



**HAL**  
open science

# Comment comprendre le défi de Google et comment y répondre ?

Yannick Maignien

► **To cite this version:**

Yannick Maignien. Comment comprendre le défi de Google et comment y répondre?. 2005. sic\_00001430

**HAL Id: sic\_00001430**

**[https://archivesic.ccsd.cnrs.fr/sic\\_00001430](https://archivesic.ccsd.cnrs.fr/sic_00001430)**

Preprint submitted on 28 Apr 2005

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

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

## Comment comprendre le défi de Google et comment y répondre ?

Devant la pléthore d'effets d'annonces médiatiques, de surenchères économiques, d'appels politiques, ou de défenses culturelles ou stratégiques de nature diverse dont la presse se fait l'écho à propos des projets Google et des réactions françaises en la matière, il serait souhaitable d'éviter les conversations "café du commerce", sinon les initiatives dispersées, pour proposer un cadre conceptuel et opérationnel cohérent.

Certes chacun a le droit de s'exprimer, mais de fait beaucoup d'ignorance se révèle au coin des meilleures intentions du monde et de vérités a-historiques assénées avec beaucoup d'aplomb. L'ampleur des réactions représente déjà un corpus significatif digne d'étude !

La principale lacune réflexive semble ressortir d'une forte opacité, sinon incapacité à penser les raisons qui ont fait que la France, étant en pointe en 1990 dans les programmes de numérisation (notamment Gallica) est devenue suiviste sinon en retrait 10 à 15 ans plus tard.

### Besoin d'inventaire

Avant donc de lancer des "Faut qu'on Y a qu'à" et des affirmations stratégiques péremptoires (sans doutes à l'aune des grands desseins européens présidentiels et ministériels), il semble qu'on ne pourra pas faire l'économie (en guise d'inventaire, mais pas seulement) des "causes" techniques, scientifiques, juridiques, éditoriales, corporatistes, culturelles économiques, industrielles et politiques (dans l'ordre d'importance) qui expliquent largement le temps et l'initiative perdus en la matière.

A n'en pas douter, l'interrogation pourrait porter (toutes choses égales par ailleurs) sur l'échec relatif qu'il y a eu, depuis les années 90 à mener à bien en France une sorte de groupement d'intérêt économique où les missions utilisatrices de conservation et d'éducation des bibliothèques pouvaient se conjuguer avec les impératifs de recherche d'une part, les indispensables dimensions éditoriales (juridico-économiques) et les intérêts industriels des opérateurs informatiques et telecom d'autre part.

Il ne faut pas se le cacher, le projet de Google est d'abord de cet ordre, avant d'être une prouesse logicielle, une performance technologique ou un montage financier ambitieux.

Faute de s'interroger sur cette carence de synergie et de négociation propres à assurer des consensus productifs là où divergent les intérêts à court terme, il semble que l'appel à la puissance publique nationale, ou même européenne, comme seul appel de fonds, ne sera pas à la hauteur du défi lancé par Google. Les mêmes causes (en 1995/97) donneront les mêmes effets (en 2005/07).

### Synergies vertueuses

Certes la France (et l'Europe) n'est pas l'Amérique, mais quand celle-ci, consciente dans les années 94 /98 d'être partiellement distancée, elle a alors, via la NSF et d'autres fonds scientifiques d'Etat, lancé un vaste National Digital Project pour répondre alors au défi ...européen. Ce furent des dizaines de projets spécifiques, dans de très nombreux domaines applicatifs précis, autour des grandes bibliothèques universitaires américaines, coordonnés par celle du Congrès, expérimentant dans tous les domaines cités plus haut, avec à chaque fois des triangles avec la recherche informatique, linguistique, logique etc. du plus haut niveau et des réalisations industrielles financées par les grandes entreprises des secteurs informatiques et de la communication.

C'est dans ce contexte que Larry Page et Sergei Brin ont situé leur "garage".

D'une part prenons la mesure de l'ambition de Google. Ramenées par exemple au montant d'investissement de Gallica en 1992, les sommes annoncées par Google de 150 à 200 millions de dollars représentent en fait un effort similaire, dès qu'on module ces 13 années en fonction de la loi de Moore divisant par deux les coûts tous les 18 mois...

Mais surtout, une grande partie (essentielle) de la capacité de réponse - certes sous l'impulsion des pouvoirs publics- résidera dans la capacité à assurer les synergies "vertueuses" d'acteurs aussi différents impliqués en amont et en aval d'un vaste processus de numérisation.

Est-on sûr, par exemple en ce qui concerne la recherche, qu'elle ne pêche pas aussi (à côté d'une grande qualité) par un superbe isolement, tant en direction des milieux utilisateurs que de ceux impliqués dans des logiques économiques éditoriales, informatiques ou communicationnelles ?

Le CNRS a-t-il fait les choix stratégiques, certes pour l'instant dans le strict domaine des périodiques scientifiques, à même de fédérer des intérêts à terme antagonistes, et donner la pleine mesure de ce vecteur de travail en réseau qu'est Internet?

### Moteurs de recherche et moteur de connaissance

Les grandes bibliothèques n'avaient-elles pas abdiqué un peu vite (devant une entreprise de recherche d'information comme Google), en tous cas pendant la période citée, leurs missions régaliennes, depuis Alexandrie qui consistent

d'abord à organiser des connaissances, à structurer de l'encyclopédisme, face aux nouveaux besoins d'intelligence du monde, et dont la recherche d'occurrence ou d'information n'est qu'un sous ensemble réducteur ? A cet égard, le Ministre de la Culture a hautement raison de citer l'"Archéologie du savoir" et Foucault comme devant présider à de telles ambitions heuristiques! Comme le disait Bachelard " Quand on ne sait pas ce qu'on cherche, on ne comprend pas ce qu'on trouve" et les objectifs cognitifs de moteurs de recherche tels Google sont limités, comme l'ont déjà souligné plusieurs scientifiques.

En 1995 à la BNF, par exemple, nous avons tenté de comprendre, en partenariat avec l'Ecole 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 10 ans plus tard...

### **Modèle économique**

Les éditeurs n'ont-ils pas aussi leur responsabilité quand certains bloquent à court terme toute réflexion sur les montages juridico-économiques possibles pour préparer (comme dans le secteur musical) les répartitions nouvelles des gains de productivité liés au numérique ? Mais en amont d'une telle « négociation », il faudra bien un jour revenir au « contenu » (de Gallica) dont on parle si peu, au sens où cette offre culturelle, en fonction des critères qui ont conduit cette sélection, demande encore à identifier et rencontrer ses publics, ses usages et ses lecteurs ! En tous cas, on ne pourra faire l'économie de la réflexion sur l'extrême diversité des « sources » et des formes éditoriales à partir desquelles numériser (comme pour Gallica, où beaucoup de contenus ne proviennent pas de la BNF et n'ont pas d'équivalent physiques au sein du site de Tolbiac...).

Les opérateurs telecom européens ont-ils réellement pris la mesure des services de valeur ajoutée auxquels ils pourraient être associés, injectant en retour les indispensables investissements informatiques et technologiques ? Les trente millions de pages de Gallica sont « déchargées » gratuitement et quotidiennement dans le monde entier (y compris déjà on l'imagine par Google ...). C'est sans doute une grande victoire de la francophonie, mais avec bien peu de retour sur investissement ...

Il ne s'agit pas ici de comptabiliser des griefs, il s'agit de réfléchir ensemble, avec les acteurs si différents de cette nouvelle chaîne éditoriale, afin de ne pas réitérer les mêmes erreurs ou lacunes, et construire une vraie réponse au défi lancé par Google.

Il nous semble que c'est autant dans cette capacité organisationnelle originale que dans les montants (importants!) des moyens financiers requis qu'un tel programme se révélera efficace. C'est peut-être aussi le meilleur moyen de conjurer certains fantasmes de risques culturels (par ailleurs bien réels).

Les pouvoirs publics ont pris la mesure, au plus haut niveau de l'Etat de cet enjeu. A tous les acteurs évoqués ici d'avoir l'abnégation et l'imagination suffisante et nécessaire, à l'échelle européenne, pour donner corps à cette volonté.