

DU CONCEPT À L'OUTIL : PENSER ET INSTRUMENTER L'ÉDITION NUMÉRIQUE DE REVUES SAVANTES

GÉRARD BOISMENU
GUYLAINE BEAUDRY

G. Boismenu et G. Beaudry,
« Du concept à l'outil », dans Claire Brossaud et Bernard Reber (dir.),
Humanités numériques 2. Socio-informatique et démocratie cognitive,
Éditions Hermès, 2007, p. 181-194. © Lavoisier.

DU CONCEPT À L'OUTIL : PENSER ET INSTRUMENTER L'ÉDITION NUMÉRIQUE DE REVUES SAVANTES

INTRODUCTION

L'écran, interface physique d'accès aux contenus numériques, offre à l'utilisateur une image lissée qui aplanit les dissemblances. Celui-ci, par la fréquence de l'usage du numérique et de ses formes d'accès, n'est peut-être pas aussi candide qu'il n'y paraît. Il sait sans doute distinguer certaines manifestations du soin apporté à la préparation des contenus et à la qualité des services offerts.

On peut supposer que ses choix dans la fréquentation de sites sont intuitivement inspirés par de telles considérations. Mais il est sans doute aussi abusé par l'apparente facilité et la simplicité qu'offrent les instruments du numérique qui se décantent en une multitude de produits et d'équipements conviviaux, prêts à l'usage et prêts à servir. Or, les choix qui ont présidé à la préparation des contenus, à leur mise en forme, à la diffusion et à la préservation ne coulent pas de source. Loin de tenir de l'évidence, il importe de prendre la mesure des choix qui relèvent de la conception éditoriale et du respect des règles de l'art dans l'édition numérique. C'est cette conception qui façonne l'outil proposé à l'utilisateur. Cela prend toute son importance pour le texte savant, particulièrement pour les revues.

Il est donc riche d'enseignements de reprendre les principaux arbitrages qui sous-tendent les opérations de la plate-forme *Érudit*^{sv}. Ils

portent sur les normes de production des contenus, sur la stratégie de diffusion, ainsi que sur les conditions de préservation des documents. On comprend qu'Érudit n'est pas seul en piste : c'est un exemple.

Dans ce chapitre, nous visons à dresser une vue d'ensemble des considérations générales au plan technique², auxquelles tout responsable de plate-forme de revues est confronté.

L'édition et la nature du document

Le travail d'édition couvre un large spectre et fait appel à divers acteurs, ce que le vocabulaire français semble gommer, alors que la langue anglaise en souligne au moins deux grands versants, en distinguant *editor* de *publisher*. Cette distinction n'épuise pas le sujet. Le processus d'édition va de la sélection des écrits à la préparation du contenu et de ses attributs (corpus, données, iconographie, etc.), en passant par la mise en forme et la définition du support et de ses caractéristiques, jusqu'à la diffusion auprès des lectorats, à la distribution en plusieurs lieux et au respect des règles du dépôt légal et des législations sur le droit d'auteur. Ces tâches ont pu s'inscrire historiquement dans des cadres organisationnels assez variés.

Même si ces cadres ne sont pas neutres, ce sont substantiellement ces tâches qu'il importe de mettre en relief. Le processus d'édition prend une couleur particulière selon la nature du document qui est traité. À travers l'histoire, le soin, les caractères physiques et les ressources mobilisées ont varié largement en fonction de l'espérance de vie utile des documents publiés³.

Dans le monde du papier, la représentation de l'information ainsi que le format physique diffèrent selon le type de document. On peut évoquer le dictionnaire, l'encyclopédie, l'ouvrage, la revue savante, le roman, l'édition de luxe ou le livre d'art. Cette diversité n'est pas l'attribut des seuls formats physiques : ces genres sont diffusés selon des modalités, des acteurs et des circuits qui leur sont propres. Avec le numérique, tout passe par le même écran, que ce soit une revue, un livre, une encyclopédie, un dictionnaire. Les modes de représentation semblent normalisés entre les différents types de document. L'impression d'une rupture avec la tradition en édition peut choquer le lecteur. Les textes sont rarement justifiés, l'apostrophe française n'est

pas souvent utilisée, pas plus que les chevrons ou les tirets cadratin. Des subtilités techniques non maîtrisées peuvent expliquer ces « erreurs », mais la constance dans le non-respect de la grammaire typographique et des normes de base de la représentation des textes scientifiques amène à poser l'hypothèse que les éditeurs maîtrisant cette expertise sont soit absents, soit peu influents dans la conception des processus d'édition numérique.

Quatre dimensions permettent de qualifier le rôle de l'éditeur du document scientifique : l'évaluation et la sélection des manuscrits⁴, le traitement et la mise en forme du texte (contenu et forme), la diffusion, la promotion et (éventuellement) la vente, et la conservation des ouvrages⁵. Ces dimensions, qui s'inscrivent dans un parcours qui va du document comme création intellectuelle jusqu'à sa diffusion auprès du lectorat, sont aussi présentes dans le processus d'édition numérique.

Le changement d'environnement renouvelle le cadre, les acteurs, les compétences, les outils et les conditions qui désormais encadrent le processus d'édition et y participent. On assiste de ce fait à une alliance de compétences nouvelles, et à une recomposition des expertises, des lieux et des modes organisationnels⁶. Le numérique suppose un investissement dans la conception du document et du traitement éditorial, tout autant que dans le développement de l'infrastructure (conception, savoir-faire, procédés) nécessaire pour une édition numérique de qualité et efficace. De plus, la publication et la diffusion doivent être pensées de façon concomitante, afin de rendre possible une exploitation optimale de l'impact du numérique dans la valorisation des revues savantes.

DES NORMES D'ÉDITION

La revue représente un type de document dont il faut prendre la mesure pour faire des choix conséquents avec sa mission, son lectorat et sa place dans le processus d'accumulation des connaissances. Ce document participe à l'avancement des connaissances, est validé par les pairs et s'adresse principalement à eux, témoigne de l'état des travaux à un temps « t » et se projette dans un processus en spirale d'affinement, de rectification ou

de controverse dans la constitution du savoir. Il a vocation à être pérenne et à être largement diffusé auprès des communautés de chercheurs dans le monde. Cela suppose des attentes élevées en termes de traitement éditorial, de flexibilité d'utilisation et de garanties de conservation.

Partant de là, les normes les plus élevées doivent être réservées aux revues, alors que d'autres documents de recherche (pensons à la littérature grise), qui n'ont pas le même statut dans la diffusion de la connaissance, n'imposent pas un traitement équivalent (comme cela était d'ailleurs le cas dans le monde de l'imprimé).

Un choix fondateur : le langage structuré

L'entrée de l'édition dans le monde numérique passe nécessairement par la reconnaissance que, in fine, la version numérique de la revue sera conçue comme la version première et fondatrice, qu'il y ait ou non en parallèle une version imprimée. Cela invite à considérer certaines conséquences des choix arrêtés pour le processus d'édition : la norme d'édition retenue détermine la qualité et la polyvalence des documents, et le souci du traitement éditorial définit la qualité des services proposés. Ces caractéristiques de l'édition vont se révéler dès que l'utilisateur dépassera la simple consultation à l'écran et cherchera à exploiter les possibilités de la recherche ou à réutiliser les données [CHA 02].

Le choix de l'implantation d'une chaîne de traitement basée sur XML⁷ s'est imposé, car il assure les meilleures conditions de traitement des revues, en fonction de la nature du document, et ce, de la publication à la conservation, en passant par la diffusion. Contrairement à l'édition imprimée, contexte dans lequel on contrôle le support de diffusion, l'édition numérique permet différents formats de diffusion, soit une variété de qualités et de caractéristiques techniques de ces formats. Les documents structurés contiennent de l'information sur leur organisation logique et sémantique, ce qui leur confère une plus grande richesse et permet de convertir une information de structure en une instruction de restitution.

Pour chaque article, nous disposons d'un document structuré suffisamment riche pour assurer toutes les exploitations envisagées, présentes et à venir⁸. Par là, il est possible d'automatiser les traitements

effectués sur le texte et de décliner cette information spécifiée qu'est l'article en différentes formes de diffusion.

Le mode de structuration de l'information doit être approprié pour les articles de revues⁹. Une « grille grammaticale et sémantique » devait être adoptée pour structurer les documents. Après une analyse très fine des DTD¹⁰ existantes pour l'édition savante et les articles de revue en particulier, la DTD Erudit Article a été élaborée. Depuis, cette DTD, qui a migré vers un schéma¹¹, s'est imposée comme modèle de données dans plusieurs projets en France et au Canada et elle sert à la mutualisation des collections de diverses plates-formes. Ces outils se sont avérés particulièrement efficaces dans le processus d'édition, car ils répondent bien aux diverses situations et cas de figure de textes tels que les articles. Notre modèle de données est en mesure de baliser autant les articles faisant l'objet de numérisation rétrospective (à partir du papier) que ceux de l'édition courante. Cette précaution se justifie par notre souci de mettre à la disposition des lecteurs ces deux types de corpus, au sens technique, par les mêmes interfaces de navigation et de recherche. Les distinctions de traitements informatiques ne devraient pas donner lieu à des modes d'accès techniquement différenciés pour l'utilisateur. Peu importe l'année de publication, ces unités documentaires sont du même genre éditorial : l'article. Le lecteur gagne à les consulter par les mêmes outils, tant en feuillant par volumes et numéros qu'en utilisant un moteur de recherche¹². L'évolution que le modèle de données connaît pour rendre compte de l'univers des « possibles » assure sa pertinence par un affinage toujours plus méticuleux¹³. Afin d'en faire un instrument au service de la communauté, des modalités précises, et largement connues dans le monde de l'*Open Source*, sont pratiquées pour encadrer cette évolution.

Le traitement et le format des textes

La structuration du texte scientifique est une condition nécessaire pour réaliser une édition qui permet une exploitation optimale du numérique. Dans la mesure où la publication numérique ne se limite pas à sa mise à disposition dans le *web*, mais est plutôt perçue comme point d'origine des multiples possibilités offertes pour l'ensemble du cycle de création,

de diffusion et de conservation du document scientifique, ce choix constitue la pierre d'angle de l'architecture. Mais ce n'est pas suffisant. L'histoire de l'édition savante est loin de suivre un parcours linéaire, mais globalement elle est caractérisée par le développement de procédures et de codes éditoriaux toujours plus riches. Le numérique a pu devenir synonyme, d'un côté, de la réplique du papier avec la « page de verre » que proposent certains formats reproduisant l'image du papier et, d'un autre côté, de l'étalement de textes quasi continus dans des formats propres au *web*. Dans ce dernier cas, le numérique en est arrivé à provoquer une régression dans la qualité du traitement des textes publiés et un appauvrissement de l'édition. Ce constat s'impose encore dans bien des cas, sans égard à la stature de la maison d'édition ni au coût parfois somptuaire du prix des abonnements. On peut difficilement se résigner à recourir à la fatalité de l'histoire, mais plutôt essayer de produire des textes possédant une édition aussi riche et témoignant d'un souci d'utiliser au mieux la puissance propre au numérique.

Pour cela, la réappropriation des compétences de l'éditeur s'impose, tout autant que l'expertise proposée par les sciences et technologies de l'information. Bien que le numérique conduise à travailler dans l'univers de l'informatique, le matériau de travail ce ne sont pas des octets, mais d'abord des textes spécifiques que sont les articles rassemblés sous un titre de revue. Force est de restituer aux textes leurs principaux éléments dans leurs propriétés propres et dans leur cadre d'existence comme genre spécifique. Avec le numérique, l'article est devenu, plus que ce n'était le cas avec la version imprimée, l'unité documentaire vers laquelle l'accès peut être direct. Cette unité est extraite de son cadre non seulement physique, mais aussi intellectuel. Il devient primordial de fournir à l'utilisateur toutes les informations permettant de situer n'importe quel article dans son cadre éditorial. Pour chaque article les données suivantes sont mises en évidence : le titre de la revue et sa signature graphique, les coordonnées de la publication, l'accès au site de la revue, les ISSN (imprimé et numérique), la référence au sommaire du numéro, le nom du directeur et du rédacteur en chef de la revue, le nom du directeur invité pour le

numéro (s'il y a lieu), le thème du numéro (s'il y a lieu) et le nom des auteurs et leur affiliation.

Le site a privilégié une présentation visuelle soignée des articles. Pensons au miroir de page facilitant la lecture à l'écran (sans compter la modification de la grosseur des caractères), à la justification du texte et à la grille typographique épurée qui rend clairement les différents niveaux de titres, de même qu'au respect de conventions d'édition telle l'utilisation de l'espace insécable ou de caractères particuliers comme l'apostrophe française. Cela sous-entend le choix d'une fonte répondant aux besoins de la langue française et d'autres langues de publication.

L'utilisateur a accès d'emblée à une « boîte à outils » qui permet de consulter le plan de l'article, la liste des figures, la liste des tableaux, l'accès aux notes. Cette version diffusée en HTML propose aussi l'accès au tiré à part, qui donne une copie du même article tel qu'il apparaît dans la version imprimée de la revue (la dualité des deux versions est très fréquente). Pour ce qui est des figures, nous proposons le format SVG aussi souvent que possible d'où une utilisation plus flexible et de qualité des figures ou images. Pour donner de la précision aux références à l'article ou à la citation lors de l'utilisation de la source, les paragraphes sont numérotés. L'intérêt pour le travail d'édition ne s'arrête pas là, il porte aussi sur l'exploitation du fonds documentaire avec l'outil de recherche.

Faire de la recherche avec le bon outil

Un bon outil de recherche doit satisfaire deux préoccupations essentielles : permettre de formuler des requêtes de recherche variées et complexes et présenter les résultats de façon claire et utile. Les choix technologiques d'*Érudit* permettent la combinaison de la recherche structurée et en texte intégral. La recherche peut se faire par mots libres dans le texte intégral, par titre, auteur, année, mais aussi à travers toutes les métadonnées (auteur, titre, sujet, mots clés et résumé), par une seule opération en excluant le texte intégral. La liste des résultats à la requête, qui donnent les sources disponibles, peut faire l'objet d'un tri de pertinence en fonction d'un des critères retenus. Afin de juger de l'intérêt de chacune des sources signalées, le texte de l'article peut apparaître à même l'interface de l'outil

de recherche. Ce service possède l'intérêt de repérer toutes les occurrences des mots recherchés, ce qui permet de juger de la pertinence de l'article en question. Ce service est disponible aussi bien pour les articles produits en publication courante que pour ceux ayant fait l'objet d'une publication rétrospective et pour lesquels nous procédons à un traitement en XML de base et à une reconnaissance optique de caractères non corrigée.

LA DISPONIBILITÉ N'EST PAS LA DIFFUSION

La diffusion dans le monde numérique consiste d'abord à rendre disponibles des corpus par des sites *web*, mais aussi à faire ressortir les corpus de la masse d'information. La qualité du contenu diffusé et l'ergonomie de l'interface, les services offerts aux utilisateurs, la fiabilité éditoriale et technique, les métadonnées liées aux textes, la notoriété du site, son référencement avantageux dans les principaux moteurs de recherche, toutes ces dimensions contribuent à ce que les corpus ne soient pas seulement disponibles au monde, mais surtout proposés par un « lieu » qui est une référence éditoriale de premier plan.

Référencement et positionnement

Outre l'organisation du site et sa qualité intrinsèque, le libre accès à l'information est un facteur d'accroissement d'impact : le taux de consultation des revues en accès libre est particulièrement important, comparativement à celles qui sont en accès filtré. Cela favorise aussi le signalement par les grands moteurs de recherche du *web*. La plate-forme *Érudit* pratique une barrière mobile, qui subordonne l'accès aux deux dernières années de publication des revues à un abonnement institutionnel (et bientôt individuel) pour les revues qui en font la demande. Tout le reste du fonds est en accès libre (plus de 86 % des 30 000 articles au printemps 2007). Cette mise à disposition est complétée par diverses actions cherchant à accroître le rayonnement d'*Érudit*.

L'utilisateur a un recours routinier aux grands outils de recherche comme canal d'accès privilégié à l'information dans le *web*. Il prend généralement connaissance des deux ou trois premières pages des résultats de sa requête, d'où l'intérêt d'occuper une place privilégiée dans

ces résultats. Le référencement est un processus circulaire et évolutif. La connaissance des modalités d'indexation permet d'optimiser les pages du site dans les divers outils. Un bilan régulier permet d'évaluer la « performance » du site en termes de visibilité relative dans les résultats des divers outils. Cela n'est qu'une avenue.

Cette question prend une nouvelle dimension avec l'installation de nouveaux outils pertinents pour une plate-forme de revues savantes, tels que *GoogleScholar*. Une indexation de la collection *Érudit* apportera un éclairage avantageux sur son contenu et ses auteurs. Le même intérêt doit être porté aux répertoires ou annuaires qui procèdent à une classification des ressources à la suite d'une évaluation faite souvent par des bibliothécaires. Là encore, il importe de connaître les critères pratiqués par les répertoires pour lesquels le site peut intéresser des utilisateurs.

Les bases de données bibliographiques

Les bases de données bibliographiques sont des outils de premier ordre pour le signalement des sources d'informations spécialisées. Lorsque le signalement se double de la fonctionnalité permettant en une seule opération d'avoir accès à la source en texte intégral, ce service est d'une très grande valeur pour le chercheur. Ces bases bibliographiques s'emploient à se doter de telles fonctionnalités. Pour une plate-forme de diffusion de revues, une collaboration qui consiste à fournir les métadonnées compatibles avec le modèle de ces bases est bénéfique, car ces dernières possèdent un effet multiplicateur pour la consultation et n'altèrent en rien les conditions d'accès qui auront été décidées par la plate-forme elle-même. Les articles de revues d'*Érudit* apparaissent dans les bases bibliographiques PubMed et ABC-Clio aux États-Unis. Une collaboration entre *Érudit* et l'INIST permet d'implanter dans la base Francis ce même service : les articles des revues d'*Érudit* indexés pourront être consultés à partir du signalement dans cette base bibliographique¹⁴.

Une autre modalité de liens entre les notices des bases de données bibliographiques et le texte intégral des articles de plates-formes est de plus en plus utilisée. L'application de la norme openURL¹⁵ par les bibliothèques à l'aide de logiciels tels que SFX¹⁶ devient pratique

courante. De cette façon, un bouton apparaît dans chacune des notices des bases de données bibliographiques auxquelles une bibliothèque est abonnée. L'utilisateur clique sur ce bouton pour lancer une requête openURL qui donne en résultat le (ou les) lien(s) du document demandé.

Des moyens qui se renouvellent

D'autres moyens contribuent à amplifier l'impact des publications scientifiques, à faciliter la manipulation des données et à réduire à presque rien les intermédiaires ou les délais dans l'accès aux sources. Sur ce plan, *Érudit* lance prochainement deux développements significatifs.

Le premier rendra disponibles, par l'interface des résultats de recherche, des formats compatibles avec les outils de gestion de références bibliographiques, tels que BibTeX, EndNote, ProCite, ReferenceManager ou RefWorks¹⁷. Les chercheurs pourront ainsi optimiser leur temps de recherche de documentation et intégrer plus facilement les documents diffusés par *Érudit* dans leur système personnel de documentation et de recherche.

Le second développement consiste à créer une interface Z39.50¹⁸ pour faciliter l'interrogation par les outils de gestion de références bibliographiques et des bibliothécaires. Ce service permettra aux utilisateurs des nombreux outils compatibles avec la norme Z39.50 d'effectuer par une requête unique des recherches dans plusieurs bases de données bibliographiques ou en texte intégral, en intégrant *Érudit*. Le développement de ce service apparaît prioritaire.

L'option du réseau

Une politique de réseau revêt un intérêt stratégique pour développer une alternative aux grands éditeurs commerciaux. Un réseau fondé sur une collaboration assez étroite de lieux de diffusion numérique de revues est d'un grand intérêt pour des plates-formes partageant un tel objectif. Cette mise en réseau devrait se fonder sur le partage des métadonnées pour les articles hébergés par les partenaires, afin de proposer, sur chaque site, une recherche dans une masse documentaire globale, tout en dirigeant les utilisateurs vers les sites où sont hébergés les articles cités dans les résultats de la recherche. Cela permet d'unifier le signalement des articles tout en

conservant l'autonomie de chaque site, notamment pour les conditions d'accès. Sans contraintes lourdes, ce type de réseau décuple l'intérêt du fonds documentaire des sites de revues, amplifie leur visibilité sur le *web* et favorise un positionnement stratégique dans la diffusion auprès des moteurs de recherche, des répertoires et des grands index.

LA RESPONSABILITÉ DE LA CONSERVATION

Une certaine pensée magique inspirée de la féerie du numérique semble mener à une sous-estimation d'un enjeu capital pour la préservation du patrimoine scientifique numérique : la conservation¹⁹. Les protagonistes de la pièce sont, semble-t-il, en train de définir leurs rôles, mais la trame est encore loin d'être établie. L'éditeur, ou ce qui en fait office, ne peut échapper à sa responsabilité concernant la conservation des documents mis en ligne. Il doit s'assurer de la pérennité des textes publiés et diffusés, et prendre des dispositions garantissant l'accès à long terme au service offert. Cela dépasse largement sa pratique traditionnelle de gestion d'« archives historiques d'entreprise », que connaît l'éditeur du monde de l'imprimé. Par ailleurs, la responsabilité des pouvoirs publics, sous l'angle des bibliothèques nationales, reste mal assurée avec des mécanismes de dépôt légal numérique encore hésitants ou pratiqués sur une base volontaire, alors que les bibliothèques publiques ou universitaires font généralement moins l'acquisition « physique » de la documentation que l'achat d'un droit d'accès. À défaut de procédures générales et standardisées, l'éditeur ou la plate-forme de diffusion fait face à une double responsabilité concernant cette question : il doit assurer un accès au fonds qui s'inscrit dans la durée, et il se doit d'assurer la pérennité de cette documentation aussi longtemps que les relais ne seront pas assurés de façon satisfaisante.

L'archivage : les maux du mot

Outre les difficultés réelles que représente le défi de la préservation à long terme des documents, un malaise est déjà inscrit dans les mots mêmes que nous utilisons pour en parler [BEA 02]. Les acceptions d'archivage diffèrent non seulement en fonction de la langue, mais aussi en fonction du domaine. Pour l'informaticien, l'archivage concerne généralement

la compression de fichiers et leur stockage en vue d'une éventuelle utilisation. En sciences de l'information, le terme archivage désigne les activités de gestion des documents en vue d'en assurer la préservation, la diffusion et l'accessibilité à long terme. L'archivage implique des règles canoniques qui renvoient à des pratiques et des processus bien définis et faisant parfois l'objet de réglementation. Ces pratiques sont bien différentes d'un stockage aveugle et systématique de tous les fichiers sur un support informatique (*archives*). Les animateurs francophones (surtout français) du mouvement des dépôts institutionnels (*institutional repositories*) ou dépôts de documents de recherche utilisent le mot archive(s) pour désigner ce mode de collecte et de diffusion de différents genres de publications scientifiques. Cette utilisation du terme est regrettable, car elle favorise le glissement de sens que nous dénonçons. La vigilance est de mise devant un certain optimisme technologique qui laisse présager que, par la simple existence des archives, les besoins justifieront bien les moyens à investir pour trouver une façon de faire apparaître à l'écran l'information qui se cache derrière un nom de fichier, qu'on ne saura pas ouvrir avec les logiciels et matériels usuels, 10 ou 15 ans après sa création. Ce raisonnement est peut-être réaliste dans certains milieux où le financement est aisé à trouver pour la récupération des données par des « archéologues de fichiers informatiques ». Pour les publications scientifiques, ce pari est loin d'être gagné.

Du défi technique au défi conceptuel

L'histoire récente du numérique fourmille d'exemples qui témoignent de mémoires orphelines en raison de la perte de documents, dont on n'arrivera plus à percer le message ou le mystère, car ils ne sont plus lisibles. Du point de vue de l'éditeur, la conservation se présente sous l'angle du défi posé par l'obsolescence technologique et des parades pour la contrer. Contrairement au monde de l'imprimé, qui présente une technologie stable et contrôlée, le monde numérique se caractérise par l'instabilité, la volatilité et la constante évolution. Cela suppose que l'on doive assurer la conservation des nouveaux documents au moment même où se créent de nouveaux modèles de diffusion, tant aux plans conceptuels, typographique qu'économique [DEW 06]. Ces nouveaux modèles doivent, dans un horizon de long terme, assurer la lisibilité,

l'authenticité, l'intégrité, la pérennité des documents numériques et cela, en considérant une contrainte de coûts toujours présente. L'enjeu de la conservation pose nécessairement la question des supports matériels et des formats, et souligne avec plus d'insistance les choix stratégiques qui doivent être faits très tôt dans les processus, soit dès la production qui prédétermine le champ des possibles pour la suite. En raison de leurs caractéristiques propres, les normes ouvertes pour les formats d'encodage s'imposent de préférence aux formats propriétaires²⁰. Les normes qui garantissent la conservation des documents correspondent également à celles qui assurent les meilleures conditions de diffusion des documents (notamment XML et XSL qui peuvent s'adapter à l'évolution des formats de diffusion « à la mode » et lisibles dans le *web*). Par ailleurs, les plateformes de revues doivent s'assurer de respecter les conditions techniques de programmes interuniversitaires de préservation à long terme de revues numériques, tels que LOCKSS²¹.

CONCLUSION : VERS UN SYSTÈME INTÉGRÉ DE PUBLICATIONS DE RECHERCHE

Le caractère innovateur d'*Érudit* consiste à réunir dans un même système d'information différents types de publications, et non uniquement des revues savantes. Sa valeur ajoutée spécifique tient à la capacité d'avoir accès à un système d'information constitué d'un fonds composé de divers corpus (articles, livres, actes, thèses et autres documents) que l'on peut exploiter avec le même outil de recherche. Ce modèle arrivera à maturité avec les développements de l'année 2007-2008²². Il s'agit d'une condition nécessaire pour la mise en œuvre de ce système qui repose sur la recherche intégrée (permettant de parcourir transversalement l'ensemble des corpus). D'où un travail à la fois sur l'interface de recherche pour l'interrogation du système d'information *Érudit*²³ et sur l'interface de résultats de recherche.

Ajoutons que le parti pris conceptuel cherche à offrir la recherche en texte intégral. Nous avons déjà développé certains outils comme le schéma *Érudit Corpus* qui est offert en licence *Open Source*. Cet outil est le point de départ et non le point d'arrivée. L'édition numérique, avec toute la palette des fonctions que cela suppose, est au confluent

de savoir-faire et d'expertises qui se renouvellent, bien que s'inspirant d'une mission générale inscrite dans une longue histoire.

Deux grands défis sont posés : d'abord, la mise en phase de compétences pluridisciplinaires dans le but de publier et de diffuser selon les règles de l'art, tout en exploitant au mieux les potentialités gigantesques du numérique, et, ensuite, malgré une apparente facilité qui lui est associée, la détermination des choix qui relèvent d'une vision panoramique permettant de s'inscrire dans des normes internationales riches et présentant les meilleures possibilités de développement d'un système d'information dont on sait qu'il est appelé à connaître des bouleversements.

L'un et l'autre de ces défis demandent un travail de conception préalable et continu en mesure de baliser les choix opérationnels diversifiés qui ponctuent le cheminement d'une plate-forme de revues et, plus largement, de documents scientifiques.

Bibliographie

[AME 00] AMERICAN NATIONAL STANDARD FOR TELECOMMUNICATION, Atis Telecom Glossary 2000, US Government. General Services Administration, Alliance for Telecommunications Industry Solutions, 2000.

[BEA 02] BEAUDRY G., « Les archives électroniques : une mémoire orpheline ou en mutation? L'expérience de la documentation », *Actes du 4^e Symposium en archivistique du groupe interdisciplinaire de recherche en archivistique*, Montréal, 2002.

[BOI 02] BOISMENU G., BEAUDRY G., *Le nouveau monde numérique. Le cas des revues universitaires*, « Recherche », Éditions La Découverte, Paris, 2002.

[CHA 02] CHARTON G., GALLETTOZ G., *Les chercheurs et la documentation numérique : nouveaux services et usages*, « Collection Bibliothèques », Cercle de la librairie, Paris, 2002.

[DEW 03] DEWATRIPONT M. et al., « Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe », *Commission européenne*, Bruxelles, 2006.

[SCH 03] SCHONFELD R.C., *Jstor : A History*, Princeton University Press, Princeton, NJ, 2003.

[SCH 04] SCHONFELD R.C., KING D.W., OKERSON A., FENTON E.G., *Nonsubscription side of periodicals : Changes in library operations and costs between print and electronic formats*, Council on Library and Information Resources, Washington, <http://www.clir.org/PUBS/reports/publ27/contents.html>, 2004.

- 1 Érudit est une plate-forme d'édition et de diffusion de revues savantes, ainsi que d'autres types de documents scientifiques. Sur un autre plan, c'est un consortium interuniversitaire (Université de Montréal, Université Laval, Université du Québec à Montréal) soutenu par le gouvernement du Québec et qui compte sur des contributions importantes des universités membres (www.erudit.org).
- 2 En raison du cadre éditorial de cet ouvrage, nous avons écarté de notre propos tout ce qui touche aux caractéristiques organisationnelles, bien que nous considérons qu'il s'agit là d'un élément essentiel aux succès technique et scientifique d'une plate-forme de revues. À ce sujet, consulter entre autres [BOI 02, SCH 03].
- 3 En consultant les différents genres de documents savants publiés à travers l'histoire, le lecteur est à même de constater qu'il n'y a pas rupture, mais plutôt évolution des modes de représentation des textes.
- 4 Pour la revue savante, ce rôle est assumé par le comité de rédaction et le processus d'évaluation et de validation par les pairs.
- 5 Ne serait-ce qu'involontairement, par la dispersion géographique des exemplaires dans les diverses bibliothèques, par la participation à la procédure de dépôt légal obligatoire et par son inventaire d'entreprise.
- 6 Loin de le dépouiller, le numérique enrichit de travail d'édition qui ne se maintient pas à l'identique. Quel que soit le lieu où il se pratique (éditeur professionnel, centre d'édition, revue autonome, bibliothèque et autres), le numérique pose des défis qui peuvent être relevés en ayant recours à de nouvelles ressources, en comparaison· à celles engagées dans les processus bien connus de l'imprimé.
- 7 Le XML (Extensible Markup Language) est un langage de balisage structuré qui permet de distinguer le texte et les données qu'il contient, des représentations visuelles qu'on lui donne, papier ou électronique, selon des besoins actuels et futurs. Le texte d'un article en XML est organisé de telle sorte qu'il peut être matérialisé par plusieurs médias avec un effort minimum : papier, web, base de données, synthèse vocale et autre.
- 8 D'ailleurs les articles balisés en GSML, qui remontent à 1998, n'ont jamais eu de mal à s'adapter à l'évolution des normes et des outils de diffusion au cours des dernières années.
- 9 Érudit a mis en place une chaîne de traitement permettant d'extraire les textes préparés à l'aide de logiciels de mise en page pour produire des documents structurés. Il s'agit d'une voie de passage qui prend en compte les contraintes sociales liées à la juxtaposition des versions imprimées et numériques de la revue. Lorsqu'une revue ne publie qu'en version numérique, le processus est simplifié d'autant.

- 10 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.
- 11 XML Schéma est un langage de description de format de document XML permettant de définir la structure d'un document XML.
- 12 Le balisage par un modèle de données unique n'est qu'une des décisions facilitant la consultation des collections courantes et rétrospectives. Elle est loin d'être la seule à prendre, mais elle est la première.
- 13 La richesse du schéma Erudit Article, de même que la possibilité de le décliner dans des versions simplifiées permettent de l'utiliser comme modèle de données pour la mutualisation des fonds de plates-formes partenaires, tout autant que de générer les données en fonction des protocoles spécifiques retenus par les bases de données bibliographiques afin d'accroître la diffusion des revues.
- 14 Dans un avenir proche, des démarches seront entreprises auprès d'autres bases bibliographiques spécialisées (pensons à Current Contents et à Repères), pour que les revues indexées par leurs soins puissent bénéficier d'un même traitement.
- 15 L'Open URL est une syntaxe informatique dont l'objectif premier est de faciliter le transfert de métadonnées. Cette syntaxe a pour premier objectif de faciliter le relais d'un document d'une base de données à une autre.
- 16 SFX (Special Effects) établit un lien entre une référence trouvée dans une base de données et le texte intégral électronique de l'article référencé. Cette technologie permet d'accéder aux documents électroniques pour lesquels une bibliothèque possède des droits, directement à partir des bases de données.
- 17 Pour BibTeX : http://www.irit.fr/ACTIVITES/BQ_TCI/BIBLIO/help_bibtexJr.html, EndNote, ProCite, ReferenceManager : <http://thomsonresearchsoft.com/products/>, RefWorks : <http://www.refworks.com/>.
- 18 Z39.50 est un protocole qui définit des modalités standard de communication entre deux ordinateurs en vue de l'extraction d'information : http://www.chin.gc.ca/Francais/Normes/normes_echange.html.
- 19 Récemment, une étude [COU 04] sur la comparaison des processus de traitement des revues papier et numériques dans les bibliothèques universitaires se permettait de faire le silence sur cette question, peut-être parce qu'il n'y a pas de solution facile.
- 20 Pour notre part, Érudit a fait une série de choix qui se déclinent de la façon suivante : XML (pour les métadonnées et les documents), unicode, identifiants uniques, XLink, XSL, TIFF, PNG.
- 21 Site web du programme Lockss : <http://www.lockss.org/>.
- 22 Sur la base des acquis et du développement de la zone « revues », la zone « livres et actes de colloque » se développera en proposant pour les actes un

traitement d'édition numérique de base réalisé dans des délais très courts, mais avec l'ensemble des attributs permettant de participer à la recherche intégrée dans la plate-forme Érudit. La zone « autres documents et données », qui réunit les notes de recherche, communications, rapports ou autres documents jugés pertinents par les chercheurs, connaîtra une autre étape avec des actions visant à augmenter le volume des documents et à établir des passerelles avec les dépôts institutionnels qui voient progressivement le jour. L'ensemble de ces documents s'insérera dans l'outil de recherche. La mission d'Érudit pour les thèses est essentiellement « fédératrice » pour mutualiser le signalement de ces documents qui sont au cœur de l'avancement de la connaissance.

- 23 Pour l'interface de recherche, il faut concevoir et réaliser une interface simple et conviviale qui permettra à l'utilisateur de préciser son besoin d'information et de déterminer quels types de documents il souhaite obtenir. Pour l'interface des résultats, le lecteur doit pouvoir consulter d'un seul coup d'œil les résultats et identifier le contexte de l'unité documentaire qu'on lui propose. De la même façon, l'ordre de tri pourra être déterminé par le lecteur, bien qu'un tri selon des critères par défaut sera fait.