique viable et prometteur.

Quel que soit le mode particulier, la mise à contribution d'un « commanditaire public » répondrait, sans doute au mieux, à la situation des revues et à leurs objectifs complémentaires de production de qualité, de diffusion élargie et de stabilité organisationnelle. La grande difficulté tient à faire reconnaître le statut de bien public à la diffusion en ligne des revues savantes et à « intéresser » une agence publique pour soutenir financièrement ce bien public. C'est certainement la voie qui offre les meilleures conditions pour valoriser la publication électronique et pour échapper aux règles de la commercialisation, auxquelles on note peu d'attachement de la part des éditeurs de revues. Dans le chapitre 5, une évaluation des besoins financiers que cela suppose est présentée.

L'EFFET DE LA GRATUITÉ DE LA VERSION ÉLECTRONIQUE SUR L'IMPRIMÉ

Il vient immédiatement à l'esprit de s'interroger sur l'impact qu'une diffusion libre de la

version électronique peut avoir sur l'avenir de la version imprimée de la revue. Toutes les consultations menées à ce jour auprès des revues québécoises et la prise en compte des pratiques des revues d'importance en sciences humaines et sociales qui ont, au départ, un support papier conduisent à dire que la version papier est appelée à se maintenir pour une durée encore indéterminée. Plusieurs raisons militent en ce sens, qui sont bien loin d'une simple attitude nostalgique. On peut prendre ce constat comme une donnée et donc se questionner sur l'interaction entre les deux supports, surtout si on ne pratique pas les mêmes conditions (économiques) d'accès : accès libre dans le Web et abonnements pour l'imprimé.

Dans la mesure où l'imprimé répond à un besoin réel des lecteurs et à des situations non moins réelles se rapportant à la consultation et à la diffusion des revues, on peut penser que la seule diffusion en ligne ne ferait pas table rase des motifs favorisant le maintien de la version papier.

D'ailleurs, le bilan de certaines expériences dans le domaine semble aller dans ce sens. Après trois ans d'existence, le bilan du projet *Muse* soulignait en 1998⁸⁴ que la diffusion électronique n'avait pas fait périr le nombre d'abonnements ; chez Blackwell,⁸⁵ on note qu'il y a un intérêt pour les formats électroniques, mais pas tendance à une substitution du support papier pour le support électronique : les utilisateurs préfèrent plutôt combiner une version électronique avec la version imprimée de la revue. Pourtant, on ne considère pas moins que les abonnements pour la version papier sont appelés à disparaître lentement. Dans ces deux cas, l'accès en ligne n'est pas gratuit. Pour *Muse*, qui pratique surtout l'abonnement institutionnel donnant accès aux individus rattachés aux institutions, la consultation se rapproche, dans les faits, de la gratuité pour les individus, mais le maintien d'un abonnement pour la version papier relève d'abord des bibliothèques. Dans le second cas, l'accès est filtré et payant pour les individus et les institutions. Si on peut retenir ces exemples pour la discussion des tenants et aboutissants de l'accès filtré et payant (qui suit), les constats qui sont faits ici ne peuvent être transposés sans réserves, dans la mesure où l'accès gratuit sur le Web change l'environnement général pour les utilisateurs.

On pourrait sans doute estimer que l'accès en ligne gratuit aurait une influence négative sur le niveau des abonnements pour la version imprimée. Mais jusqu'à quel point devrait-on parler d'un fléchissement ou d'une contraction significative ou d'une chute majeure ? Hypothèse particulièrement pessimiste ou résolument optimiste, on pourrait imaginer que la contraction n'irait pas au-delà, pour le proche avenir, de la moitié des abonnements actuels. Sur cette base, ce qu'il faut pouvoir apprécier, c'est si le coût direct de la version papier (papier, impression, expédition⁸⁶) peut être amorti par les revenus des abonnements pour ce format. Cela conduirait sans doute aussi à imaginer des modes différents de production de la version papier si le niveau du tirage devait être en deçà d'un certain seuil (500 exemplaires, par exemple), car, dans cette situation, l'impression numérique, et non plus l'imprimerie offset, s'imposerait pour des raisons économiques. Si besoin est, la modification de certains procédés de production devrait permettre aux abonnements pour le format papier d'amortir les coûts pour ce support. Autrement, on peut se demander si la demande pour la version papier est

⁸⁴ Marie R. Hansen dir., *Project Muse, Process Report*, The Johns Hopkins University Press avec The Milton S. Eisenhower Library, janvier 1995 à juin 1998, 40p.

⁸⁵ Richard Hodson, «The demand for journals – fact versus fiction», *Learned Publishing*, vol. 11, no 3, juillet 1998, p.205-208.

⁸⁶ Seulement ces coûts sont considérés, car la chaîne de traitement électronique produit une version mise en page Postscript destinée aux besoins de l'impression papier.

assez importante. Cela étant, les revenus actuels provenant des abonnements sont supérieurs aux coûts directs de la version imprimée : une partie, environ le tiers des revenus des ventes, supporte la préparation de la première copie. Si cette préparation de la première copie était désormais impartie au « commanditaire public », on peut réellement poser une nécessaire et exclusive adéquation entre le coût de la version papier et les revenus des ventes (abonnements et ventes au numéro). En somme, en prenant en compte l'ensemble de ces éléments et dans des termes qui restent grossiers — diminution de moitié des revenus de vente et baisse correspondante des tirages, avec modifications techniques le cas échéant —, il est permis de penser que cette adéquation est vraisemblable. À plus ou moins long terme et selon les situations particulières des revues, on pourrait d'ailleurs imaginer une seule livraison par année de la revue imprimée.

Toutes ces questions sont ouvertes et laissent place à des appréciations diverses, compte tenu du degré d'incertitude concernant la réaction à la fois des directeurs de revue, des abonnés et des lecteurs, mais aussi quant à la présence d'un « commanditaire public ». Par contre, la gratuité sur le portail ferait fondre de façon significative les revenus autonomes des revues et ceux-ci n'excéderaient plus, comme c'est le cas dans la situation actuelle, les coûts de production de la version papier. Cette perspective peut soulever certaines craintes de la part des revues dans la mesure où, si elles dépendent en totalité des pouvoirs publics pour exister au-delà de la production du format papier *stricto sensu*, cela peut être ressenti comme une position de grande vulnérabilité à l'égard des politiques publiques. D'où le besoin d'assurances fermes sur la détermination des autorités publiques d'assurer ce financement de façon durable et satisfaisante ; sans cela, les revues seraient réticentes à faire un pari qui deviendrait très risqué.

L'ACCÈS FILTRÉ ET PAYANT

Si la gratuité est impossible à pratiquer, parce que les éléments qui la soutiennent ne sont pas réunis, il faut songer à l'accès filtré et payant au serveur. La situation ne serait pas particulièrement atypique, car une forme ou une autre d'accès filtré est généralement pratiquée dans les principaux portails de revues, y compris ceux qui sont sans but lucratif (*Muse*, *HighWire* et, bientôt, *BioOne* — pour ne nommer que ceux-là). Cela a pour conséquence qu'il faut mettre en place un système de gestion intégré des abonnements et que l'on conçoive un modèle de relations fonctionnelles entre les abonnements pour la version imprimée et la version électronique de la revue. On peut aborder sur cette base quelques cas de figure afin de traiter plus spécifiquement de l'infrastructure transactionnelle.

LE PORTAIL GÈRE LES ABONNEMENTS INSTITUTIONNELS ET INDIVIDUELS

Le premier cas de figure consiste à donner au portail un rôle premier dans la gestion des abonnements, que ce soit pour les versions papier ou électronique. Dans cette perspective, il faut garder à l'esprit que les revues, à l'heure actuelle, gèrent elles-mêmes leurs abonnements sous un mode « in house » ou les font gérer par un éditeur ou par une agence d'abonnement spécialisée, commerciale ou universitaire. En clair, cela signifie qu'il y a un nombre élevé de lieux de gestion des abonnements pour le nombre total de revues. Les protocoles et le nombre de services offerts sont différenciés ; de même, les grilles tarifaires ont des niveaux de complexité très variables. Ajoutons, ce qui n'est pas accessoire, que dans plusieurs cas, la gestion des abonnements, qui exige des ressources particulières, est bien souvent pour

la revue un coût caché, c'est-à-dire dont on voit mal les contours ou qui est assumé par un tiers, soit l'institution universitaire ou la société savante, par exemple.

Si le portail s'occupe des abonnements pour les deux supports, cela exige nécessairement un arrimage entre ces divers lieux et le portail : cela suppose, en effet, de créer au moins deux écrans de saisie des données de la commande, ainsi qu'un double mécanisme de perception des paiements (individus et établissements) et, enfin, une redistribution des recettes aux revues clientes du portail. La modalité la plus simple pour le portail, ce serait d'être la porte d'entrée principale des abonnements : cela éviterait de devoir composer avec des protocoles et des logiciels de bases de données assez disparates et permettrait de réduire les délais à sa plus simple expression pour l'accès en ligne.

Mais le portail doit-il devenir une agence d'abonnement, y compris pour la livraison de la version papier ? Pour répondre à cette interrogation, il faut avancer plusieurs considérations. *A priori*, les tâches d'une agence d'abonnement sont complexes et, sans doute, dépasseraient-elles les compétences propres d'un portail de production et de diffusion de revues ; cela explique que, bien souvent, la responsabilité de cette partie de la chaîne est laissée à une agence spécialisée par entente contractuelle. Par ailleurs, l'expédition des exemplaires aux abonnés, reviendrait aux revues, si bien qu'elles assumeraient la vente des numéros individuels, par la poste et en librairie — dans ce dernier cas, par le biais d'un distributeur —, et prendraient en compte les campagnes de publicité.

Dans la logique du portail, la gestion intégrée des abonnements et de première ligne devrait être privilégiée. D'abord, cela signifie que la porte d'entrée des abonnements, quel que soit le support, devrait être le portail. Pour le portail, il n'y aurait qu'une base de données et qu'un protocole. Pour les revues, cette hypothèse signifie qu'elles se retirent des ententes contractuelles qu'elles ont en ce domaine avec des éditeurs ou des agences spécialisées, au profit du portail ; cela signifie aussi que, pour les nombreuses revues qui gèrent elles-mêmes les abonnements, elles se mettent désormais en deuxième ligne, ce qui resterait particulièrement utile pour l'expédition de la version imprimée et pour les campagnes de publicité. Si cela peut soulager certaines revues, cette nécessité devrait en heurter d'autres. D'ailleurs, la gestion des abonnements par le portail, dont le coût doit être établi et assumé, aura pour effet de rendre visibles des coûts qui, dans bien des cas, sont actuellement cachés. Cela apparaîtra comme une dépense supplémentaire pour la revue dont le budget est déjà serré. Dans la logique de la revue, il pourrait paraître plus simple de voir le portail comme un intervenant de deuxième ligne qui s'ajoute à la structure actuelle. Dans ce cas, la multiplicité des sources et la grande difficulté d'imposer un protocole unique, dans la gestion des données et dans une logique de tarification, donnent au portail une responsabilité de standardisation et de gestion d'information extrêmement lourde et coûteuse en raison des conditions dans lesquelles il doit œuvrer. Les délais de transmission et de gestion de l'information, qui ne sont pas techniques mais organisationnels, risquent de nuire au service de diffusion en ligne.

Ensuite, dans la perspective où les abonnements individuels et institutionnels sont maintenus, cela impose deux modes de filtrage de l'accès : un accès par adresse IP pour les institutions et un accès par mot de passe pour les individus, le premier étant assez simple alors que le deuxième demande une logistique plus lourde. L'abonnement individuel impose aussi l'infrastructure pour la transaction en ligne alors que les institutions n'utilisent généralement pas ce genre de paiement. Il y a, entre ces deux systèmes de filtrage d'accès, une

différence significative pour les coûts d'implantation et de fonctionnement. Celui qui permettrait de traiter les abonnements, qu'ils viennent des individus ou des institutions, coûte quelque 50 000 \$ pour l'implantation (contre 16 000 \$ pour le système ne s'adressant qu'aux institutions), alors que le coût de fonctionnement, par la suite, est nettement décalé : 100 000 \$ pour le premier contre 30 000 \$ pour le second (voir les estimations budgétaires au chapitre suivant).

Ces remarques sont faites sans compter l'effet que peut avoir l'abonnement institutionnel pour la version en ligne sur l'abonnement individuel. L'accès aux revues qui est donné par l'abonnement institutionnel souffre de moins en moins de restrictions : avec le service *proxy*, nul n'est besoin de consulter une revue depuis le campus, par exemple. En ce sens, on peut se demander s'il y aura une forte demande pour l'abonnement individuel en ligne ; l'abonnement individuel se concentrera peut-être sur la version imprimée pour ceux et celles qui n'auraient vraiment pas accès à la version électronique par le biais des bibliothèques, notamment. Il y aura certainement glissement du nombre d'abonnés individuels, pour ne pas dire une contraction. Dans cette mesure, il y a lieu de s'interroger sur l'utilité de mettre en place un système de filtrage de la version électronique pour les individus. En d'autres termes, les abonnements individuels ne devraient-ils pas être prévus que pour la version imprimée ?

LE PORTAIL SE CONCENTRE SUR LES ABONNEMENTS INSTITUTIONNELS

Les remarques précédentes supportent cette hypothèse voulant que le portail pratique l'accès filtré sous le mode de l'abonnement institutionnel. Cette formule est assez répandue : elle est d'une grande simplicité pour l'utilisateur, elle répond bien aux consultations ponctuelles d'un grand nombre de revues que font les chercheurs — consultations qui ne justifient pas des abonnements à chaque fois ou qui imposeraient des transactions à l'article —, elle est assez simple à mettre en place, sur les plans technique et organisationnel, et elle correspond de plus en plus aux usages dans les milieux de la recherche. Après la gratuité, c'est la formule qui favorise le plus la consultation des documents en ligne. Les individus se verraient dans la position de s'abonner à la version imprimée, en communiquant avec la revue ou avec son agence, ou de faire pression, au besoin, pour que leur institution souscrive un abonnement au moins pour la version électronique (et/ou papier).

Dans ce cas de figure, la gestion des abonnements ne serait pas intégrée, puisque partagée et par le portail et par chacune des revues. Cela suppose l'usage de deux banques de données différentes, et une double comptabilité pour les revues, surtout pour suivre la progression de leurs divers types d'abonnements et de leurs revenus. La situation se complique surtout au moment de produire des rapports — et ils sont nombreux —, à commencer par ceux qu'exigent les organismes subventionnaires.

LE PORTAIL PRATIQUE DES ABONNEMENTS PAR PANIER

Selon ce cas de figure, le portail s'occuperait seulement d'abonnements institutionnels mais en privilégiant la vente par panier, c'est-à-dire la vente de l'ensemble de la collection des revues diffusées. Ce modèle, qui a été pratiqué par les groupes commerciaux oligopolistiques comme méthode de vente, est de plus en plus utilisé par les bibliothèques qui forment des

consortiums et qui y trouvent leur intérêt en tant qu'acheteurs.

Ce modèle peut être considéré rigide, si on s'y limite, pour l'acheteur institutionnel spécialisé — pensons à un centre de recherche et à une université à l'étranger, et ils sont nombreux — qui ne verrait pas l'intérêt de l'abonnement à l'ensemble de la collection. Il faudrait voir à l'introduction d'aménagements en temps voulu. On peut penser que ce modèle doit être panaché avec une autre formule pour atteindre la souplesse désirable. Par contre, il faut voir une plus grande simplicité dans les transactions entre le portail et les institutions qui s'abonnent, soit à titre individuel ou comme participant à un consortium. Par ailleurs, cette formule accroît le rayonnement de plusieurs revues, dans la mesure où elles peuvent être accessibles à des clientèles de bibliothèques pour lesquelles il n'y avait pas nécessairement d'abonnement jusque-là.

QUI DIT ABONNEMENT DIT COMMERCE ÉLECTRONIQUE

L'accès tarifé impose l'implantation d'un système comprenant au minimum la gestion des abonnés, l'accès contrôlé, la grille de tarifs et le catalogue des revues en vente. D'où un système de diffusion des revues qui prévoit la gestion des abonnements, le contrôle d'accès aux documents, la « livraison immédiate » des articles, les multiples formats d'un même article, des statistiques d'accès, la gestion du catalogue des titres et la grille des tarifs, le paiement en ligne et la gestion des comptes, y compris le dépôt des paiements et la répartition des recettes. Deux avenues peuvent être pratiquées. Il est possible, d'une part, d'utiliser un système de base qui permet une programmation de session Web simple ou, d'autre part, acheter un système déjà existant de catalogue. En effet, les systèmes de commerce électronique sont nombreux, mais sont habituellement destinés à la vente de nombreux produits et non à une activité correspondant à celle d'un portail. Il serait en ce sens plus approprié d'utiliser un environnement de programmation simple et de faire la programmation pour l'adapter aux besoins d'un portail (voir l'annexe 4.1).

L'évaluation qui a été faite des ressources nécessaires pour implanter un tel système a pris en compte deux paramètres, soit l'environnement technologique actuel du lieu d'hébergement du serveur et un souci d'économie. Sur cette base, et en n'utilisant pas de serveur dédié mais plutôt un espace sur le serveur déjà prévu pour héberger le site, il semble nécessaire de réunir près de 5 000 \$ pour l'achat de logiciels — serveur Web, base de données, gestion du catalogue et des prix, gestion de la prise d'abonnements et de la transaction, et statistiques d'utilisation — et 50 jours de développement et d'installation. Avec un professionnel jouissant de plus de trois ans d'expérience dans la technologie utilisée, l'ensemble coûterait environ 45 000 \$. Il s'agit d'un coût de démarrage ou d'implantation non récurrent. Cela correspond à l'infrastructure pour les abonnements individuels et institutionnels, alors que pour celui adressé exclusivement aux institutions coûterait environ 16 000 \$. Les coûts récurrents seraient particulièrement faibles pour ce qui est des composantes « logiciel » et système

À cette estimation, il faut ajouter le coût de la gestion des abonnements et donc d'utilisation et de mise en service du système technique mis en place.⁸⁷ Évidemment, ce coût dépend des types de responsabilité du portail pour les abonnements. S'il doit s'occuper tout autant des abonnements individuels que des abonnements institutionnels, et pour les deux versions, on peut estimer, dans l'état des connaissances actuelles, qu'il en coûterait vraisemblablement 10 dollars par abonnement pour gérer les abonnements d'une trentaine de revues, ayant en moyenne 400 abonnements. Cela exclut l'expédition des copies papier aux abonnés et les exemplaires gratuits, de même que la vente de numéros individuels courants et anciens, par la poste et en librairie (là par le biais d'un distributeur), qui incombent aux revues. Par contre, si le portail ne s'intéresse plus qu'aux abonnements institutionnels, le volume d'opérations diminue et le système de filtrage d'accès se simplifie. Dans ce cas, on peut estimer à 7 dollars le traitement d'un abonnement. Dans ce cas, les revues ont à maintenir un système parallèle, qui a des frais. Ces estimations ne peuvent compter sur l'expérimentation d'une informatisation de l'ensemble de la chaîne d'abonnement, car celle-ci nous fait défaut. Il s'agit d'une projection, basée sur une connaissance fine des procédures actuelles, transposée pour les prochaines années dans un nouvel environnement. Une utilisation poussée de procédures informatisées devrait permettre une plus grande optimisation des ressources mobilisées à cette gestion, du moins on peut le croire.

QUI DIT ABONNEMENT DIT AUSSI REVENUS

Il est assez difficile d'anticiper les réactions des utilisateurs à la mise en place d'un portail de diffusion avec accès au plein texte, sous condition d'abonnement. En principe, ce service devrait susciter un certain engouement, mais, en même temps, on doit constater une certaine lenteur dans l'assimilation des nouvelles données dans les comportements individuels et institutionnels. Là-dessus, les interprétations peuvent être partagées. Dans le cadre de cette réflexion, on se doit de ne pécher ni par trop grand optimisme ni par l'attitude contraire. Des prémisses doivent être clairement énoncées afin de dégager quelques perspectives. Dans cette rubrique, ce qui est en cause, c'est l'estimation qui peut être donnée des revenus d'abonnement.

Pour ce qui est des prémisses, retenons que le portail se concentrerait sur les abonnements institutionnels et qu'il pourrait panacher la formule de l'abonnement revue par revue et la formule du panier ; le prix de l'abonnement à la version papier serait égal à 100 % du prix actuel⁸⁸ et celui de l'abonnement jumelé de la version électronique et de la version papier s'élèverait à 130 % du prix actuel. Dans l'immédiat, près de 90 % des abonnés institutionnels opteraient pour un abonnement combiné en raison de la faible marge supplémentaire pour obtenir les deux supports⁸⁹; autre élément : aucune augmentation du nombre d'abonnements n'a été imputée en raison de l'incertitude à cet égard et de la compensation qu'une augmentation pourrait avoir sur un fléchissement selon les créneaux (types d'institutions ou catégories d'individus). Les revenus des abonnements institutionnels

⁸⁷ Pour avoir une idée des opérations et des étapes que recouvre la gestion des abonnements dans un environnement de revues sur support papier, on peut se référer à l'annexe 4.2. Ces étapes et opérations s'inscrivent dans un processus de travail assisté par ordinateur ; avec le portail, l'environnement change et se présentera en termes de processus électronique de travail.

⁸⁸ De façon générale, le prix des abonnements des revues en sciences humaines et sociales est très modeste. Le prix pour l'abonnement institutionnel oscille entre 40 \$ et 70 \$.

⁸⁹ De même, en raison de cette faible marge et de la présence de la collection de la revue déjà sur les rayons, l'incitation à ne retenir que l'électronique n'est pas très appuyée dans l'immédiat.

seraient divisés en donnant aux revues l'équivalent du prix actuel des abonnements et le 30 % de majoration au portail. Les abonnements individuels seraient destinés aux revues pour la version imprimée. Finalement, quelque 80 % abonnés individuels maintiendraient leur abonnement papier⁹⁰. Sur cette base, on peut établir des projections pour un parc de 30 revues en ajustant les données agrégées de notre échantillon de 27 revues ayant leur direction au Québec.

À partir de ces données, les abonnements institutionnels assureraient des rentrées financières de 409 500 \$ au portail, dont 315 000 \$ seraient remis aux revues. Les revues bénéficieraient de rentrées venant des abonnements individuels de l'ordre de 112 000 \$. En somme, alors que le portail recevrait 94 500 \$, les revues cumuleraient 315 000 \$, 112 000 \$ et 35 000 \$ pour un total de 462 000 \$. Cela signifie qu'en moyenne, chaque revue recevrait 15 400 \$. Dans la situation actuelle, chaque revue encaisse en moyenne 16 353 \$ de revenus venant des ventes et paie, en moyenne, 12 305 \$ pour les coûts du papier, de l'impression et de l'expédition. Ces nombres donnent une appréciation des ressources venant de la vente et qui servent à d'autres dépenses que la seule production du papier ; ils révèlent aussi un manque à gagner entre les revenus actuels et ceux anticipés. Pour le portail, les revenus de 94 500 \$ viennent de la vente d'abonnements, ce qui reste assez modeste — une autre répartition se ferait au détriment des ressources servant aux revues — et nécessite une contribution significative d'une autre source pour absorber les coûts de diffusion.

LES CONDITIONS DE LA DIFFUSION POUR LE FONDS EN RÉTROSPECTIF ET POUR LES PRÉPUBLICATIONS

Deux éléments sont à distinguer pour le fonds rétrospectif. Lorsqu'une revue est publiée en version électronique, elle n'a qu'à offrir, dans cette version, les numéros ou volumes courants. Cela prend du temps avant de pouvoir constituer un corpus ayant un certain volume. Il devient très tôt d'un grand intérêt de pouvoir offrir en version électronique, par la numérisation rétrospective de la collection, sinon tout le fonds de la revue du moins la portion la plus récente du fonds. Par ailleurs, la publication courante en version électronique est, de fait, le rétrospectif de demain. Or, dans l'un et l'autre cas, le fonds rétrospectif devrait pouvoir être consulté gratuitement. Revenons aux deux facettes qui viennent d'être évoquées.

Afin d'offrir aux utilisateurs un matériel plus substantiel que les parutions courantes pour une revue nouvellement disponible en version électronique, la numérisation rétrospective devient un atout. Plusieurs techniques sont disponibles. Nous avons déjà indiqué des choix à faire dans le chapitre sur la production, compte tenu des contraintes de ressources. Le format image permet assez rapidement et dans des conditions acceptables d'offrir une assez grande partie de la collection d'une revue à des prix acceptables. Un financement *ad hoc* serait nécessaire. Dans l'état actuel des programmes pour les revues, un tel financement peut se trouver. Ce financement permettrait d'assumer en totalité le coût de la numérisation et de l'indexation. Le coût de sa diffusion serait lié aux coûts généraux de gestion du site et à l'entretien des fichiers et du moteur de recherche. Amorti sur l'ensemble, ce coût serait assez faible pour chaque revue. Par ailleurs, dans un environnement commercial, la durée de vie

⁹⁰ Les exemples en pareil domaine semblent montrer que l'effet est à peu près nul sur les abonnements ; aussi, avons-nous imputé aux individus rattachés aux institutions une propension limitée à ne pas renouveler leur abonnement dans un proche avenir.

marchande active des parutions d'une revue est relativement courte. On peut estimer que les ventes d'un numéro sont très faibles de 18 à 24 mois après la date de parution. En ce sens, pour les parutions courantes en version électronique, il faut envisager un mode de passage dans le rétrospectif. Il appartiendrait aux directions des revues d'établir le délai qui leur convienne, mais on peut imaginer qu'il serait souhaitable que les parutions courantes disponibles dans le Web soient versées dans le fonds rétrospectif au-delà de ce délai (restant à déterminer). Dans la mesure où ce basculement dans le rétrospectif consacre la fin de l'espérance de vie marchande de la revue, il conviendrait de donner un accès libre à ces corpus.

Le fonds rétrospectif serait accessible sans restrictions, parce que les coûts auraient été absorbés soit au moment de la numérisation rétrospective, soit lors de la parution courante, et ce fonds se composerait, d'abord, de la collection numérisée (le cas échéant) et, ensuite, des publications courantes après un délai (nécessairement). Il faut voir que le fonds rétrospectif constitue une valeur ajoutée à un portail qui se met en place et que, même dans un environnement commercial, il permet l'accès à un corpus de plus en plus considérable en accès gratuit. C'est la pratique qui est largement retenue sur le site de *HighWire*, par exemple. D'un autre côté, tous les efforts de diffusion de la revue, par le référencement ou autrement, rejouent sur la revue, dans son ensemble, à la fois pour le courant et pour le rétrospectif. Aucune démarche spécifique ne serait engagée pour le fonds rétrospectif seul.

Par ailleurs, dans la mesure où un service de prépublication serait mis en place, il devrait pouvoir trouver son financement intégral. Dans l'étude des coûts, devront être retenus les frais liés non seulement à la production, mais aussi à la diffusion. L'accès à de tels services est généralement gratuit, les stratégies de diffusion sont limitées ainsi que les services. En tout état de cause, tous ces éléments devraient faire l'objet d'estimations budgétaires et être appuyés financièrement de façon conséquente.

CHAPITRE 5

L'ARCHIVAGE

La conservation des périodiques savants a, depuis toujours, été assurée par différents organismes : bibliothèques, centres d'archives et éditeurs, suivant des procédures et méthodes, polies par les ans. Si toutes les règles de l'art sont suivies, on peut affirmer sans crainte qu'un document papier sera encore lisible dans 500 ans.

Il est difficile de faire une telle affirmation pour les documents électroniques. Le portail devra donc mettre en place des procédures permettant de donner des assurances quant à la pérennité de l'accès aux revues électroniques. Il s'agit d'une question fondamentale non seulement pour les producteurs d'information mais également pour tous les acteurs du processus de la communication scientifique. D'abord, parce que les revues savantes représentent un investissement important en temps et en argent. Ensuite et surtout, parce que ces revues constituent une espèce de « procès-verbal » de l'évolution d'une discipline et qu'elles font partie d'un patrimoine intellectuel et culturel que tous les acteurs ont le devoir moral de protéger et de transmettre à la postérité.

Le problème peut paraître simple à première vue. Il ne s'agit en somme que de faire transiter des bits d'information vers le futur. Des techniques, bien connues et maîtrisées, de rafraîchissement⁹¹ et de migration⁹² pourront être employées sans difficulté avec les documents en format structurés (XML, SGML). Ces documents ne contiennent en effet que du texte « pur ».

Mais nous devons également assurer l'intégrité de ces documents. Or, ils font aussi référence à d'autres objets numériques (images, sons, modèles, formules, hyperliens, etc.) qu'il importe de conserver. Ces différents objets liés peuvent être de formats propriétaires, incompatibles ou simplement de différentes versions.

On le voit, la conservation des revues savantes est un problème technique qui peut être complexe. En s'engageant dans la durée, ça devient également un problème économique car les coûts de rafraîchissement et de migration sont récurrents. C'est enfin un problème organisationnel qu'on oublie souvent de considérer.

Le portail devra tenir compte de tout cela dans sa planification et son fonctionnement quotidien. Nous verrons dans cette section du rapport tout les aspects à considérer dans le développement d'une politique de conservation et d'archivage des revues diffusées par le portail.

DISTINCTION ENTRE DIFFÉRENTS TERMES

Avant toute chose, il est important de bien distinguer certains termes français et anglais qui peuvent porter à confusion suivant la langue ou la profession exercée.

⁹¹ Rafrâichissement : action de copier de l'information numérique d'un médium de stockage à long terme vers un autre (source : projet CEDARS – traduction libre)

⁹² Migration: La migration est un élément d'une stratégie globale de conservation qui consiste à mettre en œuvre une série de tâches visant à convertir de façon périodique des objets numériques d'une configuration (matérielle/logicielle) vers une autre ou d'une génération de technologie informatique vers une nouvelle génération. Le but de ces opérations est de préserver l'intégrité de l'objet numérique et de conserver la capacité de le récupérer, l'afficher et l'utiliser en dépit des technologies en changement constant. (source : projet CEDARS – traduction libre)
<http://www.leeds.ac.uk/cedars/documents/PSW01.htm>.

La Bibliothèque nationale du Canada donne les définitions suivantes de l'archivage et de la conservation.

« *Archivage* : Les documents sont versés sur un serveur institutionnel dans le but de préserver leur contenu intellectuel de façon permanente. L'archivage s'entend dans son sens le plus large, et couvre les concepts de compilation, de conservation et de mise en disponibilité à long terme. »⁹³

Notons au passage que l'activité de conservation de documents électroniques signifie beaucoup plus que la simple création d'une copie de sauvegarde.

« *Conservation* : Activité [...], qui garantit la longévité des collections. Les activités de préservation comprennent la conservation des collections, la création de substituts, le contrôle des conditions ambiantes et l'établissement des conditions d'utilisation. (Politique de conservation de la BNC, 1989). En ce qui concerne les publications électroniques, les activités de conservation englobent l'organisation, la description, la mise à jour et la migration de l'information électronique afin d'assurer l'accessibilité à long terme des publications. »⁹⁴

« *Preservation* : storage, maintenance and access to digital objects/materials over the long term. »⁹⁵

Le but de la conservation des publications électroniques est donc de s'assurer que l'information subsiste et soit accessible sous un format lisible aussi longtemps que possible, même après plusieurs changements technologiques. La conservation inclut aussi toutes les stratégies permettant d'arriver à cet objectif, par exemple le transfert de l'information de supports en supports, etc. On retrouve dans cet objectif une distinction souvent faite en archivistique. D'une part, il y a la notion de préservation de l'information contenue dans un document, soit la protection de cette information, d'autre part, il y a la notion de conservation du support sur lequel se trouve l'information, soit le maintien en bon état physique de ce support. Les deux notions s'avèrent complémentaires pour assurer une stabilité et un accès à long terme aux publications électroniques.

En anglais, *archives* correspond à une « collection » regroupant des documents dans un domaine particulier. Les sites de prépublications dans Internet ont souvent le terme *Archive* dans leur appellation, et il s'agit bel et bien d'une collection de prépublications sur un domaine.

Le terme *archiving* ne correspond pas au terme « archivage ». En anglais, l'accent est mis sur la sélection (identification et regroupement) des documents. La notion d'accessibilité reste

⁹³Bibliothèque nationale du Canada, Groupe de coordination des collections électroniques. *Politiques et directives relatives aux publications électroniques diffusées en réseau*, octobre 1998.

< <http://www.nlc-bnc.ca/pubs/irm/fneppg.htm#annexe> >

⁹⁴ *Ibid.*

⁹⁵ Russell, Kelly. « CEDARS : Long-term Access and Usability of Digital Resources. The Digital Preservation Conundrum ». *Ariadne*, no 18, décembre 1998. < <http://www.ariadne.ac.uk/issue18/cedars> >

présente mais pas celle de conservation. La définition de Margaret E. Phillips le montre bien : « First, the materials have to be identified, collected and made accessible in their current format (the archiving process). »⁹⁶.

Finalement, le terme *preservation* correspond au terme français conservation.

Ces définitions étant posées, il est évident que, dans un monde numérique, la nuance est subtile entre archivage et conservation. Toutes les opérations de conservation nécessaires à son bon fonctionnement devront naturellement être assumées par le portail afin qu'il demeure en activité. Ces activités du portail peuvent donc également être considérées comme de l'archivage. Le portail ne peut toutefois prétendre être aussi durable que des institutions telles que les bibliothèques nationales.

OBSOLESCENCE TECHNOLOGIQUE

Toutes les mesures de conservation visent à contrer le problème de l'obsolescence technologique, en premier lieu l'obsolescence physique rapide d'abord. Pour comprendre le problème, on n'a qu'à évoquer les disquettes de 8 pouces, introduites sur le modèle 3310 d'IBM en 1979 ou même les disquettes de 5¹/₄ pouces introduites en 1980 sur les premiers ordinateurs personnels qui ne sont plus utilisables avec les ordinateurs d'aujourd'hui et donc illisibles. 20 ans à peine! on ne parle pas ici de textes écrits par des Sumériens!

Le rafraîchissement routinier des données nous assure de conserver le flot de bits en bon état mais nous devons également lutter contre l'obsolescence des logiciels : pensons ici au logiciel *WordStar*, fort populaire il y a à peine 15 ans, pensons aux multiples versions d'un même logiciel qui paraissent à intervalle régulier. Bref, s'ils ne sont pas rendu inaccessibles en raison de l'obsolescence technologique du matériel, les documents électroniques peuvent quand même être endommagés et amputés de certains éléments caractéristiques. « À titre d'exemple, il est possible de conserver le contenu d'un document électronique tout en perdant la mise en page. Il est possible de conserver la présence physique (p. ex., un fichier) sans toutefois arriver à conserver la lisibilité. »⁹⁷

Depuis les cinq dernières années plusieurs projets expérimentaux cherchent à trouver et définir les meilleures solutions, et proposer plusieurs modèles relatifs à la question de l'archivage et la conservation des publications électroniques sur le Web. Ces projets proviennent de divers organismes, tels les éditeurs, les bibliothèques nationales, les instituts de recherche, les universités, les archives nationales, les programmes gouvernementaux. Dans notre revue de la littérature, nous avons porté une attention particulière aux projets impliquant des publications scientifiques, à la question des formats et des supports, ainsi qu'aux stratégies employées pour l'archivage et la conservation.

⁹⁶ Phillips, Margaret E. « Ensuring Long-Term Access to Online Publications », *The Journal of Electronic Publishing*, vol. 4, no 4, juin 1999. < <http://www.press.umich.edu/jep/04-04/phillips.html> >

⁹⁷ Bullock, Allison. « La conservation de l'information numérique : ses divers aspects et la situation actuelle ». *Flash réseau* no 60, 22 octobre 1999. < <http://nlc-bnc.ca/pubs/netnotes/fnotes60.htm> >

LES FORMATS D'ARCHIVAGE ET DE CONSERVATION LES PLUS UTILISÉS

Dans les diverses expériences étudiées, plusieurs s'accordent pour dire qu'il est primordial de garder l'apparence, le cachet et l'originalité des publications scientifiques électroniques en plus, évidemment, du contenu informationnel.

En faisant le tour des projets, on constate que les plus utilisés sont les HTML, SGML et PDF. Le format XML est encore peu utilisé pour le moment mais il doit dorénavant être considéré.⁹⁸ JSTOR, pour ne citer qu'un exemple, envisage de migrer ses métadonnées dans une structure XML⁹⁹.

Projet	Formats
Pandora (Preserving and Accessing Networked Documentary Resources of Australia)	PDF, SGML, HTML
HighWire (États-Unis)	SGML, PDF
Muse (États-Unis)	HTML, PDF
Allen Press (États-Unis)	SGML
Institute of Electrical and Electronic Engineers (États-Unis)	SGML, PDF
American Astronomical Society (États-Unis)	SGML
American Institute of Physics (États-Unis)	PDF, SGML
Danemark, projet de dépôt légal des publications électroniques	ASCII Text, format d'image (par exemple TIFF)

Le PDF est un format propriétaire, largement utilisé et accepté par le milieu de l'édition. Son accessibilité à long terme demeure toutefois source d'inquiétude pour les archives nationales et les bibliothèques.¹⁰⁰

Pour le portail, le format de conservation sera le XML. La priorité du portail est d'assurer la conservation de ce type de document électronique. La garantie d'accès à long terme aux fichiers PDF est probable mais ne peut être affirmée d'une façon aussi certaine.

LES SUPPORTS D'ARCHIVAGE DE CONSERVATION

Les supports d'archivage et de conservation utilisés par les différents projets étudiés ne sont que très rapidement mentionnés, souvent ils ne le sont pas du tout. La raison est simple, l'accent est mis sur l'augmentation des contenus et, comme le coût des supports magnétiques ne cesse de décroître, on continue de les garder en ligne. Les projets procèdent évidemment à

⁹⁸ Hodge, Gail et Carroll Bonnie C. *Digital Electronic Archiving : The State of the Art and the State of the Practice*. International Council for Scientific and Technical Information, Information Policy Committee, 26 avril 1999, p. 60.

< http://www.icsti.org/icsti/99ga/digarch99_MainP.pdf >

⁹⁹ Kenney, Anne R. et Rieger Oya Y.. « Developing a Digital Preservation Strategy for JSTOR ». *RLG DigiNews*, vol. 4, no 4, 15 août 2000. < <http://www.rlg.org/preserv/diginews/diginews4-4.html> >

¹⁰⁰ Hodge, Gail et Carroll Bonnie C. *Digital Electronic Archiving : The State of the Art and the State of the Practice*. International Council for Scientific and Technical Information, Information Policy Committee, 26 avril 1999, p. 60.

< http://www.icsti.org/icsti/99ga/digarch99_MainP.pdf >

des copies de sécurité routinières sur divers supports magnétiques (ruban, cartouches, disques miroirs) ou optiques (cédérom, le DVD étant encore peu utilisé).

La question des supports mérite réflexion car leur durée est comptée en dizaine d'années et non plus en siècle comme c'est le cas pour le papier de très haute qualité et les microfilms. Les supports magnétiques subissent une double détérioration. D'une part, il y a l'affaiblissement progressif du champ magnétique, qui doit être rafraîchi. D'autre part, il y a la détérioration relative aux conditions environnementales (le taux d'humidité, les variations de température, la pollution, la poussière, etc.) Les supports optiques, sont plus durables mais ils restent sujets à la détérioration due à l'environnement (conditions ambiantes), aux matériaux utilisés pour leur fabrication, à la corrosion des différentes couches de métal, etc.)

LES STRATÉGIES EMPLOYÉES DANS LES DIVERS PROJETS

Pour pallier la détérioration rapide des supports, les différents projets que nous avons analysés utilisent différentes techniques telles que la redondance des données, la dispersion géographique, les copies de sécurité de routine, sur une base quotidienne, de façon à ce qu'aucun événement, tels qu'une panne ou un bris du matériel, une attaque de hackers, une catastrophe naturelle, etc., ne puisse détruire toutes les données.

Les bibliothèques ont aussi créé un modèle distribué d'archivage pour le matériel en ligne : LOCKSS (Lots of Copies Keeps Stuff Safe)¹⁰¹.

LOCKSS is a self-organizing, freeware-based, low-cost, voluntary approach to archiving online material, self-selected by participating institutions, that relies on consensus among several linked servers to determine authoritative states of files and restore lost or damaged files automatically.¹⁰²

Cet outil permet aux bibliothèques de conserver les publications en ligne sur les disques des ordinateurs locaux. Les publications qui ont été téléchargées localement ne sont pas effacées et sont continuellement confrontées aux mêmes publications qui sont en ligne pour s'assurer que leur contenu n'a pas été détérioré ou perdu. Si c'est le cas, les publications sont restaurées. Plus de bibliothèques utiliseront ce modèle distribué assurant une répartition géographique des fichiers, plus il y aura de copies des publications conservées.

Il s'agit d'une approche intéressante, différente de celles des copies de sauvegarde, mais encore trop récente pour tirer des conclusions. Toutefois, le portail devrait s'assurer de produire des documents compatibles avec le système LOCKSS.

L'ÉMULATION COMME STRATÉGIE DE CONSERVATION À LONG TERME

Différentes approches ont été suggérées pour conserver les documents électroniques à long terme : 1) l'application des normes et des standards pour les formats, 2) la conservation de

¹⁰¹ Pour en savoir plus sur LOCKSS: <<http://lockss.stanford.edu/projectdescbrief.htm>>

¹⁰² *Ibid.*

la technologie (les logiciels et le matériel informatique), 3) la migration des documents dans une forme accessible pour les générations futures, 4) l'impression des documents électroniques sur papier.

Aucune de ces solutions n'est totalement satisfaisante. Certains spécialistes préconisent une cinquième option : l'émulation.

L'émulation est une opération de simulation qui consiste à imiter le fonctionnement d'un ordinateur ou d'un logiciel sur un autre ordinateur généralement plus puissant d'une génération subséquente. On recrée virtuellement l'environnement matériel et logiciel d'origine. Les documents électroniques sont ainsi accessibles et lisibles sous leur forme originelle.

Cette stratégie peut s'avérer intéressante car elle n'altère pas les données et que l'émulation n'est effectuée que lorsque il y a un besoin réel d'accéder aux données, « à la volée ». Elle permet la conservation de l'aspect, du cachet et de l'originalité du document électronique aussi bien que de son contenu. Aucune autre opération n'est requise si ce n'est celle du rafraîchissement du support sur lequel sont les données.

Toutefois, cette solution demeure encore largement théorique et on peut s'attendre à ce que les coûts de recréation d'environnements technologiques complexes soient faramineux. Enfin, il y a encore peu d'études sur le sujet. Nous ne recommandons pas cette stratégie pour le portail dans l'immédiat.

GARANTIR L'INTÉGRITÉ ET L'AUTHENTICITÉ DES TEXTES

L'authenticité et l'intégrité des textes sont deux éléments essentiels de la communication savante. Les auteurs autant que les lecteurs veulent s'assurer que les documents électroniques n'ont pas été manipulés, altérés ou encore falsifiés après leur création et leur publication.

Afin que les publications électroniques soient protégées et garanties, plusieurs solutions existent tel le cryptage. Le cryptage est l'« opération par laquelle est substitué, à un texte en clair, un texte inintelligible, inexploitable pour quiconque ne possède pas la clé permettant de le ramener à sa forme initiale. »¹⁰³ « Essentiellement, il s'agit de coder un message de façon à ce que seul un interlocuteur connaissant la « clé » puisse le décoder. »¹⁰⁴ Cette problématique fait l'objet d'une analyse détaillée dans le chapitre 5 du rapport Érudit.¹⁰⁵

Le cryptage apparaît comme une solution adéquate pour la préservation de l'authenticité et de l'intégrité des publications électroniques scientifiques. Cependant, certains auteurs, croient qu'il peut y avoir certaines difficultés pour la conservation à long terme par exemple si un éditeur cesse ses activités et que la clé est perdue.

Signalons une autre solution, présentement en développement, par un groupe de travail du World Wide Web Consortium, le *Digital Signature Initiative*¹⁰⁶ (Dsign). Il s'agit d'un groupe de travail dont la mission est de développer une syntaxe XML représentant la signature des

¹⁰³ Office de la langue française. < <http://www.olf.gouv.qc.ca/> >

¹⁰⁴ Rapport Érudit. < <http://www.erudit.org/erudit/rapport/chap5/chap5.htm> >

¹⁰⁵ Rapport Érudit. < <http://www.erudit.org/erudit/rapport/chap5/chap5.htm> >

¹⁰⁶ <<http://www.w3.org/Signature/>>

ressources Internet (tout ce qui peut être référencé par une adresse URL). Le développement de ces signatures permettra d'assurer l'intégrité des données, de même que l'authentification. Cette initiative vise aussi le développement de procédures pour créer et vérifier de telles signatures. À terme l'objectif du Web est d'établir le « Web of trust ».

Le portail devrait opter pour la simplicité et l'élégance de cette dernière solution et chercher à implanter la signature numérique dès que la norme aura atteint le statut de recommandation du W3C.

LA RESPONSABILITÉ DE LA CONSERVATION ET DE L'ARCHIVAGE

Dans le rapport du *Task Force on Archiving of Digital Information*, il est suggéré que les éditeurs doivent être les premiers responsables de l'archivage et de la conservation de leurs publications. Ce point de vue est récurrent dans plusieurs autres textes sur le sujet. En raison de la complexité de ces tâches, et de leurs coûts inhérents, les éditeurs de revues savantes du Québec préféreront recourir aux services du portail et/ou des Bibliothèques nationales.

Généralement, par la loi du dépôt légal, les bibliothèques nationales s'assurent que toutes les publications de leur pays, soient acquises, conservées et rendues accessibles. Toutefois la législation sur le dépôt légal canadien (et québécois) ne couvre pas encore les publications diffusées en réseau. Comme il n'y a pas encore d'infrastructures nationales reconnues pour assurer l'archivage et la conservation à long terme de ces publications électroniques, qui doit assumer cette responsabilité ?

Le Canada étudie activement cette question et il est évident que la loi sera éventuellement modifiée pour inclure les publications en réseau. Dans l'attente d'une nouvelle législation, le dépôt des publications électroniques diffusées en réseau se fait sur une base volontaire dans de nombreux pays, par exemple au Canada, en Australie, en France, en Suisse, au Royaume-Uni et en Allemagne.¹⁰⁷

Plusieurs portails assument actuellement des responsabilités d'archivage, mentionnons JSTOR, dont les frais d'abonnements¹⁰⁸ couvrent ce service. Pour les petits éditeurs, l'avantage majeur de cette approche est qu'ils n'ont plus à se soucier de l'archivage et de la conservation et que le tout est fait à meilleur coût.

EN RÉSUMÉ

La conservation et l'archivage du patrimoine culturel des revues est au cœur du projet de portail. En premier lieu, le site d'hébergement devra offrir des assurances de sécurité à la fine pointe des technologies disponibles. Également, une redondance des données devra être assurée avec des collaborateurs choisis et répartis géographiquement (autres portails, bibliothèques nationales du Canada et du Québec). De plus, comme dans le cas du projet

¹⁰⁷ Martin, Libby. *Gestion des publication sélectroniques diffusées en réseau : état de la question dans divers pays*, Bibliothèque du Canada, 31 décembre 1999.

¹⁰⁸ Ces frais peuvent être divisés en deux : d'une part il y a les frais pour accéder aux archives, d'autre part les frais relatifs aux coûts de fonctionnement et de mise à jour des archives.

Muse, les abonnés institutionnels devraient recevoir une copie sur cédérom de la version XML de la base de données (sans le moteur de recherche) à chaque année.

Le portail devra poursuivre sa veille technologique en matière d'intégrité et d'authenticité des textes électroniques. Les groupes de travail du W3C sur la signature numérique et les métadonnées devraient donner sous peu des résultats tangibles, permettant d'assurer la confiance envers les documents distribués électroniquement.

Pour le portail, le format de conservation sera uniquement le XML. La priorité du portail est d'assurer la conservation¹⁰⁹ de ce type de document électronique. La garantie d'accès à long terme aux fichiers PDF est quant à elle probable mais ne peut être affirmée d'une façon aussi certaine. Le portail ne peut s'engager à faire migrer le format PDF vers les versions 5, 6 ou 7 car cela représenterait d'éventuels coûts supplémentaires inutiles. En effet, puisque le PDF est dérivé automatiquement de la version Postscript, il pourra être recréé à volonté. Le mandat de protéger la version Postscript pourra être confié aux Bibliothèques nationales.

En attendant que la loi sur le dépôt légal des documents électroniques soit adoptée, le portail devrait participer, dans les plus brefs délais, au dépôt volontaire de ses publications électroniques à la Bibliothèque nationale du Canada, en prenant des ententes afin que la diffusion ne se fasse qu'à partir du site du portail afin, notamment, de simplifier la compilation des statistiques de consultation.

L'archivage des périodiques du portail est donc vu comme une responsabilité partagée entre le portail et les Bibliothèques nationales. Toutefois, il ne faut pas oublier que la conservation des données telle que définie plus haut est essentielle pour l'utilisation (et donc la survie !) du portail. La conservation est donc au cœur des préoccupations du portail et devra être prise en compte dans les budgets de fonctionnement.

¹⁰⁹ Dans le sens large, défini en introduction.

CHAPITRE 6

PRÉVISIONS BUDGÉTAIRES

Les estimations des coûts de production et de diffusion s'appuient sur une analyse fine des différentes activités que nécessite cette partie du travail. Pour ce qui est des revenus, les postulats sont précisés dans le cours du texte. Pour faciliter l'identification des coûts des différentes fonctions du portail, nous distinguons par modules la production, la diffusion, les coûts de démarrage du portail et le budget de fonctionnement. Finalement, une estimation des ressources nécessaires pour l'aide à la transition des revues vient compléter cette section.

Compte tenu du mandat de cette étude et du temps à y consacrer, nous n'avons pu estimer les coûts d'implantation de la partie serveur de prépublications comprenant éventuellement les zones prépublications et littérature grise du portail. L'architecture n'étant pas aussi bien définie, il apparaissait difficile de faire une estimation des coûts d'implantation de ces zones du portail. La définition de ce service devrait se préciser à l'occasion du colloque de la fin octobre ainsi qu'à l'occasion d'autres rencontres avec des chercheurs, des directeurs de centres de recherche et des responsables des universités pour s'assurer de l'intérêt de tel service et de répondre aux besoins dans l'implantation de ce mode de communication dans le milieu des sciences humaines et sociales. Compte tenu des besoins des revues en ce qui a trait à la transition vers un mode de production et de diffusion électronique, nous estimons que cette partie du projet devrait faire l'objet d'un financement *ad hoc*.

COÛTS DE PRODUCTION D'UNE REVUE

Les coûts de production d'une revue sont évidemment reliés aux caractéristiques du document à produire telles que le traitement de tableaux complexes, une iconographie abondante, des formules mathématiques complexes ou des caractères non-latins. Pour les fins de cette étude, nous avons imaginé une revue inexistante : la revue moyenne. Cette revue imaginaire publie trois numéros d'environ huit articles de vingt pages par année. On retrouve en moyenne deux à trois tableaux par article et autour de six figures par numéro.

La chaîne de traitement proposée visant la plus grande automatisation possible, une étape importante avant même la production du premier numéro consiste à analyser la publication et à faire des ajustements dans la programmation des différents outils de conversion et de traitement, autant pour la production que pour la diffusion. Cette réalité explique la différence entre la première année de publication et la deuxième, le premier numéro demandant une intervention plus importante de la part d'un professionnel. Par la suite, la presque totalité du traitement est effectuée par un technicien qui pourra avoir recours au professionnel en place pour les situations plus complexes.

Coût de production d'une revue (XML/HTML/PDF)	
Premier numéro	4 200\$
Numéros subséquents	2 200\$
Total pour une première année de publication	8 600\$
Total pour les années subséquentes	6 600\$

Les coûts associés au traitement initial d'une revue couvrent les activités d'analyse de la publication, de création de la feuille de styles Word, de formation d'une personne de l'équipe de la revue à l'utilisation de la feuille de styles et la programmation. L'estimation des coûts

de production de la revue comporte la préparation et la vérification des fichiers « Word++ », la numérisation des images et des figures, la préparation des épreuves, la saisie des corrections, la préparation de la version XML de référence pour la production des autres formats et pour l'archivage, la version HTML de diffusion, le fichier Postscript pour l'impression des exemplaires papier, le fichier PDF pour la diffusion en ligne d'un format d'impression sur demande, la préparation des informations à transmettre au graphiste pour la réalisation de la couverture, la gestion du site Web de la revue et la préparation du sommaire pour la version électronique.

DÉMARRAGE DU PORTAIL

Le tableau de cette section donne un aperçu des coûts de démarrage du portail. Deux scénarios sont évoqués. La différence entre les deux scénarios se justifie d'une part, par des questions de choix d'équipement. Les choix techniques sont invariablement faits en fonction d'un cadre institutionnel, des ressources dont on dispose pour la réalisation du projet et de l'expertise de l'équipe chargée de la mise en œuvre. Ne disposant pas de cette information, il nous était impossible de proposer un budget précis. Toutefois, nous soulignons l'importance de faire des choix de technologies et d'équipements de qualité pour le développement à long terme du portail. Il ne faut pas ici sous-estimer les possibilités d'acquisition en collaboration entre les différentes institutions et agences concernées par la diffusion des résultats de la recherche.

D'autre part, le scénario 2 se distingue par quelques économies qui pourraient être réalisées par la collaboration avec d'autres portails de revues savantes, déjà à la phase d'implantation du projet. Nous pensons notamment à l'implantation du *Digital Object Identifier* et du service *CrossRef* qui pourraient se faire en collaboration avec le site canadien CyJournal/CyberRevue, en phase de conception également. Dans cette éventualité, le portail québécois n'aurait pas à développer une expertise aussi poussée pour la mise en route et la gestion complète de pareils systèmes.

Les prévisions suivantes sont basées sur une collection de 30 revues.

Démarrage du portail – Scénario 30 revues

	Scénario 1	Scénario 2
Infrastructure		
Serveur	45 000 \$	25 000 \$
Hébergement	12 000 \$	12 000 \$
Équipements informatiques (matériels et logiciels)	35 000 \$	35 000 \$
Démarrage des services de production et de diffusion		
Outil de recherche	30 000 \$	20 000 \$
Réalisation et test de la chaîne de traitement et traitement initial des revues	90 000 \$	90 000 \$
Création et gestion du site Web		
Conception du site Web	50 000 \$	20 000 \$
Organisation logique des périodiques		
Création de l'interface de navigation		
Conception et réalisation graphique du site		
Création de l'interface de consultation		
Choix d'un modèle de métadonnées	25 000 \$	10 000 \$
Programmation pour la création des métadonnées		
Création du système de fiches descriptives		
Achat d'un DOI		
Implantation du DOI		
Achat Services CrossRef		
Implantation de CrossRef		
Référencement	10 000 \$	5 000 \$
Bases de données bibliographiques		
Autres portails		
Outils de recherche Web		
Répertoires		
Gestion des droits d'accès et mesure de l'usage		
	50 000 \$	
H 1 – Tous les abonnements, institutionnels et individuels		
H2 et 3 – Abonnements institutionnels seulement		16 000 \$
Total :	347 000 \$	233 000 \$

BUDGET DE FONCTIONNEMENT DU PORTAIL – SCÉNARIO POUR 30 REVUES

Les dépenses de fonctionnement du portail sont essentiellement de trois ordres : les ressources humaines dédiées à la production et à la diffusion, et l'infrastructure matérielle. La gestion des activités reliées à la production et la diffusion électroniques de 30 revues justifie la création d'un poste d'administration assisté d'un secrétariat pour coordonner la production et s'assurer du respect des calendriers. Cette fonction assurerait un lien entre les équipes de production, les pigistes chargés de la mise en page et les revues tout en étant la courroie de transmission de l'information vers le conseil d'administration du portail.

Une fois les traitements initiaux des revues terminés, un poste de programmeur/spécialiste de l'information devrait suffire pour d'une part, répondre aux questions des techniciens et ajuster les programmes de conversion aux différents cas de figure nouveaux qui se présente pour la production technique des revues et d'autre part, assurer le développement et la gestion du système de diffusion. Les deux postes techniques assureraient pour leurs parts la production courante des versions électroniques des revues et la préparation des fichiers à transmettre aux pigistes chargés de la mise en page des revues.

Les ressources estimées pour l'infrastructure matérielle seraient consacrées à l'hébergement des données dans un environnement sécurisé donnant accès à une large bande passante ainsi qu'au renouvellement des matériels et logiciels informatiques des équipes de production technique et de diffusion.

Le total des dépenses de fonctionnement avant la gestion des abonnements s'élève à 344 000 \$ par année. De cette somme, 182 000 \$ représentent les coûts de production techniques qui seraient assumés par les revues à même les subventions éventuellement ajustées qu'elles obtiennent. La différence, soit 162 000 \$, serait ainsi dédiée à la diffusion.

Pour ce qui est de la gestion des abonnements, deux scénarios sont présentés. Le premier correspondant à la centralisation au portail de toute la gestion des abonnements des revues, institutionnels et individuels. Nous évaluons le coût de la gestion centralisée à 9 \$ par abonnement. Selon l'information transmise par le CRSH, le nombre d'abonnements aux revues québécoises s'élève à près de 11 000, ce qui porterait le coût de l'implantation d'une telle mesure à 99 000 \$ par année. Toutefois, dans l'éventualité où le portail ne gèrerait que les abonnements institutionnels à la version électronique, les coûts de gestion des 6 000 abonnements seraient de 42 000 \$, à raison de 7 \$ par abonnement.

Dépenses	
Ressources humaines pour le portail	
Un poste d'administration (production et diffusion)	60 000 \$
Un poste de secrétariat	35 000 \$
Un poste de programmeur/spécialiste de l'information (production et diffusion)	70 000 \$
Deux techniciens (production et diffusion)	80 000 \$
Pigistes – Mise en page et préparation du PDF	72 000 \$
Infrastructure matérielle	
Équipements	15 000 \$
Hébergement	12 000 \$
Total des dépenses avant la gestion des abonnements :	344 000 \$
Gestion des abonnements	
Abonnements électroniques institutionnels	42 000 \$
Total des dépenses :	386 000 \$

Le premier scénario est celui, déjà développé dans le chapitre 4, qui ferait des revues savantes québécoises un bien public, donc diffusées gratuitement en version électronique dans le Web. Selon ce scénario, on pourrait compter sur les revenus des activités de production technique par les revues à même leurs subventions et, d'autre part, sur une subvention à la diffusion qui serait accordée au portail. Ce scénario ne nécessite pas la mise en place d'un système de gestion des accès au serveur, pas plus d'un service de gestion des abonnements électroniques.

Le deuxième scénario est celui de l'accès payant. Ici encore, la production technique des différentes versions doit être assumée par les revues, comme c'est présentement le cas pour la version papier. Toutefois, il nous faudrait augmenter les revenus pour permettre de financer le système de diffusion. Compte tenu de la situation financière des revues québécoises, on peut difficilement penser à une solution qui ferait en sorte que leurs revenus en seraient diminués.

Revenus – Scénario 1 – revues comme bien public

Production des revues (assumée par les revues avec les subventions)	182 000 \$
Subvention au portail	162 000 \$
Total des revenus pour le scénario 1 :	344 000\$

S'appuyant sur des expériences diverses¹¹⁰, les pratiques se situent autour d'une augmentation de 30% des tarifs d'abonnement reliée à l'accès à la version électronique des revues. Les revenus générés par cette augmentation pourraient être alloués au portail. Les prévisions de revenus impliquent que 90% des bibliothèques qui sont abonnées présentement aux revues s'abonneraient à la version électronique, malgré l'augmentation de 30%. Cela peut sembler optimiste comme prévision. Toutefois, il ne faut pas oublier que la moyenne des tarifs d'abonnement institutionnel à nos revues est autour de 60 \$. Une augmentation de 30% pour avoir la version électronique ne représente pas une somme faramineuse, précisément 18 \$, portant donc le coût de l'abonnement à 78 \$.

Ce scénario implique l'implantation d'outils de filtrage des accès et la création d'un service de gestion des abonnements institutionnels électroniques. Les dépenses de fonctionnement de ce scénario, malgré les revenus qu'il permettrait éventuellement de générer, ne permettent pas de présenter des prévisions budgétaires équilibrées. Le manque à gagner se situe ici à 109 500\$ qui devraient être financés par ailleurs.

Revenus – Scénario 2 – accès électronique filtré

Production des revues (assumée par les revues avec les subventions)	182 000 \$
Augmentation de 30% du prix des abonnements pour l'accès aux versions électroniques	94 500 \$
Total des revenus pour le scénario 2 :	276 500\$
Manque à gagner :	- 109 500\$

¹¹⁰ Notamment, celle du projet MUSE de Johns Hopkin University Press.

COÛTS DE TRANSITION POUR LES REVUES

La transition du mode de production entraîne des impacts pour la revue qui sont précisés à la fin du chapitre 3 de cette étude portant sur la production. Les coûts prévus pour le traitement initial de la revue comprennent une journée de formation pour l'utilisation de la feuille de styles. Cependant, il importe de s'assurer que chaque revue dispose d'un équipement adéquat en fonction des nouvelles procédures de production technique. L'acquisition d'un poste de travail bureautique performant et bien relié à Internet ainsi qu'une formation d'appoint contribueraient dans plusieurs cas à faciliter la transition.

Coûts de transition – Revues

Acquisition d'un poste de travail performant et formation d'appoint	5 000 \$
Total pour 30 revues :	15 000\$

CHAPITRE 7

CAHIER DES CHARGES

- **Infrastructure organisationnelle**

Être une entité institutionnelle et légale, sans but lucratif, éventuellement basée sur une formule coopérative

Réunir dans cette entité institutionnelle des acteurs de milieu universitaire ayant une expérience en publication électronique

Être la propriété du milieu universitaire

Infrastructure intellectuelle

Spécifier les types d'expertises et les profils de compétences requises pour le portail

Spécifier les procédures de veille technologique

Développer une stratégie afin d'être un acteur majeur dans l'adoption et le développement de normes en matière de publications électroniques à l'échelle nationale et internationale

- **Infrastructure matérielle**

Héberger le portail en s'assurant de la qualité de la sécurité, de la stabilité et de la constance des services

Opter pour un serveur dédié et autres équipements matériels et logiciels de qualité, c'est-à-dire, permettant la croissance et la stabilité des opérations du portail

Implanter le portail sur une partie du réseau donnant accès à une large bande passante

S'assurer de bénéficier d'une surveillance continue des équipements et adopter des procédures de sauvegarde selon les normes

SYSTÈME DE PRODUCTION

Réaliser une chaîne de traitement XML et autres normes affiliées pour la production des revues

Opter pour des procédures permettant la production de formats de diffusion Web et mise en page à partir d'une même chaîne de production

Définir une procédure d'automatisation optimale de la chaîne de traitement pour la production de tous les formats électroniques de la revue

Privilégier l'utilisation des logiciels gratuits ou du domaine public

Créer ou adopter un modèle de métadonnées, pour la description des articles (courant et rétrospectif), des prépublications et autres documents diffusés par le portail, qui est compatible avec les modèles utilisés sur le plan international (par exemple : DC et OAMS)

Intégrer dans la chaîne des outils permettant la production automatique des métadonnées

Intégrer le DOI et le service CrossRef pour l'identification permanente et le référencement des documents et des références bibliographiques

Organiser en collaboration avec des entreprises spécialisées, un service de numérisation rétrospective pour les revues

À la suite de la définition précise de la zone prépublications du portail, créer l'infrastructure nécessaire pour accueillir et diffuser des prépublications et autres documents de la littérature grise

SYSTÈME DE DIFFUSION

Créer le site Web du portail

Créer l'interface de consultation comprenant les services de base suivants :

Outil de recherche parcourant les métadonnées et le plein texte structuré

Structure de navigation claire

Gestion des utilisateurs et droits d'accès selon le mode de financement retenu

Mesure de l'usage et de la fréquentation du portail

Planifier la conception et le développement du portail en fonction de l'intégration éventuelle des services complémentaires suivants :

Diffusion sélective de l'information

Aide à la lecture (annotations, liens et signets électroniques)

Personnalisation de l'interface de consultation

Développer et implanter une stratégie de diffusion sur la Toile par le référencement du portail dans les outils de recherche et les répertoires appropriés

Développer et implanter une stratégie visant à augmenter la visibilité des revues et prépublications dans les autres portails ou serveurs de prépublications existants

GESTION DE L'INFORMATION

Mettre au cœur du projet la gestion, la conservation et l'archivage du patrimoine scientifique confié au portail

Utiliser des formats normalisés, non propriétaires, pour garantir l'archivage et la conservation des documents électroniques

Adopter un modèle de gestion des documents électroniques qui répond aux normes internationales en ce qui a trait aux formats d'encodage de l'information, aux supports utilisés et à la redondance et la migration des données.

Participer au dépôt légal électronique volontaire de la Bibliothèque nationale du Canada et éventuellement de la Bibliothèque nationale du Québec

PLAN D'AFFAIRES

Présenter un calendrier de mise en œuvre des activités sur un période de trois ans

Établir un budget de dépenses selon ce calendrier

Élaborer une politique de financement du portail, en fonction des paramètres qui seront précisés par le Fonds FCAR

CONCLUSION

L'étude de faisabilité a permis d'illustrer que le projet d'un portail de production, de diffusion et de gestion des publications électroniques est un projet majeur pouvant avoir une portée considérable. Le projet est porté par une tendance mondiale où différents acteurs participent à la transformation de la publication et de la communication scientifique. Le Québec, se dotant d'une infrastructure pour la production et la diffusion de l'information scientifique, participera à ce mouvement en développant un modèle, qui, tout en étant relié aux pôles importants au niveau international, correspondra à sa situation particulière et aux besoins de ses communautés scientifiques.

La transition vers la publication électronique est bien engagée, sans pour autant que les publications imprimées disparaissent. La simultanéité des versions électronique et papier durera encore un temps qu'il est difficile d'apprécier. L'aspect central dans le processus de transition, c'est que la version électronique devienne la version première, pour l'auteur, lorsqu'il conçoit l'exploitation et la mise en forme de ses résultats de recherche, pour le chercheur, lorsqu'il procède à l'investigation des sources, pour les directions de revue lorsqu'elles arrêtent leur politique éditoriale et réfléchissent au contenu, à la forme et à la place de la revue dans la communication scientifique et, enfin, pour l'éditeur lorsqu'il orchestre le processus de publication. Lorsque la version électronique deviendra le cadre de référence premier, le centre de gravité sera décisivement déplacé. Une mutation profonde s'est engagée et le portail québécois pourrait l'accompagner, y participer et l'amplifier, pour peu que l'action publique soit avisée et que les ressources nécessaires soient mobilisées.

Pour cette dernière dimension, le portail de production et de diffusion constituerait une infrastructure au service de la valorisation de la recherche menée au Québec et des institutions de publication savante qui contribuent à sa diffusion. Ce devrait être, à terme — et il faut le prendre en compte concrètement dès le départ —, le lieu de convergence et le carrefour des différents genres d'écrits scientifiques, avec l'aménagement de zones distinctes pour les revues savantes, pour les prépublications et pour la littérature grise. Afin de répondre à la situation des revues et pour consolider ces institutions de publication savante, il est judicieux de s'appuyer, dans un premier temps, surtout sur les revues pour bâtir le portail. Cette structure d'accueil devrait agréger d'autres modes de communication scientifique et établir des passerelles à l'interne entre les corpus, comme avec des partenaires externes. La mission du portail devrait déborder le Québec par son rayonnement, il va s'en dire, mais aussi par son contenu. Ce qui importe c'est de doter le Québec d'une infrastructure de stature et de dimension internationale. D'ailleurs, on voit mal comment on pourrait imposer des restrictions ou des exclusives sur l'origine de la documentation, sans régionaliser le portail, ce qui serait de mauvais aloi.

Il s'agit, à n'en pas douter, d'un projet d'envergure, non seulement pour sa mission, mais aussi pour les ressources nécessaires. Nous avons dégagé des ordres de grandeur qui montrent qu'un portail semblable, même en se concentrant dans un premier temps sur les revues savantes, exige une injection de ressources significatives, non seulement pour la période de démarrage, mais aussi pour les dépenses de fonctionnement pour l'avenir prévisible. Cette conclusion est conforme aux expériences étrangères. Il s'agit d'un investissement dans une infrastructure au service de la valorisation et du rayonnement de la recherche. Même après la phase d'implantation, on peut difficilement imaginer, avec les informations dont on dispose aujourd'hui, que le portail se financerait grâce aux économies que dégagerait l'introduction des technologies de l'information dans la publication et de la diffusion électroniques. Il y aura un déplacement des dépenses et une réorganisation des pratiques, certes, mais

aussi nécessité de ressources nouvelles pour les cinq premières années tout au moins. Par ailleurs, cette infrastructure de diffusion électronique de la recherche peut, en toute logique, être considérée comme un bien public qui devrait trouver un autre mode de financement que les recettes venant de la vente du produit de la recherche. Les différentes estimations montrent que la mise en place de mécanismes de filtrage d'accès, qui impose des frais non négligeables, donne un résultat net plutôt faible pour ce qui est des rentrées, si bien que le soutien public serait très peu affecté (à la baisse) par la présence de modes de paiement pour l'accès aux documents en ligne.

On comprend aisément que le Fonds FCAR, en raison de sa mission dans le développement de la recherche au Québec, soit l'initiateur de ce projet, mais on voit mal comment il pourrait suffire seul à supporter ce projet. Le Fonds FCAR peut être un fantastique levier, en étayant et en supportant les objectifs et les conditions de réalisation de ce projet, et en intéressant les autres agences gouvernementales, qui ont une responsabilité dans ce secteur ou qui peuvent y apporter un appui sous une forme ou une autre. Les récentes réformes substantielles de prestations de services dans le secteur public montrent qu'il est périlleux de s'engager dans cette voie en l'absence de ressources humaines et financières suffisantes. Non pas que l'immobilisme soit préférable, mais qu'il faille ajuster l'ampleur des transformations provoquées aux capacités réelles de les mener à terme. La sous-estimation des besoins réels est un trompe-l'œil qui ne résiste pas longtemps à l'épreuve de la mise en œuvre de réformes et qui peut donner des résultats bien éloignés de ceux qui sont escomptés. On peut penser qu'il faudrait une concertation et une collaboration fructueuses entre agences gouvernementales pour supporter cette transition vers la publication électronique et, ce faisant, pour mettre en place le portail. Pour aller vite, on peut parler de trois composantes dans le financement du portail et de la transition des revues à la publication électronique : 1° les coûts d'implantation du portail qui ne sont pas récurrents, 2° le financement adéquat des revues pour faire face à leurs obligations, en général, et pour s'adapter aux procédures de la publication électronique, 3° la participation aux frais de diffusion du portail. Ces trois composantes pourraient correspondre à des partenaires distincts, mais dont l'action complémentaire permettrait la réalisation de ce projet. En tout état de cause, le Fonds FCAR devrait pouvoir être le levier pour rendre possible ce montage.

On ne peut minimiser l'importance et la signification des transformations que suppose la transition vers la publication électronique pour les revues savantes. Fort heureusement, cette transition est supportée par les changements qui s'opèrent dans l'environnement général de la recherche et du milieu universitaire. Il s'agit d'une tendance lourde favorable aux actions à entreprendre dans les diverses sociétés. Pour une institution publique qui veut susciter ce mouvement dans une société particulière, il importe d'impulser cette transition sur un mode pro-actif, sans pour autant céder à la tentation volontariste. La politique de la table rase est mal avisée. Il convient plutôt, partant d'une bonne connaissance du milieu visé, de fixer des objectifs définis et spécifiques, qui pointent résolument dans la direction retenue. Parce qu'on ne peut engager, sur la base du plus faible dénominateur commun, une transition majeure mettant en cause les façons de faire, l'action publique exercera nécessairement une tension auprès du milieu. Il reste à savoir jusqu'où cette ligne de tension peut être poussée sans être rompue. La connaissance du milieu et une communication directe et franche avec les acteurs sont des guides précieux, mais une vision stratégique, prenant en compte la direction générale, les objectifs particuliers, les ressources disponibles, l'étalement des actions particulières et l'évaluation des résultats, possède de grands

avantages pour clarifier les intentions, élaborer la politique d'intervention et définir les actions à entreprendre.

ANNEXE 1.1
COMPTE RENDU DES GROUPES
TÉMOINS
DE QUÉBEC ET DE MONTRÉAL

6 ET 11 SEPTEMBRE 2000

REVUES PARTICIPANTES

QUÉBEC

MONTRÉAL

Anthropologie et sociétés	L'Actualité économique
Les Cahiers de droit	Annales des sciences mathématiques du Québec
Cahiers de géographie du Québec	Cahiers québécois de démographie
Écosciences	Circuit
Ethnologies	Cinéma
Études internationales	Criminologie
Études littéraires	Frontières
Laval théologique et philosophique	Géographie physique et quaternaire
Protée	Meta — Journal des traducteurs
Recherches sociographiques	Politique et sociétés
Relations industrielles	Recherche amérindienne au Québec
	Regiliologiques
	Revue de droit de McGill
	Revue des sciences de l'éducation
	Revue québécoise de linguistique
	Voix et images

MISE EN SITUATION

Ce compte rendu cherche à faire ressortir les principales idées qui ont été mises de l'avant par les directeurs de revues lors des rencontres du 6 septembre, à Québec, et du 11 septembre à Montréal.

LE PASSAGE À L'ÉLECTRONIQUE

Les revues sont conscientes que la tendance actuelle se dirige vers l'électronique. Plusieurs revues ont déjà une page Web avec les tables des matières de leurs numéros ainsi que des résumés des articles. Les directions des revues sont ouvertes à cette réalité et sont prêtes, pour la plupart, à faire le saut.

Plusieurs revues sont donc prêtes à aller de l'avant avec une version électronique de leurs revues et le projet du portail. L'électronique est une voie d'avenir et la transition est essentielle pour conserver et augmenter la visibilité des revues québécoises. Cependant, certaines conditions sont essentielles pour assurer cette transition vers l'électronique.

Certaines revues craignent que le passage ne soit pas si facile qu'il puisse le sembler. Il implique des changements dans les façons de faire à tous les niveaux et, à cet effet, une stratégie en plusieurs étapes serait salutaire. Le changement doit se faire progressivement en tenant compte de la réalité du milieu des revues qui en est un d'artisans et de bénévoles.

MAINTIEN DU PAPIER

Les revues sont unanimes en ce qui concerne la nécessité du papier. Elles croient que le développement d'une version électronique ne doit pas se faire au détriment de la version papier. Selon les directions, la version électronique ne doit pas être opposée à la version papier.

Elles sont bien contentes et rassurées d'entendre parler de la version papier dans les hypothèses du portail. Le Fonds FCAR, dans son rapport, donnait l'impression que le papier n'était plus nécessaire et les revues sont loin d'être en accord. Elles croient, au contraire, que le papier a encore sa place, et ce, pour longtemps. Plusieurs raisons les amènent à penser de cette façon.

- 1) Les habitudes des lecteurs changent lentement. Encore aujourd'hui, ils sont intéressés à recevoir 3 à 4 fois par année un numéro papier qu'ils pourront feuilleter, annoter et conserver dans leur bibliothèque.
- 2) Plusieurs des abonnements proviennent de l'étranger où l'accessibilité à un ordinateur et à Internet est difficile. Pour les pays d'Afrique et de l'Europe de l'Est, notamment, la suppression de la version papier équivaut à la suppression pure et simple de la revue. C'est donc contraire au vœu d'une meilleure

diffusion et visibilité des revues québécoises à l'étranger.

- 3) Une proportion non négligeable des abonnements individuels provient hors du milieu universitaire. Il s'agit de professionnels, qui travaillent sur le terrain et qui se tiennent au courant de ce qui se passe dans leurs domaines respectifs. Pour ces gens, la version papier est plus pratique, plus facilement consultable.

LE PORTAIL COMME INFRASTRUCTURE COMMUNE

La majorité des revues croient que le portail est un projet intéressant dans la perspective d'une infrastructure commune. Le portail permet de regrouper les forces, d'offrir des services que les revues seraient incapables de s'offrir individuellement.

Selon les revues, la production d'une version électronique ne peut se faire artisanalement comme c'est le cas avec la version papier. Des compétences particulières, des ressources communes, des moyens techniques et financiers sont nécessaires. Le portail permet de donner forme à cette infrastructure commune.

Dans cette optique, les revues considèrent qu'il est essentiel de tenir compte de la complexité de leur situation. Les revues savantes québécoises ne constituent pas un tout cohérent, mais plutôt un ensemble de revues avec des situations et des besoins différents.

Cette infrastructure commune apporte des conséquences importantes au niveau des façons de travailler pour les revues. Cela amène des modifications dans les façons de faire et peut aussi appeler à une révision complète du mode de fonctionnement actuel des revues et du rôle des collaborateurs.

Pour les revues qui s'occupent directement de toutes les étapes de la production, il y a une peur de perte de contrôle de la production. Étant habituées à tout faire elles-mêmes, ces revues craignent que les normes communes de production et les nouvelles façons de faire les amènent à perdre le contrôle de la production de leur revue.

Ces directions de revue craignent également une perte de contrôle de la qualité de leur revue, puisque la production technique des revues sera assurée par le portail. Elles se demandent qui fera le suivi, qui s'assurera de la qualité ? Cette crainte s'explique du fait que les directions croient qu'elles ne seront plus aussi présentes à chaque étape de la production.

La différenciation faite entre le travail intellectuel assumé par les revues et le travail technique fait par le portail a été bien perçue. Elles ont aussi bien reçu la reconnaissance par le portail de l'autorité éditoriale des revues.

Néanmoins, certaines revues gardent quelques inquiétudes. Certaines appréhendent une centralisation de la production qui peut amener une standardisation. Tant au niveau des contenus que des présentations, les directions des revues désirent conserver leur contrôle. Elles considèrent comme primordial de rester indépendantes, originales, différentes des autres revues et de conserver leur équipe de production. Pour bien des revues qui ont mis

sur pied une équipe de production au fil des années, il est essentiel de pouvoir conserver cette équipe.

Quant à la diffusion, l'idée d'une infrastructure commune et d'un portail présente des avantages évidents pour les revues.

Certaines directions de revue considèrent, néanmoins, qu'il ne faut pas se limiter à un portail "local". Le portail doit être une porte d'entrée pour le Web et ne doit pas limiter la diffusion internationale, qui est importante. Des liens avec d'autres portails doivent être établis ; le référencement dans les grandes bases de données et les index est également important pour la diffusion et l'impact des revues québécoises.

Enfin, certaines revues s'interrogent sur la nature du portail. Sera-t-il un portail exclusivement pour les revues financées par le Fonds FCAR ou bien un portail pour toutes les revues québécoises ?

LE FINANCEMENT

La situation financière actuelle des revues savantes québécoises est précaire. Dans la plupart des cas, elles font des déficits année après année et, avec leurs équipes qui sont réduites, composées bien souvent de bénévoles et sous-financées, c'est leur capacité de maintenir une grande qualité de contenu et une reconnaissance dans le milieu scientifique québécois et étranger qui est compromise.

Dans cette perspective, la proposition d'un portail et l'aide du Fonds FCAR sont appréciées par les directions des revues. Elles considèrent qu'autrement elles n'auraient pas les possibilités de se positionner dans le Web avec tous les services nécessaires pour y parvenir.

Cependant, le financement demeure une question primordiale pour les revues. Il est essentiel que le portail ne soit pas financé à même l'enveloppe actuellement disponible pour les revues. Un financement indépendant pour le portail est capital afin d'assurer la survie des revues québécoises. En fait, le portail ne doit pas être financé au détriment des revues qui éprouvent déjà des difficultés sur le plan financier.

Certaines directions de revue perçoivent même le portail comme une voiture de luxe qu'elles doutent fort d'avoir les moyens de s'offrir. Le portail est une offre de services, mais ces revues s'interrogent sur leur capacité de s'offrir ces services.

D'autres croient aussi que le portail peut les amener à accroître leur déficit, en raison des dépenses supplémentaires qui seront occasionnées : nécessité d'une équipe compétente, besoins techniques, etc.

Les directions des revues perçoivent aussi un manque de réalisme économique de la part du Fonds FCAR. Il va devoir investir plus d'argent qu'il ne le fait actuellement pour rendre un projet tel que celui-là possible. Dans la situation actuelle, plusieurs revues s'interrogent puisque le Fonds FCAR a plutôt tenu un discours de restrictions budgétaires ces dernières années en réduisant notamment le nombre de revues financées au dernier concours.

Les directions des revues considèrent comme essentiel qu'il y ait une proposition de financement des institutions publiques qui soit durable et sérieuse. Le passage doit être garanti pour assurer l'avenir des revues. Selon elles, la continuité est importante.

En ce qui concerne le modèle économique, l'hypothèse de la gratuité fait peur à plusieurs revues, car celles-ci se retrouveraient complètement dépendantes des organismes subventionnaires, n'ayant plus de sources de revenus externes. Elles envisagent donc l'hypothèse de l'accès payant plus positivement.

SERVICE D'HÉBERGEMENT DE SERVEURS (CO-LOCATION)

Le RISQ inc. offre un service d'hébergement de serveurs qui permet aux entreprises d'exploiter un service d'information sur Internet (HTTP, HTTPS, FTP, Real, Usenet, autres) à partir de leurs propres serveurs, *colocalisés* dans les locaux du RISQ. Les serveurs du client sont installés en permanence sur la dorsale du RISQ, et peuvent disposer d'une bande passante et d'un temps réponse Internet très avantageux. En effet, le RISQ offrant à ses membres et clients abonnés une interface Internet **sans contention**, les serveurs disposent des mêmes fonctionnalités et avantages.

LES BÉNÉFICES DE LA 'COLOCALISATION' D'UN SERVEUR AU RISQ SONT :

ACCÈS :

- l'accessibilité du serveur à très large bande passante, à des coûts abordables;
- des performances réseau supérieures (débit, latence);
- l'accessibilité à très large bande aux requêtes provenant du réseau canadien de la recherche CA*net 3, et de tous les réseaux de recherche internationaux;
- la proximité (au sens de l'adressage) du QIX, le centre d'interconnexion Internet québécois, où sont interconnectés les principaux fournisseurs d'accès Internet du Québec;
- un serveur hébergé au RISQ peut être visible ou non à partir d'Internet. Par exemple, on peut héberger un serveur Web (visible d'Internet), et un serveur de base de données (auquel seul le serveur Web peut accéder).

GESTION :

- la surveillance continue des équipements par le personnel du RISQ;
- l'assurance de la continuité du service grâce aux installations de sécurité du RISQ.

DESCRIPTION DU SERVICE

- Largeur de bande disponible : selon vos besoins, minimum T1 (1.5 Mbps) -Inclus dans le tarif de base);
- Installation de l'appareil ou des appareils clients sur un châssis; 400 W de puissance électrique sont compris;
- Création d'un réseau local virtuel privé (VLAN) pour assurer l'intégrité et la sécurité des données;
- Surveillance 7 /24 des appareils clients par des logiciels de surveillance;
- Climatisation et d'alimentation garantie (UPS et génératrice);
- Prise de copies de secours selon l'horaire du RISQ ou selon un horaire à déterminer avec le client;

- Possibilité de redémarrer le ou les appareils clients sur demande;
- Maintenance des DNS pour établir les noms logiques des serveurs;

OPTIONS

- Maintenance du système d'exploitation des appareils;
- Plan de relève.

COÛT DU SERVICE

Le tarif 'Clients et Abonnés' pour l'hébergement d'un serveur est de : CAN\$ 15 000 / an

Le tarif 'Membres' de l'hébergement d'un serveur est de : CAN\$ 12 000 / an.

ANNEXE 2.2

CAHIER DES CHARGES TECHNIQUE

Dans le but de formuler un cahier des charges technique pour un projet de portail québécois de revues savantes, on trouvera ci-dessous une description technique succincte des fonctionnalités envisageables pour un tel portail, une description des contenus et de l'espace disque nécessaire, une description de l'architecture capable d'assurer la permanence du service, sa surveillance et sa sécurité, enfin une description des systèmes et des infrastructures recommandés garantissant que le portail sera en mesure de remplir sa mission.

DESCRIPTION TECHNIQUE DU PORTAIL

FONCTIONNALITÉS (SYSTÈME DE DIFFUSION)

Le portail souhaité est un portail de type vertical qui devrait pouvoir permettre (pas nécessairement dès la phase d'implantation) :

- l'adaptation de l'affichage du poste client en fonction du navigateur et de son système d'exploitation (OS);
- la personnalisation des accès des visiteurs en fonction de leurs intérêts (profil lecteur) et éventuellement de leurs droits d'accès;
- de créer et/ou s'abonner/désabonner à des zones d'intérêts personnels (channels), mis à jour dynamiquement (push) ou par demande de l'utilisateur (pull); avec possibilité de paramétrer la zone (taille, bordure, fond, police, etc.) ;
- de s'abonner/désabonner à des bulletins d'information; à l'envoi d'avertissements; à des calendriers;
- de montrer des 'caméos', i.e. de petits aperçus de données ou statistiques, d'application, ou de page Web;
- de choisir et de modifier ses propres paramètres et formats, d'ajouter et d'ôter des liens;
- d'enregistrer des habitudes d'usage;
- d'utiliser un outil de recherche plein texte et exploitant des métadonnées;

ARCHITECTURE

L'architecture du portail est fonction de la vision de son rôle et de l'étendue des services qu'on désire qu'il rende. Nous prenons pour acquis dans ce qui suit que dans la phase initiale du projet, le Portail devra se considérer le dépositaire, le “gardien” de la copie numérique officielle de ses corpus, à la façon d'une Bibliothèque Nationale.

En principe l'archivage n'est pas du ressort de l'éditeur ou du portail. Mais ce serait ignorer qu'il n'y a pas pour l'instant au Québec de système d'archivage établi avec des responsabilités claires pour les documents numériques. Le portail ne peut faire comme si la question de l'archivage était réglée. Il doit donc prendre sur lui d'assumer cette responsabilité. Avant le transfert de la responsabilité de l'archivage à un organisme ad hoc, le portail doit s'assurer de la conservation à long terme de ses corpus, ce qui correspond à sa mission dans la mesure où il s'engage à fournir un service sur la durée, et se préparer à les faire éventuellement migrer. Cette décision a naturellement un impact important sur la sécurité et elle entraîne des coûts additionnels qui peuvent doubler la facture des infrastructures matérielles, et augmenter les besoins en personnel.

Pour assurer la permanence du service et la résistance aux pannes, on recommande d'habitude de dédoubler les installations :

- Connexion réseau dédoublée (redondance) ;
- Alimentation électrique dédoublée ;
- Installation de deux machines identiques. La tâche des services est répartie entre les deux machines (load balancing). De plus ces deux machines surveillent tous les signes vitaux l'une de l'autre (processus logiciels, connexion réseau, alimentation électrique). Si l'une des machines vient à défaillir, l'autre prend immédiatement la relève. Entre les deux machines on installe un système de copies de sécurité de niveau RAID 5 (avec disque “hot spare”). Cette option n'est pas retenue dans le devis technique ci-dessous.
- Les installations doivent pouvoir supporter une augmentation de trafic et de traitement important (scalability). Pour prévenir les augmentations temporaires, on peut utiliser les deux machines mentionnées ci-dessus pour se partager et équilibrer la tâche (load balancing), mais quand la charge augmente de façon continue, on recourt à des “grappes” (clusters) de machines.
- *Caching* : Mettre en antémémoire les données les plus demandées pour en faciliter l'accès est une solution aux augmentations de charge ponctuelles ou continues. La cache peut être associée au proxy (voir plus bas) ;
- Des copies de sécurité (backups) autovérifiants (intégrité et lisibilité des données) doivent être prises selon une périodicité à établir (système Véritas ou Legato) ;

- L'archivage est discuté dans une autre partie de l'étude. Soulignons seulement ici qu'il s'agit d'assurer que toutes les données soient disponibles, mais pas nécessairement sur disque. On peut installer un système étagé. On peut disposer d'un archivage local, sur une machine dédiée, et/ou sur des bandes placées dans un coffre fort, voire à la banque. Plus intéressant serait un système éloigné dédié à l'archivage (dans une institution ad hoc), accessible par réseau numérique en tout temps. Dans tous les cas, on doit installer un système de synchronisation des archives. Autre aspect souvent négligé : on doit recopier les données régulièrement en s'assurant de leur intégrité et de leur lisibilité. Ceci a pour but non seulement d'assurer la permanence du service, mais aussi d'assurer la permanence des données à travers les changements technologiques.
- Les questions de sécurité informatique et de protection du serveur contre les attaques sont la responsabilité de l'administrateur du système (*sysadmin*). Les configurations du système et des divers outils sont copieusement documentées sur le site du World Wide Web Consortium (<http://www.w3c.org/Security>) ;
- L'architecture la plus sécuritaire et protégeant le mieux les données est une architecture étagée. Idéalement le serveur Web (HTTPD) et la base de données (corpus) devraient être sur des machines différentes. Les données sont sur une machine inaccessible de l'extérieur, et à laquelle seul le serveur Web a accès. Seul le serveur Web est accessible de l'extérieur (Internet), mais pas directement : on place devant lui un *proxy*, machine qui reçoit toutes les requêtes, les filtre, résiste aux attaques du genre "dénier de service" et pourrait permettre, par exemple, de différencier les requêtes venant du réseau des universités québécoises, de celles venant du reste du Canada, ou du reste du monde. Le proxy peut comporter une *cache* qui conserve les données les plus demandées. Entre le proxy et le serveur Web, on peut installer un PC sous linux ou NetBSD, petite machine qui ne s'occupe que d'enregistrer le trafic des accès. C'est un instrument fort utile pour les mesures de fréquentation et les statistiques.
- Enfin une machine spécialisée pourrait être installée pour effectuer la surveillance (*monitoring*) du serveur Web et de tous ses signes vitaux (messages d'erreurs, espace disque, etc.), conserver tous ces signes vitaux (*syslogs*) et même expédier ces signes vitaux à l'extérieur, pour permettre, par exemple, le post-mortem d'un événement de sécurité, ou des attaques de "dénier de service" ; cette machine pourrait aussi enregistrer les signes vitaux de la salle des machines, les chutes de tension électrique, etc.

L'architecture logicielle sous-jacente devrait utiliser les technologies récentes les mieux éprouvées :

- Java/servlets /Orienté-objet, OODBMS
- Architecture ouverte / Multi plates-formes

- dans la mesure du possible, tout l'information permanente devrait résider sur le serveur, minimisant ou éliminant l'utilisation de " cookies " et de scripts CGI (souvent non sécuritaire, et non orienté-objet) ;
- Interface des zones d'intérêt personnels (channel interface) : CDF ou RDF
- Utiliser dans toute la mesure possible XML : Structuration/description des données ; Format de bases de données commun ; Description des zones d'intérêts personnels (Channel description) ; CDF (Channel Definition Format) [Microsoft] ; RDF (Ressource Definition Framework) [Netscape] ; Knowledge Management ; Commerce électronique (abonnements, copies d'articles).

Il est possible -- et moins coûteux -- d'utiliser des infrastructures logicielles déjà programmées pour bâtir un portail. Quelques fournisseurs : Plumtree, Viador, Epicentric, SUN iPlanet, Hummingbird, Sybase Entreprise Portal, Jetspeed (gratuit!), ou encore IBM WebSphere, et Oracle Portal Framework.

D'une manière générale, il est plus coûteux d'engager une équipe d'analystes et de partir de zéro. Cette option exclue, les autres choix sont les suivants :

- soit acheter un logiciel de construction de portail ;
- soit conclure une entente avec un partenaire commercial ;
- ou encore faire équipe avec les institutions d'enseignement supérieur, comme le font quelque vingt universités américaines dans le cadre du groupe JA-SIG (Java in Administration Special Interest group – <http://www.ja-sig.org>). Ce consortium construit en partenariat une infrastructure logicielle de portail universitaire et de services accessibles via le Web à la communauté universitaire (professeurs et étudiants).

En tout état de cause, le système doit être extensible, pouvoir s'adapter à la demande et à la croissance, conférer à ses administrateurs beaucoup de contrôle et de souplesse, permettre aisément l'intégration de nouveaux modules et fonctionnalités.

DESCRIPTION DES CONTENUS

Les contenus à héberger et à diffuser seraient de deux types : articles de revues savantes et prépublications.

- ARTICLES " COURANTS " DES REVUES SAVANTES

Les articles " courants " sont les articles produits en vue de la diffusion électronique, tout autant que sur papier, c'est-à-dire, des articles qu'on met en ligne au même moment, ou à peu près, qu'on lance le numéro imprimé.

La production des articles de revues savantes est contrôlée, ce qui veut dire que les formats sont normalisés, autant pour les données textes que images fixes, en

mouvement ou sonore (Quoique ces deux derniers types d'information, sont à peu près absents pour le moment de la production des chercheurs en sciences humaines et sociales).

Pour chaque article, un format SGML serait produit duquel serait dérivé deux formats de diffusion : HTML et PDF. Le HTML sert à la lecture en ligne ainsi qu'à la recherche tandis que le format PDF sert de format d'impression. De plus en plus, et bientôt définitivement, le fichier XML prendra la place du SGML et sera éventuellement utilisé pour la diffusion (XSLT). Finalement, des fichiers images (GIF ou JPEG) composent parfois l'article sous sa version électronique.

- ARTICLES “ RÉTROSPECTIFS ”

Par “ rétrospectif ” il faut entendre les articles qu'il faudra numériser à partir du format papier. Pour ce type de document, des fichiers TIFF sont produits aux fins de la conservation et des fichiers PDF - image pour la diffusion.

- PRÉPUBLICATIONS

Contrairement aux articles de revues savantes, les formats des documents ne seront pas normalisés. Dépendant de la fonction conférée au site de Prépublication, les auteurs devraient en principe pouvoir déposer un document pour fins de diffusion dans la section du site qui sera consacrée à la prépublication. Comme il s'agit essentiellement de documents touchant les sciences humaines et sociales, on peut présumer que les formats propriétaires usuels de traitement de texte seront les plus souvent rencontrés. Les documents devront être inspectés aux fins de la sécurité.

ESTIMATION DE L'ESPACE DISQUE NÉCESSAIRE POUR LES CONTENUS

Un article scientifique peut occuper entre 0.02 Mo et 8.7 Mo (Mégaoctets) d'espace disque, comme le montrent les deux exemples extrêmes suivants :

drwxr-xr-x	2	webadmin	netra	512	Apr 29	1999	.
drwxr-xr-x	16	webadmin	netra	512	Feb 10	1999	..
-rw-r--r--	1	webadmin	netra	283	Feb 10	1999	arseneault.ent
-rw-r--r--	1	webadmin	netra	102647	Feb 10	1999	arseneault.html
-rw-r--r--	1	webadmin	netra	8009297	Feb 10	1999	arseneault.pdf
-rw-r--r--	1	webadmin	netra	131365	Feb 10	1999	arseneault.sgm
-rw-r--r--	1	webadmin	netra	125474	Apr 29	1999	arseneault.xml
-rw-r--r--	1	webadmin	netra	1080	Feb 10	1999	catalog
-rw-r--r--	1	webadmin	netra	246	Feb 10	1999	entityrc
-rw-r--r--	1	webadmin	netra	19261	Feb 10	1999	fig01.gif
-rw-r--r--	1	webadmin	netra	90147	Feb 10	1999	fig02.jpg
-rw-r--r--	1	webadmin	netra	57681	Feb 10	1999	fig03.jpg
-rw-r--r--	1	webadmin	netra	65498	Feb 10	1999	fig04.jpg
-rw-r--r--	1	webadmin	netra	18543	Feb 10	1999	fig05.gif
-rw-r--r--	1	webadmin	netra	3235	Feb 10	1999	t_fig01.gif
-rw-r--r--	1	webadmin	netra	17289	Feb 10	1999	t_fig02.gif
-rw-r--r--	1	webadmin	netra	20143	Feb 10	1999	t_fig03.gif
-rw-r--r--	1	webadmin	netra	20206	Feb 10	1999	t_fig04.gif
-rw-r--r--	1	webadmin	netra	3497	Feb 10	1999	t_fig05.gif

et :

drwxr-xr-x	2	webadmin	netra	512	Sep 23	1999	.
------------	---	----------	-------	-----	--------	------	---

```

drwxr-xr-x 28 webadmin netra      512 May 21 1999 ..
-rw-r--r--  1 webadmin netra     1056 May 21 1999 catalog
-rw-r--r--  1 webadmin netra     4903 May 21 1999 dionne.htm
-rw-r--r--  1 webadmin netra     6595 May 21 1999 dionne.pdf
-rw-r--r--  1 webadmin netra     4080 May 21 1999 dionne.sgm
-rw-r--r--  1 webadmin netra     1284 Mar 28 11:47 dionne.websearchresults
-rw-r--r--  1 webadmin netra      185 May 21 1999 entityrc

```

En tenant compte des trois types de documents (articles courants, articles rétrospectifs et prépublications), l'estimation des besoins d'espace disque pour les 5 prochaines années est (en Go) :

```

Année 1 = 4.000 GigaOctets
Année 2 = 8.000 Go.
Année 3 = 15.000 Go.
Année 4 = 20.000 Go.
Année 5 = 30.000 Go.

```

Ceci ne comprend pas l'espace requis pour le système d'exploitation, les divers ensembles logiciels nécessaires pour la diffusion, y compris l'outil de recherche.

DESCRIPTION DES SYSTÈMES

- Système d'exploitation : Solaris 2.6 ou 8
- Serveurs
 - . HTTP
 - . Applications
 - . Index server
 - . "Personalization" / "Customization"
 - . authentification
 - . etc.
- Logiciels de programmation et de gestion : A déterminer
- Serveur http : Apache ou Java Web Server / OODBMS
- Serveurs d'applications : Personalisation Manager / Customization Manager / Channel Descriptions / Integration broker (pour permettre aux applications de partager les mêmes données)
- ASP (Active Server Pages) et PHP (Hypertext PreProcessor : server-side, cross-platform, HTML embedded scripting language), deux façons de créer des pages dynamiques ;
- Stats (mesures de fréquentation et gestion des résultats)
- Langages de développement : C, C++, Visual Basic, Perl 5, etc
- Outil de recherche performant permettant des recherches plein texte, des recherches multiformats/multisites et sachant exploiter les métadonnées.

- ESTIMÉ DE L'ESPACE DISQUE SYSTÈME :

Dépendant de l'installation, le système d'exploitation peut nécessiter entre 600 Mo et 2 Go.

En conséquence et au vu de ce qui précède, une configuration possible de bonne qualité serait la suivante :

INFRASTRUCTURES MATÉRIELLES

- OPTIONS MACHINE :

Marque / OS : Sun / Solaris
Modèle : Ultra Entreprise 220R Server
CPU : 2 x UltraSPARC II 450 MHz
RAM : 1 Gb
Disque(s) / miroirs : 1 x UltraSCSI 18,2 Go (systèmes)
1 x UltraSCSI 18,2 Go (données)
Contrôleur(s) disque(s) : 1 contrôleur de disque pour 2 disques
Bloc d'alimentation redondant
RAID niveau 5 'Hardware' : Sun StorEdge A1000

Le coût total approximatif de cet ensemble est compris entre CAN \$ 42 000 et 45 000. Ce montant inclut un contrat de service de 3 ans au site (lundi-vendredi, 9 à 5) ; il peut être élevé au niveau "Silver", qui comprend des interventions en moins de 4 heures sur le site pour environ 5% du coût des équipements.

- LOGICIELS : (COÛT APPROX. : MOINS DE \$ 2000)

ENGIN DE RECHERCHE :

Un outil de recherche performant permettant la recherche multiformats/multisites et sachant exploiter les métadonnées.

- SALLE DES MACHINES : (Coûts compris dans le contrat d'hébergement de serveurs)

Génératrice UPS
Système suppression Incendie

HÉBERGEMENT :

Serveur hébergé avec contrat de service de base (\$ 12,000/année)

Le Portail contenant des données sensibles et stratégiques, il est recommandé qu'il ne soit pas installé sur un serveur partagé où d'autres sites web et applications pourraient résider et compromettre son fonctionnement; il est recommandé que le Portail soit installé sur sa propre machine.

D'autre part, étant une entité interinstitutionnelle, il est recommandé que le portail soit hébergé dans un lieu interinstitutionnel neutre, ayant directement l'accès le plus direct à la dorsale de l'enseignement supérieur et de la recherche au Québec : le RISQ, qui offre ce service et qui est relié à très haut débit à tous les réseaux de recherche nationaux et internationaux (CA*net 3, Internet 2, Renater 2, etc.) est l'endroit tout indiqué.

- EXIGENCES RÉSEAU :

Bande passante : Min. 1.5 Mbps
 Opt. 10 Mbps
redondance (path diversity)
Peers / upstream (fournisseur(s))
Membre du QIX (Quebec Internet Exchange)
Sécurité : VLAN
 Pare-feu

PERSONNEL : A DÉTERMINER

FORMATION : A DÉTERMINER

ANNEXE 3.1

SITES VISITÉS DANS LE CADRE DE CETTE ÉTUDE

- Allen Press inc. ; <http://www.allenpress.com/>
- Blackwell publishers ; <http://www.blackwellpublishers.co.uk>
- CatchWord ; <http://www.catchword.com/>
- Conseil national de recherche Canada
http://www.nrc.ca/cisti/cisti_f.shtml/journals/rp_home_f.html
- Elsevier Science ; <http://www.elsevier.com/homepage/>
- HighWire Press ; <http://highwire.stanford.edu/>
- International Consortium for Alternative Academic Publication ;
<http://www.icaap.org/>
- MCB University Press ; <http://www.mcb.co.uk/>
- Link ; <http://www.link.springer.de/>
- Project MUSE Scolarly journals online ; <http://muse.jhu.edu/>
- Wiley Interscience ; <http://www.interscience.wiley.com/>