

Rapport publié le 10 septembre 2007

23 octobre 2006

**POUR UNE INITIATIVE EUROPÉENNE DANS L'INTERNET EN FAVEUR DU
DÉVELOPPEMENT DE L'ÉCONOMIE DE LA CONNAISSANCE**

Plus qu'au contrôle des outils de traitement et de gestion de l'information (Microsoft, Intel, Cisco), c'est à la maîtrise de l'accès à l'information et de sa diffusion que la France et l'Europe sont confrontées aujourd'hui. L'enjeu ne se limite plus aux investissements dans les réseaux et les supports (ADSL, fibre optique, etc.), mais s'étend aux contenus numériques et impose d'être acteurs dans la formation de ces derniers.

L'ensemble des moteurs de recherche et d'archivage de ce qui transite sur le Web peut-il être sous la responsabilité d'un seul État, fût-il ami ? Est-il économiquement et éthiquement acceptable de dépendre, pour une fonction aussi importante que l'accès à l'information, quasiment d'une seule société ? Comment encadrer par les règles habituelles du marché des prestations mises gratuitement à la disposition de centaines de millions d'utilisateurs ?

L'émergence d'un nouveau modèle économique d'opérateurs, prêts à offrir aux internautes tous les services d'accès gratuitement à condition que ceux-ci mettent à leur disposition leurs informations personnelles, place les décideurs nationaux et européens face à des défis juridiques et politiques majeurs.

1. UNE DÉPENDANCE GRANDISSANTE VIS À VIS D'UN PETIT NOMBRE D'OPÉRATEURS

Google est l'archétype de ce nouveau modèle d'opérateur. Non seulement Google représente le premier moteur de recherche et de loin le plus performant, mais il constitue une entreprise hors normes, d'une puissance économique considérable. Google pèse aujourd'hui environ 100 milliards d'euros, et son modèle d'affaires, fondé sur une offre a priori gratuite pour l'utilisateur, constitue une approche qui ne peut que déstabiliser l'ensemble des acteurs en place.

1.1. LES SERVICES DE GOOGLE

Son image de spécialiste de la recherche sur Internet est trompeuse : Google ne se contente pas, loin de là, d'un positionnement apparemment inoffensif sur ce qui pourrait apparaître comme une niche de services d'Internet. En louant les espaces publicitaires¹ que son moteur de recherche génère, Google dispose d'une rente de situation qui lui donne les moyens d'être présent sur l'ensemble des services Web. Le modèle d'affaires est reproduit dans ces nouveaux services, qui emportent une petite dose de publicité, mais sont de plus en plus éloignés du moteur de recherche qui a fondé la puissance économique de Google.

En effet, dès aujourd'hui, en plus de sa fonction de moteur de recherche, Google offre gracieusement les fonctions décrites en annexe², pour la plupart très performantes.

[Google Base](#) propose aux particuliers et aux entreprises d'y déposer toutes les données qu'ils souhaitent rendre publiques, quelles que soient leur structure ou leur complexité : petites annonces en tous genres ou catalogue complet des produits d'une entreprise ...

¹ Le marché publicitaire en France a été en 2004 de l'ordre de 31 Milliards d'euros soit pour chacun des 26 millions de ménages une dépense de 1.200 € annuellement! Si seulement 5% de cette somme passe aujourd'hui par Internet, on comprend que réinvestir une partie de cette somme dans le média Internet va devenir rapidement rentable.

² Version anglaise car plus complète

C'est ainsi la possibilité de mise en œuvre et de contrôle de véritables canaux de vente et d'échanges qui est visée. C'est aussi déjà Google Checkout : c'est-à-dire un système de micro-paiements qui risque de devenir une norme de fait, en l'absence d'entente et de proposition du système bancaire traditionnel.

De surcroît, Google Drive, c'est-à-dire la mise à disposition d'un immense disque dur virtuel, est annoncé pour bientôt. Au départ, il s'agira d'offrir à chacun une sauvegarde des données stockées sur son PC. Mais, de fait, il apparaîtra très rapidement que ce disque sera le disque de référence des clients. Il sera en effet extensible à l'infini et accessible à partir de n'importe quel poste (PC, téléphone, etc.).

1.2. LES ATOUTS TECHNOLOGIQUES DE GOOGLE

Pour mettre en œuvre son outil de recherche, Google a fait un choix technique très significatif : équiper ses serveurs de processeurs, issus du commerce, et les empiler en se donnant comme unique contrainte de pouvoir redistribuer la puissance de calcul et de stockage sur ses différentes machines en fonction des besoins, en rajoutant « à la volée » autant de machines que le trafic dans une région donnée l'exige. Les experts les évaluent à quelques 450.000 serveurs, répartis dans une vingtaine de « data centers ». Ce faisant Google a construit une architecture informatique propriétaire sans équivalent, très souple et très robuste et une phénoménale capacité de calcul qui permet de répondre à plus d'un milliard de requêtes par jour, chacune interrogeant huit milliards de pages Web en moins d'un cinquième de seconde...

Il s'agit à l'évidence d'un atout important de Google. La puissance informatique installée permet une adaptation à la variation de la demande, le développement de nouveaux services et une qualité de service de bon niveau. Les équipes techniques de Google ont su maîtriser une architecture d'informatique distribuée mâtinée de calculateurs centralisés (les « data centers »). Elles ont développé un savoir-faire allant de la programmation des instructions élémentaires des micro-processeurs à l'optimisation des fonctions logicielles de management des systèmes en passant par la maîtrise d'une architecture évolutive et standardisée, qui conduisent à une articulation particulièrement efficace entre algorithmique et architecture. Ce savoir-faire pourrait bien constituer la véritable avance technologique de Google.

Les coûts correspondants sont remarquablement bas (450.000 machines soit environ 200 M€). Google peut ainsi loger sur ses serveurs non seulement les index et les copies du Web, mais aussi tout le reste : c'est-à-dire l'intégralité des données importantes (10 % des disques durs) des ordinateurs de ses utilisateurs !

1.3. LE POTENTIEL DE GOOGLE

Ainsi :

- Google a mis en œuvre une architecture et un système d'exploitation qui lui permettent à moindre coût de « tout » stocker ;
- Il offre gratuitement cet espace à tous les utilisateurs ;
- Il offre également les outils pour consulter mais aussi exploiter ces données de manière simple et partageable dans la sphère privée ou professionnelle ;
- Google finance sa démarche avec les profits de la publicité pour laquelle il développe de nouveaux marchés et attire de nouveaux annonceurs (cf. la théorie de la « longue queue »³) ;
- Son modèle économique basé sur la publicité et la gratuité l'expose moins que Microsoft au droit de la concurrence ;
- Il est important de noter que le fait que ces fonctionnalités soient mises gratuitement à la disposition des utilisateurs s'accompagne d'une exigence de qualité de service de bon niveau, qu'il s'agisse de l'ergonomie des écrans, des temps de réponse ou de l'étendue des fonctions annexes et des liens offerts. Google met ainsi largement en œuvre le principe qui voudrait que, dans une économie de la connaissance, l'accès à l'information soit gratuit ;

³ « LongTail » fait référence à la forme d'une courbe commentée par le [management de Google](#) montrant qu'en touchant des millions de niches, Internet peut automatiquement satisfaire des micro-marchés dont la somme est un énorme marché nouveau

- Il accélère son emprise économique avec un rythme exceptionnel de sortie de produits quasi hebdomadaire. Certes, avec une granularité des services qui reste fine, les fonctionnalités offertes sont largement indépendantes les unes des autres. La perspective d'une intégration et d'une mise en cohérence de ces services reste lointaine, ce qui facilite cette profusion de nouveautés et rend moins sensible la perception des risques intrusifs et invasifs dans la vie personnelle des internautes. Cette stratégie est à l'œuvre et il est fascinant de voir qu'elle peut avancer aussi vite, sans aucune contrainte. Or les enjeux, notamment de souveraineté, sont immenses !

2. DES CAPACITÉS EUROPÉENNES MAIS DES RÉPONSES PONCTUELLES ET NON COORDONNÉES

2.1. DES COMPÉTENCES RECONNUES

La France dispose de compétences reconnues sur le plan international, qu'il s'agisse du domaine des principes et algorithmes de recherche d'informations par le contenu, ou de la construction de très grandes infrastructures informatiques capables de fournir les performances d'un site comme celui de Google.

2.1.1. RECHERCHE PAR LE CONTENU

Déjà au cours des années 70 et jusqu'à la fin des années 80, le logiciel Mistral de la société CII/Bull était un leader mondial en matière de recherche documentaire.

L'école française de Recherche se distingue en majorité de l'école américaine par un fort recours à la linguistique pour améliorer la pertinence des réponses aux requêtes. Des équipes structurées existent dans les grands organismes comme le CEA, le CNRS ou l'INRIA, localisées principalement à Grenoble, à Nancy, en région parisienne et à Toulouse. Au-delà du problème de la recherche d'informations textuelles, les laboratoires s'orientent vers la recherche d'informations multimédia (image, son, vidéo). Des équipes du GET ou de l'école des mines de Paris travaillent ainsi sur les aspects de l'image, de la vidéo voire de la reconnaissance biométrique.

A côté de ces équipes de recherche plusieurs « jeunes pousses » se sont créées, qui conservent dans leurs produits cette caractéristique d'une approche linguistique, comme Lingway, New Phoenix, GO-Albert, Sinequa, La société Exalead quant à elle se positionne directement face à Google avec des techniques de moteur de recherche similaires.

En matière de recherche d'images les sociétés LTU ou encore New-Phoenix ont développé des offres. Certains grands groupes ont constitué de fortes équipes sur le sujet, tels Thalès, France Télécom, ou Thomson.

2.1.2. INFRASTRUCTURE INFORMATIQUE

Le rôle de l'infrastructure dans la fourniture de services du type de ceux de Google est essentiel. L'architecture à développer est celle dite d'un cluster qui permet de regrouper de manière cohérente au sein d'une même infrastructure plusieurs dizaines, centaines, ou milliers de serveurs interconnectés avec ce qu'il convient de modularité de croissance (scalabilité), de haute disponibilité, de sécurité et de « manageabilité ».

Deux types de traitements sont mis en œuvre :

- des traitements de type « portail » où il s'agit d'écouler un maximum de requêtes indépendantes en même temps, chaque requête étant simple ;
- des traitements ou requêtes dont la complexité nécessite de paralléliser les algorithmes correspondants pour obtenir un temps de réponse raisonnable.

Pour les traitements de type « portail », les solutions relèvent des constructeurs informatiques ou de quelques grandes sociétés de services qui disposent des compétences techniques nécessaires. Il faut aussi mentionner l'existence de jeunes pousses très innovantes comme Kewego ou Dailymotion dans le domaine en émergence des portails de blogs vidéo.

Pour les traitements parallèles, l'infrastructure est plus complexe et très peu d'acteurs ont les compétences techniques pour aborder les architectures de très grande taille. Exalead a démontré ses compétences « système » sur des configurations moyennes et Bull sur de très grandes configurations dans le domaine du calcul à haute performance. Dans les deux cas des logiciels Open Source peuvent être utilisés, même si l'importance des configurations à mettre en œuvre nécessite une très grande expertise technique pour optimiser ces logiciels et les rendre performants à l'échelle requise.

2.2. DES INITIATIVES LIMITÉES ET NON COORDONNÉES

Il existe en France et en Europe un foisonnement d'initiatives privées relevant des champs décrits en point 1.

On peut citer à titre d'exemples :

- les banques d'images Kewego (<http://www.kewego.fr/>) et Wat (<http://www.wat.tv/>) ;
- la téléphonie sur Internet Wengo (<http://www.wengo.fr/>) ;
- les données géographiques, plans et itinéraires Mappy (<http://www.mappy.com/>) ;
- la traduction automatique avec Systran (<http://www.systran.fr/>).

Cependant, hormis ces deux dernières qui ont acquis une stature internationale, elles restent quelque peu confidentielles et sans lien entre elles. Leur diffusion et leur croissance sont limitées par le manque de capacités matérielles et financières leur permettant d'asseoir leur développement en Europe et par l'absence de grands vecteurs fédérant leur promotion.

Les quelques initiatives publiques européennes et françaises :

- la bibliothèque numérique européenne (<http://www.theeuropeanlibrary.org/>) ;
- le système d'information géographique Geoportail : <http://www.geoportail.fr/> ;
- les archives audiovisuels de INA (<http://www.ina.fr/>) ;
- le moteur de recherche Quaero : <http://www.quaero.fr/> ;

souffrent des mêmes maux et s'ignorent entre elles. Il suffit de se connecter aux sites de la Bibliothèque numérique européenne ou de Quaero pour constater combien les résultats souffrent de la limitation des moyens.

Dans le même esprit, la récente décision (du 29 septembre 2006) du Ministère de l'économie, des finances et de l'industrie de confier à Thalès la mise en place pour fin 2007 d'un service de consultation du plan cadastral pourrait être analysée à l'aune de la même carence de vision d'ensemble.

3. QUE FAIRE ?

3.1. LES MOTIVATIONS POUR AGIR

Les questions de souveraineté sont importantes, même si leur perception par le grand public est variable. En effet, si les États apparaissent légitimes pour le contrôle des informations de nature régaliennne (comme l'état civil par exemple), la situation est toute autre pour les données non publiques. De fait, Google a démontré qu'il avait une légitimité implicite à contrôler des informations de nature privée : d'aucuns préféreront mettre leur agenda personnel sur Google que de le placer sur un site du ministère de l'Intérieur.

La défense de la langue française est une motivation importante, même s'il faut bien admettre que pour des raisons essentiellement mercantiles, Google en tient compte. Si Google Books propose des livres en français, c'est parce qu'ils sont consultés en France, et que cela permet de drainer de la publicité en France.

L'innovation technologique est un élément fort de développement de savoir-faire et de développement de l'emploi. Google ne s'y trompe d'ailleurs pas, qui embauche de jeunes talents français, suisses ou allemands pour étendre son influence sur l'Europe.

La compétitivité économique pourrait finalement constituer la meilleure motivation. Google a inventé un modèle économique fondé sur une petite dose de publicité dans son moteur de recherche, multipliée par un effet de masse. Google est difficilement attaquant sur son terrain : il est peu probable qu'un meilleur moteur de recherche que Google puisse être construit, en démarrant aujourd'hui. Par ailleurs, le suivisme est sans intérêt et ne répond pas aux défis soulevés. Mais il semble important d'éviter le monopole : si rien n'est fait, Google sera présent sur tous les nouveaux services web.

3.2. FAVORISER EN FRANCE LA DEMANDE DE NOUVEAUX SERVICES EN DÉVELOPPANT UNE OFFRE ADAPTÉE

Il paraît essentiel de construire une argumentation en faveur d'un modèle économique favorisant le développement d'une offre française de services constitutifs de la société de l'information.

La question se pose de l'émergence d'un consensus⁴ autour de ce qui pourrait être qualifié⁵ de « Service Public de l'Énergie Informatique ». Il s'agirait de développer des *Centres de ressources de contenus numériques, associant offre de services innovants et puissance de traitement et d'archivage*.

Les enjeux vont bien au-delà des interventions publiques existantes, comme le projet QUAERO, car c'est de l'émergence de véritables services de contenus numériques à l'image de ceux de l'énergie et de même importance pour la société et l'économie qu'il est question.

3.2.1. UNE INITIATIVE PUBLIQUE POUR LE DÉVELOPPEMENT DES INFRASTRUCTURES

Au même titre que pour les sources d'énergie ou les ressources rares, l'intervention du secteur public semble indispensable en phase initiale avant d'envisager une intervention des entrepreneurs privés.

Celle-ci peut prendre diverses formes (financement public intégral⁶, projet de l'Agence de l'Innovation Industrielle, pôle de compétitivité sur une région pilote, partenariat public-privé, etc.).

A cet égard, l'implication des régions pourrait être favorisée : la mise en œuvre de centres de serveurs régionaux consacrés aux problématiques régionales ou familiales pourrait être une contribution concrète et visible au développement de la société de l'information. Ces centres serveurs pourraient constituer autant de plates-formes à partir desquelles différentes collectivités et administrations, des associations ou des entreprises pourraient proposer des services ciblés, des médias, etc., et innover.

Les entreprises informatiques et notamment les constructeurs pourraient également être mobilisés et les investisseurs être intéressés par des avantages fiscaux à la clé.

⁴ Parvenir à convaincre les décideurs qu'il est nécessaire de construire un centre de traitements de données équivalent à 30.000 serveurs, alors que la demande n'est pas encore identifiée, participe de la construction de ce consensus.

⁵ Par analogie avec le Service Public de l'Électricité.

⁶ Le modèle d'une centrale de traitement informatique entièrement financée par le secteur public peut être étudié.

3.2.2. LA DIFFUSION DE LA COMPÉTENCE DES LABORATOIRES PUBLICS

L'initiative publique pourrait comporter un volet s'appuyant sur les compétences des laboratoires de recherche publics (INRIA, CNRS,...).

Leur apport serait articulé autour :

- des principes de mise en œuvre des logiciels de base de serveurs informatiques, de leur modélisation tant qualitative que quantitative ;
- de l'organisation des applications elles-mêmes, et notamment de leur parallélisation⁷.

Toutefois, il ne faudrait pas conclure du fait que, les composants nécessaires à la constitution de puissants serveurs informatiques existant sur le marché, et même en Open Source pour ce qui concerne les éléments logiciels, la mise en œuvre d'un « data center » performant, tolérant aux pannes, facilement opérable et présentant une bonne garantie de service serait à la portée de tout un chacun, à partir du moment où il disposerait d'un « kit ressource informatique » qui serait issu des laboratoires publics.

L'intervention des industriels spécialisés, leur savoir-faire, leur méthodologie et leur expérience, sont en tout état de cause essentiels. Il s'agit de problèmes d'ingénierie complexes qui sont affaires de spécialistes, compte tenu notamment de la taille des infrastructures à construire et du fait que les solutions peuvent varier d'une application à une autre.

Ces préoccupations correspondraient à un changement de paradigme des axes de recherche français ou même européens : ont été privilégiés dans le passé récent les sujets de recherche relevant des couches élevées du modèle ISO, c'est à dire globalement des logiciels et applications. Le besoin exprimé ici croise étroitement les questions matérielles et les questions de logiciels de base.

3.2.3. L'ACCÈS AUX DONNÉES PUBLIQUES

Les administrations devraient donner l'accès aux données publiques sous leur forme électronique afin de permettre aux services sur le web d'atteindre une taille critique (exemple : les listes de propositions d'emploi de l'ANPE, les fonds numérisés de la Bibliothèque Nationale⁸,...).

En effet, il devient urgent que les données numériques de l'État soient effectivement ouvertes au public, ce qui signifie que les administrations concernées ne les utilisent et ne les rendent publiques plus seulement à travers leurs sites institutionnels (selon des calendriers et des logiques qui leur sont propres), mais aussi que tout tiers puisse extraire et republier tout ou partie de ces données (sous certaines conditions minimales de garantie de non-manipulation). Il faut rendre les données publiques vraiment publiques.

En tout état de cause, il faut favoriser l'accession par des entrepreneurs, même tout petits, aux données publiques - administratives, cartographiques, économiques, culturelles, de transport, etc.

3.2.4. LA QUALITÉ DANS LA GRATUITÉ POUR FAVORISER L'ADHÉSION DES UTILISATEURS

Il est important, comme cela a été souligné dans le cas de Google, que l'objectif d'une performance optimale des services fasse partie intégrante du cahier des charges d'une offre publique de ressources informatiques.

Il faut ainsi que l'ergonomie d'utilisation de ces ressources informatiques soit facile, complète et à l'état de l'art, que le système soit capable d'absorber les pointes de charge, et que le service rendu soit de qualité.

⁷ La loi de Moore, qui se traduit par une multiplication par 2 de la puissance des ordinateurs tous les 2 ans, ne trouvera bientôt plus à s'appliquer par une augmentation de la fréquence d'horloge des micro-processeurs, du fait des problèmes de dissipation de chaleur dans les puces. Aussi, la piste de l'accroissement de performances par le parallélisme des tâches représente un vrai challenge.

⁸ La Bibliothèque nationale de France doit numériser le maximum d'ouvrages, et rendre ceux-ci accessibles en ligne à tous les fournisseurs de contenus qui le souhaitent (Google, mais aussi la bibliothèque numérique virtuelle, etc.).

3.2.5. DE NOUVELLES RÈGLES DU JEU POUR LES ENTREPRENEURS

Quelques pistes de réflexion :

- Innover vite, tester avec les clients finaux, raffiner ensuite. Ne plus être en position de suivisme par rapport à Google ;
- Être à l'affût des bonnes idées nées dans les laboratoires et les garages, et les racheter vite ;
- Considérer ses actifs essentiels comme des plates-formes d'innovation ouvertes (sous conditions) aux autres innovateurs. Des millions de petits entrepreneurs se développent grâce aux publicités et au moteur de Google, à Google Earth, au catalogue et au système de prise de commande d'Amazon, etc.

3.2.6. LE CHOIX DES SERVICES PILOTES

La mise en place de serveurs de traitement et d'archivage de données doit s'accompagner de la définition de services susceptibles d'intéresser un large public.

À titre d'illustration, une expérimentation sur l'un ou l'autre de ces deux services pourrait être entreprise :

- la sauvegarde des disques durs personnels ou d'entreprises ;
- le paiement en ligne de petites sommes.

3.3. TRAVAILLER À L'ÉMERGENCE D'ALTERNATIVES EUROPÉENNES

Quelques exemples de mise en œuvre :

- Exiger de projets tels que QUAERO, en contrepartie de l'aide publique, d'aller plus vite (de lancer des choses par étape), de travailler d'une manière plus ouverte, de publier non pas leur code source ou leurs secrets, mais leurs interfaces de programmation ; considérer ces projets comme des plates-formes d'innovation, et non des tuyaux dont doivent un jour sortir des innovations géniales entièrement maîtrisées par les membres du consortium ;
- Faciliter des projets d'entreprise autour des concepts du "Web 2.0" mais en insistant toujours sur l'ouverture et la création de valeur par agrégation et coopération ;
- Favoriser l'émergence de normes "sémantiques" d'échange qui permettront aux personnes qui stockent des données en réseau, partagent des profils et des contacts, publient des blogs, numérisent leurs livres, etc..., de ne jamais être prisonnières d'une plate-forme donnée, et pourraient même contraindre à une certaine interopérabilité entre plates-formes, qui constitue un assez bon antidote aux monopoles.

3.4. NÉGOCIER, DÉVELOPPER DES PARTENARIATS AVEC GOOGLE

L'Europe pourrait aussi agir directement vers Google. Par exemple :

- elle pourrait faire valoir ses préoccupations en matière de souveraineté pour, par exemple, s'assurer que les données personnelles des européens sont stockées en Europe, sur des architectures sécurisées et auditables. Elle pourrait pour ce faire définir un nouveau registre juridique protégeant ses souverainetés ;
- elle pourrait aussi favoriser le développement des partenariats⁹ à partir de ses domaines d'excellence, et investir fortement pour en être une référence incontournable. Le sujet bancaire pourrait être un des thèmes de prédilection.

⁹ Ses préoccupations premières pourraient être d'initier des partenariats public/privé de type GALILEO en s'appuyant sur le caractère stratégique des projets et en promouvant l'impératif de non-dépendance de l'accès à l'information, de fédérer et de rendre interopérables les différentes initiatives européennes, de s'assurer de leur robustesse matérielle et logicielle. Elle pourrait également les rendre éligibles au financement par le PCRD et les réseaux trans-européens.

3.5. PROPOSITIONS D' ACTIONS

En conclusion, il est essentiel de répondre au double défi posé par Google, en mettant à profit les compétences techniques existant en France, qui sont efficaces sur ces sujets :

- le défi technique, par la mise en place de centres de ressources informatiques d'initiative publique, en faisant le pari que les acteurs seront capables de créer du contenu et de développer des services ;
- le défi de l'accroissement de valeur des contenus, en inventant de nouveaux services à valeur ajoutée, à destination, comme Google l'a fait, du grand public, pour développer une économie de la connaissance de masse.

Il est proposé de mener pour cela deux actions complémentaires :

- Construire un *démonstrateur* et proposer rapidement *des services pilotes grand public* sur des thèmes de souveraineté comme la sauvegarde des disques durs ou sur de nouveaux besoins comme les micro-paiements ;
- Mettre en place *un groupe de travail de prospective à moyen terme* (comportant dans un premier temps une mission technique vers Google). L'objectif consistera à travailler sur *la génération de services qui suivra* Google, qui pourrait nécessiter 5 ans de développement.