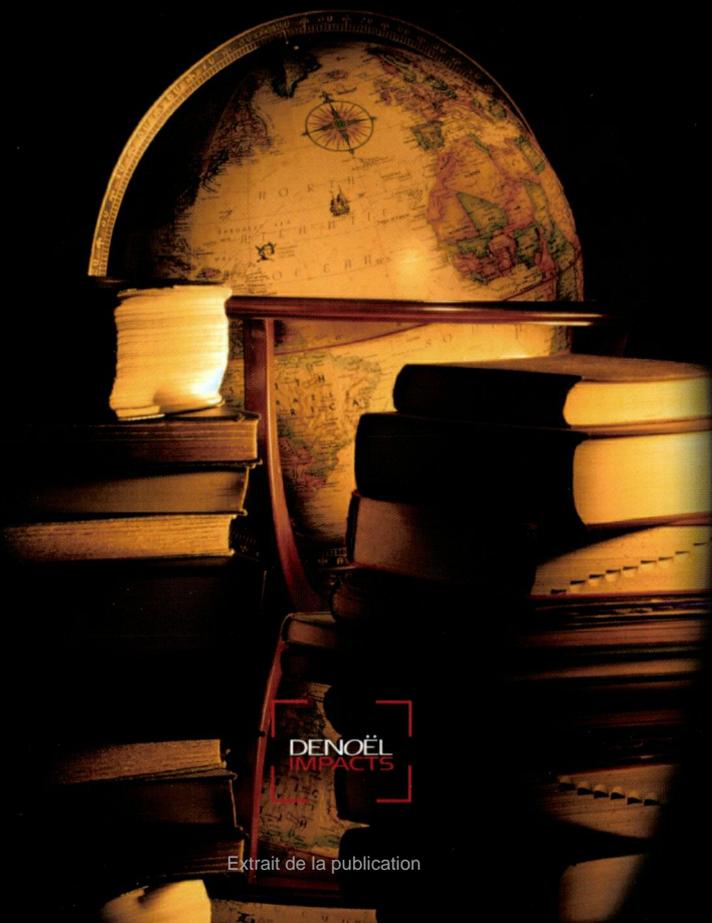


Lucien X. Polastron

LA GRANDE NUMÉRISATION

Y a-t-il une pensée après le papier?



DENOËL
IMPACTS

Extrait de la publication

La Grande Numérisation

DU MÊME AUTEUR

*Livres en feu, Histoire de la destruction sans fin
des bibliothèques, Denoël, 2004*

*Le Papier, 2000 ans d'histoire et de savoir-faire, Imprimerie
nationale, 1999 (Actes Sud)*

Lucien X. Polastron

La Grande Numérisation

Y a-t-il une pensée après le papier ?

**DENOËL
IMPACTS**

Cet ouvrage est imprimé sur papier permanent ISO9706 blanc naturel issu d'une pâte TCF, entièrement dépourvue de tout produit chloré.

© Éditions Denoël, 2006

Extrait de la publication

Que trouvent ici l'expression de la reconnaissance de l'auteur tous ceux qui ont apporté à cette enquête une partie de leur temps, de leurs lumières et de leur expérience, en particulier Alain Absire, Corinne Aribaud, Michèle Battisti, Luc Bellier, Gérard Cathaly-Prétou, Antoine Chéron, Pierre Cubaud, Boris Engelson, Michel Fingerhut, Alain Jacqueson, Catherine Lupovici, Pierre Mansiat, Guillaume Marsal, John L. Needham, Ève Netchine, Alexis Rivier, Jean-Marc Rod, Laurence Santantonios, Michel Valensi, Eric White, ainsi que bien d'autres bonnes personnes, qui se reconnaîtront.

Toutes les expressions, celles de tous les temps et de tous les espaces, se mélangent dans le même hypertexte, constamment réaménagé, et elles sont communiquées à n'importe quel moment, partout...

MANUEL CASTELLS ¹

Nous oscillons entre l'illusion de l'achevé et le vertige de l'insaisissable.

Au nom de l'achevé, nous voulons croire qu'un ordre unique existe qui nous permettrait d'accéder d'emblée au savoir ; au nom de l'insaisissable, nous voulons penser que l'ordre et le désordre sont deux mêmes mots désignant le hasard.

GEORGES PEREC ²

Parlons francs car le chiffre est plus net qu'en euros : l'entretien de la bibliothèque universitaire appelée Bibliothèque nationale de France coûte un milliard par an, soit le cinquième des subventions aux établissements publics de la culture, sept fois plus

qu'au début des années 90 pour une fréquentation à peine triple *. Or on constate que dans le même temps, l'institution acquiert de moins en moins d'ouvrages : par exemple la British Library en achète 140 000, contre 60 000 ici. Alors où passe l'argent ? Pour une bonne part en des charges liées au bâtiment de Tolbiac, qui finance plusieurs troupes de danseuses : brigade de sapeurs-pompiers, poste de police au complet, laveurs de carreaux géants, gardes forestiers et autres mouchards anti-fauche s'ajoutent à tous les corps de métiers indispensables à la survie et au développement d'une très grande bibliothèque nationale ordinaire. Quand on sait que 58 % d'étudiants de 2^e cycle occupent déjà les places prévues pour les professionnels au sous-sol mais désertées de bon cœur par les écrivains et 84 % celles qui avaient été mitterrandiennement conçues pour les « démunis » au rez-de-chaussée, on peut se demander si tant de capitaux ne seraient pas mieux répartis dans les bibliothèques universitaires du pays (150 millions d'euros représentant les trois cinquièmes de ce qu'elles reçoivent) ou alors, nous y voici, dans la numérisation en mode texte de davantage de livres utiles aux chercheurs planétaires, qui travaillent de plus en plus sur écran et dont la fréquentation de la basse-fosse du XIII^e arrondissement ne peut que stagner, tandis que « le caractère sophistiqué des équipements techniques et infor-

* Le rapport d'activité 2003 de la BnF, page 54, estimait à une moyenne d'environ 1 175 les entrées journalières qui avaient lieu naguère à Richelieu, contre 3 441 à Tolbiac, très précisément.

matiques nécessite un entretien attentif, dont le coût ne pourra aller que croissant en raison de leur obsolescence rapide et de leur durée de vie limitée », pointent les sénateurs dans un rapport publié en 2000³.

C'est donc à celui que les gens bien en cour voyaient alors comme un farfadet, en ce sens qu'il jetait en l'air des idées pouvant passer pour hasardeuses parce qu'elles ne retombaient jamais, Jacques Attali, que l'histoire donne, au moins sur un point, raison * : il fallait, en 1988, se mettre vaillamment à numériser la totalité des fonds de la Nationale. De toute évidence, si on avait consacré à cette gageure l'immense fortune engloutie dans la seule construction de la Tolbiac (1,2 milliard d'euros), l'effort appartiendrait aujourd'hui au passé. On peut parier que, restées à Richelieu, les salles de la BN seraient bientôt redevenues suffisantes pour les chercheurs encore en chair et en os, lieu poétique en harmonie avec les tendances de notre époque au fétichisme ou à la nostalgie. La communauté internationale du savoir aurait eu plaisir à se faire offrir de temps à autre un séjour instructif dans ce temple des adorateurs du papier.

Quinze ans plus tard, on découvre donc avec des rougissemments de chaisière que l'on a fait trop peu, trop mal, trop tard.

La polémique a été relancée accidentellement en janvier 2005 alors que l'oubli commençait à l'enliser :

* Ce sur quoi il semble parfaitement d'accord d'après son billet dans *L'Express* du 14 mars 2005.

pour monter dans le train du changement, Paris avait donc attendu qu'il arrive à destination. Pourtant il n'était pas passé loin : « En 1995 à la BNF, par exemple, nous avons tenté de comprendre, en partenariat avec l'École des mines et le moteur de recherche le plus prometteur de l'époque, LiveTopics d'Altavista, développé par le français François Bourdoncle, ce que signifiait "trouver de l'information" dans un fonds structuré de connaissances comme la collection Gallica. Il est dommage qu'il n'ait pas été donné suite à cette initiative. Le moteur Exalead du même F. Bourdoncle est, sans doute, toujours candidat pour des initiatives renouvelées dix ans plus tard... » rappelle Yannick Maignien, qui fut, jusqu'à 1997, responsable de la politique scientifique de numérisation à la Bibliothèque nationale de France.

Gallica, inauguré en janvier 2000, offre incontestablement une possibilité de lecture gratuite de documents anciens scannés : « Cet ensemble de romans, d'essais, de revues, de textes célèbres et d'œuvres plus rares donne les moyens à tout lecteur, du curieux au bibliophile, du lycéen à l'universitaire, d'approfondir la connaissance d'une époque dans ses aspects politiques, philosophiques, scientifiques ou littéraires. » (Document BnF en ligne.) Lancé dès 1992, le programme de numérisation devait s'enrichir de 100 000 titres tous les trois ans, la moitié seulement pouvant être consultés sans formalités de chez soi (le reste est soumis à copyright

et ne peut être lu qu'en salle). On en est loin. Sont seulement disponibles 52 764 monographies et 13 482 volumes de périodiques, le processus s'étant avéré plus lent ou plus cher que prévu : 9 millions d'euros dépensés de 1992 à 1998, puis 500 000 à 750 000 euros par an suivant les sources, des chiffres déjà bien faibles. « Les responsables de l'institution s'interrogent donc sur l'opportunité de faire payer, à terme, certains services aux utilisateurs », s'écrivait-il en 2002 *, mais peut-être pensait-on aux œuvres soumises au droit d'auteur. Rien n'est vraiment clair dans ce dossier : quand on lit par ailleurs que les textes en ligne représentent seulement 55 % du travail réalisé, un rapide calcul permet d'imaginer que 130 000 ouvrages ont aujourd'hui un double virtuel, information qui n'apparaît nulle part, notamment dans le rapport d'activité de l'établissement en 2003, où Gallica est à peine cité, ni dans la « charte documentaire » publiée à l'automne 2005 ⁴ pour apporter bien des précisions.

Or excepté les 1 200 ouvrages en « mode texte » fournis par le CNRS, qui en aurait déjà 3 737 sur son *Frantext* et dont l'usage, même en payant, est interdit aux citoyens douteux dépourvus d'un tampon institu-

* Pierre Le Hir, dans « La BnF numérise ses collections rares pour les mettre en valeur » (*Le Monde*, 7 mars 2002), cite le chiffre de 750 000 € ; on serait retombé à 500 000 par la suite. Mais soudain en 2005 l'institution se plaint de ne disposer que de « mille fois moins » que Google, ce qui ne ferait plus que 123 421 €. On sait les lettrés fâchés avec les chiffres ; si toutefois ce montant était confirmé, ce n'est pas une bibliothèque européenne qu'il faut organiser mais une vente de charité.

tionnel *, aucune de ces dispendieuses reproductions d'éditions admirables au demeurant n'est ni si utile au spécialiste ni si facile à déterrer : c'est une simple photographie de page qui se dévoile à un train de vieille tortue sous les yeux exclusifs de l'obstiné s'il connaît déjà le site de l'archive, et ce sans indexation ni citation possible sauf à copier à la main ou imprimer le livre au format A4 page après page pour le passer à la moulinette OCR, qui va trébucher sur chaque empatement un peu écrasé des Bodoni et compter soigneusement chaque chiure de mouche. Cet anachronisme bon marché s'appelle le « mode image ». Or, disent les réalistes froidement, « le texte ne peut exister dans la société de l'information qu'en mode texte **. Mieux vaut moins de texte, mais du texte, que davantage de texte glacé en mode image ⁵ ». Tout se passe comme si l'esprit de Gallica avait été défini par des militaires assiégés, alors que ce fut plus vraisemblablement par un consortium bibliophilique digne d'estime mais dont la vision perçante n'englobait que le passé éloigné. La France montre donc avec lenteur et majesté ses possessions

* « CNRS » serait-il une anagramme inachevée de « caserne » ? Il existe heureusement des entités moins rigides : Eulogos à Rome a réalisé depuis 1994 un remarquable travail selon les mêmes critères, à découvrir librement avec *Infratext* (toutes les adresses des collections numériques se trouvent en fin d'ouvrage).

** Un logiciel spécial, en cours de perfectionnement, serait susceptible de transformer, à la demande, la page d'un livre de la collection Gallica en page de texte. Lire à ce sujet Elisabeth Freyre, « Structuration et description automatique des documents électroniques : le projet européen METAe », *Culture et Recherche*, n° 100, 1^{er} trimestre 2004. Dans ce cas, pourquoi l'archive ne l'appliquerait-elle pas à l'ensemble du corpus sans barguigner ?

scripturaires, en commençant par des textes souvent rendus désuets par de nouvelles publications critiques, et elle en prolonge l'intangibilité même virtuelle. En un mot, le progrès fut de mettre sous verre ce qui avait été mis sous cloche.

Sanctuarisé donc, aussi pétulant que l'hôtel des Invalides, Gallica constitue un noble album souvenir que le père de famille cultivé peut électroniquement feuilleter avant d'aller peut-être commander tel ou tel texte étrange et sauvage chez son libraire éberlué. Des études précises⁶ sur ses utilisateurs — en moyenne cadres urbains mâles de quarante-huit ans — ont bien montré qu'ils se répartissent principalement en professionnels « chercheurs d'une information primaire » et en « bibliophiles anticipant un achat ». Le sondé qui avoue avoir lu du texte n'apparaît qu'une fois sur seize et encore s'agit-il moins d'une lecture que d'une « consultation/navigation ».

Bien que commencent à poindre des bandes d'impatiens bénévoles qui font le travail à sa place *, il va sans dire que la première décision saine à prendre pour la Nationale serait de mener à son

* Par exemple <http://ebooksgratuits.com/index.php> offre depuis peu un parfait accès à la *Très-plaisante nouvelle du démon qui prit femme* de Machiavel, dans les deux traductions, 1664 et 1825, avec notice sobre du cyber bienfaiteur qui a accepté de le traiter et sous les quatre formats : Mobipocket, PDF, Word et HTML, afin que nul n'en ignore le contenu quelle que soit sa machine à lire. On remarquera incidemment que la philosophie du groupe à l'origine de cette collection d'environ mille titres libres est que « le travail de numérisation, d'OCR et de correction effectué sur les textes ne donne aucun droit sur ces textes ». Une juste perception de la nouvelle technologie contribuerait-elle à restaurer le sens de l'honneur ?

terme la tâche entamée et doubler en vrais textes les textes de la collection, puis de les lancer dans le vide à se laisser débusquer par tout un chacun au bout du monde sans qu'il ait à connaître l'enseigne au préalable. Munis d'un tatouage électronique, ils ne feraient qu'en décupler la notoriété.

Là-dessus, tremblement de terre : Google fait son annonce, les morts sortent des tombeaux et notre archive de son auguste autisme, bien qu'il lui ait fallu six semaines pour affûter sa position. Six années auraient mieux valu, mais ne lui auraient peut-être pas donné plus de sagesse.

BnF contre BNF *

Déjà en 1994 les observateurs étrangers pouffaient de voir ce qu'ils avaient toujours connu et respecté en tant que « BN » devenir, dans le maelström de toutes les incompétences qu'a été sa inodernisation, la « BNF », c'est-à-dire une appellation repeinte d'une double couche de bleu blanc rouge. On imagine les gloussements si la Librairie du Congrès ajoutait soudainement à son nom « fédérale des États-Unis », et

* C'est par suite d'une coquetterie dans le logo que nous écrivons BnF. En réalité, la religion typographique voudrait que les trois lettres soient en capitales (avec des points s'il vous plaît, ajoute un correcteur). Mais comme il nous semble qu'il y a déjà assez de *national* dans cette affaire adoptons le sigle tel que voulu par la jeune institution, en espérant toutefois qu'une réforme ultérieure amènera le sobre, élégant dans sa pompe et définitif Bibliothèque de France. Il faut toujours être plus royaliste que le roi.

comment ne pas rappeler au passage que la deuxième bibliothèque du monde occidental se dit sobrement « britannique » ? La dérive parisienne, pourtant, ne devait rien au pilotage aveugle : le nom du programme de numérisation, Gallica, faisait déjà cocoricou rouillé. Aussi la réponse à Google — qui n'en attendait aucune — fut-elle des moins imprévues : la culture se trouva roulée dans le drapeau.

Le porte-parole d'un organe aussi emblématique que la BnF* ne peut accepter moins que ne le font tous les hauts fonctionnaires l'astreinte à justifier le résultat des erreurs antérieures, quelque profondes qu'elles soient. Grand serviteur de l'État depuis au moins Clemenceau, l'homme très estimable qui assumait en janvier 2005 la responsabilité de l'institution ne pouvait sans doute imaginer mieux que brandir un argument souverainiste aussi saugrenu qu'indéfendable : « Je garde en mémoire l'expérience du Bicentenaire de la Révolution, en 1989, quand j'en dirigeais les manifestations. Il eût été délétère et détestable pour l'équilibre de la nation, pour l'image et la connaissance qu'elle avait d'elle-même, de son passé, des événements, lumineux ou sombres, qu'il nous revenait de commémorer, d'aller chercher dans les seules bases de données anglaises ou américaines un récit et une interprétation qui y étaient biaisés de

* Il en est « président », mais le chercheur français a beaucoup de mal à expliquer à ses confrères chinois ou canadiens que ce pays est riche d'un très grand nombre de présidents en dehors de l'Élysée et que, non, ce n'est pas seulement là la source de bien des problèmes et surtout de notre incapacité à les régler.

multiples façons : *Le Mouron rouge* écrasant *Quatre-vingt-treize*, les vaillants aristocrates britanniques triomphant des jacobins sanguinaires, la guillotine occultant les droits de l'homme et les intuitions fulgurantes de la Convention⁷. » Il était alors facile de rétorquer à cela — mais quel journal, quelle radio, quel adversaire politique aurait eu les connaissances et le culot pour le faire ? — que d'une part c'était une vision fautive, étroite et paternaliste du passé, et que par ailleurs les universités américaines regorgent au contraire d'impeccables et incroyablement complets documents numérisés où l'étudiant peut puiser dans la réalité même de notre histoire, comme la *quasi-totalité* des actes de la Révolution française⁸. Quant à la « domination écrasante de l'Amérique » sur la langue, chacun d'un peu informé sait que moins de la moitié des fonds de la Library of Congress sont en anglais, de même que ceux des *Google Five*, comme les bibliothèques partenaires de ce projet se trouvent dorénavant surnommées, se composent de seulement 49 % de matériel dans cette langue contre 23 % en français, allemand et espagnol, le solde se répartissant en quatre cents autres idiomes.

Si Montesquieu est né à La Brède et Perrault à Paris, si grâce à Richelieu ou au voyou Napoléon un trésor de livres s'est amoncelé sur ce territoire, leur nationalisation n'en autorise aucunement un usage nationaliste, bien au contraire : quand les révolutionnaires en font le bien de tous, c'est à la race qu'ils pensent, et il n'échappe à aucun apprenti de l'éru-

tion que la véritable connaissance ignore les frontières (parfois même certains de nos meilleurs classiques en conchient les hymnes, puisqu'il faut encore le rappeler); si elle est synonyme de liberté c'est bien parce qu'elle garantit l'accès à une vision sans restrictions de l'existence. « Le mot de *Littérature nationale* ne signifie pas grand-chose aujourd'hui; nous allons vers une époque de *Littérature universelle*, et chacun doit s'employer à hâter l'avènement de cette époque. » Goethe, venant de lire un roman chinois, disait cela à table, le mercredi 31 janvier 1827⁹. N'avons-nous pas avancé d'un pas depuis cent quatre-vingts ans ou souhaitons-nous aujourd'hui reculer de deux siècles ?

S'ajoute à cette affaire assez pénible une déplaisante pincée de juridique. On enrégimente assez le public aujourd'hui sur les questions de droit pour qu'il n'ignore plus que le contenu des livres dont l'auteur est mort avant 1930 — en gros ce qui s'est fait de mieux dans l'histoire de l'humanité et pourrait largement suffire à son salut — n'appartient plus à personne. En conséquence, rien n'autorise les employés payés par le contribuable pour être les dépositaires de leur apparence physique — et *faciliter leur lecture* — à s'exprimer en leur nom. Une telle tentative de mainmise rhétorique sur les textes est donc infiniment moins tolérable que celle des rapaces capitalistes du web et, comparée à leur logique simple et franche, ressemble à un vol de ciboire par le bedeau. Mais voici qu'il existe un travers pire encore : « La Bibliothèque nationale de France est titulaire des droits

d'auteur sur le site Gallica. Pour un usage strictement privé, la reproduction du contenu de ce site est libre. Dans le cadre de communication, d'édition ou autres actions à caractère professionnel, ne sont autorisées que les courtes citations sous réserve de la mention BnF/Gallica. Toute autre reproduction ou représentation, intégrale ou substantielle du contenu de ce site, par quelque procédé que ce soit, doit faire l'objet d'une autorisation expresse de la BnF. » Comment des fonctionnaires, spécialistes du livre, en appelant si facilement à la notion de nation, qui devraient donc être à triple titre les parangons de l'éthique et de la générosité intellectuelle, osent-ils parler de « droits d'auteur » à propos d'une besogne technologique inhérente à la mission ordinaire qui leur est impartie ? On revient sur cette question terrible un peu plus loin, mais il serait critiquable que le service public se laisse entraîner à l'état d'esprit nouveau et peu ragoûtant qui consiste à créer un droit privé sur un bien commun par la simple application d'une couche de logiciel. Moins répugnante — c'est dire — est encore l'image du soldat de fortune plantant le drapeau en haillons de son pays lointain sur une île jusque-là paradisiaque.

Moteur de recherche, web *, « digitalisation »,

* D'aucuns préfèrent écrire « la Toile », comme la matière des matelas et des torchons, une appellation propre à exprimer l'appréhension légèrement dégoûtée que leur inspire cette nouvelle composante du quotidien. Pour mémoire, le *Journal officiel* a suggéré en 1999 « Toile d'araignée mondiale » ou Tam, espérant sans doute que cela évoquerait le tam-tam ou bien, peut-être, la revue TAM, dont se délectaient naguère les Bat'd'Af.

Entre Google et le nouveau droit d'auteur, voici le livre pris en otage, menacé de pillage et même de disparition.

La numérisation de l'écrit est une évidente bénédiction pour l'accès au savoir, mais elle déchaîne les convoitises commerciales et multiplie les barrières juridiques. Perçue comme un immense danger pour les patrimoines culturels, elle suscite de nombreuses réticences politiques, notamment en France, tandis que le bibliothécaire serre les dents et que l'intellectuel regarde ailleurs.

Entre faux prophètes et vrais profiteurs, la littérature est-elle réellement menacée? Risque-t-on l'appauvrissement des ressources culturelles mises à la disposition de chacun? Va-t-on vers la confiscation par quelques-uns de ce qui appartient à tous? Le livre en papier est-il mort, et le libraire, et la lecture publique? Va-t-on voir le non-lucratif réprimé par la loi?

Ou bien, au contraire, la dématérialisation vient-elle à point pour rénover l'écrit, lui offrir d'échapper aux filières qui peuvent étouffer la liberté d'enseigner, de dire, de rêver...

Avec lucidité, ces pages dissipent les malentendus accumulés autour de la gigantesque révolution que nous vivons.

Historien du papier et des arts de l'écriture, Lucien X. Polastron a publié *Livres en feu. Histoire de la destruction sans fin des bibliothèques* (Denoël, 2004).

DENOËL
www.denoel.fr

B 25669.8  03.06
ISBN 2.207.25669.3
17 €

