

Schriften des Instituts für Dokumentologie und Editorik — Band 3

# Kodikologie und Paläographie im digitalen Zeitalter 2

---

## Codicology and Palaeography in the Digital Age 2

herausgegeben von | edited by

Franz Fischer, Christiane Fritze, Georg Vogeler

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Malte Rehbein, Patrick Sahle

2010

BoD, Norderstedt

**Bibliografische Information der Deutschen Nationalbibliothek:**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

© 2011

Online-Fassung

Herstellung und Verlag der Druckfassung: Books on Demand GmbH, Norderstedt 2010

ISBN: 978-3-8423-5032-8

Einbandgestaltung: Johanna Puhl, basierend auf dem Entwurf von Katharina Weber

Satz: Stefanie Mayer und L<sup>A</sup>T<sub>E</sub>X

# La numérisation du patrimoine livresque médiéval : avancée décisive ou miroir aux alouettes ?

Ezio Ornato

## Résumé

La numérisation de la totalité du patrimoine livresque médiéval, qui pouvait paraître une utopie il y a quelques années, tend à devenir une perspective raisonnable dans un avenir relativement proche. Bien que de telles opérations doivent être nécessairement promues et gérées par les bibliothèques elles-mêmes, on peut se demander jusqu'à quel point les modalités de la numérisation seront compatibles avec les exigences des chercheurs : on peut craindre, en effet, un certain nombre de distorsions en ce qui concerne le choix des livres à numériser, la définition et la disponibilité effective des images et l'accès direct aux livres qui auraient été numérisés. Cela dit, la numérisation n'est pas une fin en soi. Elle devrait être considérée comme le moment clé d'une initiative plus vaste, ayant pour objectif la constitution d'une véritable *Bibliotheca universalis*. Or, cette bibliothèque virtuelle ne saurait fonctionner de manière satisfaisante en l'absence d'un arrière-plan descriptif adéquat, ce qui soulève beaucoup de problèmes. Cette constatation renvoie tout naturellement à l'interrelation entre la numérisation de l'objet et sa description catalographique qui, lorsqu'il s'agit de définir des priorités, pourrait donner lieu des choix difficiles. Cet antagonisme potentiel pourrait être dépassé grâce à la mise en œuvre sur le web d'un nouveau type de catalogue interactif qu'on pourrait appeler « grand-ouvert ». Il ne faut pas oublier, cependant, que le concept de *Bibliotheca universalis* présuppose que l'on puisse naviguer à loisir dans l'ensemble des images numérisées, et donc que les bases descriptives sous-jacentes soient interconnectées. Pour que cela soit possible, il faudrait que les données fondamentales soient harmonisées grâce à l'existence d'une métabase d'« autorités ». Nous en sommes loin. Il n'est pas sûr, enfin, que le chercheur puisse tirer pleinement parti de l'ensemble des données textuelles et iconographiques dans la perspective d'un véritable traitement de l'information : en effet, la structure et le contenu des bases ne sont pas assez rigoureux et les buts poursuivis se limitent le plus souvent à faciliter le repérage et la sélection d'objets possédant une ou plusieurs caractéristiques communes. A tout cela, il convient d'ajouter des empêchements qui découlent, malheureusement, de points de vue moins scientifiques qu'administratifs, voire lucratifs, chez les institutions.

## Zusammenfassung

Die Digitalisierung sämtlicher mittelalterlicher Buchbestände, vor einigen Jahren noch utopisch, nimmt mehr und mehr die Gestalt einer realistischen Zukunftsperspektive an. Ein solches Vorhaben müsste notwendigerweise von den Bibliotheken selbst getragen und vorangetrieben werden. Es stellt sich dann aber die Frage, wie weit sich die Modalitäten einer solchen Digitalisierung auf die Bedürfnisse der Forschung zuschneiden ließen. Dabei könnte es zu einer Reihe von Zielkonflikten kommen: bei der Auswahl der zu digitalisierenden Bücher, der Auflösung und der Verfügbarkeit der Bilder sowie bei der Frage, welchen Benutzungseinschränkungen die Originale, die bereits als Digitalisat vorliegen, unterworfen werden. Davon abgesehen ist Bild-Digitalisierung kein Selbstzweck. Sie sollte vielmehr den entscheidenden Schritt einer sehr viel weiterreichenden Initiative darstellen, die sich die Errichtung einer *Bibliotheca universalis* zum Ziel setzt. Eine solche virtuelle Bibliothek würde allerdings nicht zufriedenstellend funktionieren, solange das zugrunde liegende Konzept nicht in angemessener Weise modelliert ist. Und hier fangen die Probleme an: Gerade im Verhältnis von Digitalisierung und Katalogisierung wären schwierige Entscheidungen für eine Prioritätensetzung zu fällen. Eine mögliche Lösung wäre dadurch zu erreichen, dass man im Internet einen neuartigen, interaktiven Katalog ins Leben rief, der sich als "grand-ouvert", als "weit offen" bezeichnen ließe. Dabei darf nicht außer Acht gelassen werden, dass das Konzept der *Bibliotheca universalis* voraussetzt, dass der Benutzer unbeeinträchtigt zwischen allen Digitalisaten navigieren kann und die zugrunde liegenden Beschreibungen sinnvoll verknüpft sind. Um das zu erreichen, müssten die Daten über einen noch zu schaffenden Thesaurus standardisiert werden, wovon wir jedoch noch weit entfernt sind. Es bleibt ungewiss, ob Forscher vollen Nutzen aus den angesammelten Text- und Bilddaten im Sinne einer wirklichen Informationsverarbeitung werden ziehen können, denn Struktur und Inhalt der Datenbanken sind nicht klar genug, und Zielvorgaben beschränken sich meist auf Kennzeichnung und Auswahl von Objekten mit einem oder mehreren einfachen Merkmalen. Hinzu kommen für gewöhnlich noch Zugriffsbeschränkungen, die leider weniger aus wissenschaftlichen denn aus ökonomischen Interessen der bewahrenden Institutionen erlassen werden.

## Abstract

The complete digitisation of all collections of medieval books has long been regarded as an utopian goal. Now, however, this ambition is developing more and more into a realistic perspective for the future. Of course, such an enterprise needs to be initiated and driven by the libraries themselves. But it has to be clarified to what extent research and its needs and requirements have to be taken into account in modeling

this digitisation effort. A couple of conflicts of interest need to be considered: selecting the objects for digitisation, resolution and accessibility of the images, usage of the originals after digitisation etc. Furthermore, mere digitisation is not an end in itself but should become the core of a far-reaching initiative aimed at the establishment of a *bibliotheca universalis*. Such a virtual library would, however, be insufficient without a well designed underlying data model. And this is where the story gets really challenging: the relationship between digitisation on the one hand and cataloging on the other forces us to make severe decisions. A possible solution for this area of tension might be the establishment of a new kind of interactive catalogue in the internet: the “catalogue grand-ouvert”. Yet, one should not forget that the model of the *bibliotheca universalis* allows its user to freely navigate between all available digital objects and that this requires meaningful linkages among the descriptive data. To accomplish this, all data would need to be standardised by means of an ontology that has yet to be defined, and we are still far away from this.

## 1. Numériser l’intégralité du patrimoine manuscrit médiéval : est-ce faisable ?

Dans un article rédigé il y a quelques années, je préconisais la création d’une *Bibliotheca manuscripta universalis* dont le point fort serait la numérisation et la mise à disposition sur le Web de la totalité du patrimoine manuscrit conservé dans nos bibliothèques.

Bien sûr, je me suis vite aperçu que le terme *manuscripta* était à la fois trop large — car il s’agissait essentiellement de la production médiévale — et trop réducteur, car il excluait la production imprimée : l’expression *Bibliotheca universalis librorum Medii Aevi* aurait exprimé le concept de manière plus pertinente. Toutefois, malgré quelques remarques plutôt sarcastiques sur la faisabilité d’une telle entreprise, et bien que le mot « utopie » figurât prudemment dans le titre, je dois avouer que ce vœu me paraissait empreint de bon sens et somme toute relativement modeste dans ses ambitions. Cette étrange manifestation d’optimisme était, à mes yeux, justifiée par une série de considérations.

En premier lieu, les performances des systèmes informatiques, sur le plan aussi bien matériel que logiciel, ne cessent de progresser. Ces systèmes sont aujourd’hui capables de mémoriser et de traiter efficacement en un temps très court une immense quantité d’information : Google, Facebook, Youtube et bien d’autres initiatives, quelles que soient les réticences que l’on peut manifester à leur égard, nous font la démonstration quotidienne de ce que la technologie est déjà mille lieues en avance sur toute initiative que le médiéviste le plus hardi et le plus mégalomane serait capable de caresser ; et même, disons-le, sur toutes les initiatives auxquelles il n’a pas encore songé... et ne songera peut-être jamais. Aussi, ne s’agit-il plus d’appliquer ironiquement la célèbre devise des étudiants parisiens en révolte : « Soyons réalistes, demandons l’impossible »,

mais plutôt de mettre en œuvre une démarche mentale que dans notre pauvre (au sens financier du terme) univers des « humanités » nous sommes trop habitués à refouler : « Soyons irréalistes, imaginons le possible ». N'oublions pas que le corpus de livres dont Google Books — le vrai, pas la pâle caricature qui nous fait miroiter l'appât de quelques pages d'un livre au milieu d'un océan de trous et de liens vers les sites marchands — envisageait la numérisation est sans commune mesure avec le nombre, pourtant élevé, d'éditions incunables et de manuscrits du Moyen Âge occidental que conservent nos bibliothèques.

Ceux qui ont fréquenté autrefois assidument les anciens « palais » de l'informatique — où la « salle machines », comme dans les anciens paquebots, était interdite au personnel non autorisé — peuvent seuls se rendre compte de la transformation radicale, à tous points de vue, que le rapport « homme/ordinateur » a subie en l'espace de quarante ans. A l'époque, c'était le chercheur, surtout en sciences humaines, qui devait se plier à des exigences draconiennes auxquelles il était parfaitement étranger. Faire imprimer un titre en Fortran, quelle prouesse ! Lancer un calcul en Cobol, quelle performance ! Codifier strictement toute information pour gagner quelques kilooctets de mémoire, demander au moyen d'une carte perforée que l'on monte notre bande magnétique, attendre fiévreusement pendant des heures le retour peu glorieux des listings parsemés d'erreurs de programmation, quelle apothéose de la frustration ! Aujourd'hui, au contraire, c'est l'informatique qui vient au-devant de nos moindres caprices... Mais prenons garde : *Timeo Danaos, et dona ferentes*.

En deuxième lieu, la supériorité des technologies numériques par rapport aux dispositifs analogiques est telle, tant en termes de résultats qu'en termes de coûts, qu'elle pourrait être comparée à celle de l'imprimé par rapport au manuscrit, d'autant plus que le passage du manuscrit à l'imprimé sanctionnait le triomphe du noir et blanc, alors que le passage du microfilm à l'image numérisée coïncide précisément avec une réappropriation de la couleur. Or, qui peut imaginer ne serait-ce qu'un seul instant que le manuscrit aurait pu gagner la bataille ? Ou que le 78 tours aurait pu survivre au microsillon ? Et le microsillon au CD-ROM ? Et le CD-ROM au téléchargement ou au streaming ? Et si le téléphone fixe coexiste encore avec son homologue portable, c'est uniquement parce que les tarifs de ce dernier sont maintenus à un niveau artificiellement élevé grâce à des procédés plus ou moins occultes et, surtout, plus ou moins justifiés.

En troisième lieu, je presentais que l'idée de numériser les fonds manuscrits ne tarderait pas à faire tâche d'huile, une fois qu'un certain nombre d'obstacles seraient levés et que les modalités de l'opération évitent soigneusement de porter atteinte aux intérêts et aux privilèges de qui que ce soit. D'une part, parce qu'une initiative telle que Google Books, encensée par les usagers virtuels et dénoncée en chœur par les ayants droit et les tenants de l'identité culturelle — et pour ces raisons largement médiatisée — allait montrer qu'un projet de numérisation globale pouvait faire partie de la réalité concrète et fonctionner à la perfection ; de l'autre, parce que, dans une

société se proclamant soucieuse de la préservation de son environnement et de plus en plus conditionnée par le respect du principe de précaution, toute initiative ressemblant de près ou de loin à une sauvegarde trouverait assez vite un terrain propice à son développement.

Lorsque quelques grammes de CO<sub>2</sub> en moins deviennent un argument de vente — certes « bidon », mais ô combien révélateur — pour les industriels de l'automobile, on se dit que les temps sont mûrs pour que même la numérisation des livres anciens rentre de plein droit dans l'aéropage du « politiquement correct ». Ainsi, ce qui pouvait être au départ le rêve de quelques esprits visionnaires s'est concrétisé peu à peu en un petit nombre d'initiatives d'avant-garde, puis dans des projets de grande envergure, et bientôt les grandes bibliothèques qui n'auront pas mis en chantier leur programme de numérisation seront taxées de passésistes ; car les images des manuscrits sont capables de livrer aux yeux de tout un chacun le spectacle « magique » d'une partie de notre passé, et puisque l'existence d'un public potentiellement réceptif à un spectacle quelconque suscite nécessairement l'intérêt des pouvoirs de toute espèce, l'argent qui va avec ne se fera pas attendre.

Enfin, il fallait tout de même considérer l'éventualité que les bibliothèques — qui devraient être tout naturellement le centre moteur de la numérisation — rechignent à s'aventurer dans un chemin qui leur paraîtrait trop onéreux en termes de finances et de ressources humaines. Pour venir à bout de cette difficulté, il suffisait que :

- a) La numérisation soit financée par un budget spécifique sans préjudice pour la dotation courante de l'institution ;
- b) La bibliothèque conserve la gestion intégrale et exclusive des images ;
- c) L'opération soit entièrement externalisée grâce à des appels d'offre prévoyant le respect d'un cahier des charges précis et rigoureux ;
- d) La numérisation soit clairement découplée de toute initiative de catalogage.

Puisque ces conditions tendent désormais à être satisfaites, la numérisation déjà bien avancée du fonds « Plutei » de la Bibliothèque Laurentienne dans le cadre du projet « Teca digitale » et le lancement préliminaire d'un projet analogue concernant à terme les 80000 manuscrits de la Bibliothèque vaticane laissent présager que « les jeux sont faits » : dès lors que l'Église elle-même vient inopinément se placer dans un créneau d'avant-garde, il n'y a vraiment plus rien à craindre.<sup>1</sup>

---

<sup>1</sup> Et pourtant, ce n'est pas la première fois ! Dès 1991, en effet, la Bibliothèque vaticane avait réalisé un vidéo-disque iconographique (Baryla ; Baschet).

## 2. La numérisation, les bibliothèques et les aspirations des chercheurs

Cependant, le concept de *Bibliotheca universalis librorum Medii Aevi* ne se réduit pas à la numérisation des livres anciens d'une, de plusieurs ou même de chaque bibliothèque de conservation. Il englobe bien sûr l'accès virtuel à toute la production livresque conservée du Moyen Âge occidental (et d'autres aires géographiques si l'on veut), mais requiert aussi :

1. La création de bases de données pourvues de multiples fonctions, à savoir :
  - a) Fournir des informations aptes à guider préalablement la navigation à travers l'ensemble des livres numérisés.
  - b) Fournir les informations sur les caractéristiques codicologiques et textuelles des livres que l'on ne peut extraire à partir d'une reproduction.
  - c) Fournir, autant que possible, les informations qui, du fait qu'elles requièrent des compétences spécifiques, sont habituellement le fait des spécialistes des diverses disciplines : localisation, datation et provenance des livres, identification des copistes et des ateliers d'enluminure ; identification des auteurs et des textes.
  - d) Permettre la sélection de livres, ou de groupes de livres, qui possèdent en commun une ou plusieurs propriétés.
  - e) Pouvoir s'intégrer, le cas échéant, dans un réseau de bases de données relatives à l'histoire de la culture écrite.
  - f) Fournir des résultats structurés, permettant à tout utilisateur de construire par ses propres soins de nouvelles bases dans la perspective d'une recherche ciblée.
2. Le libre accès à toute la littérature scientifique concernant l'histoire de la culture écrite.

Le fait que les programmes en cours ou en projet mettent systématiquement l'accent sur la numérisation, et que les autres points soient tout aussi systématiquement ignorés, n'est pas anodin : cela signifie que le point de vue des institutions qui ont la charge de conserver le patrimoine est plus restreint que celui des chercheurs qui voudraient pouvoir y accéder à leur gré ; cela signifie, aussi, que les chercheurs concernés ne forment pas une communauté scientifique suffisamment active et consensuelle pour pouvoir formuler et justifier de manière claire et organisée un certain nombre d'exigences, et encore moins pour donner le jour à des initiatives concertées.

Pour les bibliothèques, en effet, l'objectif premier de la numérisation du patrimoine n'est pas celui d'en faire un outil performant et irremplaçable pour l'historien de la culture écrite. En fait, l'opération est porteuse à la base de trois connotations différentes : elle est à la fois un service rendu aux « usagers », une vitrine destinée au « public » et un dispositif de régulation dans le contexte de la conservation.



En tant que service rendu aux « usagers », elle remplace avantageusement le microfilm destiné autrefois aux lecteurs potentiels qui ne pouvaient se rendre sur place ou y séjourner pendant un laps de temps suffisamment long. L'avantage réside dans le fait qu'une image en couleurs immédiatement accessible, pouvant être multipliée à loisir et à moindre coût à partir d'un original, surclasse sur tous les plans l'image analogique en noir et blanc dont la reproduction demande du temps, du personnel et... de l'argent.

En tant que vitrine destinée à un « public », la numérisation s'inscrit à merveille dans la nouvelle philosophie qui a été imposée aux bibliothèques depuis une trentaine d'années et qui règne en souveraine à l'heure actuelle : l'institution doit justifier auprès du grand public l'argent que celui-ci dépense en tant que contribuable ; dans cette perspective, elle doit produire constamment des « événements », et donc exposer le plus possible ses « trésors ». Inversement, elle doit empêcher la dégradation du patrimoine dont elle a la charge ; patrimoine qu'il est donc commode de maintenir à l'abri dans d'obscurs magasins. Problème : entre la vitrine et les magasins, il y a tout de même une salle de « médiation » ; celle qui devrait mettre les témoins du passé au contact de ceux qui voudraient en prendre connaissance et qui ne devraient pas pour autant être considérés comme un « facteur de dégradation » supplémentaire. En substance, est-ce trop prétendre que le rôle d'une bibliothèque de conservation ne soit pas confondu avec celui d'un musée ?

Dans le domaine de la conservation, la numérisation joue un double rôle : direct d'une part, dans la mesure où, si l'objet matériel venait à disparaître suite à un événement catastrophique d'origine naturelle ou humaine, l'image numérisée en garderait malgré tout un souvenir beaucoup plus fidèle que ne pouvait le faire le pâle simulacre microfilmé ; indirect d'autre part, dans la mesure où la qualité de l'ersatz pourrait rendre plus acceptable le refus de l'accès à l'original.

En d'autres termes, la numérisation, vue par les bibliothèques — ou du moins par certaines d'entre elles — apparaît comme le point d'arrivée d'une opération stratégique, alors que, dans la perspective d'une *Bibliotheca universalis*, elle représente le point de départ indispensable d'une ouverture sans précédent. Dès lors, cette opération risque d'être envisagée comme un aspect de la *gestion* du patrimoine plutôt que comme l'une des étapes essentielles de sa *fructification* dans l'intérêt de la recherche historique.

Dans ces conditions, il y a lieu de craindre que les modalités de l'opération puissent être en quelque sorte « détournées » en fonction de finalités qui dans le meilleur des cas seraient étrangères aux aspirations des chercheurs et, dans le pire, incompatibles avec elles.

Un certain nombre de distorsions porte sur les principes inspirateurs du programme de numérisation et les choix qui en découlent. Le résultat est toujours une perte d'exhaustivité, ce qui est l'antithèse même du concept de *Bibliotheca universalis*.

La distorsion la plus répandue — et sans doute la plus difficile à éradiquer — pourrait être caractérisée, pour utiliser une expression imagée, par la célèbre répartie

« *Haec sunt ornamenta mea* ». <sup>2</sup> Elle consiste à restreindre la numérisation aux seuls manuscrits connus ; en général, ceux qui brillent par la richesse de l'apparat décoratif et illustratif. Cette solution représente le point de rencontre spontané entre le concept de « bibliothèque vitrine », l'engouement très compréhensible des bibliophiles pour les « beaux livres » et la tendance — beaucoup moins compréhensible — des historiens de l'art à ne regarder dans les livres que l'habillage iconographique.

Cette option, déjà peu enthousiasmante en tant que principe général, présente à son tour des variantes qui le sont encore moins : la numérisation des seules pages enluminées, la numérisation de quelques-unes parmi les pages enluminées ou pire, la numérisation des seules parties enluminées de quelques pages.

Il ne faut pas négliger non plus la distorsion de type identitaire qui consiste à restreindre la numérisation aux livres « autochtones » : éditions originaires d'un pays ou d'une ville, exemplaires uniques, manuscrits dont les textes sont écrits dans la langue nationale.

Citons également une solution bancale à tous points de vue : la numérisation des microformes en lieu et place des originaux ; opération qui devient encore plus incompréhensible lorsqu'on poursuit en même temps le microfilmage systématique des fonds au lieu d'en entreprendre la numérisation.

Il convient malgré tout de rappeler que ces distorsions ne relèvent pas toujours d'une position de principe : il s'agit souvent d'un compromis imposé par une dotation financière insuffisante. Dans ce cas, il faut nécessairement établir des priorités, dont le caractère provisoire risque malheureusement de s'éterniser.

D'autres distorsions, tout aussi redoutables, concernent non pas la numérisation elle-même, mais la gestion des images numérisées. Ainsi, ce dont on pourrait à première vue craindre la généralisation, c'est un accès aux images entièrement payant et, par conséquent, strictement réglementé.

Heureusement, il ne semble pas, désormais, que l'on s'achemine vers ce genre de solution. Jusqu'à il y a peu de temps, la politique ultralibérale menée par quelques pays dans le domaine de la culture avait contraint les bibliothèques à rentabiliser au maximum les services rendus. Mais aujourd'hui, personne n'accepterait de payer pour avoir accès sur le Web à des banques de données ou d'images dont la création et la mise en œuvre auraient été financées par de l'argent public. Aussi, n'ai-je trouvé sur le Web qu'un seul exemple de ce type de gestion : le projet « Irnerio », relatif aux manuscrits du Collège d'Espagne à Bologne, dont le mode d'utilisation, de plus, est particulièrement contraignant : un contrat d'abonnement, en effet, n'est valable que pour un seul numéro d'IP. <sup>3</sup>

<sup>2</sup> « Voici mes bijoux ! ». Phrase prononcée, selon la tradition, par Cornelia, la mère des Gracques.

<sup>3</sup> Dans des domaines voisins, nous pourrions citer des bases de données à vocation culturelle dont la consultation est malheureusement payante, comme celles qui sont regroupées dans des portails tels que

Une distorsion plus subtile, mais qui pourrait se révéler beaucoup plus dangereuse, pourrait bien nous rappeler une fable fort connue : « Le renard et la cigogne ». <sup>4</sup> Il s'agit d'un type de gestion où l'accès libre et gratuit serait réservé à des images de qualité délibérément inférieure, insuffisante pour une utilisation scientifique, l'accès au matériel exploitable étant soumis à conditions : accès payant, accès *in loco*. Les motivations réelles de cette restriction ne sont pas toujours explicitées mais, en tout cas, ne semblent pas relever exclusivement de la sphère financière.

Enfin, une dernière distorsion est étroitement liée à la problématique de la conservation : le libre accès aux images sur le Web pourrait impliquer, en contrepartie, que l'accès aux originaux soit réduit au minimum et, à la limite, totalement interdit, du moins pour certains manuscrits. Cette possibilité est explicitement mentionnée par les gestionnaires du projet de numérisation de la Bibliothèque Laurentienne : « L'elevata risoluzione delle immagini master consentirà di limitare, laddove necessario, il ricorso diretto agli originali (favorendone così la tutela e la conservazione) [c'est moi qui souligne] » (Degl'Innocenti 109).

Cet aspect de la question est absolument crucial, surtout si la politique de restriction est appliquée à large échelle. Peut-on considérer que des livres faits pour défier les siècles — et dont les pertes sont moins imputables à l'usure matérielle ou à un quelconque processus d'interaction avec le milieu ambiant qu'à leur obsolescence à la fois fonctionnelle, culturelle et idéologique qui les a envoyés au rebut — doivent être gérés comme la grotte de Lascaux ? Personnellement — mais je ne suis pas le seul — je ne vois aucune différence, sur le plan cognitif, entre un livre détruit et un livre enfermé à jamais dans un coffre-fort.

Cette contradiction entre les exigences de la conservation et les aspirations de la recherche n'est pas nouvelle : « Dans le domaine qui nous concerne, celui du manuscrit, on voit se développer dangereusement une tendance qui s'apparente de près à cette proposition [celle d'enfermer définitivement les œuvres d'art dans un coffre-fort qui, à l'époque, avait été assimilée à une simple boutade] : on voudrait que les chercheurs se contentent de consulter les microfilms et laissent tranquilles les précieux originaux » (Rizzo 14). Ces quelques lignes datent du printemps . . . 1984 (précision nécessaire : il s'agit d'une vraie date, et pas du titre du roman homonyme), lorsqu'une historienne de la culture humaniste s'était vue refuser l'accès à un manuscrit, d'abord parce qu'il fallait

---

*Brepolis* ou *Mirabile*. Leur gestion n'est toutefois pas assurée par des organismes publics, mais par des entreprises commerciales ou par des institutions privées.

<sup>4</sup> Le renard invite la cigogne à partager son repas, mais celui-ci est servi dans des assiettes plates sur lesquelles le long bec de l'oiseau n'a aucune prise. La cigogne invite le renard à son tour, et ce dernier se révèle incapable d'introduire son museau dans l'étroite ouverture du bocal qui contient le repas.

préservé l'intégrité du précieux objet, puis, une fois arrachée l'autorisation, parce que le volume était exposé dans une salle à côté!<sup>5</sup>

C'est par ailleurs en 1989 que les rédacteurs de la *Gazette du livre médiéval* proposaient, certes sur un mode plaisant, une « Déclaration des droits du manuscrit, du lecteur et du conservateur » dont l'article premier était ainsi rédigé : « Le manuscrit est fait pour le lecteur.<sup>6</sup> Le traitement qui lui est réservé ne peut avoir pour but que d'en assurer le libre accès à tous ». <sup>7</sup>

Il faut également souligner que la « pression » qu'exerce le lecteur sur l'intégrité du patrimoine manuscrit est largement surestimée. Sur ce point, nous ne possédons qu'une seule statistique, datée de 1985, portant sur la Bibliothèque nationale de Turin. Elle est pourtant éloquente : il en résultait, en effet, qu'un manuscrit n'avait été consulté, en moyenne, qu'une fois tous les quatre ans et que la durée de consultation moyenne n'était que d'un jour. 37% des volumes n'avaient jamais été consultés pendant une période de cinq ans, et 37% n'avaient été consultés qu'une fois. Donc, les ¾ du patrimoine avaient été extraits des magasins au plus une seule fois en cinq ans. Seuls 1% des manuscrits avaient été consultés plus d'une fois par an.<sup>8</sup>

Cela dit, nul ne conteste l'idée selon laquelle la numérisation des manuscrits pourrait réduire la nécessité d'accéder directement à l'objet original. En effet, il y a deux situations qui donnent lieu à des consultations, alors qu'elles pourraient être évitées. La première est celle où l'on connaît l'existence d'un manuscrit potentiellement « intéressant » — ou pertinent a priori dans le cadre d'une recherche quantitative — mais où l'on ne sait rien ou presque sur lui. Dans ce cas, il arrive qu'on consulte à l'aveuglette, simplement pour

<sup>5</sup> La bibliothèque en question était... la Bibliothèque Laurentienne. On doit néanmoins souligner que cet épisode ancien ne préjuge en rien du présent et de l'avenir : la politique concernant l'accès aux originaux présente, en effet, beaucoup de variations, dans la diachronie et dans la synchronie, en fonction du point de vue du personnel dirigeant, ce qui introduit une variable aléatoire extrêmement dommageable. Un autre passage de ce petit article pourrait avoir été écrit aujourd'hui : « On a souvent l'impression qu'aujourd'hui, dans l'utilisation des biens culturels, prévaut une politique qui tend à s'en servir comme d'un œillet à la boutonnière ou d'un moyen de publicité, plutôt que comme une source éternelle d'enrichissement de notre humanité et comme le témoignage irremplaçable d'un passé que nous devons étudier pour qu'il revive en nous dans toute sa richesse. Il n'est pas rare que les choix culturels d'une administration locale, d'une bibliothèque, d'un musée, et même d'un ministère n'aient pas, en dernière analyse, des motivations bien différentes de celles d'une grosse industrie ou d'une banque qui patronnent une restauration ou une publication » (Rizzo 16–17).

<sup>6</sup> Ce terme doit être entendu au sens large : dans un objet archéologique, on peut « lire » autre chose que le texte.

<sup>7</sup> L'article 4 mérite également d'être cité : « Toute restriction apportée à la consultation du manuscrit ne peut être fondée que sur la nécessité impérieuse d'en assurer l'intégrité *en vue des études à venir* [souligné par moi]. Sauf pour un temps limité, le manuscrit ne sera jamais privé totalement de sa finalité première, qui est d'être lu » (2).

<sup>8</sup> Voir Vitale Brovarone, qui remarque d'ailleurs, à juste titre, que la consultation d'un manuscrit peut permettre de mettre en évidence les symptômes de certains processus de dégradation qui passeraient inaperçus dans le cas contraire.

vérifier à quel type d'objet, de texte ou d'image l'on a affaire. La seconde est celle où l'on doit transcrire ou collationner entièrement le texte d'un ouvrage, ou encore analyser un programme iconographique dans son ensemble. Dans le premier cas, la consultation *de visu* pourrait être évitée ; dans le second, elle demeure de toute manière nécessaire, mais sa fréquence pourrait baisser radicalement du fait qu'une bonne image en couleurs fournit tout de même une quantité non négligeable de détails. Il s'agirait, quoi qu'il en soit, d'un effet automatique, et non d'une contrainte imposée en échange de l'accès aux images : personne n'est assez stupide pour échanger définitivement un partenaire en chair et en os avec une poupée gonflable, aussi ressemblante soit-elle.

Ces réserves étant formulées, il serait bon que les chercheurs établissent une sorte de « charte de la numérisation » définissant les critères minimaux pour que les résultats de l'opération correspondent réellement à leurs aspirations. Pour ma part, j'insisterais surtout sur les points suivants :

1. La numérisation doit se faire directement sur les originaux, et non à partir de microformes.
2. Pour chaque bibliothèque, l'opération doit concerner à terme la totalité des fonds, et la totalité des manuscrits de chaque fonds.
3. Chaque livre doit être intégralement numérisé.
4. Toutes les pages doivent être numérisées avec la même définition.<sup>9</sup>
5. La totalité des images doit être librement et gratuitement disponible sur le Web.
6. Les images doivent être disponibles sur le Web sous une forme exploitable. Cela signifie qu'elles doivent pouvoir être agrandies sans « pixelliser », afin de pouvoir mettre en évidence tous les détails de la page écrite (structure du parchemin, piqûres, réglure, ductus de l'écriture, etc.) et ce, quelles que soient les dimensions du volume original. En d'autres termes, puisque cela est techniquement possible, il faudrait, à certains égards, que l'on puisse examiner une page dans des conditions plus favorables que celles dont on jouit lors de la consultation directe en salle de lecture.
7. Les images doivent pouvoir être aisément comparées et retravaillées à l'intérieur d'un logiciel de retouche. Cela présuppose qu'elles puissent être téléchargées à loisir, soit isolément, soit, éventuellement, en bloc (par exemple comme fichier PDF), et qu'elles aient été créées dans un format universellement reconnu.
8. Les images doivent pouvoir être librement utilisées et reproduites dans le cadre de la normale activité scientifique, pourvu que les règles concernant la propriété intellectuelle (notamment l'obligation de citation) soient rigoureusement respectées.

Combien, parmi les entreprises achevées, en cours ou en projet, satisfont toutes les exigences exprimées ci-dessus ? Il n'y a pas lieu, dans le cadre de cette contribution,

---

<sup>9</sup> Ainsi, les manuscrits de la Bibliothèque du *Sacro convento* d'Assise ont été numérisés selon deux standards différents : « I parametri qualitativi indicati dalla Biblioteca Digitale Italiana, la risoluzione digitale adottata va dai 300 ai 600 dpi per i manoscritti non miniati (da 400 a 1000 dpi per i miniati) ».

de distribuer de bons ou de mauvais points, mais il serait tout de même intéressant de passer au crible de ces critères tout ce que l'on trouve à l'heure actuelle sur le Web.<sup>10</sup> Je ne crois pas me tromper en affirmant que très peu d'initiatives résisteraient à ce test.

### 3. Numérisation, catalographie et bases de données

Toutefois, à supposer même que tous les fonds soient numérisés avec des standards satisfaisants et que toutes les images soient diffusées auprès de la communauté scientifique dans les meilleures conditions, il n'en reste pas moins que les albums d'images ne sauraient être livrés sans le moindre support descriptif ; et c'est sans doute ici que le bât blesse.

Ce n'est pas pour rien, soyons-en sûr, que la numérisation des fonds a été d'emblée déconnectée de toute initiative dans le domaine de la catalographie. Compte tenu de l'état d'avancement du catalogage du patrimoine manuscrit dans certains pays et de son état prévisible dans l'avenir proche et lointain, c'était le seul moyen de faire démarrer rapidement la nouvelle démarche et de l'achever dans des délais raisonnables.

Face à l'absence d'un support descriptif conçu *ad hoc*, la solution généralement adoptée consiste à numériser les catalogues imprimés déjà parus et à instituer un lien hypertextuel entre les images et la notice correspondante. Or, ces catalogues sont souvent anciens, et parfois très anciens : dans le cas de la Bibliothèque Laurentienne, ils remontent au XVIII<sup>e</sup> siècle. De plus, bien des bibliothèques — et non des moindres — ne possèdent même pas de catalogue imprimé de tout ou partie de leurs fonds.

Or, il est évident que cette situation ne saurait perdurer et ce, pour deux raisons.

À partir du moment où le nombre de livres numérisés atteint un niveau conséquent, toute navigation efficace dans le corpus devient impossible si l'on ne dispose pas *au préalable* de quelques données de base sur chaque volume. À titre d'exemple, supposons que, dans la base numérisée d'une grande bibliothèque, on veuille sélectionner l'ensemble des bibles dites « de poche ». Puisque les manuscrits de la Bible sont nombreux et leurs typologie très variée, une simple liste très laconique de « Biblia sacra » obligera à cliquer sur chaque élément pour voir les images et établir de quoi il s'agit. Dans ce cas précis, il est évident que la présence des dimensions dans le résultat de la requête initiale serait la bienvenue, ce qui permettrait de ne pas tenir compte des manuscrits dont les dimensions sont supérieures à une certaine limite. Dans d'autres cas, on pourrait avoir besoin de la mention du support ou du nombre de feuillets, ou de la langue du texte. On voit bien qu'à terme on ne saurait faire l'économie d'une base de données, même très squelettique, où seraient enregistrées ces quelques données fondamentales.

---

<sup>10</sup> On pourra en trouver la liste sur le portail *Digitale Handschriften*. Pour une description raisonnée de chaque initiative, voir les commentaires de Marc Smith dans le portail *Menestrel*.

Par ailleurs, un simple lien entre les images et le texte d'un catalogue, même si ce dernier était récent et les notices extrêmement fouillées, ne constituerait que le prolongement de la situation imposée par la rigidité de la forme imprimée, avec l'avantage toutefois de pouvoir disposer d'images en couleurs. Or, cette rigidité empêche d'établir une liaison automatique entre deux ou plusieurs notices et entre deux ou plusieurs jeux d'images relatives à des objets différents.

Il en serait de même si, à partir des images numérisées, on construisait *ex nihilo* un catalogue calqué sur le modèle des catalogues imprimés. On conviendra aisément, en effet, que, dans ce domaine, l'apport du Web ne saurait se réduire à la production à bas coût d'un catalogue, certes illustré, mais irrémédiablement figé. En d'autres termes, pour donner une nouvelle vie à ce qu'on appelle couramment « catalogue », il faudrait que celui-ci soit structuré comme une base de données.

Avant d'aborder ce problème, il convient de se pencher sur l'interrelation étroite qui lie l'existence d'images numérisées et la démarche descriptive.

Lorsque j'avais envisagé l'existence d'une *Bibliotheca universalis*, l'une des objections à ce projet « chimérique » mettait en exergue une opposition de fait entre numérisation et catalogage. La véritable procédure cognitive étant représentée par le catalogage, il était à craindre que l'appât des images et le différentiel de rapidité entre les deux entreprises conduise inévitablement à détourner les financements vers la numérisation, sacrifiant ainsi le savoir au profit du « voir ». En d'autres termes, les images numérisées auraient fini par constituer une sorte de palliatif à l'absence de catalogue ; absence qu'elles auraient de fait contribué à pérenniser.

Ce point de vue n'est pas tout à fait injustifié, mais il a le défaut d'être statique. Il faut réfléchir, en effet, sur le fait qu'un catalogue est toujours porteur de deux fonctionnalités différentes : d'un côté, il est un *médiateur descriptif* entre l'objet et ceux qui, en étant éloignés, se trouvent dans l'impossibilité d'y accéder ; de l'autre, il est un *médiateur cognitif*, en ce sens qu'il fournit des informations complexes dont l'acquisition ne serait pas immédiate même en présence de l'objet, du fait qu'elle requiert des compétences spécifiques et souvent pointues.

La disponibilité des images sur le Web a comme effet de changer la donne, en ce sens qu'une bonne partie de la médiation descriptive n'est plus strictement nécessaire à partir du moment où les pages du manuscrit sont devant nos yeux. Dès lors, la tâche ingrate qu'est le catalogage s'en trouve considérablement allégée. Bien entendu, cette affirmation optimiste doit être nuancée : elle n'est valable que pour les utilisateurs « séquentiels » ; ceux qui, comme c'est le cas le plus fréquent, se contenteraient de prendre connaissance de chaque manuscrit pris singulièrement et de naviguer dans la *Bibliotheca* à partir de quelques paramètres essentiels (date, localisation, support, etc.). Cela dit, le conflit de priorité entre la numérisation et le catalogage peut être atténué en étalant le catalogage dans le temps et en le distribuant dans l'espace. Le dispositif

qui permettrait de répondre simultanément à ces critères est celui que j'appellerais volontiers « catalogue grand-ouvert ».

Nous connaissons déjà le concept de « catalogue ouvert » dont l'exemple autoproclamé le plus connu est le *Catalogo aperto* de la Bibliothèque Malatestienne de Cesena.<sup>11</sup> Ce catalogue ouvert consiste en un ensemble fait de notices descriptives, des pages entièrement numérisées de chaque manuscrit décrit, ainsi que de contributions scientifiques — anciennes, récentes ou nouvelles — relatives au fonds manuscrit de la bibliothèque, numérisées sous des formes variées. L'ensemble ainsi défini est dynamique, puisqu'il est sans cesse augmenté, et les notices descriptives ne sont pas figées dans le marbre.

Un catalogue « grand-ouvert » est, lui, quelque chose de plus : les notices sont structurées en base de données et comportent donc des champs de recherche plus ou moins détaillés ; mais, surtout, l'ensemble est géré de manière interactive. En effet, grâce à la présence des images, le rôle du catalogue en tant que médiateur descriptif comporte désormais deux aspects qui, sans être différenciés sur le plan épistémologique, peuvent subir, dans la pratique, deux traitements différents : il y a d'un côté les caractéristiques qui sont visibles sans ambiguïté sur l'image numérisée et, de l'autre, celles qui ne le sont pas. On aboutirait ainsi à une tripartition ergonomique qui serait transversale par rapport à la stratification conceptuelle qui, dans une description traditionnelle, distingue les aspects matériels, les aspects textuels et les aspects historiques.

Dans cette nouvelle perspective, toutes les informations, quelle que soit leur nature, que l'on peut acquérir par la simple observation visuelle n'ont plus besoin d'être explicitées (l'utilisateur les acquiert lui-même) ou peuvent être insérées directement dans le catalogue/base de données sur le Web sans nécessité de qualification scientifique particulière. Celles qui sont tout aussi évidentes à relever, mais qui requièrent un examen de l'objet en chair et en os, peuvent l'être *in loco* par les bibliothécaires, les chercheurs ou un personnel recruté temporairement à cet effet. Enfin, tout ce qui relève d'un travail d'expertise — datation, localisation, identification de copistes, d'artistes ou de possesseurs — ne peut être réellement pris en charge que par des chercheurs expérimentés dont la compétence porte sur des périodes, des aires géographiques et des typologies textuelles spécifiques et ce, quel que soit le milieu où ils travaillent (bibliothèques, universités, institutions de recherche). Bien entendu, l'existence d'un catalogue « grand-ouvert » en cours d'avancement ne doit pas empêcher d'utiliser, faute de mieux et provisoirement, les données provenant des catalogues existants.

---

<sup>11</sup> Cartelli *et al.* Un autre exemple de « catalogue ouvert », différent dans son fonctionnement de celui de la Bibliothèque Malatestienne, est constitué par la *Nuova Biblioteca Manoscritta*. Cette initiative est consacrée aux bibliothèques de la Vénétie et rassemble surtout du matériel d'époque moderne (Bernardi *et al.*). Il s'agit en fait d'un catalogue collectif où la saisie et la mise à jour interactive des données sont effectuées par les divers collaborateurs directement sur le Web.



On comprendra donc que des paramètres immédiatement « catégorisables » ou quantifiables, tels que le support, la disposition du texte, le nombre de lignes par page, la présence de réclames ou de titres courants, les rubriques, les incipits et ainsi de suite, peuvent ne pas être jugés prioritaires (mais dans ce cas ils ne pourront être la cible d'un formulaire de « recherche avancée ») du fait que, quel que soit leur intérêt dans le cadre d'enquêtes particulières, ils ne fournissent pas, comme c'est le cas pour la datation et la localisation, d'informations indispensables à la grande majorité des utilisateurs. On peut tout aussi bien concevoir, inversement, que les informations indispensables, mais qui malheureusement sont bien souvent fondées sur une évaluation synthétique subjective, fassent l'objet d'appréciations divergentes qui pourraient être rendues publiques dans un forum de discussion couplé à chaque manuscrit, afin que les utilisateurs de la base puissent en tirer profit et, le cas échéant, les mettre en balance.

#### **4. Les bases de données en ligne : peuvent mieux faire. . .**

Il n'y a pas lieu d'entrer dans les détails de la structure et du fonctionnement de ce type de catalogue dans le cadre de la présente contribution. Ce qui importe de souligner, c'est en revanche la contradiction dont tout catalogue de manuscrits disponible sur le Web est inévitablement porteur.

Malgré les difficultés inhérentes à leur création et à leur exploitation, les bases de données conçues à l'époque de l'informatique « lourde » possédaient un certain nombre de qualités. Puisqu'elles étaient destinées au calcul statistique, elles avaient le plus souvent une structure tabulaire dont les lignes — correspondant à des « individus » — constituaient les enregistrements et les colonnes les champs. L'information relative à chaque individu était rigoureusement catégorisée en « objets » ; chaque objet était pourvu de plusieurs propriétés distinctes, et chaque propriété présentait plusieurs modalités rigidement codifiées. Par ailleurs, chaque modalité de chaque propriété était rigoureusement codifiée de la même manière.

C'étaient-là des évidences pour tout concepteur de bases de données. Malheureusement, de ce point de vue, la situation a beaucoup empiré depuis. En effet, les bases de données disponibles sur le Web n'obéissent que très rarement à ces principes. Elles ne sont pas destinées au calcul statistique (et ne sont donc pas confrontées à ses verdicts impitoyables), mais à fournir une réponse adéquate à des interrogations — des requêtes — provenant d'utilisateurs tiers ; et une réponse n'est jamais le résultat organisé d'un calcul, à savoir une statistique descriptive : il s'agit toujours d'une sélection, c'est-à-dire d'une liste des individus de la base dont les caractéristiques correspondent aux paramètres de la requête. La sélection est l'œuvre d'un « moteur de recherche » qui utilise des opérateurs booléens d'une manière plus ou moins complexe, selon que la recherche peut ou non englober simultanément plusieurs champs et que les opérateurs

booléens élémentaires peuvent faire l'objet de combinaisons plus ou moins élaborées. Une fois la réponse obtenue, un simple clic permet de visualiser (copier, imprimer) l'ensemble de l'information qui concerne un ou, dans le meilleur des cas, plusieurs individus préalablement retenus dans la sélection.

Dans cette nouvelle perspective, ce que l'on peut uniquement prétendre d'une base, c'est la pertinence de la sélection obtenue par rapport à la requête envoyée. Cela revient à dire que la sélection doit contenir *tous* les individus concernés par la requête et *aucun* intrus.

Cet objectif est d'autant plus facile à atteindre que la structure de la base remplit au mieux les critères définis ci-dessus. Or, plus une structuration est performante du point de vue analytique, moins elle est satisfaisante du point de vue synthétique. Autrement dit, ce que l'on gagne en précision dans la requête est perdu en visibilité dans la réponse. Pour s'en rendre compte, il suffit de parcourir la description du système de codification TEI (*Text Encoding Initiative*) relatif à la description des manuscrits :<sup>12</sup> on voit nettement que, à partir d'une description sommaire provenant d'un catalogue imprimé, la structuration de l'information peut se faire à des niveaux de plus en plus détaillés dont le plus simple est celui qui permet de reproduire tel quel le texte d'origine et le plus détaillé correspond à un état de l'information extrêmement « décomposé ». Bien évidemment, si l'on souhaite maximiser les performances du moteur de recherche, le balisage du texte en champs doit non seulement être très fin, mais également s'accompagner d'une normalisation stricte du contenu de chaque champ, ce qui implique la perte du tissu discursif, dont la réception est aisée et immédiate, au profit d'une présentation sèchement énonciative.

Le travail de structuration, lorsqu'on le pousse à l'extrême, est assurément très onéreux, surtout lorsque la base n'est pas créée *ex nihilo* mais reconfigure l'information puisée dans un matériel préexistant. De plus, le résultat va à l'encontre des attentes de l'utilisateur moyen : celui-ci, en effet, aime retrouver sur l'écran ce qu'il a d'ordinaire devant les yeux lorsqu'il parcourt les pages d'un catalogue imprimé ; il s'accommode donc mal d'une présentation excessivement structurée, jugée trop « décharnée » et perçue, par conséquent, comme dépayssante.

Les solutions adoptées par les concepteurs de bases de données sur le Web représentent en général un compromis entre les deux exigences opposées. Toutefois, il arrive souvent que la structuration en champs se résume à la simple stratification typologique de l'information, en l'absence de toute formalisation — et *a fortiori* de toute normalisation — à l'intérieur de chaque champ. La notice ci-contre, tirée du catalogue des manuscrits enluminés de la British Library, disponible sur le Web, constitue un bon exemple de cette manière de procéder : son contenu ne diffère en rien de ce qu'on pourrait lire

---

<sup>12</sup> Ce système de codage n'est cité ici qu'à titre d'exemple ; il ne s'agit en aucun cas d'en évaluer les qualités intrinsèques.

sur une page d'un bon catalogue imprimé, et la mise en paragraphes est à peine plus accentuée.

Cependant, comme on peut le voir dans la capture d'écran, ce catalogue est également une base de données que l'on peut interroger grâce à un formulaire de « recherche avancée » portant sur un certain nombre de champs dont quelques-uns sont indexés. Lorsque les informations dans la base ne sont pas rigoureusement structurées, le moteur de recherche opère une exploration contextuelle du champ concerné par la requête (recherche « full text »). C'est une sorte de « pêche à la ligne » dont le résultat dépend à la fois du degré d'élaboration du moteur, du degré de promiscuité des « objets » et de leurs propriétés à l'intérieur du même champ, et, *last but not least*, de la variabilité des formulations employées dans la base pour définir une même modalité. En général, le mélange de ces trois sources potentielles d'inconvénients engendre un résultat décevant : la sélection obtenue pêche par défaut et/ou par excès, et de plus certaines caractéristiques dont la formulation dans la base n'est pas assez rigoureuse ne peuvent faire l'objet d'aucune requête ayant une chance quelconque de succès.

Le catalogue/base de données en question — qui au demeurant peut rendre d'excellents services — n'échappe pas à ce genre de travers, et malheureusement il n'est pas le seul.

Le tableau ci-dessous (tableau 2) montre que le résultat de certaines requêtes diverge selon que la requête est effectuée dans le formulaire de recherche avancée ou à partir des index correspondants. La recherche avancée semble ignorer toutes les subdivisions de l'Italie, sauf l'Italie centrale. « Italy, Central, Rome » et « Rome » fournissent quatre résultats différents. L'index comporte deux entrées différentes pour le nord-est de l'Italie. L'expression « Italy or France » donne lieu à une sélection de 1887 manuscrits, alors que « Italy » + « France » prises séparément donnent lieu à 1972 manuscrits (1895 par l'index). Enfin, dans la requête « Italy or France », « or » est systématiquement interprété comme un opérateur booléen ; il est donc impossible de sélectionner les volumes désignés dans la base par l'expression « Italy or France » lorsqu'elle signifie qu'il y a doute sur le lieu d'origine.<sup>13</sup>

Dans certains cas, l'obstacle pourrait sans doute être contourné grâce à un usage plus sophistiqué des opérateurs booléens. Mais bien souvent le moteur de recherche, trop fruste, interdit toute subtilité logique. Par ailleurs, l'utilisateur ordinaire est le plus souvent incapable de mettre en œuvre des manipulations savantes, d'autant que, s'il ne procède pas à des requêtes comparatives, il ne peut même pas s'apercevoir de l'existence d'un problème.

---

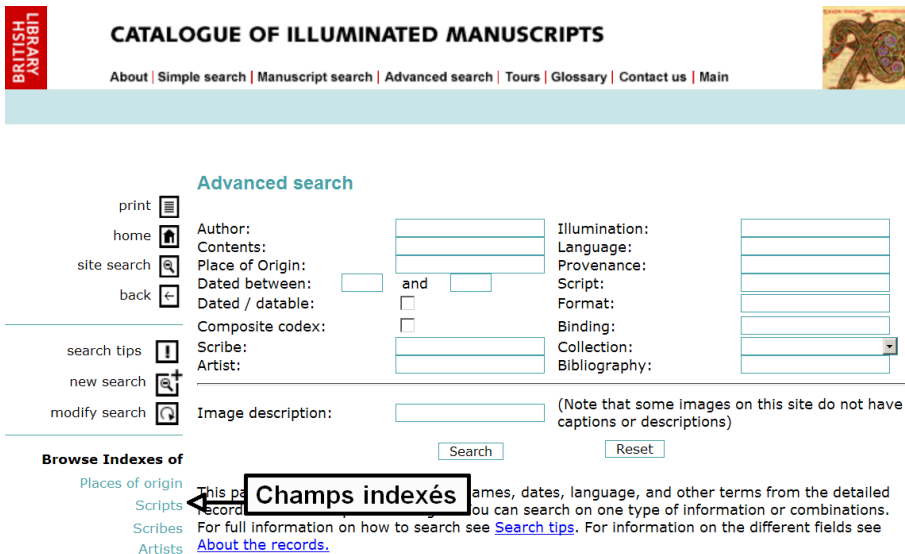
<sup>13</sup> L'effectif de la base étant en augmentation constante, il convient de préciser que la situation décrite dans le tableau correspond à des requêtes effectuées le 15 mai 2010.

---

Author	Pseudo-Augustine ; John of Fécamp
Title	Meditationes, Soliloquia (ff. [28–74v]), and Manuale (ff. [74v–97v]); Liber Supputationum (suspiria) (ff. 97v–[114]); and various prayers or meditations
Origin	Netherlands, N.
Date	2 <sup>nd</sup> half of the 15 <sup>th</sup> century
Language	Latin
Script	Gothic
Decoration	5 large initials in blue with penwork decoration and colour washes with leaves and acorns, extending to form a partial border at the beginning of texts and major divisions (ff. [5], [28], [75], [98], [114]). Large initials in red. Paraphs in red. Highlighting of letters in red. Rubrics in red. Spaces left for initials.
Dimensions in mm	115 × 75 (85 × 60)
Official foliation	Unfoliated (ff. [148])
Form	Parchment codex
Binding	Pre-1600. 15 <sup>th</sup> -century Netherlandish brown calf over wood boards, with blind tooling including of the letter 'T', a dragon and leaves, and remnants of 1 pair of clasps
Provenance	The Cistercian abbey of Mount St Bernard, near Leicester : its book-plate, 'Ex Libris Abbatiae De Monte Sti. Bernardi.' (inside upper cover). Mrs. Constance Goetze, wife of the popular painter Sigismund Goertze (d. 1939) and former owner of a number of manuscripts now at the Fitzwilliam Museum ; sold at Sotheby's, London, 2 December 1942, lot 326. Henry Davis (b. 1897, d. 1977), businessman and book collector : his book-plate (inside upper cover) ; and manuscript number-plate 'HD M27' (first folio). The Henry Davis Gift of book-bindings was donated to the British Museum in 1968.

---





TABLE 1. Manuscrits enluminés de la British Library. Description du ms Henry Davis Collection 597.






**CATALOGUE OF ILLUMINATED MANUSCRIPTS**

About | Simple search | Manuscript search | Advanced search | Tours | Glossary | Contact us | Main

### Advanced search

print  home  site search  back 

search tips  new search  modify search 

Author:  Illumination:   
 Contents:  Language:   
 Place of Origin:  Provenance:   
 Dated between:  and  Script:   
 Dated / datable:  Format:   
 Composite codex:  Binding:   
 Scribe:  Collection:   
 Artist:  Bibliography:

Image description:  (Note that some images on this site do not have captions or descriptions)

**Browse Indexes of**

Places of origin  This page contains names, dates, language, and other terms from the detailed records. You can search on one type of information or combinations. For full information on how to search see [Search tips](#). For information on the different fields see [About the records](#).

Scripts  **Champs indexés**

Scribes

Artists

FIGURE 1. Manuscrits enluminés de la British Library : écran « recherche avancée ».

Cela dit, il vaut mieux réserver à une meilleure occasion l'illustration approfondie des inconvénients engendrés par les champs « fourre-tout » et la mise en garde quant aux aléas imprévisibles qui accompagnent inévitablement l'exercice difficile de la recherche contextuelle ; aléas qui, d'ailleurs, ne relèvent nullement du... hasard mais sont, bien au contraire, l'expression d'une conception plutôt « laxiste » des bases de données dont les causes sont multiples, et les effets sous-estimés.

L'idéal, ce serait que l'on puisse concilier le besoin d'une présentation « traditionnelle », plus conforme à ce qu'attendent en général les chercheurs de nos disciplines, avec la possibilité de mettre en œuvre des requêtes plus performantes et dont les résultats seraient plus fiables. Cela ne pourrait se faire que si l'information enregistrée sous une forme familière à la majorité des utilisateurs, et donc destinée à l'écran, était couplée avec une base de données sous-jacente dotée d'une structure plus rigide mais plus robuste qui, elle, pourrait fournir des réponses à un moteur de recherche élaboré. Il ne semble pas, hélas, que l'on s'oriente dans cette direction ; et non seulement parce que cette tâche monopoliserait beaucoup d'énergie, mais aussi parce que la formalisation des phénomènes les plus complexes, tels que la décoration, présente des difficultés

requête	Recherche par l'index	Recherche avancée
Italy	863	911
Italy, N.	222	911
Italy N. E.	12	911
Italy, N. E.	131	911
Italy, N. W.	7	911
Italy, S.	30	911
Italy, Central	293	296
Italy, Central, Rome	3	69
Rome	80	84
France	1032	1061
Italy or France	nd	1887
Italy and France	nd	21

TABLE 2. Manuscrits enluminés de la British Library : quelques résultats de requêtes<sup>14</sup>.

conceptuelles difficiles à surmonter qui peuvent paraître réhibitoires dès lors qu'on n'entrevoit pas clairement la nécessité de les affronter.

## 5. L'interconnexion des catalogues et des bases de données

Cependant, même si la numérisation du patrimoine manuscrit se faisait selon les vœux des chercheurs ; même si elle était associée à des bases de données solides et fiables, issues de l'élaboration progressive d'un catalogue « grand-ouvert » collectif et interactif, et aptes, par conséquent, à faciliter la navigation dans les « magasins » virtuels d'une bibliothèque, nous serions encore loin d'avoir affaire à une véritable *Bibliotheca universalis*.

Ce que nous aurions, ce serait une liaison bidirectionnelle hautement perfectionnée, une sorte de télévision où la souris, telle une télécommande, nous permettrait de « zapper » très rapidement et d'une manière efficace d'un manuscrit à l'autre à l'intérieur d'une seule et même bibliothèque. Or, la télévision a été inventée il y a bien longtemps et le Web représente quelque chose de profondément différent du fait de sa structure en réseau, car il permet, en théorie, d'accéder simultanément à des informations puisées dans des sources multiples et de les rassembler sous une forme exploitable.

Pour illustrer la portée de ce propos, songeons à deux opérations qui seraient largement profitables à la démarche érudite : la recherche des *membra disiecta* –

<sup>14</sup> Sélection obtenue en cliquant sur le champ indexé. « Recherche avancée » = sélection obtenue en insérant la même chaîne de caractères dans le champ approprié du formulaire « recherche avancée ».

à savoir la reconstitution de manuscrits autrefois dépecés et dispersés aujourd'hui aux quatre vents — et la reconstitution des bibliothèques anciennes à partir des premiers et derniers mots du deuxième et de l'avant-dernier feuillet des manuscrits inventoriés. Il est évident que ce type de recherche serait accéléré d'une manière décisive s'il existait un moteur capable de puiser ces informations dans les bases de données de toutes les bibliothèques et de les fournir rassemblées à l'utilisateur.

Comment mettre en œuvre des dispositifs permettant d'obtenir ce résultat ? Bien sûr, il n'est guère besoin que l'information contenue dans toutes les bases soit strictement uniformisée. L'uniformisation ne devrait concerner que les données « élémentaires » et/ou directement quantifiables. Pour le reste, il suffirait de créer une ou plusieurs « métabases intégratives ». Une métabase intégrative n'est rien d'autre qu'une liste d'autorités pourvues de toutes leurs variantes attestées. Le choix de la forme faisant autorité n'a aucune importance : à la limite, on pourrait se contenter d'un simple numéro. Un tel dispositif est techniquement très facile à réaliser et pourrait être appliqué aux champs qui, dans une hypothétique *Bibliotheca universalis*, constitueraient à coup sûr les cibles de la très grande majorité des requêtes : l'auteur et le titre.

Sur le plan scientifique, la situation apparaît plus compliquée. Si de telles listes d'autorités sont proposées depuis longtemps aux bibliothécaires de plusieurs pays afin de faciliter le catalogage, il n'existe, à ma connaissance, qu'une seule base publique potentiellement intégrative : le *Thesaurus* du *Consortium of European Research Libraries* (CERL) qui, cependant, concerne essentiellement le livre imprimé.<sup>15</sup> Lorsqu'on introduit dans le formulaire de recherche le nom d'un auteur — par exemple « Thomas de Hibernia », l'auteur du *Manipulus florum* — la forme vedette s'affiche, suivie de toutes les variantes repérées :

*Headings* : Thomas <Palmeranus>

*Variant names* :

Hibernicus, Thomas

Hibernia, Thomas de

Hybernicus, Thomas

Palmer, Thomas

Palmeranus, Thomas

Thomas <Hibernicus>

Thomas <Hybernicus>

Thomas <Palmer>

Thomas <Palmerstonensis>

Thomas <Palmerstonus>

Thomas <aus Palmerston> [(VD-16)]

<sup>15</sup> Pour les arts plastiques, il faut signaler la base *Union List of Artist Names Online* (ULAN) du Getty Museum.

Thomas <d'Irlande>  
 Thomas <de Hibernia>  
 Thomas <de Hybernia>  
 Thomas <de Palmerstown>  
 Thomas <of Ireland>  
 Thomas <von Irland>

Pour pouvoir sélectionner tous les manuscrits des œuvres composées par l'auteur du *Manipulus florum*, il suffirait que chacune des variantes repérées dans les sources primaires (manuscrits) et secondaires (catalogues) pointe systématiquement vers « Thomas <Palmeranus> » dans la métabase.

Trop simple ? Sans doute, car, si la requête était envoyée à partir de « Thomas Hibernicus » plutôt qu'à partir de « Thomas de Hibernia », la sélection des auteurs-cibles ne se limiterait pas à « Thomas Palmeranus » ; en effet, dans le *Thesaurus* du CERL, ce nom renvoie également à un franciscain mort en 1270, ainsi qu'à un « Thomas <aus Palmerston> » qui, ayant vécu au XVI<sup>e</sup> siècle, n'a rien à voir avec l'auteur du *Manipulus florum*. Le même problème se poserait si la requête portait sur « Thomas <Palmer> », car ce nom est aussi celui d'un dominicain qui a vécu entre 1371 et 1413.<sup>16</sup>

Les cas d'homonymie étant assez nombreux dans la littérature médiévale, on comprend qu'il est impossible d'échapper à un travail préalable de désambiguation, faute de quoi la recherche se révélerait particulièrement hasardeuse : si une même variante peut renvoyer à deux ou plusieurs autorités différentes, il est nécessaire que les correspondances correctes soient établies au moment de la création de la base.<sup>17</sup>

Quelle est la situation actuelle sur le terrain ? Il existe une base informatisée — *Manuscripta Mediaevalia* — qui regroupe 63000 manuscrits issus de catalogues imprimés concernant tout particulièrement les régions germanophones. Une requête sur l'auteur du *Manipulus florum*<sup>18</sup> a abouti aux résultats suivants :

Aucune sélection n'englobe la totalité des manuscrits du *Manipulus florum* contenus dans la base — qui sont vraisemblablement au nombre de 37 —, ce qui signifie qu'aucune sélection n'est un sous-ensemble de la sélection la plus riche. Ce tableau sommaire n'a pour but que de mettre l'accent sur les difficultés méthodologiques et pratiques de toute entreprise d'intégration qui rassemble des sources préexistantes d'origine disparate ; le terme « disparate » devant s'entendre comme « élaboré par des personnes différentes »,

<sup>16</sup> La base du CERL renvoie également à d'autres homonymes qui toutefois, ayant vécu après le XVI<sup>e</sup> siècle, peuvent être négligés a priori. Sous la vedette « Thomas Palmeranus », elle met d'ailleurs en garde contre les confusions possibles avec ses homonymes connus : « Nicht identisch mit Thomas <Hibernicus> (OFM ; gest. 1275) und Thomas <Palmer> (OP ; um 1319/ 1413), mit denen er bisweilen gleichgesetzt wurde ».

<sup>17</sup> Ce travail est encore plus nécessaire — et surtout beaucoup plus lourd — lorsqu'il s'agit de noms de lieu, car les cas d'homonymie y sont extrêmement nombreux.

<sup>18</sup> À savoir : Thomas de Hibernia (ou autre) UND *Manipulus florum*, car Thomas d'Irlande a écrit autre chose que le *Manipulus*, et d'autres auteurs ont écrit des *Manipulus florum*.



type de recherche	requête	réponses
recherche standard	Thomas de Hibernia	31
	Thomas Hibernicus	26
	Thomas Palmeranus	30
	Thomas Palmer	24
	Thomas Hybernicus	24
	Thomas de Hybernia	24
recherche avancée	Thomas <Palmeranus>	21
	Palmeranus	23
	Thomas de Hibernia	5
	Hibernicus	25

TABLE 3. Manuscripta mediaevalia : résultats de la recherche sur Thomas d'Irlande.

même dans le cadre d'une entreprise organisée. Les responsables de la base *Manuscripta mediaevalia* se montrent d'ailleurs tout à fait conscients de ces difficultés.<sup>19</sup>

## 6. Les obstacles extrascientifiques

Le parcours conduisant à la réalisation d'un arrière-plan scientifique solide, apte à maximiser les performances d'une *Bibliotheca universalis*, est donc parsemé d'embûches, ce qui n'est pas étonnant et, à la limite, pourrait même paraître stimulant. Quoi qu'il en soit, on peut prévoir, sans risque de démenti, que ce parcours ne peut être que long : il s'agit, en effet, d'un travail critique patient et obscur qui doit être assumé par des personnes hautement qualifiées et ce, dans un domaine dont l'utilité sociale n'est évidente qu'aux yeux de quelques passionnés. Or, ce qui est à la fois cher, obscur et « inutile » a toutes chances d'être relégué au plus bas dans l'échelle des

<sup>19</sup> « Bitte beachten Sie bei Ihren Recherchen, dass die Qualität der Daten und die Systematik der Register in den einzelnen Katalogen sehr unterschiedlich sein kann. Diese Varianzen spiegeln sich trotz aller Bemühung um Normalisierung auch in der Datenbank. Die Zentralredaktion arbeitet kontinuierlich an der Klassifizierung der Registereinträge nach Personen- und Körperschaftsnamen, Orten, Datierungen, Werktiteln, Sachschlagworten etc., allerdings ist die Datenmenge schneller gewachsen als diese Arbeiten voranschreiten konnten. Daher sind (noch) nicht alle Daten über die Suche in den inhaltlich spezifizierten Indexfeldern zu finden, wohl aber – mit etwas Phantasie – in der Freitextsuche im "Gesamtregister" ». Notons par ailleurs que certains programmes de catalogage et/ou de numérisation ont créé en interne leur propre métabase de noms d'auteurs : c'est le cas pour *e-codices*, site consacré à la numérisation des manuscrits des bibliothèques suisses. Le fait que ce programme concerne jusqu'à présent un nombre limité de manuscrits et que le travail de normalisation peut donc être effectué au fur et à mesure de l'avancement de l'entreprise facilite bien évidemment les choses.

priorités, dans une dynamique négative qui présente depuis quelque temps des signes indiscutables d'accélération, du moins sur le plan institutionnel (suppression de chaires, marginalisation des « sciences auxiliaires » dans les plans d'études des universités).

Comme si cela ne suffisait pas, il faut également tenir compte d'obstacles d'une toute autre nature qui, je le crains, vont se révéler encore plus redoutables. Celui qui s'impose avec le plus d'évidence a trait à l'emprise presque « obsessionnelle » des copyrights éditoriaux sur la diffusion de la littérature scientifique.

Dans la mesure où la quasi totalité des catalogues de manuscrits se présente actuellement sous la forme imprimée, leur visualisation et leur utilisation sur le Web en tant que support normalisé pour les notices descriptives des manuscrits appartenant à des bases interconnectées suscitera inévitablement l'hostilité des éditeurs dont les droits — rappelons-le — couvrent une période de 70 ans. Si l'on songe au fait que bon nombre de ces publications à faible tirage n'ont pu voir le jour sans le recours à des subventions publiques, on peut s'étonner de ce que, comme l'impose la loi, les droits demeurent en vigueur même si les coûts de production ont été largement amortis et l'éditeur n'envisage aucune réimpression dans un avenir prévisible. Il s'agit, en fait, d'un véritable « droit de glaciation » qui de plus, dans la plupart des cas, ne profite à personne.

Bien entendu, les ayant droit s'opposeront également avec succès à toute initiative visant à diffuser gratuitement sur le Web la littérature scientifique. En réalité, il serait temps de comprendre que, dans le domaine de la recherche, en dehors de quelques opérations d'envergure où l'intervention de professionnels hautement qualifiés se révèle indispensable, la figure de l'éditeur commercial n'a plus de raison d'être. Autrefois indispensable, encore utile dans un passé récent, le système de production éditorial est désormais devenu nuisible, dans la mesure où il n'est plus un moteur, mais un obstacle au processus de diffusion des connaissances.<sup>20</sup> On ne peut rester indifférent au fait que, si une contribution scientifique était librement consultable sur le Web, le nombre de ses lecteurs serait multiplié par cent, voire par mille ; et la finalité d'une contribution scientifique n'est pas de rémunérer un investissement lourd, mais bien de donner la plus grande diffusion à la connaissance dans les plus brefs délais avec une dépense minimale d'argent et d'énergie.

Aujourd'hui, quelques clics suffiraient à n'importe-qui pour transformer un fichier Word mis en page de manière autarcique en un fichier PDF, en annoncer la parution en quelques instants à des centaines de collègues et d'institutions dans une « mailing

---

<sup>20</sup> Précisons que la disparition de l'édition commerciale n'a rien à voir avec la survie de la forme imprimée qui, elle, relève d'une problématique tout à fait différente : il ne faut pas oublier que, si la transformation d'un texte imprimé en un outil informatisé performant est une opération longue et pénible, l'opération inverse ne soulève aucune difficulté. Sans doute, s'inspirant de la pratique dominante d'il y a quelques siècles, nos descendants les plus nostalgiques se rendront-ils chez le relieur pour faire assembler les centaines de pages virtuelles téléchargées en quelques secondes et à un coût pratiquement nul.

list » parfaitement ciblée, enregistrer le fichier sur son propre site ou celui de l'institution d'appartenance pour qu'il puisse être librement téléchargé, encaisser éventuellement par l'intermédiaire de Paypal les contributions volontaires destinées à l'amortissement des frais, ouvrir un forum de discussions destiné à accueillir critiques et suggestions, introduire des liens vers d'autres sites qui pourraient héberger d'éventuels comptes rendus. Pourquoi ne le fait-on pas encore massivement ? Paradoxalement, parce que la forme imprimée... requiert un investissement lourd, et cette nécessité représente une ligne de partage entre la littérature « blanche » — celle qui a été jugée digne d'un tel investissement — et la littérature « grise » ; celle qui, autrefois, se distinguait par la forme artisanale des caractères non proportionnels et la présence de pages sans justification à droite.

Mais ces obstacles ne sont pas les seuls. D'autres, en effet, relèvent à la fois, d'une part de la nature intrinsèque du Web et de son mode de fonctionnement ; d'autre part, de la politique des institutions culturelles en ce qui concerne le patrimoine dont elles ont la charge et les rapports qu'elles entretiennent avec le monde de l'enseignement et de la recherche.

Qu'est-ce que le Web ? Très vaste question à laquelle on ne peut certes essayer de répondre en quelques lignes et qui, comme tout le monde sait, peut recevoir quantité de réponses différentes, voire contradictoires. Mais « contradictoire » ne signifie pas « incompatible » : en fait, toutes les réponses possibles comportent une part de vérité, selon qu'on privilégie tel ou tel angle d'observation. Ainsi, le Web est sans conteste un espace de liberté, puisqu'il ouvre la voie à une expression sans entrave et à la mise en place de réseaux de communication à la fois multiples et rapides qui permettent de diffuser toute sorte d'informations, opinions, créations et, en retour, de prendre immédiatement connaissance de nouvelles et d'agissements qui, autrefois, seraient demeurés confidentiels. Il n'empêche qu'il s'agit également d'un outil d'intoxication qui ouvre la voie à toute sorte de propagande douteuse et permet de déguiser en vérités les rumeurs les plus invraisemblables ; de plus, il constitue aussi un outil de surveillance qui permet à une autorité politique ou à une entité commerciale de recueillir toute sorte d'informations d'ordre privé qui sont immédiatement utilisées ou pourraient l'être par la suite avec des mobiles peu rassurants. Le Web est un outil de connaissance, puisqu'on peut obtenir en l'espace d'un éclair des milliers de renseignements sur n'importe quel sujet ; mais c'est aussi un piège d'ignorance car, comme la bibliothèque de Babel jadis imaginée par Jorge Luis Borges, il contient, dans un mélange indissociable, tout ce qui est vrai, tout ce qui est presque vrai, tout ce qui est approximativement vrai ou faux, tout ce qui est faux.

La définition partielle que je retiendrais ici revient à considérer le Web comme une sorte de « foire de Babel ». La différence avec la bibliothèque de Babel est fondamentale : la bibliothèque imaginaire est secrète, ou du moins silencieuse et discrète, alors que la foire est une exposition bruyante et permanente, une course perpétuelle à la visibilité :

au début, il était bon d'être sur le Web pour avoir une longueur d'avance sur les concurrents ; maintenant, il est indispensable d'y être, et d'y être avec un battage suffisant pour pouvoir émerger de la masse. Dans un élan d'imagination, j'assimile parfois le Web à ce que pourrait être la grand-rue de Trantor, la capitale de l'empire galactique, au temps de sa splendeur la plus éclatante.

Le Web est donc l'une des réalisations les plus abouties de ce que des esprits clairvoyants ont appelé la « société du spectacle ». C'est un spectacle sur écran, à l'instar de la télévision. Mais dans le Web, le spectacle se veut interactif. Cela donne une fausse impression de puissance, alors que le parcours de l'internaute est en fait soigneusement encadré : dans la plupart des cas, il n'est qu'un « client », au sens très large du terme, auquel on ne montre que ce que l'on veut bien montrer, c'est-à-dire une série de chalands de toute nature<sup>21</sup> ; et le client qui essaye de pénétrer dans le serveur, donc derrière la vitrine, est considéré à juste titre comme un pirate. C'est pourquoi il ne faut jamais oublier la double fonction de l'écran et la double connotation de ce terme : un écran, c'est certes quelque chose qui affiche, expose et invite au dialogue, mais en même temps c'est quelque chose qui sépare, filtre et cache.

Le Web est tout cela ; il est inutile de s'en plaindre, car il ne pouvait se développer que selon les lignes de force qui sont actuellement les siennes. Seulement, ce qui apparaît tout à fait compréhensible dans un mode de fonctionnement marchand, l'est beaucoup moins dans un contexte culturel ou scientifique. Or, dans presque toutes les institutions qui relèvent de ces domaines, l'utilisateur est traité malgré tout comme un client : il n'a pas accès aux données de base ;<sup>22</sup> il ne peut que prendre connaissance des résultats d'une requête spécifique présentés de manière parcellaire et dont l'usage lui-même est expressément soumis à condition. Le moins que l'on puisse dire, c'est qu'il s'agit d'une conception extensive des droits de propriété intellectuelle et, dans ces conditions, il y a lieu de se demander si, dans certaines institutions publiques, la notion de « mise en valeur » n'est pas plus importante que celle de « service rendu ».

Ce point mérite quelques réflexions, et tant mieux si leur contenu représente le point de vue d'un vieux chercheur qui, tout en reconnaissant sa partialité, ne se considère tout de même pas comme un voleur à la fois potentiel et « potentiel » (c'est-à-dire comme

<sup>21</sup> Ce terme est employé ici dans une acception qui n'implique pas nécessairement un but lucratif.

<sup>22</sup> Exemple (tiré du site suisse *e-codices*) : « Toute modification, reproduction, publication, octroi de licence, vente d'une image, de métadonnées (descriptions des manuscrits) ou de tout autre information du *e-codices* sont formellement interdits. Pour des raisons techniques le téléchargement automatique de l'ensemble ou d'une partie du site web n'est pas autorisé. De telles tentatives sont enregistrées, sous réserve d'éventuelles poursuites ultérieures [c'est moi qui souligne]. Le téléchargement manuel de pages isolées est en revanche autorisé » (<<http://www.e-codices.unifr.ch/fr/info/terms>>). Le « siphonage » de la base est donc interdit, ce qui est le droit le plus strict de ses gestionnaires. Mais pourquoi prévoir des poursuites éventuelles si le téléchargement n'est impossible que pour des raisons techniques ? Dans un registre ironique, on pourrait plutôt imaginer que les « pirates » devraient être récompensés pour avoir su venir à bout d'un problème technique malencontreux.

un « gibier de potence »). Cette prise de position ne concerne, bien entendu, que les données de nature scientifique créées et rassemblées par un organisme public dans le cadre d'un programme de documentation dont le produit s'adresse à la communauté des chercheurs. Il s'agit, d'autre part, de considérations d'ordre moral qui dessinent les contours de ce qui serait souhaitable, tout en prenant acte de la situation telle qu'elle existe.

Lorsqu'une base de données<sup>23</sup> à finalité documentaire et destinée à une utilisation publique a été créée par un organisme public dans le cadre d'un programme financé par des fonds publics (ou privés à titre inconditionnel), il n'y a que quatre sortes d'opérations qui doivent être rigoureusement interdites pour des raisons déontologiques :

1. Le plagiat, soit l'appropriation pure et simple de la base et de sa paternité. Ce serait de la piraterie.
2. La substitution, soit l'appropriation de la gestion de la base, même s'il n'y a pas usurpation de paternité ; peu importe si le but de l'appropriation est lucratif ou non. Les humains ne sont pas des coucous.
3. La dénaturation, soit l'altération de tout ou partie du contenu de la base sans que la responsabilité des concepteurs soit expressément dérogée. Des modifications malencontreuses ou erronées ne doivent pas faire perdre la face à autrui.
4. L'oblitération, soit l'utilisation des données sans en citer la source. La paternité d'un travail, que celui-ci ait fait ou non l'objet d'une rémunération spécifique, doit toujours être reconnue et identifiée. Il faut rendre à César ce qui est de César.

En dehors de ces opérations interdites, la disponibilité sous une forme brute tabulaire des données et des résultats des requêtes devrait être entière et globale. On ne voit pas, en effet, en quoi et comment cette disponibilité lèserait les intérêts des concepteurs et des gestionnaires de la base, même lorsqu'un aspect important, sinon primordial, de l'existence de cette dernière réside dans la mise en valeur d'une institution ou d'un site. Toute crainte à ce sujet serait absurde, car il faut bien se rendre compte de ce que les données brutes ne serviraient à rien pour la grande majorité des utilisateurs. Pour les usages courants, il serait de toute manière plus avantageux pour tout le monde, car plus rapide et plus ergonomique, de se connecter au site en ayant recours à son interface et à son moteur de recherche.

Mais — pourrait-on alors objecter — à quoi bon disposer des données brutes ? Simple, pour pouvoir les utiliser dans la constitution de nouvelles bases, aussi bien privées que publiques (dans ce dernier cas avec l'accord, bien entendu, des gestionnaires des bases sources).

Cette nécessité peut être illustrée par un exemple. Supposons qu'un historien de la culture écrite ait l'intention de procéder à une analyse bibliométrique de la production

---

<sup>23</sup> Ce terme désigne ici l'ensemble formé par un corpus de données structurées et les dispositifs permettant de l'interroger et de fournir des réponses à des utilisateurs.

d'incunables dans une aire géographique donnée. Ce dont il aurait besoin, c'est d'une liste complète des éditions imprimées dans ce territoire, comportant le nom des auteurs et les titres des œuvres, le lieu d'impression, la date, l'atelier typographique, le format et sans doute d'autres renseignements d'ordre matériel. Ces renseignements pourraient être puisés dans les deux répertoires actuellement disponibles sur le Web : l'*Incunabula Short Title Catalog* (ISTC) et le *Gesamtkatalog der Wiegendrucke* (GW). En revanche, l'historien devrait chercher ailleurs — ou plus vraisemblablement rassembler par ses propres moyens à partir d'informations très éparpillées — des données concernant les auteurs (tel que la date de mort, le statut social et professionnel) et les œuvres, afin d'élaborer une typologie permettant de suivre l'évolution des besoins et des goûts des lecteurs, ainsi que des politiques éditoriales mises en œuvre par les producteurs.

Il est évident que l'accès direct aux données brutes, lorsqu'elles existent, serait de nature à faciliter grandement la tâche de notre historien. Non seulement, mais les champs « typologie textuelle » de la base nouvellement créée pourraient enrichir en retour les répertoires existants : la possibilité de sélectionner d'un clic toutes les éditions de droit canon, ou de médecine, ou de grammaire, constituerait sans conteste une avancée non négligeable aux yeux de tout utilisateur.

La disponibilité des données brutes n'est pas un vain mot : sur les sites gouvernementaux des États-Unis — pays où l'économie marchande jouit pourtant des plus grandes faveurs — il est possible de télécharger sans problème des fichiers « texte » contenant les coordonnées géographiques de tous les lieux peuplés du monde ou l'historique des observations météorologiques pour toutes les villes américaines, et de les utiliser sans restriction.<sup>24</sup>

Quelles sont les perspectives d'avenir ?

Dans le domaine de la culture écrite, le cas de l'incunable peut être envisagé avec un certain optimisme. La nature même de l'objet interdit le repli sur soi : il ne nous reste, en effet, qu'environ 28000 éditions, et la description établie à partir d'un seul exemplaire peut s'appliquer automatiquement à tous les autres, sauf, bien entendu, pour ce qui relève du chemin parcouru par chaque exemplaire après parution.<sup>25</sup> C'est pour cette raison que le répertoriage total des éditions a pu être entrepris depuis longtemps, que le recensement exhaustif des exemplaires survivants a été achevé dans la presque totalité des pays et que les grandes bibliothèques de conservation ont pu établir le catalogue de leur fonds dans des délais raisonnables.<sup>26</sup> Dans ce contexte, le processus

<sup>24</sup> C'est le cas pour la *National Geospatial Intelligence Agency* (NGA) qui gère le *GEOnet Names Server* (GNS) : « Geographic names data is freely available. A suitable citation note is : "Toponymic information is based on Geographic Names Database, containing official standard names approved by the United States Board on Geographic Names and maintained by the National Geospatial-Intelligence Agency ».

<sup>25</sup> Il est vrai, aussi, qu'il n'existe pas d'édition dont tous les exemplaires sont parfaitement identiques, mais les différences sont le plus souvent négligeables.

<sup>26</sup> Celui de la *Bayerische Staatsbibliothek* de Munich a été mis en ligne. La numérisation des éditions est en cours.

de numérisation, déjà bien entamé en Allemagne<sup>27</sup>, peut s'appuyer sur des principes rationnels : le fait que les livres les plus remarquables sur le plan esthétique sont ceux qui ont été conservés en plus grande quantité interdit de privilégier les « trésors », et la numérisation prioritaire des éditions autochtones ou de celles qui contiennent des ouvrages écrits dans la langue nationale n'apparaît plus comme l'expression de velléités identitaires extrascientifiques, mais comme un partage du travail dicté par des critères somme toute cohérents, puisqu'il est à peu près certain que chaque pays assumera sa part du fardeau.

Pour ce qui est du manuscrit, en revanche, rien n'incite à être résolument optimiste. À vrai dire — mais c'est peut-être une impression personnelle — je constate que la plupart des institutions, surtout les plus grandes et prestigieuses, tendent à jouer en solo ; leur mode de fonctionnement me paraît singulièrement « monadique », au sens leibnizien du terme : elles ne veulent pas se dessaisir de leurs précieuses données, même à titre provisoire, même pour permettre la constitution d'ensembles harmonisés et intégrés pouvant être interrogés simultanément.<sup>28</sup> Dès lors, il faut s'attendre à ce que ces comportements monadiques produisent une série de chants monodiques, sans doute dissonants, et tout ce que l'on peut espérer, c'est que la recherche d'un minimum d'harmonie ne se fera pas au détriment des aspirations légitimes des chercheurs.

## Bibliographie

- Baryla, Christiane. « La banque d'images des manuscrits de la Bibliothèque Vaticane. I. La réalisation des vidéo-disques ». *Gazette du livre médiéval* 20 (1992) : 16–26.
- Baschet, Jérôme. « La banque d'images des manuscrits de la Bibliothèque Vaticane II. L'indexation iconographique ». *Gazette du livre médiéval* 20 (1992) : 26–29.
- Bernardi, Francesco, Paolo Eleuteri et Barbara Vanin. « La catalogazione in rete dei manoscritti delle biblioteche venete : Nuova Biblioteca Manoscritta ». *KPDZ* 1. 3–11.
- Brepolis. The hometown of Brepols' online publications*. Turnhout : Brepols Publishers, 2010. <<http://www.brepolis.net/>>.
- British Library Catalogue of Illuminated Manuscripts*. Londres : British Library, 2010. <<http://www.bl.uk/catalogues/illuminatedmanuscripts/welcome.htm>>.

<sup>27</sup> Cette entreprise (la «Verteilte digitale Inkunabelbibliothek» ou «VDLB») est gérée par la Herzog August Bibliothek (HAB) de Wolfenbüttel, ainsi que par la Universitäts- und Stadtbibliothek (USB) de Cologne.

<sup>28</sup> A côté de leurs indéniables mérites, il convient par ailleurs de souligner les incontestables limites des opérations visant à l'intégration sur le Web de bases de données ayant une thématique commune, ce qui est tout de même autre chose que la simple normalisation, grâce à une métabase conçue *ad hoc*, des noms d'auteurs et des titres d'œuvres. Bien souvent, la nécessité impose de se contenter d'un simple « coup de peinture » qui aboutit à la création d'un portail permettant une interrogation simultanée des bases intégrées. Il ne s'agit en aucun cas d'une véritable harmonisation, car on ne modifie pas le contenu, et encore moins les principes — souvent fort différents — qui ont présidé à la conception de chaque base. Dans ces conditions, l'interrogation simultanée se restreint aux seuls champs, en général peu nombreux, communs à toutes les bases, à condition que leur contenu ait préalablement été uniformisé.

- Cartelli, Antonio et al. « Il catalogo aperto dei manoscritti Malatestiani ». *KPDZ 1*. 13–23. *Catalogo aperto dei manoscritti Malatestiani*. Cesena : Biblioteca Malatestiana, 2002–2010. <<http://www.malatestiana.it/manoscritti>>.
- Collectif de rédaction. « Déclaration des droits du manuscrit, du lecteur et du conservateur ». *Gazette du livre médiéval* 14 (1989) : 1–4.
- Consortium of European Research Libraries (CERL) *Thesaurus*. Londres : CERL, 2007–2010. <[http://www.cerl.org/web/en/resources/cerl\\_thesaurus/main](http://www.cerl.org/web/en/resources/cerl_thesaurus/main)>.
- Degl'Innocenti, Emiliano. « Il Progetto di digitalizzazione dei Plutei della Biblioteca Medicea Laurenziana di Firenze ». *DigItalia* 1 (2007) : 103–113. <[http://digitalia.sbn.it/upload/documenti/digitalia20071\\_DEGLINNOCENTI.pdf](http://digitalia.sbn.it/upload/documenti/digitalia20071_DEGLINNOCENTI.pdf)>.
- Digitale Handschriften. Large Digital Libraries of Western Manuscripts*. Éd. par Klaus Graf. Lüneburg : Netbib.de, 2009–2010. <<http://wiki.netbib.de/coma/DigitaleHandschriften>>.
- Digitalizzazione dei manoscritti del fondo antico presso il Sacro convento di Assisi*. Assisi : Società Internazionale di Studi Francescani. <<http://www.sisf-assisi.it/digitalizzazione.htm>>.
- e-codices. Bibliothèque virtuelle des manuscrits en Suisse*. Fribourg : Université de Fribourg, 2008–2010. <<http://www.e-codices.unifr.ch/fr>>.
- GEOnet Names Server (GNS)*. Washington : National Geospatial Intelligence Agency (NGA). <<http://earth-info.nga.mil/gns/html/index.html>>.
- Gesamtkatalog der Wiegendrucke (GW)*. Berlin : Staatsbibliothek Preussischer Kulturbesitz. 2007–2010. <<http://www.gesamtkatalogderwiegendrucke.de/>>.
- Incunabula Short Title Catalog (ISTC)*. Londres : British Library, 1980–2010. <<http://www.bl.uk/catalogues/istc/index.html>>.
- Inkunabelkatalog der Bayerischen Staatsbibliothek (BSB-Ink online)*. Munich : Bayerische Staatsbibliothek, 2010. <<http://www.bsb-muenchen.de/Inkunabeln.181.0.html>>.
- Digitale Sammlungen* : <<http://www.bsb-muenchen.de/Digitale-Sammlungen.72.0.html>>.
- KPDZ 1 : Kodikologie und Paläographie im Digitalen Zeitalter – Codicology and Palaeography in the Digital Age*. Éd. par Malte Rehbein, Patrick Sahle et Torsten Schaßan. Schriften des Instituts für Dokumentologie und Editorik 2. Norderstedt : Books on Demand, 2009. En ligne : <urn:nbn:de:hbz:38-29393>, <<http://kups.ub.uni-koeln.de/volltexte/2009/2939/>>.
- Manuscripta Mediaevalia*. Éd. par Zentralredaktion mittelalterlicher Handschriftenkataloge. Berlin, Marburg, München [1996–2010] <<http://www.manuscripta-mediaevalia.de/>>.
- Mirabile. Archivio digitale della cultura latina medievale*. Firenze : Società internazionale per lo studio del Medioevo latino (SISMEL) – Fondazione Ezio Franceschini (FEF). <<http://www.mirabileweb.it/>>.
- Nuova Biblioteca Manoscritta (NBM)*. Venise : Regione del Veneto et Università Ca' Foscari di Venezia, 2006–2010. <<http://www.nuovabibliotecamanoscritta.it/>>.
- Ornato, Ezio. « 'Bibliotheca manuscripta universalis'. Digitalizzazione e catalografia : un viaggio nel regno di Utopia ? ». *Gazette du livre médiéval* 48 (2006) : 1–13.
- Progetto Irnerio*. Bologna : Università di Bologna, Centro interdisciplinare in storia e filosofia del diritto e informatica giuridica (CIRSFID). <<http://irnerio.cirsfid.unibo.it/>>
- Rizzo, Silvia. « Conservation et jouissance du patrimoine manuscrit : quelques réflexions ». *Gazette du livre médiéval* 4 (1984) : 14–17.



- Smith, Marc. « Fac-similés de manuscrits ». *Ménestrel*, 19.01.2010.  
<<http://www.menestrel.fr/spip.php?rubrique731>>.
- Teca digitale*. Florence : Biblioteca Medicea Laurenziana.  
<<http://teca.bmlonline.it/TecaRicerca/index.html>>.
- Text Encoding Initiative* (TEI). TEI Consortium, 2010. <<http://www.tei-c.org/>>. Manuscript Description : <<http://www.tei-c.org/release/doc/tei-p5-doc/html/MS.html>>.
- Union List of Artist Names Online* (ULAN). Los Angeles : J. Paul Getty Museum.  
<[http://www.getty.edu/research/conducting\\_research/vocabularies/ulan/](http://www.getty.edu/research/conducting_research/vocabularies/ulan/)>.
- Verteilte Digitale Inkunabelbibliothek* (VDLB). Wolfenbüttel : Herzog August Bibliothek – Cologne : Universitäts- und Stadtbibliothek Köln. <<http://inkunabeln.ub.uni-koeln.de/>>.
- Vitale Brovarone, Alessandro. « Lector cavat codicem ? ». *Gazette du livre médiéval* 6 (1985) : 13–16.