

L'Internet savant de l'avenir : l'édition électronique en sciences humaines entre portail, revue, et hypertexte

Nathalie Ferrand, Paolo D'Iorio

▶ **To cite this version:**

Nathalie Ferrand, Paolo D'Iorio. L'Internet savant de l'avenir : l'édition électronique en sciences humaines entre portail, revue, et hypertexte. http://www.msh-paris.fr, Maison des Sciences de l'Homme. 2002. edutice-00000021

HAL Id: edutice-00000021
https://edutice.hal.science/edutice-00000021

Submitted on 18 Jun 2003

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Internet savant
L'Internet savant de l'avenir. L'édition électronique en sciences humaines entre portail, revue, et hypertexte

Rapport final du projet d'action concertée de recherche du PNER : *La recherche en sciences humaines face au numérique. Observations et études prospectives sur les usages et pratiques en cours*

Paolo D'Iorio et **Nathalie Ferrand**

Préambule

À travers l'analyse des usages dans des cas paradigmatiques, cette étude s'efforce de parvenir à une série de réflexions et de recommandations concernant les nouvelles perspectives ouvertes par la publication sur le Web dans le domaine des sciences humaines : numérisation ; publication électronique ; évaluation par les pairs ; rapports avec les maisons d'édition, les bibliothèques et les archives ; constitution de communautés savantes en ligne.
Quatre lieux d'expertise principaux ont été retenus :
·SATOR – Base de données littéraire en ligne, constituée depuis 1989 par la Société Internationale d'Analyse de la Topique Romanesque et première liste de discussion littéraire en France (fin 1994).
·Cromohs – Cyber Review of Modern Historiography, première revue européenne d'historiographie en ligne (depuis juillet 1995, accessible sur le site de l'Université de Florence).
·Fabula – Premier portail en langue française consacré à la recherche littéraire et en rassemblant l'information scientifique (appels et programmes de colloques, postes offerts, nouvelles parutions, etc.).
·HyperNietzsche – Modèle d'un Hypertexte de recherche en Open Source.
Dans ce rapport final, nous chercherons à rendre compte des résultats des travaux engagés dans ces lieux d'expertise et à résumer les réflexions auxquelles sont parvenus les experts chargés de l'observation des usages. Mais nous voudrions également esquisser les scénarios qui se dessinent pour ce domaine dans un avenir très proche, en faire ressortir les enjeux et préfigurer un cadre de développement possible.
Notre idée est que les revues en ligne actuelles constituent le point de passage entre les revues savantes d'hier et l'Internet savant de l'avenir. Nous prendrons donc comme pivot de notre analyse le débat actuel autour des revues savantes évoquant les conflits qui existent dans les sciences de la nature entre maisons d'édition, bibliothèques et communauté scientifique et en nous interrogeant sur la transposition de cette problématique dans le domaine des sciences humaines, à propos de la numérisation des revues existantes et du développement des nouvelles formes de publications électroniques dans leur rapport avec les portails et les centres de services électroniques (bases de données, listes de discussion, bibliothèques numériques, newsletters...), avec les revus savantes en ligne et avec le développement d'hypertextes de recherche.

1.Lieux d'expertise

1.1 Sator

Le rapport que Jean-Marc Ramos, sociologue à l'Université de Montpellier, a rédigé au sujet de l'expérience de la SATOR (« Étude sociologique des pratiques émergentes d'une communauté savante face aux nouvelles instrumentations numériques : le cas de la Société Internationale d'Analyse de la Topique Romanesque 1986-2001 ») tisse un récit passionnant et une analyse convaincante de l'enthousiasme, des hésitations, des difficultés et de la conversion finale d'un groupe international de chercheurs littéraires à l'utilisation des nouvelles technologies de l'information comme instruments principaux de leur activité sous la forme d'abord d'un thesaurus informatisé, puis d'une liste de discussion électronique et finalement d'un site Web.
L'histoire de la SATOR – écrit J. M. Ramos – « commence avec l'introduction de l'informatique dans le projet scientifique. Le passage de la fiche au fichier ne s'est pas fait sans hésitations, ni difficultés. Après avoir suscité de vives discussions au sein de la Société, la conversion instrumentale des pratiques de recherche a débouché sur la création d'un logiciel original dont la mise au point a pris plusieurs années. Cette période est très intéressante à étudier car elle permet de remonter aux origines de la décision. Quelles sont les raisons du changement de méthode et comment s'est-il opéré dans une communauté peu préparée à l'usage des nouvelles technologies ? »
Dans le contexte de notre expérience donc, cette étude très soignée forme, du fait de l'ancienneté du groupe de recherche par rapport au développement de l'informatique, une sorte de préhistoire de la thématique que nous étudions. En plus des renseignements précieux fournis sur les réactions d'un vaste groupe de chercheurs, de différents pays et de différentes générations, face à l'introduction de l'informatique, il nous livre une analyse des usages tels qu'ils se manifestent à travers les déclarations publiques de certains des protagonistes recueillies lors d'entretiens privés ou, surtout, d'après le riche matériel formé par les 14 volumes d'actes des colloques annuels de la SATOR ou encore par les bulletins de liaison et la documentation présente sur les sites Web de Montpellier et de Toronto.
J.-M. Ramos couvre toute la période d'existence de la SATOR de 1986 à aujourd'hui. Le XVe colloque de la SATOR organisé par Nathalie Ferrand à l'ENS du 26 au 29 novembre, lui a donné la possibilité d'exposer les premiers résultats de cette recherche et de procéder à une série d'entretiens des membres de la SATOR de différents pays pour étudier l'état actuel de la question et tirer une série de conclusions qui ont été intégrées au rapport final.
Les trois parties par lesquelles le rapport est articulé traitent d'abord des fondements méthodologiques de la recherche topologique, en faisant ressortir directement à partir des fondements du projet SATOR la nécessité d'une « intelligence collective » pour le repérage des topoi littéraires qui ne pouvait que se manifester dans un projet communautaire et qui suggérait dès le début la possibilité de recourir à l'informatique pour l'exploitation du contenu des fiches compilées par les chercheurs.
Mais les premiers pas vers l'informatique, objet du deuxième point, ne sont pas très aisés. De premières inquiétudes se manifestent face à l'instrument informatique, dont les grilles rigides « nous obligeraient à décrire les topoi de façon très stéréotypée » (Henri Coulet, fondateur et président d'honneur de la SATOR, en 1988). En ceci aussi, la communauté satorienne témoigne d'un certain nombre de craintes destinées à se généraliser dans les années suivantes aux autres communautés de recherche en sciences humaines.
Troisièmement, de façon très logique et typique, se manifeste à l'intérieur de cette communauté pionnière ce que JM Ramos appelle « la tentation lexicométrique qui fait miroiter la possibilité d'un repérage topique par une procédure automatisée ». En effet, les études et les outils de lexicométrie ont été l'une des premières applications de l'informatique aux études littéraires et l'attitude d'une certaine communauté de chercheurs face à ces instruments pourrait être même symptomatique du degré de conscience méthodologique et d'innovation présentes en son sein. Bien trop de projets d'informatique littéraire se sont limités par la suite à des entreprises de numérisation, de codage et d'analyses statistiques. Outils indispensables, mais assez standard et loin de répondre aux exigences multiformes de différents domaines de recherche.
Face à l'offre d'un logiciel générique d'analyse statistique, la SATOR – c'est son originalité et sa faiblesse à la fois –, cherche à se doter d'un logiciel programmé sur mesure qui s'appellera

TOPOSATOR. Le problème est que, comme l'écrit Ian Lancashire en défendant l'approche lexicométrique : « no litterary critic has so far tried to describe the problem clearly to a programmer », et par conséquent la bonne idée de se forger de nouveaux outils de recherche en sciences humaines s'accompagne d'une incapacité de communication entre les littéraires et les informaticiens. Cela sera caractéristique de nombreux projets en sciences humaines, au moins jusqu'à la période récente où certains représentants de la nouvelle génération de littéraires commencent enfin à maîtriser les instruments logiciels de base (qui sont devenus plus conviviaux) ou au moins deviennent aptes à comprendre les logiques fondamentales des projet informatiques plus complexes, étant ainsi en mesure de se passer des informaticiens ou en tout cas d'engager avec eux un dialogue fructueux. En effet, il est bien connu (mais souvent oublié) que l'éventail des disciplines qu'un lettré ou un érudit doit savoir maîtriser change selon les époques. Il y a seulement cent ans, il aurait été impossible de se former à la littérature sans apprendre le latin, auparavant c'était à la rhétorique d'être indispensable, ou le grec ancien ou les arts de la mémoire. Aujourd'hui la capacité d'utiliser un ordinateur fait désormais partie, de façon encore officieuse mais généralisée, des présupposés fondamentaux de toute production de textes. Reste encore à intégrer dans le curriculum des études littéraires la possibilité de comprendre certains concepts fondamentaux de l'informatique pour pouvoir dialoguer de façon harmonieuse tout d'abord avec l'outil qu'un littéraire utilise chaque jour, et pour pouvoir exprimer d'une manière compréhensible à un programmeur le fondement théorique ou les nécessités pratiques de la propre discipline.
SATOR, dans cette phase, a fait les frais de ce manque de communication typique des équipes littéraires les plus anciennes en passant six ans dans le développement de son logiciel de gestion du thésaurus des topoi, TOPOSATOR, qui après avoir atteint un degré assez élevé d'abstraction et de pertinence par rapport à son objet d'étude, a été finalement abandonné en faveur de SatorBase. SatorBase, moins ambitieux sur le plan des fonctionnalités, a toutefois le grand avantage d'être conçu pour une utilisation sur le Web. Comme il l'écrit Jean-Marc Ramos : « Le logiciel SatorBase a été conçu pour la saisie et le traitement des topoi ; ce qui ne le différencie guère, sur ce point, de son prédécesseur. Son auteur, Stéfan Sinclair, l'a toutefois doté d'une ergonomie et d'une interactivité qui le rendent plus facilement manipulable par des non initiés. Mais surtout, comme son développement a été pensé pour Internet, il permet au tout venant d'accéder au thésaurus et de consulter ses données mises en lignes ; en revanche, seuls les Satoriens accrédités par le Conseil scientifique sont habilités à faire entrer de nouvelles informations dans la base ».
Mais cette mise en ligne des données a fait réapparaître, chez certains membres de la SATOR, « le thème de la crainte » : « le parti pris du développement technologique qui s'est traduit assez rapidement par des invitations répétées à changer d'habitudes pour travailler, échanger et communiquer, a suscité quelques inquiétudes et parfois même des résistances au sein de la communauté. Dans les premières années, lorsque l'usage de l'ordinateur n'était pas encore entré dans les mœurs, la conversion instrumentale (le passage de la fiche au fichier) et la réduction lexicométrique (le traitement informatique des textes) furent souvent au cœur du débat pendant les colloques. Le thème de la crainte109 s'est ainsi installé dans le discours des membres de la Société au point de réapparaître récemment à propos de la mise en ligne de la base de données. »
Le cas de SATOR, de ses usages et de sa communauté scientifique, est une source de réflexion très importante pour comprendre d'autres cas similaires et plus en général le rapport des chercheurs en sciences humaines face à la transposition informatique de leurs instruments de travail.

1.2 Fabula

L'étude écrite par Marin Dacos avec la collaboration d'Alexandre Gefen et François Lermigeaux : Fabula.org : usages d'un site scientifique littéraire (1999-2001), est la description d'une réussite. Ce « site de ressources et de recherche en ligne destiné aux chercheurs et visant à

la diffusion de l'information scientifique et de la connaissance en matière de théorie et de poétique littéraire » est devenu en peu de temps une référence dans le domaine des portails littéraires francophones.
L'historique du groupe de recherche est très différent par rapport à la genèse du projet SATOR. Ici l'outil informatique ne vient pas aider à résoudre les exigences d'une communauté déjà formée, mais c'est le moyen avec lequel on forme une communauté dans l'espace virtuel du réseau Internet. Le rapport à la fois des créateurs et des utilisateurs du site n'est donc pas du tout conflictuel ou perplexe par rapport à l'utilisation de l'informatique : à différents niveaux de compétence, tous ceux qui ont à faire avec Fabula sont ipso facto déjà à l'intérieur d'un environnement numérique. L'enquête sur SATOR et celle sur Fabula permettent donc de couvrir dans notre champ d'analyse à la fois les utilisateurs qui évoluent dès le départ dans un contexte numérique que ceux qui, tout en ayant conscience de la nécessité de l'informatique pour leur domaine d'étude, y résistent ou manifestent, à des degrés divers, de la perplexité face à une telle instrumentation.
Fabula n'est pas un logiciel conçu comme une aide spécifique aux problèmes d'un groupe particulier de chercheurs ou ceux d'un domaine de recherche défini. Le site est certes en cohérence avec son approche théorique et constitue un excellent outil d'information concernant la théorie et la poétique littéraire, mais le moteur logiciel sur lequel il est bâti pourrait très bien s'appliquer à la diffusion d'autres contenus. Il s'agit d'un outil logiciel de partage des connaissances et surtout d'informations en réseau qui s'inscrit dans la mouvance du logiciel libre et est disponible sous licence GPL. Les textes publiés, par contre, « relèvent du régime juridique commun à toutes les parutions écrites » et nous ne trouvons donc pas, dans Fabula comme dans SATOR, de réflexion juridique spécifique sur les questions du droit d'auteur sur support électronique en général et sur Internet en particulier.
Après avoir esquissé la genèse du projet Fabula, le rapport de Marin Dacos se concentre sur une analyse des usages tels qu'ils ressortent des statistiques de fréquentation du site Web. En ayant renoncé à tout moyen intrusif de monitorage des accès (par exemple l'utilisation des cookies) et n'ayant pas mis en place de système de monitorage qualitatif ou de focus group comme par exemple celui utilisé par Mnemosine pour l'étude des usages de Cromohs, le fichier log reste le principal outil de monitorage. Comme M. Dacos le précise en exposant de manière détaillée les caractéristiques techniques de cet outil très souvent employé par l'analyse statistique des accès aux sites Web, les résultats offerts par le fichier log doivent êtres soigneusement traités et interprétés pour parvenir à des résultats intéressants. Et la dernière partie de l'étude de M. Dacos se consacre donc au problème de l'évaluation du nombre des visiteurs et de la mesure des usages et de la fréquentation du site, en parvenant à des considérations intéressantes.
Trois exemples de visite du site sont cités en guise d'exemples pour une analyse détaillée, et permettent d'esquisser le portrait de trois figures de visiteurs et de leurs usages. En lui-même, le simple inventaire des pages consultées associé au temps passé sur chaque page, nous donne déjà d' intéressantes informations sur trois profils d'utilisateurs et sur leurs intérêts spécifiques.
Ensuite, M. Dacos nous fournit également « quelques éclairages factuels sur l'interactivité du support électronique [ou plutôt d'Internet ou du site Fabula], si légendaire qu'on finit par ne plus savoir si elle existe et quelle forme concrète elle prend », en prenant en compte les « réactions des visiteurs transmises par courrier électronique, qui sont l'expression d'une communauté de lecteurs ». Cet aspect du rapport, où sont confrontées les attentes des utilisateurs et leur rigidité à l'égard des possibilités inexploitées du système, qui traduisent les attentes de l'équipe gestionnaire du site, est très importante et mériterait d'être approfondi. Par exemple des phrases comme : « On constate d'abord que les internautes préfèrent l'usage du courrier électronique aux formulaires qui lui sont proposés. Même pour soumettre une nouvelle, la plupart utilisent la formule non formatée qu'est le courrier électronique. C'est sans doute le signe d'une incompréhension de la puissance de certains formulaires, qui font gagner du temps à l'équipe. [...] Cette relative rigidité de l'internaute se retrouve dans ses formulations, où l'on constate qu'il a parfois du mal à cerner ce nouveau support scientifique ». Ou alors tout un ensemble d'attentes qui s'expriment dans des questions diverses ou de demandes d'informations

que le rapport considère « totalement hors-sujet » par rapport aux objectifs de Fabula, mais qui, du point de vue de l'analyse des usages dans un contexte plus général, gardent toute leur valeur.
Nous ne savons pas si, comme le disent les auteurs du rapport, le site Fabula « incorpore au cours de sa croissance les édifices traditionnels de la communauté scientifique », mais il est certain qu'il s'agit d'un site vivant et fréquenté, qui démontre toute une problématique d'usages et d'usagers à la fois caractéristique des portails Web et propre à un site littéraire.

1.3 Cromohs

L'étude des usages de la revue on-line Cromohs a été confiée à la société Mnemosine qui a rédigé un rapport riche et intéressant, fruit d'un monitorage quantitatif et qualitatif du site pendant six mois, qui se fonde sur plusieurs éléments : les propos des fondateurs de la revue ; les fichiers log concernant l'accès des usagers du site ; un focus group ; un entretien avec les responsables du projet.
Il a été également mise en place un système de fenêtres de dialogues spécialement conçues pour cette expérience, qui ont permi un monitorage qualitatif des accès au site plus approfondi : « Il s'agit d'une interface qui peut être activée pendant la navigation de l'usager sur le site sur la base des actions par lui effectuées, et qui lui pose de simples questions, en général à réponse fermée : une interface rendue le moins intrusive possible du fait que l'usager est prévenu de la possibilité de ses apparitions lorsqu'il se connecte au site, et il peut choisir de la désactiver complètement et préalablement, comme il peut choisir de ne pas répondre quand elles apparaissent au cours de la navigation. Cependant son caractère décidément interactif, l'étroite corrélation temporelle avec les activités sur lesquelles elle enquête et la correspondance précise des questions qu'elle pose avec les actions qui sont l'objet de l'enquête, sont la garantie de résultats cognitifs d'un grand intérêt ».
Avec cette panoplie d'instruments de travail, l'analyse de Mnemosine a pu mettre en lumière non seulement la fréquentation et les usages du site, mais aussi l'interaction qui s'est créée entre la logique de configuration du site adoptée par les responsables scientifiques du projet et celle des lecteurs/usagers. L'autre interaction soigneusement étudiée est celle entre la revue électronique Cromohs (périodique en ligne, avec banque de données bibliographique et section de news) et sa bibliothèque numérique, Eliohs, contenant des textes numériques et des liens vers d'autres ressources Web du domaine disciplinaire de l'histoire de la culture européenne à l'âge moderne.
Assez rapidement, pourtant, Cromohs et Eliohs ont gagné en réputation et en respectabilité dans le monde académique international. Ils figurent dans les principaux index internationaux de ressources électroniques pour l'histoire, mais ils sont, en même temps, indexés dans les répertoires traditionnels comme Historical Abstracts et America History and Life, publiés par ABC-Clio.
Une des raisons possibles de ce succès est le soin particulier que les éditeurs ont apporté à deux aspects de leur activité, c'est-à-dire du point de vue de l'offre (la tutelle des intérêts du public) et du point de vue des intérêts des auteurs. Au niveau de l'offre, ils se sont efforcés de garder un standard de qualité dans le choix des collaborateurs de tous niveaux, dans le contenu des contributions et, enfin, dans la rigueur des éditions originales de textes dans Eliohs. Ce qui a permis de démontrer, au delà des doutes, que le moyen électronique n'entamait en aucun cas la qualité scientifique.
Parallèlement, les éditeurs ont reconnu l'importance d'assurer aux auteurs la pleine publication de leurs contributions aux termes de la législation italienne. Sur ce plan-là, la solution qui a été envisagée dès le début a été de suivre avec le plus grand scrupule les procédures d'enregistrement du périodique et de dépôt légal auprès des bureaux judiciaires compétents. Le débat et l'abondance des expériences en cours dans le monde de l'édition ont permis toutefois de réaliser une importante innovation à ce propos. Cromohs et Eliohs ont donné naissance à une société à but non lucratif (une « Onlus » selon la législation italienne), qui est devenue un sujet légal propriétaire des deux entreprises et qui, de ce fait, a pu devenir partenaire d'un contrat d'édition avec la nouvelle maison d'édition de la Firenze University Press, accédant grâce à cela aux garanties assurées par le dépôt légal dont se charge la maison d'édition de la Bibliothèque

Nationale de Florence chargée du dépôt légal à l'échelle nationale. Ce progrès est important surtout en vue d'une légitimation complète du produit électronique qui puisse offrir aux auteurs une pleine sûreté pour ce qui concerne le status juridique de leurs œuvres publiées à travers Cromohs et Eliohs. Il s'agit d'une innovation absolue dans le panorama de l'édition italienne, qui promet des développements très significatifs pour ce qui concerne les modes de diffusion et d'utilisation scientifique et professionnelle des publications académiques.
Un regard rétrospectif sur ces cinq années ouvre sur un panorama somme toute confortant et qui restitue l'image d'une production en mouvement et, le plus souvent, en mouvement progressif. Si on considère Cromohs comme le lancement d'une nouvelle revue spécialisée dans la recherche en histoire de l'histopriographie, il est indéniable que l'opération a eu un certain succès. Et c'est avec satisfaction que l'on constate que les indicateurs ne manquent pas. La revue a publié dans six macrocontenants annuels (un septième, presque prêt, va bientôt s'ajouter aux autres) une trentaine d'essais originaux ainsi que de vastes comptes rendus/critiques. Certains de ces essais – tous d'origine académico-professionnelle – sont en réalité le fruit de recherches et de réflexions bien plus proches d'une monographie que qu'un article de revue. D'autres essais contiennent des textes inédits d'une absolue valeur scientifique.
Mais au delà de l'analyse de la réussite de Cromohs en tant que revue spécialisée, l'étude a permis de réfléchire au différents aspects de la gestion d'un projet de communication scientifique et au rôle des processus de communication à l'intérieur de la communauté scientifique.
« Bien qu'il soit communément reconnu aujourd'hui que les processus d'élaboration et de validation de la connaissance scientifique sont aussi, et dans certaines phases surtout, des processus de communication, et qu'ils sont donc régulés par les normes relatives à la transmission de signes et aux échanges symboliques, il n'en est pas moins vrai qu'historiquement, les détenteurs de position d'autorité et de pouvoir dans les communautés scientifiques définissaient leur propre autorité et leur propre rôle, non sur la base de leurs fonctions de communication, mais sur la base de rôles, réels ou supposés, de type "fondamental", liés uniquement à la production de la connaissance et à son élaboration et évaluation critique. Autrement dit, la fonction de communication, bien que réelle, était perçue comme subalterne et purement accessoire.
Dans la gestion traditionnelle (et aujourd'hui encore très ancrée) des revues scientifiques, on distinguait et on distingue encore les rôles, les rôles scientifiques en l'occurrence, en fonction essentiellement de leurs liens avec les processus de sélection et de validation des contenus, des rôles plus strictement liés à la communication (de la sélection des canaux pour atteindre le public aux choix considérés généralement comme marginaux, comme le graphisme), délégués aux maisons d'édition ou à des responsables de rédaction ayant souvent un rôle inférieur dans les hiérarchies académiques.
Il se peut que tout cela soit l'effet d'une compréhension déformée des vrais processus de la production scientifique, voire même d'une fausse conscience, et que l'accroissement actuel de l'attention des chercheurs envers les processus de communication constitue une forme de "prise de conscience" positive. Mais il est certain que le cadre s'est modifié, dans la mesure où aujourd'hui, un nombre croissant de chercheurs se penchent sérieusement sur la question de la gestion des médias et des canaux de communication, dont ils tirent des positions de pouvoir et des identités professionnelles spécifiques. Et que bien d'autres voient dans leur compétence médiatique, non seulement un outil de travail, mais une des clés pour se positionner dans la communauté scientifique; vice versa, certains voient dans leur maigre compétence médiatique un problème qui pourrait peser dans leur positionnement académique et scientifique.
Il s'agit ici aussi d'un processus encore en cours, et qui pour le moment ne touche que certains segments des communautés scientifiques. Mais c'est ce qui rend notre étude significative, car elle met en lumière les premiers signes de ce passage. »

1.4 HyperNietzsche

Faute d'utilisateurs, car le site n'est pas encore ouvert au public, l'expertise sur l'HyperNietzsche s'est concentrée sur les attentes que ce projet suscite chez les utilisateurs potentiels et les commentateurs, ainsi que sur les questions d'ordre économique, sociale et de politique culturelle qu'un modèle de ce genre suscite.
Evoquons l'une des nombreuses réactions de la part de la presse. Il s'agit de l'article d'un jeune commentateur du magazine 01net (le 26 octobre 2000, dix jours après la publication du projet HyperNietzsche sous forme de livre). Nous la citons largement, car elle nous semble particulièrement significative des attentes des usagers par rapport à un projet de publication savante sur le Web :
« La société modifiée. Nietzsche, version Open Source. Un rêve d'étudiant : un site avec les copies des manuscrits de Nietzsche, gratuitement consultables. Regardons mieux : y figurent aussi des transcriptions, des traductions, des dictionnaires spécialisés, des essais, des biographies, des articles de revues, des thèses... Le tout relié par une forêt de liens hypertextes qui éviteront le recours aux index, bibliographies et autres catalogues.
En comparaison, l'ancien étudiant que je suis se souvient des heures de transport en commun pour se rendre dans des bibliothèques lugubres et surpeuplées à la recherche du document rare. Le document en question renvoyait toujours à un autre article, un mémoire archivé à Rouen, ou à Baden-Baden. Je vous passe les files d'attentes, les photocopies hors de prix, les autorisations, les laissez-passer et attestations de recherche dont il fallait se pourvoir. Un truc à devenir fou. […]
Un engin numérique enfin adapté au travail des chercheurs en sciences humaines, et qui se nourrit du travail des participants pour devenir plus complet et plus efficace. Une paillasse de laboratoire qui grandirait au fur et à mesure que les gens y travaillent ! […] Hypernietzsche jouera autant que possible la carte de l'Open Source : les textes seront gratuitement consultables, enregistrables, réutilisables, du moins tous les textes que le droit d'auteur permettra de faire passer sous ce statut. Ça fera des oeuvres complètes vraiment complètes dans toutes les langues, évolutives mais aussi gratuites ! Vous savez combien coûte une intégrale de Nietzsche digne de ce nom aujourd'hui ? Beaucoup trop cher.
Pure philanthropie ? Non, simple logique […]. Les chercheurs qui publient en tirent rarement des revenus substantiels. Il importe plus pour eux de donner à voir leur travail, d'assurer leur notoriété et de diffuser leurs idées. Ici, ils rendent leurs travaux publics sur la plus vaste plate-forme de diffusion jamais conçue : Internet.
Et puis ce n'est pas tout, l'équipe d'HyperNietzsche explique comment développer un tel site, parce que, aux frontières d'HyperNietzsche, pourquoi pas un HyperSchopenhauer ? Et puis un HyperWagner qui appellera un HyperBach renvoyant sur un HyperBoulez, et pas d'hypermarchés...
L'équipe de l'HyperNietzsche parle à notre tête, pas à notre portefeuille. Il s'agit d'installer dans le réseau un espace de partage universel de tous les savoirs juridiquement disponibles. Les éditeurs traditionnels vont s'en plaindre, leur privilège de diffusion est si confortable. Qu'importe ! Étudiants, chercheurs, amateurs liront, sans solliciter leur avis et sans frein à leur boulimie. Bon courage, messieurs les universitaires. Et prenez votre temps. » [Renaud Bonnet, http://www.01net.com/rdn ?oid=125782]
Les éléments évoqués par cet article : Open Source, généralisation du modèle, logique économique, caractère de partage à l'intérieur de la communauté savante, se retrouvent, précisés, finement analysés et commentés dans la vaste et intéressante étude rédigée par Yann Moulier Boutang, professeur de sciences économiques à l'université de Paris I. Les conclusions auxquelles parvient Yann Moulier Boutang après cette première période d'analyse du modèle HyperNietzsche sont les suivantes :
A. L'hyper Nietzsche est un véritable modèle alternatif de division cognitive du travail distinct de la division du travail hérité d'Adam Smith et de Durkheim. Il est à la fois créatif mais aussi innovant, c'est-à-dire susceptible d'un fort développement hors de son foyer de naissance. Pour qu'on puisse parler en effet d'un modèle, il faut avoir la réunion de trois caractéristiques :

a) sa consistance ou cohérence qui en fait un ensemble de définitions claires, de règles institutionnalisables ;
b) son caractère créateur qui établit quelle est la nature de l'apport net qu'il constitue par rapport aux protocoles auxquels il entend se substituer ;
c) son caractère innovant, c'est-à-dire sa puissance de diffusion au sein de la communauté scientifique, donc son pouvoir de contagion, de généralisation.
B. À la question de la viabilité économique de ce type de projet, il convient de répondre qu'elle ne présente pas de problèmes de financement particulier. Le retour sur investissement public est formé de plusieurs composantes qui se renforcent mutuellement :
a) ce mode d'organisation de la mise en ligne des connaissances activent la formation sans coût de transaction de réseaux scientifiques à l'échelle internationale intéressant aussi bien des équipes que des chercheurs isolés, des étudiants de différents niveaux ;
b) il a un effet sur la qualité de l'information mise en ligne et sur la pertinence des réseaux créés. Or la qualité de l'information et de son actualisation ainsi que celle des réseaux est l'un des principaux problèmes de fond que rencontre le recours à l'Internet comme outil de travail ;
c) le troisième avantage économique que comporte le travail coordonné par le recours à un réseau spécifié d'accès à des données en lignes sur le modèle de l'HyperNietzsche est qu'il permet d'établir à très faible coût de nouveaux critères objectifs pour l'évaluation de la production de la recherche nouvelle. La question de l'évaluation de la recherche, dans des domaines multiples et sur des programmes de recherche scientifique de plus en plus complexes est aussi un des problèmes majeurs de conduite et de gouvernance des programmes tant elle finit par réclamer une portion importante de ressources totales pour être menée à bien.
C. Ce modèle est généralisable techniquement et conceptuellement à d'autres auteurs et à d'autres disciplines scientifiques, mais aussi artistiques. L'on nommera ce modèle la production cognitive en réseau spécifié. Par réseau spécifié, on signifie que ce réseau en open source et open share est soumis dans l'élaboration de son contenu en ligne à un comité scientifique. N'importe quel article, document portant sur Nietzsche, n'est pas mis en ligne automatiquement. Ce point soulève la question des modalités de désignation ou de sélection de l'autorité régulatrice du site en ligne. Mais d'un point de vue de l'analyse formelle de ce type de modèle, il est important de souligner ses particularités par rapport au réseau de l'Internet en général, que nous nommons l'Internet générique.
D. Ce modèle contient aussi des indications fortement innovantes dans la perspective de la mise en place d'une organisation productive recourant aux NTIC (nouvelles technologies de l'information et de la communication). Du même coup, il menace par son existence même les autres modèles d'organisation de la recherche et de la production cognitive, si bien qu'il est en but à une hostilité croissante du modèle du marché comme du modèle de la hiérarchie qui veulent son cantonnement à un espace réduit à l'avance. En ce sens l'arbitrage des pouvoirs publics sera non seulement souhaitable mais, il revêtira rapidement une urgence qu'on ne saurait sous-estimer.

E. Les problèmes juridiques que pose la généralisation de l'accès libre et quasiment gratuit à des banques d'hypertextes font partie d'un problème plus général que constitue la redéfinition des droits de propriété en fonction des NTIC et de la transformation de la nature de la valeur et de ses mécanismes de captation. Deux options différentes s'ouvrent pour un système de numérisation et de placement en ligne du même type que l'HyperNietzsche : soit un aménagement à partir de l'état actuel, largement hérité du passé, du fonctionnement des financements publics à la recherche et des règles régissant la propriété intellectuelle. Cette solution, propre à rassurer les opérateurs privés tels les détenteurs des droits d'auteur, se contente de "bricoler" les financements publics afin de garantir la gratuité et l'universalité de l'accès. Elle risque toutefois d'être onéreuse et donc peu généralisable à d'autres auteurs, et en même temps, par trop dépendante des aléas qui entourent l'avenir de la propriété intellectuelle, donc dépassée par l'évolution de la réglementation et l'instauration de nouvelles règles du jeu. La deuxième solution, plus volontariste, consisterait pour les pouvoirs publics à affronter de

façon novatrice la question des garanties de revenu des producteurs de connaissance et de recours à un fonds de connaissance patrimonialisé, rendu gratuit d'accès, quitte à ce que les compensations financières pour les auteurs ou les ayant droits recourent à des solutions juridiques différentes du droit d'auteur ou du brevet telles que les licences de copyleft en ont donné l'exemple dans le domaine des logiciels. [Voir également notre appendice juridique]
Les conclusions auxquelles parvient l'étude de Yann Moulier Boutang s'inscrivent dans une réflexion plus générale sur la production des connaissances à l'intérieur d'un réseau cognitif spécifié de type hypertextuel :
« L'Internet générique a pour effet immédiat de brouiller la séparation entre les savoirs tacites et les savoirs codifiés. Nombre de savoirs tacites ou implicites reposent en effet sur leur caractère dissimulé, non transparent ou restreint à un petit nombre d'agents qui se placent en position de monopole (délivrance d'un label qualité) en restreignant ou retardant la publication et la codification des critères de sélection ; ou en situation de rente en monnayant l'accès à des savoirs qui restent non écrits (par exemple les indications sur les marottes d'un jury qui ne figurent jamais dans le rapport). L'Internet offre un moyen d'objectivation, de diffusion très large et dans délais très courts donc de démocratisation. Il provoque un chevauchement entre les savoirs tacites qui sont partiellement révélés on line sans perdre toutefois leur caractère de complémentarité avec les savoirs codifiés faute d'accord préalable sur la procédure de dévoilement. Cet effet se retrouve dans le cas du réseau spécifié de l'HyperNietzsche, mais l'existence de procédures de validation rigoureuse de ce qui est mis en ligne sur le site, aboutit à réduire le rôle des savoirs implicites dans la mesure où les renvois internes (voire externes lorsqu'il s'agit de définir des termes) permettent de présenter et d'expliciter un grand nombre d'informations, de restituer des séquences temporelles qui faute d'avoir été mises à la disposition équitable du plus grand nombre de chercheurs ou d'usagers, figuraient dans les savoirs implicites d'une élité très restreinte libre de dévoiler ou pas ces connaissances et de transmettre ou pas les règles d'interprétation. C'est pourquoi, à notre avis, un dispositif hypertextuel tel que l'HyperNietzsche, place les savoirs implicites dans une position essentiellement de substituabilité par rapport aux savoirs codifiés qu'avant les NTIC et à la différence d'un site "amateur" créé autour de tous les textes comprenant le mot Nietzsche ou des termes du lexique nietzschéen. Certes une partie des savoirs implicites liés à l'activité cérébrale, à la culture philosophique d'ensemble de l'usager du site demeure irréductible à des savoirs codifiés ou à des informations, mais l'érudition "d'autorité" souvent implicite parce qu'invérifiable faute de temps et d'accès aux sources originales (à la langue d'origine du texte noatamment) se trouve résorbée dans le corpus créé.
Dans la mesure où les instances de validation des savoirs produits ne sont pas impliquées dans la gestion des sites créés ou de l'information portée à la connaissance de l'ensemble des usagers de la Toile, cette technologie générique de l'Internet a, au mieux, un effet neutre par rapport aux modes de communication antérieurs (avis dans la presse, à la télévision à la radio), au pire un effet délégitimant des instances de régulation des savoirs si elle met en évidence des comportements de rétention de savoirs implicites. La cohésion scientifique que promeut le recours à l'Internet générique relève plutôt de "l'anarchisme méthodologique" brillamment défendu par Feyerabend dans son plaidoyer pour l'invention tandis que la cohésion du régime cognitif précédant les NTIC est de l'ordre du paradigme kuhnien. Ce dernier, mettant en jeu la "constellation complète des croyances, valeurs, techniques partagés par les membres d'une communauté donnée " (1962) inclut les savoirs implicites partagés y compris à leur insu. Dans le cas de l'HyperNietzsche, et de tout réseau "autorisé" qui contrôle le contenu du site hypertextuel, on a un double effet : le réseau ainsi constitué, véritable rasoir d'Occam, rend superflues les instances de légitimation diverses qui s'étaient constituées à la faveur de l'incomplétude même du corpus de l'œuvre et de l'ensemble des gloses et travaux (maison d'édition, traducteurs, interprétations philosophiques déjà validées dans l'histoire de la philosophie parce qu'ayant été le fait de grands philosophes eux-mêmes, de penseurs, de créateurs, ou "d'autorités"), mais de l'autre il instaure une nouvelle norme d'exigence scientifique, un véritable devoir d'érudition qui, pour figurer dans les déclarations d'intention

restait le plus souvent lettre morte. Toutefois, il ne s'agit pas du simple remplacement d'une pluralité d'instances d'autorité, par une méga autorité fonctionnant sur le mode des précédents (c'est-à-dire soit avec une forte dose d'implicite, soit avec une forte dose d'arbitraire dans le cas d'un site Internet non spécifié constituant un réseau d'individus dont le seul lien est la recherche d'information de tout ordre sur Nietzsche). Un réseau cognitif spécifié ayant la responsabilité d'assurer la qualité de l'information et la validité des savoirs (cette validité pouvant résulter d'une pluralité de critères), doit présenter ses règles de fonctionnement, sa composition, ses débats en cours. Chargé de la mission d'assurer une transparence de la mise en ligne de l'ensemble des connaissances sur Nietzsche, il doit aussi rendre transparent le méta niveau de son fonctionnement. C'est pourquoi, à notre avis, sans perdre de vue l'objectif Feyerabandien de priorité accordée à l'invention, à la production de savoirs nouveaux ("je ne cherche pas, je trouve" selon le mot fameux de Picasso), en mettant en ligne son propre mécanisme de sélection, ressort bien du programme de recherche scientifique en quoi Imre Lakatos voyait la véritable cohérence d'une vision historisée et non-historiciste d'une discipline scientifique.»

Effets des NTIC sur les composantes principales de la DCT (division cognitive du travail scientifique)

Axes principaux de la DCT	Division cognitive sans réseau Internet	Division cognitive avec Internet	Division cognitive avec site de type HyperNietzsche
Connaissances tacites/codifiées	Séparation et complémentarité	Chevauchement et complémentarité	Chevauchement et Substituabilité
Rôle des instances formelles de validation de la communauté scientifique	Fort rôle des structures implicites	Effet neutreoudélégitimant	Simplification et objectivation des procédures validantes
Type de cohésion scientifique	Effet Kuhn	Effet Feyerabend	Effet Lakatos
Divulgation et publicité	Conditionnelle et possible	Inconditionnelle et suffisante	Conditionnelle et nécessaire
Biens privés	Recherche appliquée et inférieure à la durée des brevets et droits d'auteur	Marchandisation des biens connaisssance en général (information payante)	Financement public de l'accès et règlement des problèmes de droits d'auteurs
Biens publics	Patrimoine public ancien et fondamental	Tranformation des biens connaissance en biens publics sauf limitation de l'accès et de la mise en ligne	Mise en ligne gratuite de la totalité des savoirs validés et pratiquées par la communauté scientifique

2. Pour une numérisation des publications savantes

La numérisation et la publication sur Internet des résultats de la recherche est aujourd'hui un des enjeux les plus importants pour la communauté scientifique. Pour ce qui concerne les sciences de la nature, cette problématique concerne les revues scientifiques et notamment la question de leur prix exorbitant dû à des pratiques de monopole exercées par des maisons d'édition commerciales grâce à leur emprise sur les procédures d'évaluation par les pairs et sur la maîtrise de processus de création de core journals, obtenue en exploitant et détournant partiellement le mécanisme du Science Citation Index. Nous partageons entièrement sur ce sujet l'analyse précise et éclairante de Jean-Claude Guédon, ainsi que sa force de dénonciation et son invitation

à une prise de conscience de la part de chercheurs et des autres acteurs publics dans ce domaine [Jean-Claude GUÉDON: In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing, publié dans Proceedings of the ARL - The Association of Research Libraries - Membership Meeting, 2001, disponible à l'adresse: http://www.arl.org/arl/proceedings/138/guedon.html]
Pour les sciences humaines, la question se pose sous des aspects partiellement différents. Elle n'affecte pas seulement les revues, mais également les monographies ou les thèses, et accompagne le problème de l'accès numériques aux sources primaires (livres, manuscrits, fonds d'archives, œuvres d'art ou objets archéologiques conservés des bibliothèques, archives, musées publics).
À cette double problématique, concernant à la fois les modalités de publications sur Internet des contributions des chercheurs et l'accès aux matériaux de base pour la recherche, a été consacré le colloque organisé par le PNER en janvier 2002 à l'École normale supérieure sur L'Open Source dans les sciences humaines. Modèles ouverts de recherche et de publication sur Internet.
Dans le cadre de cette étude, nous voudrions nous occuper plus particulièrement de l'Open Publishing, c'est-à-dire des pratiques de publication des articles universitaires sur Internet ou de numérisation des revues savantes publiées sur papier.
Les projets de numérisation et de publication sur Internet en sciences humaines se trouvent aujourd'hui face à deux question fondamentales :
1.Est-ce que la numérisation doit être la pure transposition numérique non seulement du document papier (article, revue, livre), mais de tout le système de publication savante que s'exprime en lui (distribution, modèle économique, mécanismes d'évaluations scientifiques et plus en général rapports de force entre chercheurs, éditeurs et organismes qui gèrent la recherche) ?
2.Deuxième question étroitement liée à la première, mais concernant la publication sur Internet de nouvelles contributions scientifiques : est-ce que nous devons laisser la révolution numérique s'absorber dans les formes de la communication savante du passé, ou, au contraire, devons-nous chercher à nourrir les systèmes et les formes de publication numériques de l'avenir avec la richesse de contenus exprimée par l'histoire de nos disciplines scientifiques ?
Dans trois des quatre cas analysés au cours de notre étude, le sentiment que le numérique représente une rupture profonde comportant une nouvelle distribution des rôles parmi les acteurs de la recherche et de l'information scientifique est très fort.
Il est affiché par les fondateurs de Fabula, de Cromohs et par le projet HyperNietzsche, et clairement ressenti par les utilisateurs de leurs sites ainsi que par les commentateurs [Même pour ce qui concerne SATOR, parmi les chercheurs qui participent au projet, le sentiment d'une avancée majeure dans leur discipline est assez sensible]
Dans la conclusion de son rapport, Marin Dacos parle de renouvellement complet du « modèle vertical de l'édition classique ; les fondateurs de Cromohs, interviewés par les experts de Mmemosine, parlent à propos du système de publication actuel en Italie de logique « fossile et bureaucratique » et de « liturgies quasi-féodales d'approbation et d'investiture » ;
Yann Moulier Boutang à propos de l'HyperNietzsche met l'accent sur la présence ici d'un « modèle alternatif de division cognitive du travail », qui à son avis « menace par son existence même les autres modèles d'organisation de la recherche et de la production cognitive, si bien qu'il est en but à une hostilité croissante du modèle du marché comme du modèle de la hiérarchie qui veulent son cantonnement à un espace réduit à l'avance ».

2.1 Vitesse, diffusion et évaluation

Ce qui frappe dans ce processus est tout d'abord la vitesse de publication et la diffusion planétaire permises par le réseau Internet. Effectivement ces facteurs marquent une profonde différence par rapport à la publication sur papier. Mais l'exemple des revue de science de la nature est là pour nous rappeler que le facteur le plus important, pour éviter la pure transposition des rigidités du monde du papier dans le numérique, est la question de l'évaluation par les pairs.

Comme le démontre l'analyse de M. Guédon, les scientifiques ont brillamment résolu la questions d'une circulation rapide et universelle de l'information scientifique (par exemple grâce à la base de données à leur disposition depuis désormais dix ans sur le site de Paul Ginsparg, avec son physics preprint server au Los Alamos National Laboratory). Mais ce qui contraint encore aujourd'hui les laboratoires de recherche à débourser de dizaines des milliers de francs (jusqu'à 115.000 FF par an pour quelques petits numéros de Brain Research) pour s'abonner à des revues scientifiques, est le fait que les maisons d'édition gardent entre leur mains l'organisation du processus d'évaluation par les pairs. Paradoxalement, les chercheurs ne semblent pas capables d'organiser d'eux-mêmes, entre pairs, le Peer Review. Ainsi, les scientifiques perdent d'un coup tous les avantages en termes de coût et de rapidité de publication, de visibilité internationale, d'accès public à l'information scientifique qui pourraient venir de la publication sur le Web. Et le fait que leurs articles soient disponibles sur Internet, mais verrouillés dans un réseau accessible uniquement à travers une « licence de site » qui est vendue avec l'abonnement à la version papier, ne fait qu »augmenter l'ironie de l'histoire et le sens de perte de contrôle sur le fruit de leur travail scientifique – qui pourtant, dans la plupart des cas, est payé avec de l'argent public.
Les sciences humaines, pour ce qui concerne la vitesse et la diffusion de leurs publications, ont un retard d'au moins dix ans par rapport à l'expérience des sciences de la nature. Notre étude démontre combien, même dans quatre cas très avancés et pionniers dans ce domaine, les résistances de la part des usagers – et pour ce qui concerne SATOR, même à l'intérieur du groupe de recherche directement concerné – sont fortes et profondément ancrées dans les mécanismes de pouvoir sur lesquels sont bâties les différentes disciplines. Parmi les quatre cas que nous avons expertisés, la question capitale de l'évaluation scientifique n'a été prise en compte que très récemment, elle est évoquée par l'équipe de Fabula, interprétée comme l'un des deux facteurs de la réussite de la revue par l'équipe de Cromohs, radicalement repensées en termes nouveaux (élection directe tous les deux ans de la part des membres de l'hypertexte) et mis au cœur du projet par le groupe de recherche qui pilote l'HyperNietzsche.
Il serait vraiment souhaitable que les projets de publication ou de numérisation savante sur le Web, comprennent que la vitesse de publication, la diffusion de l'information et l'évaluation par les pairs doivent aller ensemble. Et que l'élément fondamental pour, d'une part, ne pas transformer le Web en une poubelle où verser tout ce qui n'a pas trouvé la possibilité d'être publié ailleurs et, de l'autre, pour ne pas tomber dans de nouvelles formes de monopole créées par les maisons d'éditions commerciales, est que les chercheurs soient poussés à prendre conscience de l'importance de l'évaluation par les pairs et qu'ils apprennent à la gérer d'eux-mêmes.

2.2 Encodage, cryptage, formats ouverts et « .doc »

L'autre question fondamentale dans le panorama scientifique des publications sur Internet et des projets de numérisation, est le problème des formats d'encodage. Deux mouvement divergents sautent aux yeux : plus les entreprises commerciales essayent d'imposer, sous tous les prétextes possibles, des systèmes de cryptage et de protection des données contre la copie, marquage électronique, traçage (avec souvent des conséquences graves pour ce qui concerne le respect des principes fondamentaux du droit d'auteur, comme la copie privée, ou du respect de la vie privée), plus les chercheurs cherchent à utiliser des formats ouverts (HTML, XML, etc.). Les formats ouverts, il faut le répéter inlassablement, sont un élément indispensable pour préserver la publicité de la recherche, la communication entre différents projets et la conservation des données.
Mais dans les sciences humaines, il existe un problème grave qui dérive de l'utilisation des formats propriétaires des logiciels Microsoft pour l'écriture de toutes sortes de documents scientifiques. Le standard de facto des documents « .doc » – utilisés (et c'est vraiment le comble !) même dans les procédures de soumission de proposition au seins des grand organismes de standardisation, comme le ISO – nuit considérablement à la diffusion et à la conservation de

l'information. D'ailleurs les « .doc », avec tous leurs différents formats tous incompatibles vers le haut, n'est fait pas pour favoriser l'échange d'informations, mais pour alimenter le monopole des logiciels Microsoft.
Un format d'interéchange de documents pour l'usage quotidien des chercheurs en sciences humaines et pour la publication sur le Web, qui soit éventuellement compatible mais non dépendant de Microsoft Word, reste encore à trouver. Nos quatre cas d'étude ne font pas exception à la règle, même s'ils s'efforcent de fournir des documents en HTML ou en XML, leur matériaux de départ sont presque toujours des fichiers « .doc » ou, dans les meilleurs des cas des « .rtf » [pour les images, en revanche, la situation est beaucoup moins verrouillée, avec des formats standards plus ouverts comme jpeg, ou png].

2.3 Logiciel Open Source

La nécessité d'utiliser des formats libres pour le stockage des informations scientifiques devrait s'accompagner de l'utilisation de logiciels libres, régis par la licence GPL ou similaires, pour les produire et pour les partager. À ce propos, notre étude représentent un échantillon assez représentatif, parce que des quatre lieux d'expertise, deux affichent une démarche radicalement Open Source, l'un est basé plutôt sur Mac et l'autre plutôt sur Windows. Il est évident que le besoin d'autonomie de la part de la communauté scientifique s'accorde mal avec l'utilisation de logiciels propriétaires, pour ne pas parler du fait que le budget normalement assez serré des projets en sciences humaines peut tirer grandement profit du coût zéro de ces logiciels. D'autre part, l'utilisation des logiciels libres et de l'Open Source représente la seule chance possible pour toute tentative de créer et de diffuser un modèle de publication savante sur Internet.

2.4 Licences Open Source

[Voir notre appendice juridique pour une présentation générale de ces questions]
L'étude de licences juridiques aptes à garantir la diffusion et la libre utilisation des textes scientifiques publiés sur le Web, avec le respect de la paternité intellectuelle, semblerait être une conséquence logique des entreprises de numérisation et de publication sur le Web. Et pourtant, notre étude témoigne plutôt qu'il y a en sciences humaines un manque de réflexion juridique, ou une volonté de ne pas innover dans ce domaine.
SATOR ne se pose pas le problème juridique, alors que le type de thesaurus qu'elle construit est un objet assez délicat du point de vue du droit d'auteur. Cromohs, qui d'ailleurs ne se pose pas dans une démarche Open Source pour ce qui concerne les logiciels utilisés, cherche surtout à garantir aux textes publiés sur son site la même valeur juridique que les articles publiés dans une revue sur support papier et, premier cas en Italie, a procédé avec succès à toutes les procédures d'enregistrement et de dépôt légal auprès des autorités compétentes. L'équipe de l'HyperNietzsche, en plus de garantir cette même exigence, a cherché dès le début à mener une réflexion juridique approfondie sur les modalités de publication sur le Web dans un contexte Open Source et a rédigé une nouvelle licence, la licence OpenKnowledge, comme modèle du statut juridique de la publication sur Internet, en sciences humaines, dans un contexte d'enseignement et de recherche. Fabula, par contre, tout en affichant une démarche Open Source pour ce qui concerne le logiciel, n'a pas jugé intéressant de l'étendre des logiciels aux textes.

2.5 Revues, centres de services, hypertextes

Le panorama de la publication savante sur le Web nous montre la permanence et la grande vitalité du modèle de la revue [Voir Paolo D'IORIO, "Per una rivista elettronica di storia della filosofia", in Umanesimo & informatica. Le nuove frontiere della ricerca e della didattica nel campo degli studi letterari, a cura di Daniela Gruber e Patrick Pauletto, Trento: Metauro, 1997, pp. 109-127]
Et pourtant, cette structure de la communication scientifique, devrait désormais être obsolète. Une revue se fonde sur quatre éléments fondamentaux : une structure périodique,

l'identification d'un champ de recherche bien déterminé, une sélection de matériaux envoyé pour publication (Review), la volonté de publier (sauf exception) seulement des contributions inédites représentant l'actualité de la recherche. La structure périodique et le rapport avec l'actualité sont des éléments qui ne devraient avoir plus de sens sur Internet. Et effectivement, les expertises conduites sur Fabula et Cromohs nous ont démontré que les utilisateurs perçoivent ces sites de plus en plus en termes de centres de service, de banque de données, et qu'ils apprécient davantage la structure de la newsletter que celle du périodique, qu'ils cherchent sur Internet des données pour répondre à des questions pointues et que donc la nouveauté de l'information s'accompagne d'un désir d'exhaustivité.

Pourquoi donc le modèle du périodique résiste-t-il ? Dans les sciences de la nature, nous avons un élément de réponse assez concret à cette question qui est la volonté des maisons d'édition de préserver leur monopole dans le domaine. La forme du périodique est donc la plus adaptée à des fins commerciales, car elle permet de faire payer un abonnement et de pouvoir compter sur des revenus réguliers.

En sciences humaines, la structure du périodique et la volonté de ne publier que des nouveautés semble plutôt l'expression de la permanence de vieilles habitudes de publication. En effet, la gestion de l'évaluation par les pairs et l'identification d'un domaine disciplinaire peuvent être réalisés de façon encore plus rigoureuse en utilisant la structure de l'hypertexte, ou plutôt de ce modèle particulier d'hypertexte savant qui a été développé par l'équipe de l'HyperNietzsche.

Le rapprochement du panorama de l'édition savante actuellement disponible sur le Web et des résultats de l'enquête sur les usages de nos quatre expériences pilotes, nous donne la sensation d'un processus inachevé et d'une permanence des vieilles structures de la communication savante avec une fonction seulement rituelle. Les nouvelles revues électroniques raisonnent très souvent avec la même logique qu'une publication périodique, à comité scientifique presque immuable et restreint, qui contient des essais inédits. Il nous semble qu'il faudrait essayer d'aller au-delà de ces structures, de rapprocher les mécanismes de Peer Review à des désignations du comité scientifique dans la dynamique de la recherche, de privilégier l'hypertextualité à la périodicité et, surtout dans les sciences humaines, d'intégrer la numérisation des sources primaires et des acquis historiques d'une discipline aux découvertes les plus récentes.

La logique qui sépare d'un côté les grands projets de numérisation rétrospective et de l'autre les périodiques électroniques consacrés aux dernières nouveautés est absurde. Surtout sur Internet ; surtout dans les sciences humaines qui vivent dans l'intégration entre passé et avenir, dans lesquelles un bon vieil article a un taux d'obsolescence très bas, surtout par rapport aux sciences de la nature.

Le couple Cromohs et Eliohs et le projet HyperNietzsche s'efforcent de dépasser cette dichotomie en mettant en place un système de publication parallèle ou hypertextuelle de sources et études, d'articles édités et inédits.

Dans le cas de l'HyperNietzsche, le système de contextualisation hypertextuelle et celui de navigation par sigle, permet d'établir de manière simple et immédiate des liens entre des vieux et de nouveaux essais utilisant comme unité de base la page pour les essais numérisés en facsimilé et le numéro de paragraphe pour les essais numérisés en mode texte.

À notre avis, donc, les projets de numérisation des revues savantes et d'édition universitaire en ligne doivent réfléchir à ces trois facteurs : les usages des revues savantes en ligne montrent une tendance vers le dépassement des vieilles formes de communication scientifique ; la coexistence plus ou moins conflictuelle avec le modèle économique marchand ; mais surtout la nécessité d'éviter que ne se créent des positions de monopole comme il est arrivé dans le domaine des sciences de la nature.

Des projets de numérisation tournés vers l'avenir, visant à la numérisation des matériaux de la recherche (fonds d'archives, manuscrits, documents historiques, anciennes revues savantes qui servent à reconstruire l'histoire d'une discipline) et des contributions des chercheurs (livres, revues plus récentes), pourraient représenter un instrument formidable de politique culturelle pour un accès numérique au patrimoine écrit et à sa valorisation scientifique.

Si en outre, ces projets de numérisation que nous appelons de nos vœux étaient accompagnés de systèmes hypertextuels au sein desquels accueillir les résultats actuels du travail des chercheurs, c'est-à-dire s'ils intégraient un système de publication d'articles et d'autres contributions en ligne, alors seraient posées les bases pour l'Internet savant de l'avenir.

Appendice juridique

Il nous semble que la situation juridique ne pose pas d'obstacles à des projets de numérisation de ce type (sources documentaires et articles des chercheurs).
Des recherches en ce sens ont été développées dans le cadre du projet HyperNietzsche, pour ce qui concerne le droit français et allemand, et par l'équipe de Cromohs, pour le droit italien. Il faudrait les approfondir, les élargir à un contexte encore plus international et surtout porter à la connaissance des chercheurs (et des archives, des bibliothèques, des musées) leurs droits et leurs devoirs.
Les questions qui concernent le régime juridique les publications scientifiques peuvent se repartir en deux catégories :
1) Les droits sur les sources documentaires qui forment l'objet d'étude du travail des chercheurs (fonds d'archives, manuscrits, objets d'art ou archéologiques, etc.). Le problème de l'accès aux données tombées dans le domaine public est particulièrement important : les chercheurs ont le plus souvent beaucoup de difficultés pour y accéder, et plus encore pour les numériser et les diffuser. Or, ces données, par principe, sont libres de droit..
2) Les droits sur les articles produits par les chercheurs et publiés sur Internet. Notamment, à quelles conditions un enseignant/chercheur peut publier, de son propre chef, sur internet (sur un site personnel, celui d'un centre de recherche ou d'une institution) ses productions ?
Ces questions juridiques sont développées par ailleurs dans le cadre du PNER : Philippe Chevet est en effet chargé d'une étude sur « la numérisation et la diffusion des fonds des chercheurs ».
Concrètement, notre groupe de travail alimente l'étude de P. Chevet, en lui soumettant des cas concrets, et d'autres études encore du groupe droit du PNER (par exemple, celle menée par Mme Mallet-Poujol pour ce qui concerne les questions de données personnelles liées aux pratiques de listes et forums de discussion,…).
Les aspects juridiques de la publication scientifique en ligne sont donc largement pris en compte dans notre réflexion, mais traités plus spécifiquement par le groupe droit du PNER.
Cette thématique, et celle plus large de l'Open Source, ont été également abordées lors d'un colloque organisé par le PNER, en partenariat avec les associations Fabula et HyperNietzsche, à l'Ecole Normale Supérieure de Paris les 21 et 22 janvier 2002.