

# The stakes of Big Data in the IT industry China as the next global challenger?

(abstract)

Stéphane Grumbach  
INRIA

Information and communication technologies have lead to a major revolution in our societies. It is hard now to imagine how we communicated, how we accessed knowledge before the Internet. It all started not long ago though, essentially in 1991 when Tim Berners-Lee defined the HTML standard for data format, and the HTTP protocol for data exchange, that made the emergence of the Web possible. This fundamental breakthrough, that some authors compare to the invention of writing, was adopted very rapidly, and the Web grew exponentially fast in all regions of the world, although with different pace.

In its *Global Information Technology Report*, whose 2012 edition is subtitled *Living in a hyper connected world*<sup>1</sup>, the World Economic Forum evaluates the penetration and the impact of information technology in the world. Top ranked are Sweden (1) and Singapour (2), Northern Europe, the USA (8), Canada (9), followed by asian countries, Taiwan (11), Korea (12), Hong Kong (13), and Japan (18). China (51), which leads the BRIC countries ahead of Russia (56), Brazil (65) and India (69), has the largest information society<sup>2</sup>, with 500 millions Internet users<sup>3</sup> and a penetration rate of 37%, higher than the world average 30%.

For the last ten years, new corporations have emerged managing Internet-scale data sets, and offering services that have attracted users in the hundreds of millions. Facebook, created in 2004, and initially accessible only to Harvard students, doubles yearly its number of users now culminating at 800 million, while doubling its revenue on the same pace as well. Twitter, created in 2006, has had similar growth patterns, with 300 million users today. Google+, which has had the fastest growth, reaching 100 million users after only seven months of operation, is predicted to reach 400 millions at the end of 2012. These corporations offer fascinating services that were out of reach only a few years ago, thanks to their powerful data-harvesting technologies.

Their success relies on an amazing business model, that could hardly be anticipated before hand. They offer free services, such as web searching, page hosting, networking, and numerous communication means, that are traded for the private data of their users that can be used for commercial purposes. They scale up to hundreds of millions of users, billions of queries a day for Twitter, tens of millions of basic actions per hour on Facebook, with an astonishing quality of service, that ensure their popularity. Their customers are advertisers that exploit users' private data to target their ads, in a somehow seamless way, users become used to.

Internet-scale corporations have the capacity to produce tremendous amount of knowledge not only on individuals, but more generally on populations, their interests or their behavior. Google first demonstrated this capacity with *Google Flu Trends*, which allows to monitor the searches on flu related terms in all the world, and produces not only real time information on the spread of the

---

<sup>1</sup>GITR 2012, living in a hyperconnected world. World Economic Forum and INSEAD, 2012.

<sup>2</sup><http://www.internetworldstats.com>

<sup>3</sup>Statistical report on internet in china. China Internet Network Information Center, 2012.

disease, much ahead of disease control administrations, but also reliable information, their results have been shown to essentially coincide with CDC measures in the US<sup>4</sup>.

“Data is a vital raw material of the information economy, much as coal and iron ore were in the Industrial Revolution” noted Steve Lohr in the *New York Times*<sup>5</sup>. McKinsey Global Institute assess the tremendous economical potential of big data in their report, *Big data: The next frontier for innovation, competition and productivity*<sup>6</sup>. They estimate for instance that there is a potential annual value of \$300 billion for the US health care system. Data is thus a resource, much like a raw material, with unlimited supplies, and whose economic potential has only started to be envisioned. But it raises serious concerns, with a trade-off between the economic potential of these data and the risks they pose to users and their privacy, as shown by *The Economist* in their special issue entitled the *Data Deluge*<sup>7</sup>.

Handling Big Data is of tremendous importance. The USA have developed the strongest industry worldwide, with most of the first online social systems accessed in the world, such as Google, Facebook, YouTube, Yahoo!, Wikipedia, Windows Live, Twitter, Amazon, etc. With these corporations, the USA harvest private data of people all around the world that can be analyzed for an unpredictable set of purposes. These sites are used in all European countries where they have dominant positions. In Italy for instance, only 7 of the top 20 sites are Italian, and this holds globally, according to Alexa’s ranking<sup>8</sup>.

China is the only country which has developed a very powerful industry which harvest the data produced by people in China, where most of the fifty first sites are Chinese. As shown in the infography produced by Ogilvy<sup>9</sup>, there is no area of the social media where a Chinese company cannot be found. Moreover China has four of the first twenty sites worldwide, the search engine Baidu (5th position worldwide), which has 60% of the local market, Tencent QQ (9), for games, microblogs, Alibaba Taobao (14) in e-commerce, and Sina (15) for news and microblogs.

China will most probably challenge the US Big Data Industry in the near future. Their R&D spending, 174,9 billion US\$ in PPP, is now second behind the US (427,2), ahead of Japan (152,1), at more than half Europe (326,7)<sup>10</sup>. Battelle estimates that the growth should be 2,1% in the USA, 3,5% in Europe, and around 9% in Asia, while it is around 20% in China. For the telecom equipment, Huawei is now 39th worldwide for its R&D spending<sup>11</sup>. For the supercomputers, China is number 2, and in 2010, it produced the most powerful computer of the world, Tianhe-1A, with 2566 Tflops<sup>12</sup>. The cloud industry is also in exponential development.

While Europe is essentially absent of this stage, it is clear that China aims at becoming a global leader in the area of Big Data, which will constitute one of the economic and political challenges of the 21<sup>st</sup> century. I will demonstrate this claim with concrete examples taken from the Chinese industry and show their very large potential.

---

<sup>4</sup>J. Ginsberg et al., *Detecting influenza epidemics using search engine query data*, Nature 457, February 2009

<sup>5</sup>Steve Lohr, *New Ways to Exploit Raw Data May Bring Surge of Innovation*, The New York Times, may 13, 2011.

<sup>6</sup>Big data: The next frontier for innovation, competition and productivity, McKinsey Global Institute, May 2011, [http://www.mckinsey.com/mgi/publications/big\\_data/index.asp](http://www.mckinsey.com/mgi/publications/big_data/index.asp)

<sup>7</sup>The Data Deluge, The Economist, Feb 25th 2010

<sup>8</sup>Alexa, the Web Information Company, <http://www.alexa.com/>

<sup>9</sup>China social media equivalents: a new infographic <http://www.asiadigitalmap.com/2011/02/china-social-media-equivalents-a-new-infographic/>

<sup>10</sup>2012 Global R&D Funding Forecast, Battelle and R&D Magazine, december 2011, <http://www.battelle.org/ABOUTUS/rd/2012.pdf>

<sup>11</sup>2011 EU Industrial R&D Investment Scoreboard

<sup>12</sup>[www.top500.org](http://www.top500.org)