

Report published on 10 September 2007

23 October 2006

**FOR A EUROPEAN INTERNET INITIATIVE TO FOSTER  
THE DEVELOPMENT OF THE KNOWLEDGE ECONOMY**

Today, France and Europe have to address more than the control of data processing and management tools (Microsoft, Intel and Cisco). They have to address the control of access to information and its dissemination. The issue is no longer limited to investment in the networks and media (broadband, fibre optics, etc.), but extends to digital content and implies involvement in the formation of this content.

Can all Internet search and storage engines be under the responsibility of just one State, even if that State is a friend? Is it economically and ethically acceptable to depend, for a purpose as important as access to information, on virtually one single company? How can services provided free of charge to hundreds of millions of users be governed by the usual market rules?

The emergence of a new economic model of operators, prepared to offer Internet users all access services free of charge provided they give them their personal information, raises major legal and political questions for the national and European decision-makers.

## **1. GROWING DEPENDENCY ON A SMALL NUMBER OF OPERATORS**

Google is the archetype of this new operator model. Not only is Google the leading and by far the most high-performance search engine, it is also an exceptional company with considerable economic weight. Google is currently worth some €100 billion and its business model, based on a supply normally provided free of charge to the user, is an approach that can but destabilise the players in place.

### **GOOGLE'S SERVICES**

Google's image of web search specialist is misleading: Google far from contents itself with an ostensibly harmless position in what might appear to be an Internet services niche. By renting the advertising space<sup>1</sup> generated by its search engine, Google is in a situation of pure economic rent that gives it the means to be present in all Web services. The business model is replicated in these new services, which carry a small amount of advertising, but are increasingly removed from the search engine that founded Google's economic power.

In fact, in addition to its search engine function, Google currently offers free of charge the functions described in the annexed pages<sup>2</sup>, which are for the most part high-performance.

Google Base offers individuals and businesses a service where they can submit all the data they wish to make public, regardless of their structure or complexity: classified advertisements of all kinds or a full catalogue of a firm's products, etc.

---

<sup>1</sup> In 2004, the advertising market in France totalled some €31 billion, equal to an annual outlay of €1200 for each of the 26 million households! Although only 5% of this sum currently goes on the Internet, it is easy to see how reinvesting part of this sum in Internet media can quickly become profitable.

<sup>2</sup> English version since it is more complete.

The aim here is the possibility of setting up and controlling veritable sales and exchange channels. This is also already the case with Google Checkout: a micropayment system that could well become the norm in the absence of agreements within and proposals from the traditional banking system.

Moreover, Google Drive, a huge virtual hard drive, is said to be coming soon. It will initially offer everyone a service to save data stored on their PCs. Yet this disk is set to quickly become the standard disk for customers. It will be infinitely expandable and accessible from any station (PC, telephone, etc.).

### **GOOGLE'S TECHNOLOGICAL ASSETS**

Google made a highly significant technical choice for its search tool: to fit its servers with commercial processors and stack them, with the sole constraint being the possibility to redistribute the computational and storage power on its different machines as and when it needs to by adding as many machines as the traffic in a given region requires. Experts estimate that Google has some 450,000 servers distributed among twenty-odd "data centers". In this, Google has built an unparalleled, highly flexible and robust proprietary IT architecture with a phenomenal computational capacity capable of processing over one billion search queries per day, each operation consulting eight billion Web pages in less than one-fifth of a second ...

This is obviously a major asset for Google. The installed computer power allows for adjustment to variations in demand, the development of new services and a good level of service quality. Google's technical teams have set up a distributed architecture combined with centralised calculators (the "data centers"). They have developed expertise ranging from the programming of basic microprocessor instructions through to the optimisation of systems management software functions and command of a standardised, evolutionary architecture, resulting in a particularly efficient link-up between algorithmics and architecture. This know-how could well constitute Google's real technological edge.

The corresponding costs are remarkably low (450,000 machines for approximately €200 million). Google servers can therefore host not only the indexes and copies of the Web, but also all the rest: i.e. all the important data (10% of hard disk space) from their users' computers!

### **GOOGLE'S POTENTIAL**

#### In brief:

- Google has set up an architecture and operating system that enables it to store "everything" at a low cost;
- It offers this space free to all users;
- It also provides the tools to consult, and process, these data in an easy and shareable way in the private or professional sphere;
- Google finances its service from advertising revenues, for which it develops new markets and attracts new advertisers (see Long Tail theory<sup>3</sup>);
- Its business model based on advertising and free-of-charge services makes it less exposed to competition law than Microsoft;
- It is important to note that the free nature of these functions for the users goes hand in hand with a requisite for good service quality, whether in the user-friendliness of the screens, response times or the extent of subsidiary functions and links provided. Google therefore largely puts into practice the principle that, in a knowledge-based economy, access to information should be free;
- It has stepped up and tightened its business hold with an exceptional, virtually weekly rate of new product releases. Admittedly, with services that are still fine-grained, the functions offered are largely independent of one another. The integration and harmonisation of these services remains a distant prospect, which facilitates this profusion of new products and tempers the perception of the risks of intrusion into and invasion of Internet users' private lives. This strategy is at work and it is

---

<sup>3</sup> Long Tail refers to the form of a curve picked up on by Google's management. It shows that by reaching millions of niches, the Internet can automatically satisfy micro-markets whose sum total represents a huge new market.

fascinating to see that it can move so fast, without any hindrance. Yet the implications, especially in terms of sovereignty, are huge!

## **2. EUROPEAN CAPACITIES, BUT ISOLATED, UNCOORDINATED RESPONSES**

### **RENOWNED EXPERTISE**

France has internationally renowned expertise in the field of content-based search principles and algorithms and also the construction of massive computer infrastructures capable of providing the performances of a site such as Google.

### **CONTENT-BASED SEARCH**

As early as the 1970s and through to the end of the 1980s, CII/Bull's Mistral software program was a world leader in document retrieval.

The French research school differs from the American school mainly in terms of a high use of linguistics to improve the pertinence of responses to queries. There are structured teams in leading bodies such as the French Atomic Energy Commission (CEA), The French National Centre for Scientific Research (CNRS) and the French National Institute for Research in Computer Science and Control (INRIA), based mainly in Grenoble, Nancy, the Paris area and Toulouse. In addition to their work on textual data, the laboratories are working on multimedia data searches (image, sound and video). Teams at the Telecommunications Schools group (GET) and the National Engineering School (*Ecole des Mines*) of Paris are working on various aspects of images, videos and even biometric recognition.

Alongside these research teams, a number of innovative start-ups have been created. Their products all take this same linguistic approach: Lingway, New Phoenix, GO-Albert, Sinequa, etc. And Exalead is positioned in direct competition with Google, with similar search engine techniques.

LTU and New Phoenix have also developed image search offerings. Certain major groups have put together strong teams in this area, such as Thalès, France Télécom and Thomson.

### **COMPUTER INFRASTRUCTURE**

The infrastructure plays a vital role in providing the type of services offered by Google. The architecture to be developed is called a cluster. A cluster is a group of tens, hundreds and even thousands of interconnected servers acting as a single infrastructure and providing what is required in the way of scalability, high availability, security and manageability.

Two types of processing are possible:

- Portal processing, where a maximum number of independent requests is dealt with at the same time, with each request being simple;
- Processing of requests whose complexity calls for the corresponding algorithms to be parallelised to obtain a reasonable response time.

In portal processing, the solutions are provided by computer manufacturers and a few major service companies with the necessary technical expertise. Note also the existence of highly innovative start-ups such as Kewego and Dailymotion in the emerging field of video blog sites.

In parallel processing, the infrastructure is more complex and very few players have the technical skills to take on extremely large architectures. Exalead has demonstrated its "system" skills on medium-sized configurations and Bull on extremely large configurations in high-performance

computation. In both cases, open-source software can be used, even though the magnitude of the configurations to be set up calls for an extremely high level of technical expertise to optimise this software and make it high-performance at the required scale.

### **LIMITED, UNCOORDINATED INITIATIVES**

In France and Europe, a whole host of private initiatives can be found in the areas described in Point 1.

Some examples are:

- Kewego (<http://www.kewego.fr/>) and Wat (<http://www.wat.tv/>) image banks;
- Wengo telephony on the Internet (<http://www.wengo.fr/>);
- Mappy geographic data, maps and itineraries (<http://www.mappy.com/>);
- Systran's translation software (<http://www.systran.fr/>).

However, aside from these last two products, which have acquired international standing, the initiatives remain fairly small-scale and stand-alone. Their distribution and growth are restricted by a lack of material and financial capacities to underpin their development in Europe and the absence of major vehicles for their united promotion.

The few European and French public initiatives:

- The European digital library (<http://www.theeuropeanlibrary.org/>);
- The Geoportail geographic information system (<http://www.geoportail.fr/>);
- The INA audiovisual archives (<http://www.ina.fr/>);
- The Quaero search engine (<http://www.quaero.fr/>);

all suffer from the same drawbacks and are stand-alone. A simple connection to the European digital library or Quaero sites shows the extent to which the results suffer from restricted resources.

Similarly, the recent decision (29 September 2006) by the Ministry of the Economy, Finance and Industry to task Thalès with setting up a land register map consultation service by the end of 2007 could be analysed in the light of the same lack of overall vision.

## **3. WHAT SHOULD BE DONE?**

### **REASONS TO ACT**

Sovereignty issues are important, even if the general public varies in the way it perceives them. Although States appear to be the legitimate administrators of sovereign information (such as civil registration data), the situation is quite different for non-public data. Google has demonstrated an implicit legitimacy in the management of private information: some would prefer to put their personal diary on Google rather than a Ministry of the Interior site.

Defending the French language is an important reason, even though Google admittedly takes this into account, even if it is essentially for mercantile reasons. If Google Books offers books in French, it is because they are consulted in France, which provides an opportunity to tap into advertising in France.

Technological innovation is a strong element in the development of know-how and growth in employment. Google makes no mistake in recruiting young French, Swiss and German talents to extend its influence in Europe.

Lastly, economic competitiveness could form the best reason. Google invented a business model based on a small amount of advertising in its search engine multiplied by a mass effect. Google is hard to tackle on its own ground: it is fairly improbable that a better search engine than Google could be built starting today. Moreover, there is no point trying to jump on the bandwagon since it would not

respond to any of the issues at stake. Yet it is important to avoid a monopoly situation: if nothing is done, Google will have a hold in all the new Web services.

### **FOSTER DEMAND FOR NEW SERVICES IN FRANCE BY DEVELOPING A SUITABLE SUPPLY**

It appears vital to build an argument for a business model that fosters the development of a range of French information society services.

This raises the question of the emergence of a consensus<sup>4</sup> around what could be called<sup>5</sup> a “Public IT Energy Service”. This would mean developing *digital content resource centres, combining innovative services with powerful processing and storage facilities*.

The implications go way beyond existing public action, such as the Quaero project, since this concerns the emergence of real digital content services like the energy services and with the same importance for society and the economy.

#### **3.2.1. A PUBLIC INITIATIVE FOR THE DEVELOPMENT OF THE INFRASTRUCTURES**

In the same way as for energy sources and rare resources, public-sector action appears to be vital in an initial phase before considering any private-sector involvement.

This action could take a number of different forms (full public financing<sup>6</sup>, French Agency for Industrial Innovation project, competitive cluster in a pilot region, public-private partnership, etc.).

In this regard, the involvement of the regions could be fostered: the implementation of regional server centres dedicated to regional and family issues could form a concrete and visible contribution to the development of the information society. These server centres could form platforms from which different authorities and administrations, associations and businesses could propose targeted services, media, etc., and innovate.

IT firms and especially computer manufacturers could also be rallied and investors interested by bonus tax breaks.

#### **3.2.2. THE DISSEMINATION OF PUBLIC LABORATORY EXPERTISE**

The public initiative could include a track drawing on the expertise of the public research laboratories (INRIA, CNRS, etc.).

Their contribution would hinge on:

- Principles for the use of basic server software programs and their qualitative and quantitative modelling;
- The organisation of the applications themselves, and especially their parallelisation<sup>7</sup>.

Yet the fact that the components required to build powerful servers exist on the market, and are even available in open source for the software components, does not mean that everyone will be able to set up a high-performance, crash-tolerant, easily operable “data center” with a good guarantee of service just because they have a “computer resource kit” produced by the public laboratories.

---

<sup>4</sup> Convincing decision-makers that a data processing centre equivalent to 30,000 servers needs to be built when demand has not yet been identified is part of the building of this consensus.

<sup>5</sup> By analogy with the Public Electricity Service.

<sup>6</sup> The model of a central computer processing centre entirely financed by the public sector could be studied.

<sup>7</sup> Moore’s Law that computing power doubles every two years will soon no longer apply to the microprocessor clock frequency due to problems of heat dissipation in the chips. So the avenue of increasing performance via task parallelism represents a real challenge.

In all events, it is vital to bring on board specialised manufacturers with their know-how, methodology and experience. These are complex engineering problems that require specialists, especially in view of the size of the infrastructures to be built and the fact that solutions can vary from one application to the next.

These concerns would represent a change of paradigm for the French, and even European, research focuses. The focus in the recent past has been on research subjects concerning the upper layers of the OSI model, i.e. generally the software and applications. The need expressed here intersects hardware concerns with basic software concerns.

### **3.2.3. ACCESS TO PUBLIC DATA**

The administrations should provide access to public data in electronic form to enable the Web services to reach critical mass (e.g. the lists of National Employment Agency (ANPE) job vacancies, the *Bibliothèque Nationale de France*'s online catalogue<sup>8</sup>, etc.).

It is becoming urgent for the State's digital data to be made really open to the public, which means that the administrations concerned no longer use and publish them solely on their institutional sites (in keeping with their own timetables and logics), but that any third party can extract and republish any or all of these data (subject to certain minimum non-manipulation conditions). Public data need to be made really public.

In any case, steps should be taken to further access by businesspeople, even the smallest entrepreneurs, to public data – administrative, cartographic, economic, cultural, transport, etc.

### **3.2.4. QUALITY IN THE FREE SERVICE TO FOSTER USER TAKE-UP**

It is important, as highlighted in the case of Google, for the goal of optimal service performance to be an integral part of the technical specifications for a public supply of IT resources.

These IT resources' ergonomics should therefore be easy, complete and state of the art, the system should be capable of absorbing demand peaks, and the service provided should be high quality.

### **3.2.5. NEW GROUND RULES FOR ENTREPRENEURS**

A few ideas for consideration:

- Innovate quickly, test with end customers and then perfect. No longer play "follow my leader" behind Google;
- Always be on the look-out for good ideas from the laboratories and garages, and buy them up quickly;
- Consider our critical assets as innovation platforms open (subject to certain conditions) to other innovators. Millions of small businesses develop due to the Google advertisements and search engine, Google Earth, the Amazon catalogue and ordering system, etc.

### **3.2.6. THE CHOICE OF PILOT SERVICES**

The setting up of data processing and storage servers should go hand in hand with the definition of services liable to be of interest to a wide audience.

---

<sup>8</sup> The *Bibliothèque Nationale de France* (French National Library) is due to digitise a maximum number of works and make them accessible online to all interested content providers (Google, but also the virtual digital library, etc.).

For example, a pilot test could be conducted on one or the other of these two services:

- Saving personal or business hard disks;
- Online payment of small sums.

### **WORK TOWARDS THE EMERGENCE OF EUROPEAN ALTERNATIVES**

A few examples for implementation:

- Require that projects such as Quaero, in return for government assistance, work faster (launch products in stages), work more openly, and publish not their source code or secrets, but their programming interfaces; consider these projects to be innovation platforms and not pipelines from which brilliant innovations should one day spring under the entire control of the members of the consortium;
- Facilitate business projects based on the Web 2.0 concepts, but with the emphasis always on openness and the creation of value through aggregation and co-operation;
- Foster the emergence of semantic exchange standards so that the people who deal in storage networking, share profiles and contacts, publish blogs, digitise their books, etc. never have to be prisoners of a given platform. These standards could even impose a certain amount of interoperability between platforms, which is a fairly good antidote to monopolies.

### **NEGOTIATE AND DEVELOP PARTNERSHIPS WITH GOOGLE**

Europe could also work directly with Google. For example:

- Europe could voice its sovereignty concerns to ensure, for example, that Europeans' personal data are stored in Europe on secure, verifiable architectures. To do this, Europe could define a new legal register protecting its sovereignties;
- Europe could also foster the development of partnerships<sup>9</sup> based on its areas of excellence and invest heavily to become a key benchmark. Banking could be one of the leading subjects.

### **PROPOSALS FOR ACTION**

To conclude, it is essential to take up the dual challenge thrown down by Google by making the most of France's existing technical skills, proficient in these areas:

- The technical challenge, by setting up public-initiative information technology resources centres, assuming that the players are capable of creating content and developing services;
- The challenge of raising content value by inventing new value-added consumer services, as Google has done, to develop a mass knowledge economy.

### **For this, it is proposed that two complementary actions be taken:**

- Build a *prototype* and promptly offer *pilot consumer services* on subjects of sovereignty such as saving hard disks or for new needs such as micropayments;
- Set up a *medium-term outlook working group* (initially comprising a Google-oriented technical task force) to work on *the generation of services that will follow* Google and which could take five years to develop.

---

<sup>9</sup> Europe's primary concerns could be to set up Galileo-type public-private partnerships based on the strategic nature of the projects and promoting the imperative of the non-dependence of access to the information, to bring together and make interoperable the different European initiatives, and to ensure their hardware and software robustness. It could also make them eligible for financing by the CFRP and the Trans-European Networks.