

# Les communautés savantes sur le Web

Paolo D'Iorio, « Les communautés savantes sur le Web », dans Danielle Londei et Matilde Callari Galli (éds.), *Multiculturalità e Plurilinguismo in Europa. Percorsi alla francese?*, Bologna, Bononia University Press, 2009.

|  |   |
|--|---|
| <i>I. Une infrastructure de recherche numérique pour les sciences humaines</i> ..... | 2 |
| 1. L'infrastructure traditionnelle et ses problèmes .....                            | 2 |
| 2. L'ancien paradigme de l'informatique littéraire .....                             | 4 |
| <i>II. Un nouveau paradigme</i> .....  | 6 |
| 1. Les conditions de possibilité de la recherche en sciences humaines. ....          | 6 |
| 2. Les communautés savantes sur le Web .....   | 8 |
| 3. Scholarsource .....   | 9 |

Je vais vous parler d'un tout petit monde, le monde de la recherche et de l'enseignement en sciences humaines. Ce tout petit monde n'est certes pas le centre du monde, même si parfois il croit l'être, mais il devrait être le dépositaire des connaissances fiables, des méthodes rigoureuses, de l'esprit critique, même si de temps à temps il l'oublie. Il est finalement le monde que je connais le plus, celui où j'habite et je travaille, mon tout petit monde à moi.

Je ne suis pas un sociologue, donc je ne vous présenterai pas une description de ce que nous faisons, nous les savants, face à l'avènement du numérique. Mais plutôt, puisque je suis un philosophe, je tâcherais de vous dire ce que les chercheurs *devraient faire* face au numérique.

## I. Une infrastructure de recherche numérique pour les sciences humaines

Il y a environ douze ans, je me suis posé la question de savoir s'il était possible d'utiliser Internet pour la recherche en sciences humaines, en particulier pour établir, commenter, analyser les grands textes de la tradition littéraire et philosophique, ce qu'on pourrait appeler le patrimoine écrit. En 2000, le livre intitulé *HyperNietzsche* s'ouvrait sur cette question, comme l'indiquait son sous-titre : *Modèle d'un hypertexte savant sur Internet pour la recherche en sciences humaines. Questions philosophiques, problèmes juridiques, outils informatiques*<sup>1</sup>. En m'efforçant de répondre à cette question à la fois conceptuellement et par des réalisations concrètes, j'ai peu à peu été conduit à la reformuler dans ces termes : comment est-il possible de créer des infrastructures de recherche numériques pour les sciences humaines ?

### 1. L'infrastructure traditionnelle et ses problèmes

En premier lieu, il ne faut pas oublier qu'il existe une infrastructure de recherche traditionnelle pour les sciences humaines, chercher à comprendre comment elle fonctionne et quels sont ses problèmes. Dans l'environnement de recherche traditionnel, non-numérique, le chercheur se rend habituellement dans des lieux d'archives ou dans des bibliothèques afin d'étudier les sources primaires nécessaires à ses travaux ; en bibliothèque il lit non seulement les sources primaires mais aussi les sources secondaires, revues, monographies, éditions critiques, qu'il peut dans un certain nombre de cas se procurer également chez les libraires ; les cours à l'université, les séminaires et les colloques lui donnent l'occasion d'échanger ses connaissances avec ses collègues et de les transmettre à ses élèves ; enfin notre chercheur publie ses propres études critiques chez des éditeurs et les diffuse sur le marché du livre selon les règles établies par le droit d'auteur. Tous ces éléments matériels et organisationnels – archives, bibliothèques, librairies, universités,

<sup>1</sup> Paolo D'Iorio (éd.), *HyperNietzsche. Modèle d'un hypertexte savant sur Internet pour la recherche en sciences humaines. Questions philosophiques, problèmes juridiques, outils informatiques*, Paris, PUF, 2000, 200 p. Premier et seul livre des Presses Universitaires de France à avoir été distribué en même temps sous forme imprimée et payante en librairie et sous forme numérique et gratuite sur le site Web de l'éditeur, il est désormais accessible à l'adresse [www.hypernietzsche.org/doc/puf](http://www.hypernietzsche.org/doc/puf). Voir également P. D'Iorio, « HyperNietzsche. Modello di un ipertesto di ricerca », in Carlo Gentili, Volker Gerhardt, Aldo Venturelli (éds.), *Nietzsche illuminismo modernità*, Firenze, Leo S. Olschki, 2003, pp. 343-367 et *id.* P. D'Iorio, « Nietzsche on New Paths: The HyperNietzsche Project and Open Scholarship on the Web », in Maria Cristina Fornari (éd.), *Friedrich Nietzsche. Edizioni e interpretazioni*, Pisa ETS, 2006, pp. 475-496 (version électronique disponible à l'adresse : [www.hypernietzsche.org/doc/files/new-paths.pdf](http://www.hypernietzsche.org/doc/files/new-paths.pdf)).

cours, colloques, maisons d'édition, droit d'auteur et autres encore – constituent l'infrastructure traditionnelle de la recherche en sciences humaines telle qu'elle s'est formée lentement au cours des siècles passés, depuis la bibliothèque d'Alexandrie, et au sein de laquelle évoluent encore aujourd'hui nos modes de recherche. Disons le tout de suite, c'est une infrastructure qui atteint son objectif : celui de produire et de diffuser de la connaissance. Mais alors, pourquoi avons-nous besoin d'une infrastructure de recherche numérique ? parce que l'infrastructure traditionnelle souffre d'un certain nombre de problèmes. J'en nommerai trois.

A) L'environnement non-numérique est lent et coûteux aussi bien pour ce qui concerne l'accès aux sources primaires que pour la diffusion des études critiques : pour étudier les manuscrits d'un seul livre de Jean-Jacques Rousseau, *La nouvelle Héloïse*, il faut se rendre dans les bibliothèques de Paris, Genève et New York, les différentes versions manuscrites étant conservées dans ces trois endroits ; et pour diffuser les résultats de la recherche il faut les imprimer en plusieurs copies et les transporter matériellement chez les libraires, à des centaines de kilomètres de distance.

B) Plus le domaine d'étude s'accroît, moins l'infrastructure traditionnelle devient capable d'assurer la collaboration des chercheurs et le partage des connaissances. Elle est capable de gérer les connaissances produites par 100 chercheurs qui travaillent sur Kant, mais quand ce nombre passe à 1 000 ou à 10 000, les techniques traditionnelles d'organisation et de transmission de la connaissance ne sont plus en mesure d'assurer que, par exemple, un chercheur français soit informé qu'un collègue allemand, brésilien ou japonais a publié un article important sur un passage ou sur un concept de la *Critique de la raison pure* qui l'intéresse. Certes les bibliographies et les index existent, qui pourraient transmettre l'information, mais quand le domaine de recherche est trop étendu, ils deviennent rapidement insuffisants. Le résultat est que des chercheurs dans différentes villes ou pays se penchent sur un problème qui a déjà été résolu, ou alors qu'ils n'utilisent pas, ni comparent ni croisent des connaissances qui pourraient être utilisées, comparées et croisées avec profit.

C) Un troisième problème est l'utilisation par certaines maisons d'édition de l'évaluation par les pairs et du droit d'auteur pour mettre en place un marché de monopole des périodiques scientifiques. Un article éclairant de Jean-Claude Guédon<sup>2</sup> retrace l'histoire de la crise du prix du périodique, comme effet pervers du mécanisme d'évaluation quantitative du travail des chercheurs inventé dans les années 1950 par Eugene Garfield: le *Science Citation Index* (SCI). Le SCI comportait la définition d'un ensemble de « *core journals* », un ensemble de périodiques fondamentaux, vers lesquels s'orientaient ensuite les pratiques d'acquisition des bibliothèques.

Pour la publication savante en sciences humaines, il n'y a pas de marché de monopole – pour la bonne raison qu'il n'y a pas de marché du tout. Premièrement parce que les bibliothèques dépensent désormais 80 % de leur budget pour les périodiques scientifiques et elles ont de moins en moins d'argent pour acheter les monographies et les périodiques en sciences humaines. D'autre part les sciences humaines sont depuis toujours dans une situation de crise pour ce qui concerne la publication des recherches érudites ou des éditions volumineuses, ou en général dans tous les cas où les exigences de la science et celles du marché du livre entrent en conflit. Le résultat est que les chercheurs sont de plus

<sup>2</sup> Jean-Claude Guédon, « In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing », in *Creating the Digital Future*, Proceedings of the 138th Annual Meeting of the Association of Research Libraries, 2001, disponible sur Internet à l'adresse : [www.arl.org/resources/pubs/mmproceedings/138guedon.shtml](http://www.arl.org/resources/pubs/mmproceedings/138guedon.shtml) ; imprimé également sous forme de livre par l'Association of Research Libraries en 2005, tr. it. *Per la pubblicità del sapere. I bibliotecari, i ricercatori, gli editori e il controllo dell'editoria scientifica*, Pisa, Plus, 2004.

en plus obligés de verser une bonne partie de leur maigre budget pour financer la publication de leurs livres. Et puis, *dulcis in fundo*, la *European Science Foundation* est en train de préparer un *European Reference Index for the Humanities* avec pour objectif d'utiliser des critères objectifs pour évaluer l'activité des chercheurs (facteur d'impact) et de « donner plus de visibilité » à la *top quality European Humanity research*, avec le résultat d'en obscurcir la plupart, sans parler du fait que la liste ne comprend aucune publication en ligne et donc constitue un frein formidable à la diffusion des publications électroniques<sup>5</sup>. Il faut comprendre que la responsabilité première est celle des chercheurs. Tant qu'ils seront incapables d'organiser une évaluation par les pairs indépendante des maisons d'édition et basée uniquement sur des critères scientifiques et non économiques utilisant le cas échéant des systèmes de publication alternatifs, ils ne pourront pas sortir de ce piège. La plupart des chercheurs en sciences humaines ne sont pas encore complètement conscients de ces problèmes. Mais ils auront peut-être remarqué que certains éditeurs ont commencé à leur demander des contributions importantes aux frais de publications pour des livres qui sont à peine distribués chez les libraires. Bientôt même les chercheurs en sciences humaines se trouveront aux prises avec le « facteur d'impact » et ils seront peut-être surpris de constater qu'il est très bas, parce qu'ils ne publient pas dans les bonnes revues, ou alors parce qu'ils publient davantage de livres que d'articles.

## 2. L'ancien paradigme de l'informatique littéraire

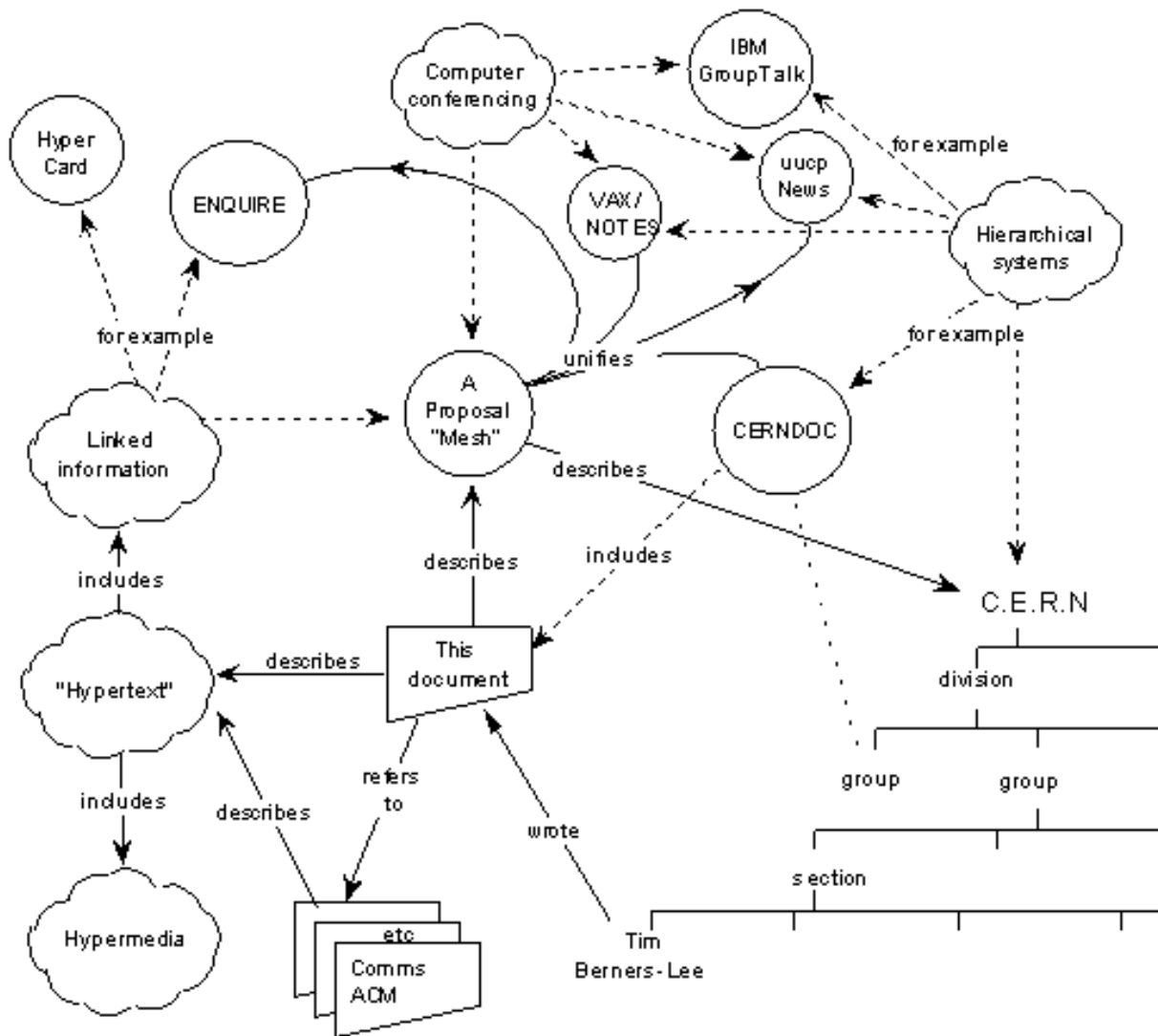
Est-ce que l'informatique littéraire (*humanities computing*) peut servir à résoudre ces problèmes ? Voyons ce qui s'est passé quand l'ordinateur et Internet ont été introduits dans l'infrastructure traditionnelle de la recherche. Au début l'ordinateur a été utilisé pour des analyses lexicales et comme aide à la création de concordances, thésaurus, etc., donnant naissance ainsi à une nouvelle discipline : la linguistique computationnelle. Ensuite, avec la diffusion de l'ordinateur de bureau, la production de textes numériques est devenue pratique courante, mais curieusement les textes numériques, au lieu d'être immédiatement publiés sur Internet, sont imprimés sous forme de livre, souvent en très peu d'exemplaire, pour être par la suite re-numérisés par des projets comme Google Books ou similaires. Ainsi les organismes publics payent une première fois pour financer la recherche (ce qui est bien), une deuxième fois pour imprimer les résultats en livre (ce qui est déjà moins acceptable) et une troisième fois pour numériser les livres (ce qui est franchement excessif).

Le résultat, dans les meilleurs des cas, est la création d'un système global "googlisé", séparé du domaine officiel de la recherche et formé par une masse de textes numérisés de différents types et avec différents niveaux de fiabilité scientifique qui ne sont consultables que par le biais d'un moteur de recherche. Cela est certainement très utile, mais en même temps c'est comme si on avait pris l'infrastructure traditionnelle de la recherche, qui était capable de produire de la connaissance, et qu'on l'avait passée dans une sorte de moulinette numérique réduisant ainsi la connaissance à une simple information déstructurée. Dans ce modèle, toute forme de structure et de sémantique est perdue. On cherche des mots et on reçoit des listes d'occurrences : c'est tout. Telle est la forme de recherche d'information typique de la linguistique computationnelle qui, à travers le modèle influent de Google, est en train de devenir l'unique forme de navigation imaginable y compris dans le domaine des bibliothèques numériques, parfois combiné avec des structures hiérarchiques comme des arbres ou des catégories.

Malheureusement (ou plutôt heureusement) la connaissance produite par les chercheurs n'est pas structurée sous la forme de listes ou d'arbres, mais plutôt sous la

<sup>5</sup> Voir [www.esf.org/research-areas/humanities/research-infrastructures-including-erih.html](http://www.esf.org/research-areas/humanities/research-infrastructures-including-erih.html).

forme d'un graphe au sens mathématique du terme, un graphe avec des arcs étiquetés. Songez un moment à un chercheur, par exemple un classiciste, au travail dans une bibliothèque de l'ère pré-électronique. Que fait-il ? Certes, parfois il consulte des listes : les catalogues de la bibliothèque, des bibliographies, des index. Mais la plupart du temps il lit. Bien sûr il lit à sa manière : il ouvre une douzaine de livres sur sa table et il passe de l'un à l'autre en suivant différents types de relations entre les textes, que d'autres chercheurs ont tracées par les références contenues dans les notes en bas de pages, selon l'ancien et vénérable instrument de la citation savante. Le système de navigation dans le graphe de la connaissance scientifique n'est pas l'arbre, ni la liste, mais un tissu de références et de citations. Tim Berners-Lee l'avait bien compris, d'après ce qu'on peut voir dans le diagramme qui accompagnait la première idée d'un système d'information avec des liens étiquetés qui allaient devenir le World Wide Web et ensuite le Web sémantique.<sup>4</sup>



<sup>4</sup> *Information Management: A Proposal*, Tim Berners-Lee, CERN, March 1989, May 1990, [www.w3.org/History/1989/proposal.html](http://www.w3.org/History/1989/proposal.html).

## II. Un nouveau paradigme

Il semble que l'approche de l'informatique littéraire en tant que recherche dans des bases de données textuelles ne soit pas en mesure de créer une infrastructure numérique équivalente à l'infrastructure traditionnelle et capable de résoudre ses problèmes. Au contraire : il a plutôt l'effet de transformer la connaissance en simple information déstructurée. Et pourtant, une véritable infrastructure de recherche sur le Web devrait permettre aux chercheurs d'effectuer la totalité ou du moins la plupart des activités scientifiques qu'ils accomplissent dans le cadre de l'infrastructure traditionnelle, de travailler de manière plus souple et efficace, et de disposer de nouveaux services et outils caractéristiques d'un environnement électronique. Elle devrait nous permettre non seulement d'avoir accès aux sources primaires et secondaires nécessaires à nos recherches, comme nous le faisons en nous rendant dans des archives et dans des bibliothèques ou chez les libraires, mais également de publier nos études critiques dans un contexte prestigieux et qui résiste à l'épreuve du temps, comme nous avons le sentiment de faire en envoyant un manuscrit à une maison d'édition, enfin l'environnement numérique devrait nous permettre de transmettre et partager la connaissance avec nos étudiants et nos collègues, comme nous le faisons quand nous enseignons à l'université.

Si tout le monde est convaincu qu'Internet va changer la recherche, malgré quelques initiatives il semble que les chercheurs en sciences humaines n'aient pas encore trouvé la manière de donner une forme au Web qui le rende utilisable pour leurs disciplines. Nous avons en effet besoin d'un nouveau paradigme. À mon avis, ce nouveau paradigme : 1) devrait tenir compte des conditions de possibilité de la recherche, 2) devrait considérer les « communautés savantes sur le Web » comme le sujet qui peut guider la transition vers le numérique, 3) pourrait profiter des modèles conceptuels et des technologies qui sont développés dans le cadre du projet *ScholarSource*.

### 1. Les conditions de possibilité de la recherche en sciences humaines.

Par conditions de possibilité, utilisant une expression kantienne dans un sens non kantien, je me réfère à ces règles générales sans lesquelles ou en cas de leur modification on se retrouverait dans un jeu complètement différent par rapport à celui de la recherche scientifique. Il me semble que l'on peut en voir au moins trois : la citation, le consensus et la diffusion / conservation.

Ceci peut sembler banal, mais la *possibilité de citer* des documents de manière stable et durable m'apparaît comme l'une des conditions essentielles de la recherche scientifique. En effet, la science est une conversation fondée sur des hypothèses, des arguments et des faits. Dans les sciences humaines, les faits sont souvent contenus dans des documents, dans des textes. Le fait qu'Emma Bovary ingère de l'arsenic est au moins aussi sûr et incontestable dans les études littéraires que l'existence des électrons en physique<sup>5</sup>, mais pour vérifier ce fait, il faut pouvoir citer l'édition originale de *Madame Bovary*. Sans la possibilité de citer, il n'est pas possible de vérifier les faits, ni de produire des interprétations ou d'en discuter publiquement. Le système éditorial traditionnel est normalement en mesure d'assurer la stabilité du texte et des données bibliographiques. Sur le Web, en revanche, la citation est une entreprise difficile voire désespérée parce que les textes aussi bien que leurs adresses sont habituellement très instables : les pages Web changent tous les jours, disparaissent, apparaissent sous d'autres noms et d'autres adresses. Mais comme qui peut le plus peut le moins, il n'est pas impossible de concevoir des applications Web qui forment un espace dans le Web, comme de petites îles, où les adresses et les textes restent immuables. Toutefois, il faut être conscient que l'existence de solutions techniques ne suffit pas à

<sup>5</sup> Voir Umberto Eco, *Les limites de l'interprétation*, Paris, Grasset, 1992.

assurer la stabilité des texte et des adresses Web si elle n'est pas accompagnée par la volonté explicite des responsables des sites sur lesquels sont enregistrées les données d'installer et de maintenir ces technologies dans le temps : il ne s'agit pas d'une question technique, mais d'une décision de politique scientifique

L'évaluation de la qualité scientifique, fondée sur le jugement et le *consensus* de la communauté des pairs, est la deuxième condition de possibilité de la recherche scientifique. Dans la science, comme dans beaucoup d'autres activités humaines, un chercheur qui aspire à faire partie de la communauté des savants et à participer aux recherches collectives, doit d'abord produire des travaux jugés dignes d'intérêt par ses collègues. Mais plus que dans d'autres activités, l'évaluation scientifique est censée être la plus impartiale et transparente possible et fondée sur des preuves incontestables. Naturellement cela est très difficile à réaliser et les procédures d'évaluation ne sont jamais exemptes de critiques. Le Web à ce propos est potentiellement plus performant que l'environnement traditionnel et pourrait offrir l'occasion d'organiser l'évaluation scientifique de manière plus simple, plus transparente et plus efficace. Sur le Web, plusieurs voies sont envisageables : créer un environnement numérique qui reproduise les procédures traditionnelles d'évaluation par les pairs ; utiliser la logique du réseau et déterminer le facteur d'impact à partir du balisage sémantique effectué par les usagers de sites comme *Citeulike* ou *Connotea*<sup>6</sup> ; ou encore avoir recours à des indicateurs de popularité comme la quantité de citations ou de téléchargements d'un article, etc. La revue *Nature*, à travers son blog *Peer-to-Peer* a commencé une réflexion sur différents modèles d'évaluation par les pairs sur le Web, tout en expérimentant concrètement un nouveau modèle de revue en ligne : *NaturePrecedings*<sup>7</sup>. Pourtant même dans ce cas, il ne s'agit pas d'un problème de technique, mais d'une décision de politique scientifique et je ne suis pas absolument sûr que la totalité de nos collègues souhaite expérimenter toutes ces possibilités, si transparents, équitables et démocratiques...

Venons à la troisième condition de possibilité : la *conservation*. Nous travaillons encore aujourd'hui avec des documents originaux qui parfois remontent au Moyen-Age et nous pouvons encore lire les textes écrits par les Grecs et les Romains. Mais pouvons-nous être sûrs que nos documents électroniques nous survivront ? Qu'en sera-t-il de notre serveur, de nos éditions et de nos articles électroniques dans vingt ans, dans cent ans, dans quelques siècles ? *Lots of Copies Keep Stuff Safe* (LOCKSS) est le nom d'un important projet international pour la sauvegarde du patrimoine numérique. Les bibliothécaires de Stanford qui ont initié ce projet considèrent que la manière la meilleure de conserver un texte n'est pas d'en déposer l'unique copie dans un endroit "sûr", mais plutôt d'en faire des milliers de copies et de les éparpiller un peu partout dans le monde. Le principe de LOCKSS est confirmé par trente siècles d'histoire culturelle : tout ce que nous n'avons pas copié, nous l'avons perdu. Pourquoi avons-nous perdu presque toutes les œuvres d'Héraclite et conservé presque toutes celles d'Aristote ? Parce qu'Héraclite avait une politique de DRM (*Digital Rights Management*, gestion numérique des droits) : il confia une copie de son œuvre au temple d'Artémis, le lieu le plus sûr de son époque, afin que ses pensées soient conservées, mais pas copiées et diffusées. Aristote, au contraire, était un tenant du *copyleft* et avait livré tout son savoir à ses élèves qui en firent des copies ; et copie après copie, l'œuvre d'Aristote est arrivée jusqu'à nous. Ce n'est qu'une boutade, mais il est vrai que les pires ennemis de la conservation des textes sont le *copyright*, les DRM, les formats

<sup>6</sup> Voir [www.citeulike.org](http://www.citeulike.org) ; [www.connotea.org](http://www.connotea.org).

<sup>7</sup> Voir [blogs.nature.com/peer-to-peer/](http://blogs.nature.com/peer-to-peer/) ; [precedings.nature.com](http://precedings.nature.com). Pour une discussion critique de ces différentes expériences, voir l'article de Francesca Di Donato, « Come si valuta la qualità nella Repubblica della Scienza? Una riflessione sul concetto di peer review », *Bollettino telematico di filosofia politica*, juillet 2007, [purl.org/hj/bfp/164](http://purl.org/hj/bfp/164).

propriétaires et tous les moyens juridiques ou techniques d'interdiction de la copie ; tandis que ses meilleurs amis sont le libre partage, le *copyleft*, les formats ouverts. Donc, pour maximiser la conservation et la diffusion, qui sont les deux faces de la troisième condition de possibilité de la recherche, nous devrions retirer tous les obstacles juridiques qui si souvent empêchent l'accès aux sources et la diffusion des résultats de la recherche. Encore un problème de politique culturelle et non pas de technologie.

## 2. Les communautés savantes sur le Web

Nous disposons des technologies qui nous permettent de créer des îles des savoirs choisis, où les textes sont stables, leur qualité est garantie et la libre diffusion et les formats ouverts assurent la conservation à long terme. Le problème n'est pas comment le faire. Le problème est *qui* va le faire, qui va prendre ces décisions de politique culturelle qui permettent de saisir pleinement la chance offerte par la diffusion d'Internet. Je pense que pour réaliser les conditions de possibilité de la recherche sur le Web, les chercheurs devraient se rassembler dans une forme de « réseau social » (*social network*) que j'ai appelée les communautés savantes sur le Web. Il s'agit de former des libres associations internationales de spécialistes qui travaillent sur un corpus ou sur un domaine de recherche spécifique. Elles collaborent avec les bibliothèques, les universités, les maisons d'édition, mais ce sont elles qui fixent leurs propres politiques et priorités, qui garantissent à la fois la qualité des publications, la paternité intellectuelle et la stabilité des textes publiés dans les serveurs informatiques qu'elles gèrent directement. Mon modèle est celui des Académies des sciences du XVIIe siècle, l'*Accademia dei Lincei*, l'*Académie Française*, la *Royal Society* étaient les réseaux sociaux qui ont accompagné la naissance de la science moderne. Après l'époque des Académies, un nouveau modèle s'est instauré qui aujourd'hui ne semble pas capable de garantir la transition vers l'époque numérique.

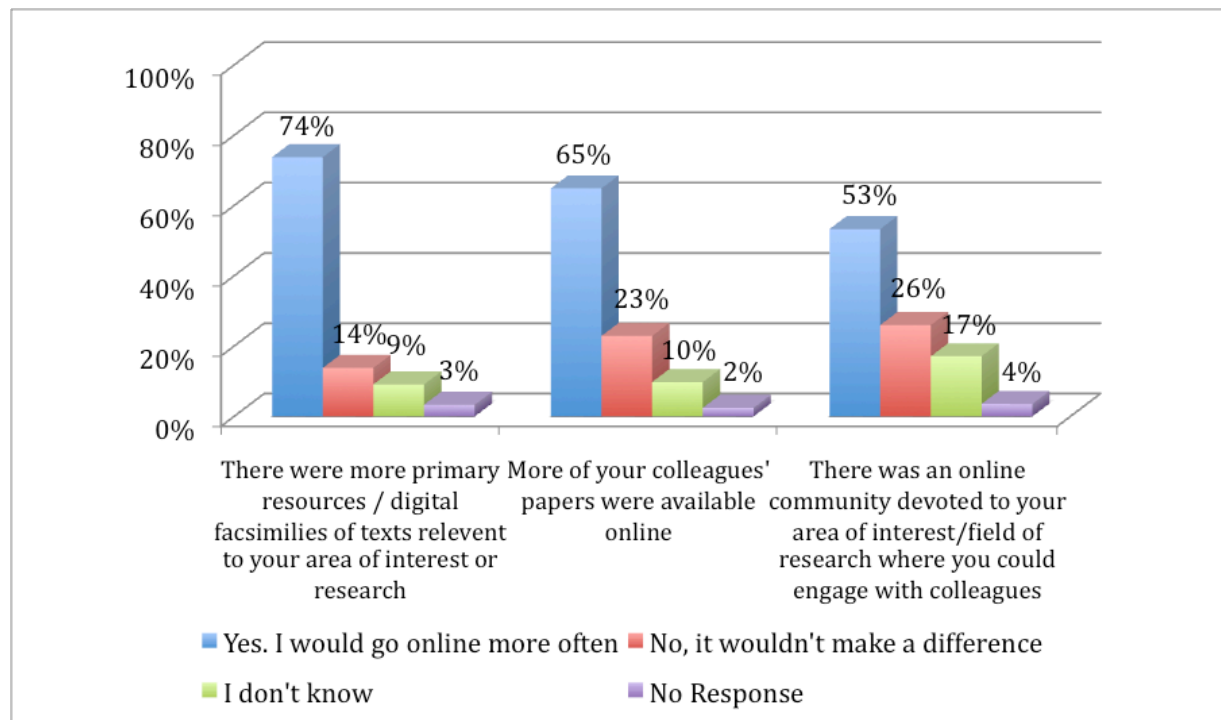
Les communautés savantes sur le Web n'existent pas. Et il sera assez difficile de leur donner naissance parce que : 1) d'une part, les chercheurs manquent souvent de compétences dans le domaine du savoir lettré numérique (*digital literacy*<sup>8</sup>) ; 2) deuxièmement parce que les institutions publiques de certains pays ne sont pas toujours ouvertes à la libre mise en réseau de leurs collections. Trop souvent, même dans le cas de fonds anciens tombés depuis longtemps dans le domaine public et conservés par des institutions publiques, les projets de publication sur le Web sont bloqués par la politique restrictive de certaines d'entre elles qui ne permettent pas aux chercheurs de publier une copie numérique de leur fonds sur Internet ou qui font payer pour cela des « droits de reproduction » exorbitants. Publier sur Internet *une* page venant des collections de la Bibliothèque nationale de France coûte 124 euros ; le tarif est de 107 livres pour 6 mois et 324 livres pour plus de dix ans à la *British Library*, 150 euros à la *Bayerische Staatsbibliothek* de Munich, mais cela ne coûte rien à la *Staatsbibliothek* in Berlin<sup>9</sup>. Il n'y a pas de règle générale ni au niveau national, ni européen, le directeur de chaque lieu d'archive étant le seul responsable de sa politique d'accès et de ses tarifs. Les infrastructures de recherche numériques pour les sciences humaines profiteraient beaucoup de l'existence d'un cadre juridique cohérent au niveau européen assurant que les institutions publiques permettent de reproduire librement au moins ce qui est dans le domaine public. 3) Troisièmement, les maisons d'édition dans leur majorité ne sont pas disposées à suivre les chercheurs dans l'ère numérique. De manière compréhensible, elles veulent préserver le plus longtemps possible leurs vieux modèles économiques, bien rodés et rentables. Moins compréhensible est le fait que les chercheurs soient pieds et poings liés aux éditeurs et heureux de l'être. Néanmoins,

<sup>8</sup> Voir Milad Doueihi, *La grande conversion numérique*, Paris, Seuil, 2008.

<sup>9</sup> Voir [www.bnf.fr/pages/accedocu/redev.htm](http://www.bnf.fr/pages/accedocu/redev.htm) ; [www.bl.uk/imaging/permission.html](http://www.bl.uk/imaging/permission.html) ; [www.bsb-muenchen.de/Reproduktionsgebuehren.793.0.html](http://www.bsb-muenchen.de/Reproduktionsgebuehren.793.0.html).



une étude statistique que j'ai demandée à l'*Oxford Internet Institute* démontre que les chercheurs en sciences humaines seraient heureux d'utiliser davantage Internet s'il y avait davantage de sources primaires et secondaires disponibles en ligne et s'il existait des communautés savantes dans leur domaine d'intérêt<sup>10</sup>.



### 3. Scholarsource

Les conditions de possibilité et les communautés savantes sont des concepts généraux, indépendants des technologies utilisées pour mettre en place une infrastructure de recherche. Dans le cadre de deux projets européens<sup>11</sup>, je me suis efforcé de traduire ces concepts et plus en général le paradigme que j'appelle Scholarsource en des solutions techniques concrètes.

En tant que *projet*, Scholarsource se divise en trois volets qui correspondent à ses trois groupes de travail. Le premier (*communities*) favorise la création de communautés savantes sur le Web, qui travaillent sur un auteur ou sur un thème particulier en utilisant les instruments techniques mis au point par le deuxième groupe de travail (*software*). Le troisième groupe de travail (*policy*) s'interroge sur les conditions juridiques et socio-économiques qui permettent aux communautés savantes de travailler sur le Web, comme la question de l'accès aux sources primaires, des licences pour la publication d'articles sur le Web qui protègent du plagiat, la question des modèles économiques qui assurent la

<sup>10</sup> Christine Madsen, *Online Survey of Humanities Scholars*, Oxford Internet Institute, Oxford, 2008, 33 p.

<sup>11</sup> Le projet eContenplus « Discovery: Digital Semantic Corpora for Virtual Research in Philosophy » ([www.discovery-project.eu](http://www.discovery-project.eu)) et l'Action COST A32 « Open Scholarly Communities on the Web » ([www.cost-a32.eu](http://www.cost-a32.eu)). Pour le passage du modèle HyperNietzsche au paradigme Scholarsource voir P. D'Iorio, « L'île des savoirs choisis. De HyperNietzsche à Scholarsource: pour une infrastructure de recherche sur le Web », in F. Leriche et C. Meynard (éds.), « De l'hypertexte au manuscrit. L'apport et les limites du numérique pour l'édition et la valorisation de manuscrits littéraires modernes », *Recherches e3 Travaux*, n° 72/2008, Grenoble, Ellug, 2008, pp. 279-301 et P. D'Iorio, M. Barbera, « Scholarsource. Designing a Digital Infrastructure in Art and Humanities. A Skype Dialog between a Philosopher and a Computer Scientist », in Thomas Bartscherer and Roderick Coover (éds.) *Switching Codes. Leading scholars, artists, and computer scientists discuss the impact of digital information technology on thought and practice in the arts and humanities*, University of Chicago Press, à paraître.

viabilité des publications électroniques ou l'analyse des usages et des comportements des internautes savants par rapport à d'autres réseaux sociaux.

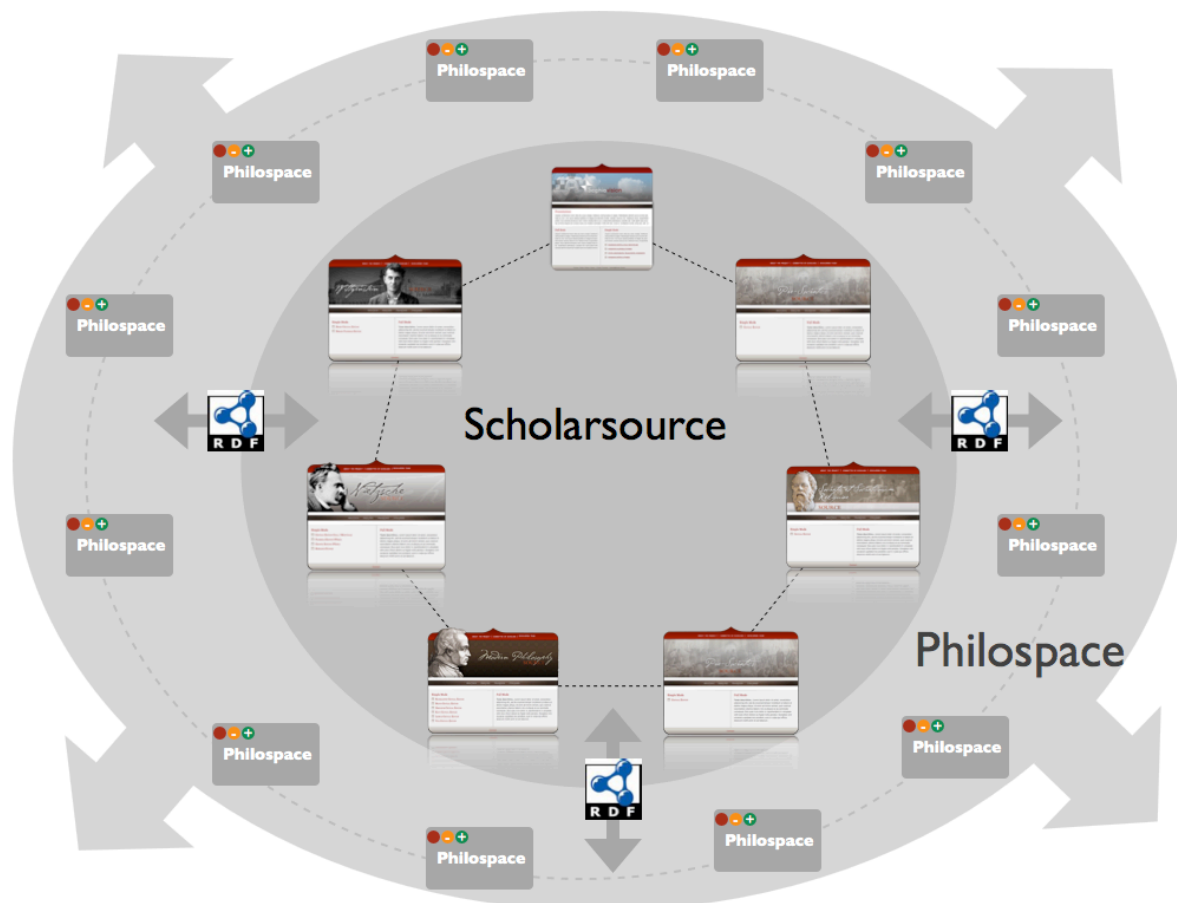
D'un point de vue *technologique*, Scholarsource se présente comme une fédération de bibliothèques de recherche numériques qui font en même temps office de maisons d'édition universitaires dans les domaines des sciences humaines : philosophie, littérature, histoire, etc. Chaque nœud de la fédération est consacré à un corpus particulier et contient des éditions critiques des textes, des éditions fac-similés des sources primaires, une revue contenant des essais, des comptes-rendus, des commentaires, une collection de monographies et une série de cours et conférences audiovisuelles. La fédération est librement consultable à travers un navigateur Web (Firefox, Safari, Explorer) et est constamment incrémentée par de nouvelles contributions scientifiques toujours évaluées par les pairs avant publication. Grâce à l'utilisation des technologies du Web sémantique, Scholarsource est complètement interopérable avec d'autres bibliothèques numériques et en particulier entend s'intégrer à *Europeana*, la bibliothèque numérique européenne. À la fédération Scholarsource s'ajoute un deuxième réseau appelé *Philospace*. Il s'agit d'un réseau fermé, qui n'est pas accessible à travers un navigateur Web, mais par le biais du logiciel homonyme que les chercheurs peuvent télécharger et utiliser sur leur ordinateur. Si Scholarsource est la bibliothèque, l'espace public de la lecture, *Philospace* est comme le calepin, c'est-à-dire l'espace privé du chercheur qui a la possibilité de lire, d'étudier les contenus publiés dans Scholarsource, mais aussi de les annoter et de partager ses notes avec d'autres utilisateurs de *Philospace*.

Concernant les *contenus*, Scholarsource est formé par une structure en cercles concentriques autour du projet pilote Nietzsche Source. Le *premier cercle*, financé par le projet Discovery, est formé par un ensemble de corpus philosophiques dont nous donnons ici un bref aperçu :

- 1) *Grèce ancienne*. Trois grands corpus seront publiés dans Scholarsource par l'*Istituto per il Lessico intellettuale europeo e Storia delle idee* du CNR de Rome : a. l'édition complète des fragments des présocratiques basée sur *Die Fragmente der Vorsokratiker* de Diels et Kranz, en grec avec traduction allemande et italienne ; b. l'édition complète de tous les témoignages sur Socrate, basée sur les *Socraticis et Socraticorum Reliquiae* de Gabriele Giannantoni, en grec ; c. l'édition complète des vies des philosophes de Diogène Laërce, en grec avec traduction italienne.
- 2) *Textes philosophiques et scientifiques de l'époque moderne*. Le CNR publiera en outre un ensemble de 52 œuvres en latin, italien, français d'auteurs modernes comme Giordano Bruno, Descartes, Spinoza, Leibniz, Vico, Baumgarten, Kant.
- 3) *Le dix-neuvième et le vingtième siècles* seront représentés par Nietzsche et Wittgenstein. Concernant Nietzsche, l'Institut des Textes et Manuscrits Modernes du CNRS publiera dans Scholarsource l'édition électronique complète des Oeuvres et de la Correspondance de Nietzsche (basée sur l'édition critique de référence éditée par G. Colli et M. Montinari), la première édition fac-similé du legs posthume de Nietzsche conservé à Weimar ainsi qu'une nouvelle édition génétique de deux œuvres de Nietzsche, *Le Voyageur et son Ombre* et *Aurore*, incluant toute la tradition manuscrite. Les archives Wittgenstein de Bergen (Norvège) publieront dans Scholarsource 5 000 pages de manuscrits du legs de Wittgenstein, comprenant des matériaux venant du *Big Typescript* (1929-1934), du *Brown Book* (1934-1936), de la *Lecture on Ethics* et des *Notes on Logic*.
- 4) *Philosophes contemporains*. La *Radiotelevisione Italiana* (RAI) publiera dans Scholarsource 300 interviews audio-visuelles de philosophes contemporains comme Hans Georg Gadamer, Gilles Deleuze, Gianni Vattimo, et autres.

Le *deuxième cercle* est représenté par une série d'autres projets de sites Web consacrés à des philosophes comme Schopenhauer, mais aussi à des écrivains comme Flaubert, Proust, Valéry, Virginia Woolf, Beckett, Yeats, qui sont réunis dans l'action COST « Communautés Savantes sur le Web ».

Scholarsource disposera d'une fonction appelée « contextualisation dynamique » qui est la forme du référencement savant à l'époque numérique, plus puissante que l'ancienne car bidirectionnelle et dynamique. Bidirectionnelle signifie que le système est en mesure non seulement de pointer vers un lieu textuel mais aussi de remonter dans le sens contraire, à l'origine de la référence. Dynamique signifie que la liste des référencements reçus se met automatiquement à jour sans besoin de dépouiller manuellement les revues, comme c'était le cas dans le *Science Citation Index*. Tandis que l'utilisateur navigue dans le site, passant d'une page à l'autre, cette forme de navigation savante rend disponibles les contributions scientifiques qui concernent la page sélectionnée. Par exemple, si l'utilisateur sélectionne une page manuscrite, le système rend immédiatement accessible, sans nécessiter de complexes recherches supplémentaires, la liste de tous les fac-similés, transcriptions et traductions disponibles concernant cette page, de même que tous les chemins génétiques et les essais critiques qui s'y réfèrent. De la même façon, si l'utilisateur sélectionne un essai critique, le masque de contextualisation présentera une liste d'hyperliens vers tous les textes et les manuscrits de Nietzsche cités par l'essai ainsi que la liste de tous les articles d'autres auteurs cités, ou qui citent l'essai consulté. En outre, les nœuds du réseau Scholarsource sont conçus de manière à pouvoir communiquer aisément entre eux. Imaginons un chercheur qui, dans un essai sur Nietzsche contenu dans Nietzsche Source, cite un passage de Wittgenstein. Le lecteur doit être en mesure de passer, par un simple clic, de Nietzsche Source à Wittgenstein Source, en suivant un lien croisé qui lui permette de retrouver le contexte original de la citation, de la voir commentée par des spécialistes de Wittgenstein ou de remonter jusqu'aux brouillons manuscrits.



Pour résumer : Scholarsource a l'ambition de reproduire toutes les fonctions de l'environnement traditionnel de la recherche en une infrastructure numérique sur le Web et de contribuer aussi à résoudre certains de ses problèmes, comme les trois que j'ai mentionnés au début (§ I.1).

A) Grâce à la puissance du Web, l'accès aux sources primaires et la diffusion des sources secondaires deviendront évidemment plus rapides et économiques. Mais à la différence du Web, les documents publiés dans Scholarsource remplissent les conditions de possibilité et sont ainsi pleinement utilisables pour la recherche : nous pourrions les citer de manière stable, nous pourrions être assurés qu'ils ont reçu le consensus de la communauté des chercheurs et nous pourrions être raisonnablement sûrs de leur diffusion et conservation. En outre, les communautés savantes seront placées dans la meilleure position pour choisir les documents à numériser au-delà des diverses collections, bibliothèques ou pays, alors que pour le moment le choix est fait par les bibliothèques ou par Google. Mais, pour reprendre l'exemple de Rousseau, si l'on devait attendre que les trois bibliothèques (Paris, Genève et New York) qui possèdent les diverses parties des manuscrits de *La Nouvelle Héloïse* se mettent spontanément d'accord pour les numériser de concert, on pourrait attendre jusqu'à la prochaine ère numérique, celle du troisième millénaire. Cela ne relève tout simplement pas leur métier, mais c'est de l'intérêt des spécialistes de Rousseau.

B) À la différence de l'infrastructure traditionnelle, plus Scholarsource s'accroît, plus il devient efficace : plus nombreux sont les articles publiés dans Scholarsource, plus il sera probable d'y trouver un commentaire au passage ou au concept de Kant qui intéresse tel ou tel spécialiste. Et il n'aura même pas besoin de le chercher, parce que grâce à la contextualisation dynamique il sera automatiquement lié au passage correspondant de la *Critique de la raison pure*.

C) La Contextualisation Dynamique fait que la liste des articles qui cite un certain passage (liste des référencements reçus) se met automatiquement à jour sans besoin de dépouiller manuellement les revues, ou les monographies, comme c'était le cas pour le *Science Citation Index*. Si plusieurs communautés savantes se mettaient d'accord pour utiliser cette technologie ou une autre similaire basée sur RDF (le standard du Web sémantique), cela permettrait entre autre d'établir automatiquement des recensements bibliométriques sans besoin d'utiliser un ensemble limité de *core journals* choisis arbitrairement et dépouillés manuellement. Si ce système se généralisait, ce serait le libre jeu des citations des chercheurs, recensés automatiquement à l'intérieur du réseau, qui déciderait de la réputation des chercheurs et non pas un petit nombre de *core journals*. Adieu marché monopoliste des soi-disant *core journals*. En plus, puisque dans Scholarsource les liens sont en réalité des graphes avec des arcs étiquetés comme dans le diagramme originaire de Tim Berners-Lee, ils peuvent exprimer toute une gamme de relations entre sources primaires et secondaires. Par exemple, on pourrait distinguer entre les mentions favorables des mentions négatives d'un article ; ou entre des analyses philologiques, rhétoriques ou philosophiques d'un passage textuel ; ou entre des considérations de type archéologique, historique ou stylistique d'un tableau ou d'un objet d'art. Aussi il serait possible de créer des logiciels qui utilisent ces relations, comme un programme de bibliométrie qui prenne en compte non seulement le nombre des référencements reçus par un article, mais également s'il s'agit de citations positive ou négative et qui, de cette façon, calcule un facteur d'impact pondéré. Et si toutes ces informations étaient encodées avec un langage standard comme RDF, tout le monde pourrait concevoir des logiciels pour l'utilisation scientifique de ces données, en les combinant selon des manières qu'aujourd'hui nous n'arrivons même pas à imaginer.

Cela peut paraître utopique. Mais si les chercheurs en sciences humaines veulent encore avoir un rôle en nos temps inquiets, ce sera peut-être celui de diffuser leurs savoirs choisis et de le faire avec le medium le plus puissant qui existe, le Web. Renoncer à utiliser Internet signifierait probablement condamner les sciences humaines non pas à la mort, mais à l'insignifiance, au silence. La recherche deviendrait alors une activité de plus en plus éloignée de la réalité de l'échange et du partage du savoir ; et inversement l'information diffusée sur le Web ne bénéficierait nullement de l'apport du savoir critique. Le Web est une grande opportunité pour nos disciplines et probablement la seule manière de préserver les valeurs et les méthodes que nous avons adoptées et perfectionnées en deux mille ans d'histoire intellectuelle. Car nous devons les préserver dans le mouvement et non pas les préserver du mouvement : nous avons certes besoin de notre île bienheureuse, et même d'un archipel d'îles bienheureuses, mais immergées dans la mer vaste et mouvante du Web.

Paolo D'Iorio  
*Maison Française d'Oxford* (CNRS-MEA)  
*Oxford Internet Institute*, Oxford  
www.diorio.info  
diorio@ens.fr