Global perspective on the information society

I. Europe at the periphery of the information society?

II Information society in China, the Beijing consensus?

May 14, 2013

Stéphane Grumbach
INRIA

Seminar for the Council of the European Union

April 17, 2013

Europe at the periphery of the information society?

Discrepancy between the

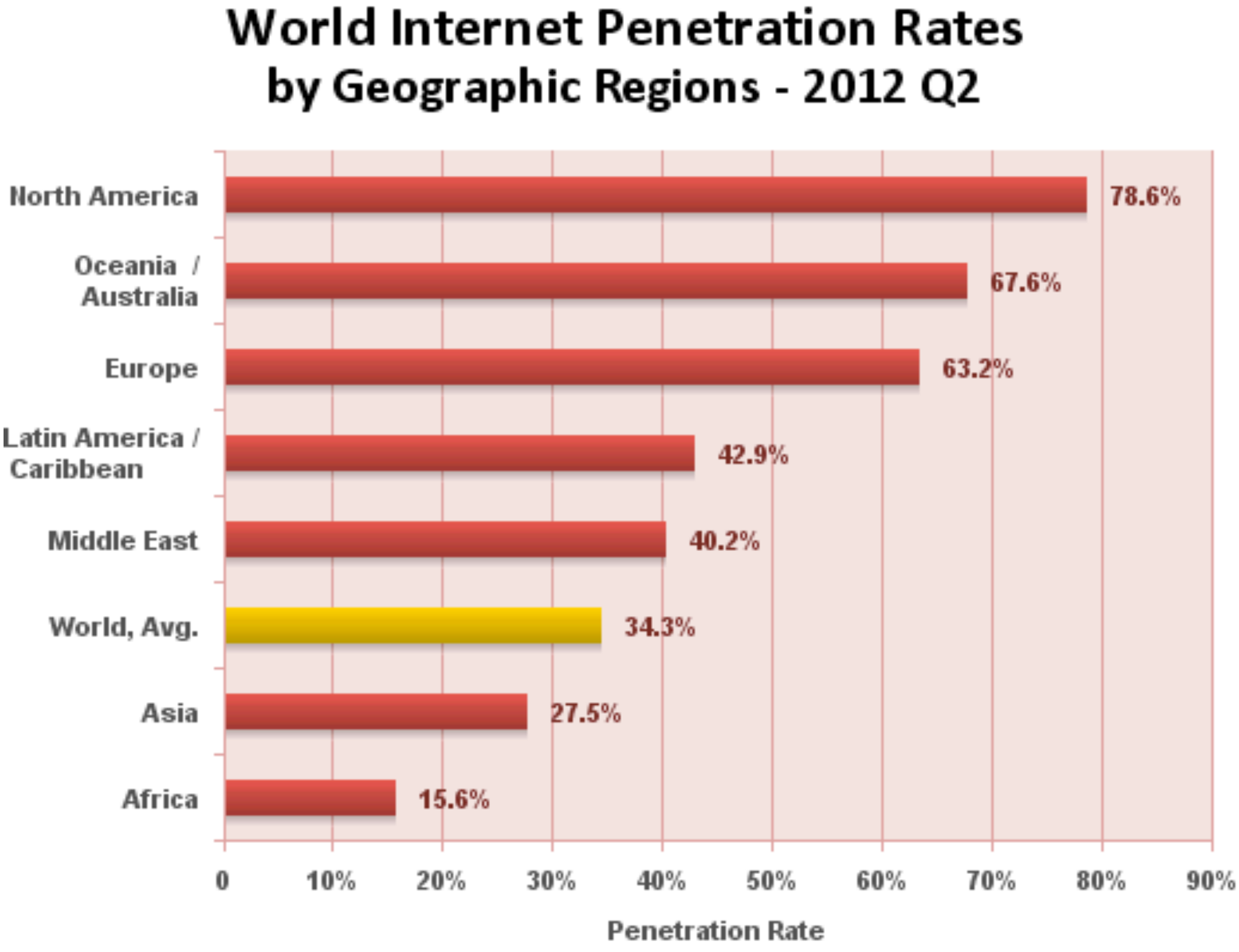
importance of Europe

cultural, economical, political, ...

its weak influence in the information society

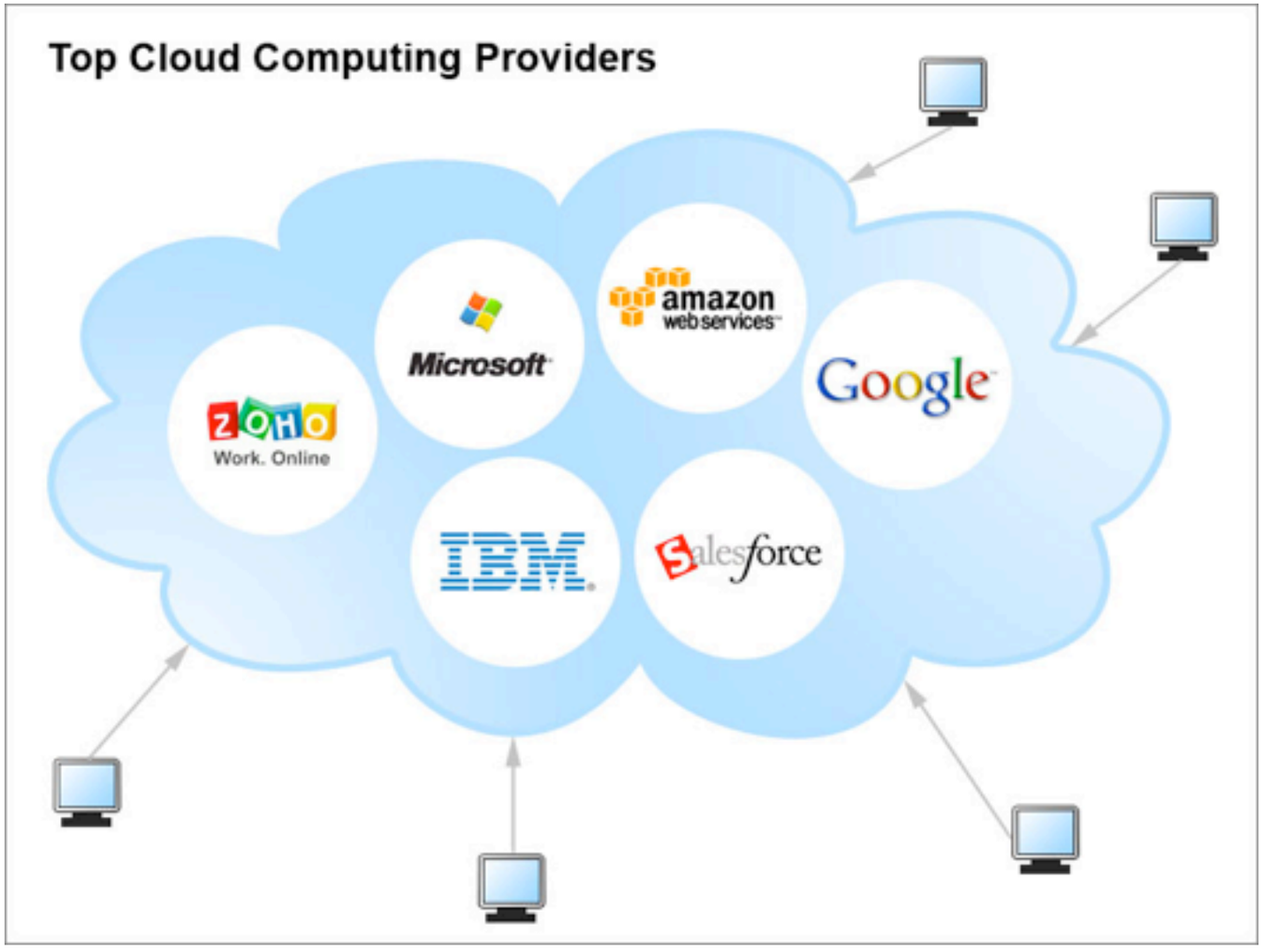
materials, systems, services, ...

Large penetration in Europe



Source: Internet World Stats - www.internetworldststs.com/stats.htm

supported by non European devices and operating systems



http://talkcloudcomputing.com/cloud-service-providers-compete-to-capture-the-cloud-market/

carrying American services

with a shallow presence on the Web

E.G. Belgium

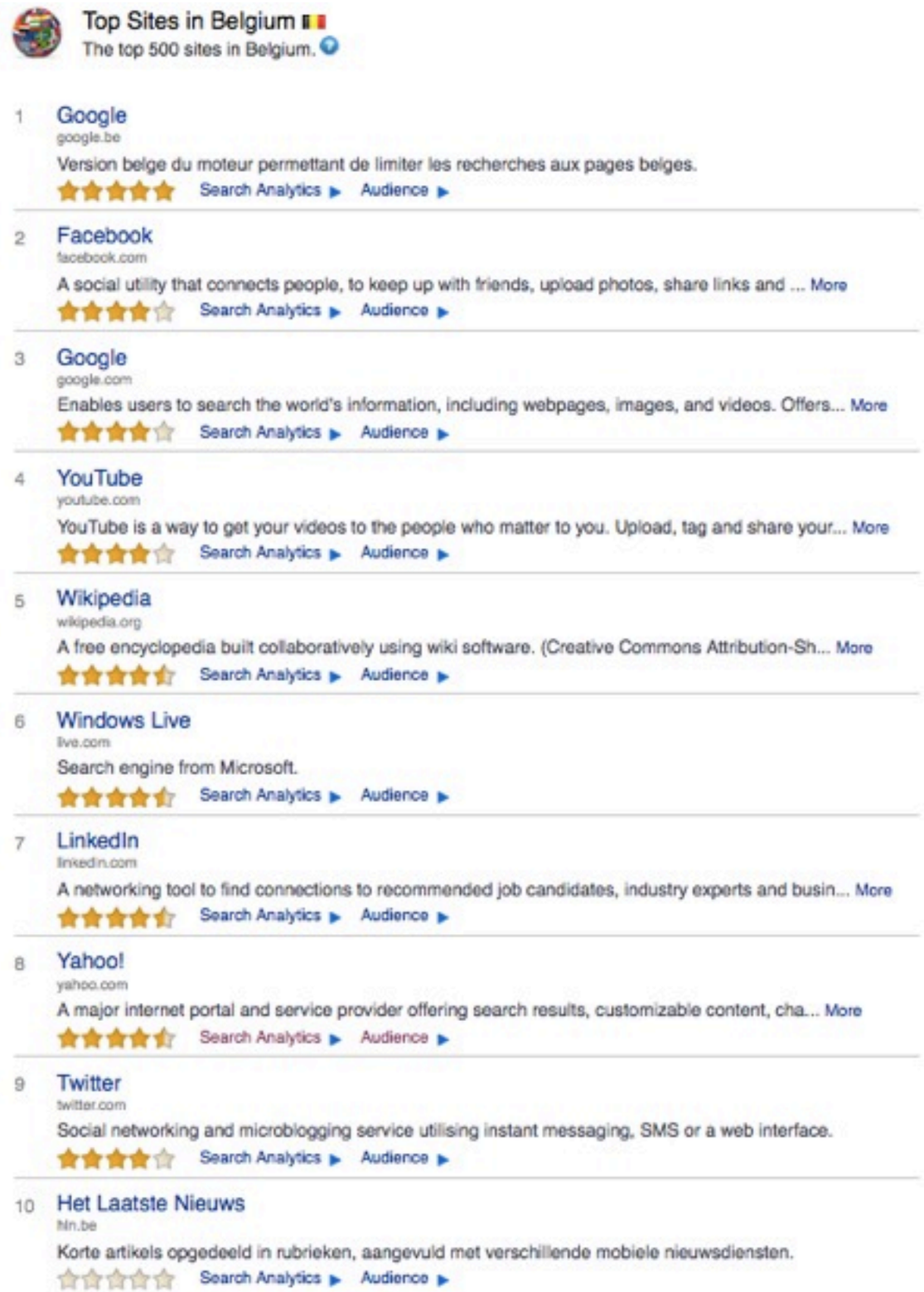
Local Web sites

among

top 10

used in Belgium

10%



while in other countries...

Percentage of national web corporations

among top 25 by country

- USA: 100% (no foreign site)
- China: 92% (only Google makes it in the top 25)
- Russia: 68% (others main US sites)

The European Internet Paradox

Why is Europe's influence so weak in the Info Soc ?

Does it matter for our future?

I Europe at the periphery of the Information Society

II Information society in China, the Beijing consensus?

May 14, 2013

Cartography of Asia's Internet
China's Web giants
Designed by China, an R&D momentum
What consensus on the Internet?

Chapters

I The age of information

II A Web of Services

III The Economy of Internet Data

IV Cartography of data flows

V Europe at the periphery

Chapters

I The age of information

II A Web of Services

III The Economy of Internet Data

IV Cartography of data flows

V Europe at the periphery

Continuous growth and complexification of information / knowledge

from writing and printing

to the Digital Revolution

Massive growth of information during the Industrial Revolutions

Complex financial instruments and banking

Industrialization, technological advances

Trade and transportation at a global scale

Intellectual property rights

Education level

Primary -> Secondary

Secondary -> Tertiary

Tertiary -> Quaternary

Information in mechanics

from direct

mechanical control

to

electronic control

steering wheel -> information

information -> control direction / vehicle balance, ...

-> many more applications (e.g. driver control)

Although useless, steering wheels are still in use (conservatism)

Digital Revolution

Turn 20th-21st century

digitalization, modeling

communication, social networking

"*Every two days we create as much information as we did up to 2003*"

Eric Schmidt

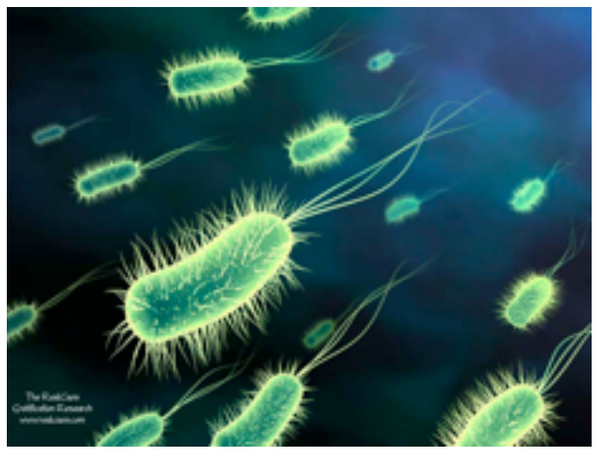
Information in natural sciences

from

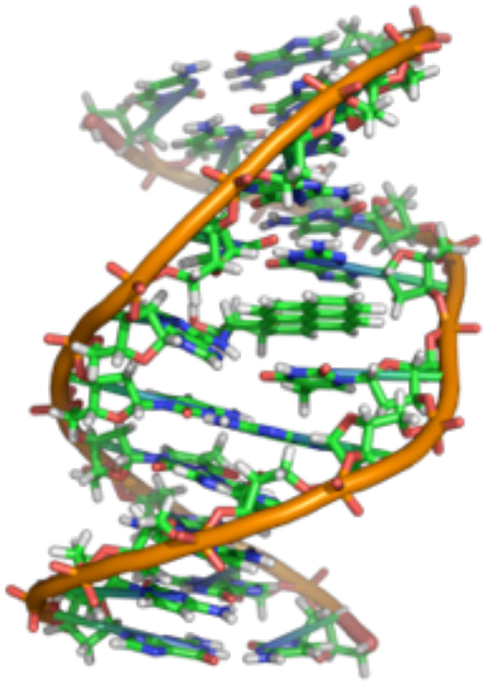
biological reactions

to

information processing



Escherichia coli



Sensors detecting chemicals attractant and repellent
Cells move after processing information from sensors

Information processing in cells: genes, protein mechanisms, dynamic adaptation to the environment

Computation: natural vs artificial

The emergence of a virtual world

Everything transformed into digital data

my position, my body

what I see, hear, smell, feel, think,...

Full imbrication of

the virtual world and

the real world

Transformation

physical to digital

digital to physical (3d printers)

Google

The intuition of a virtual world

…In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. . .

On Exactitude in Science
Jorge Luis Borges, 1946

Chapters

I The age of information

II A Web of Services

III The Economy of Internet Data

IV Cartography of data flows

V Europe at the periphery

Ranking Web pages: search engines

Complex ranking algorithm of the Web

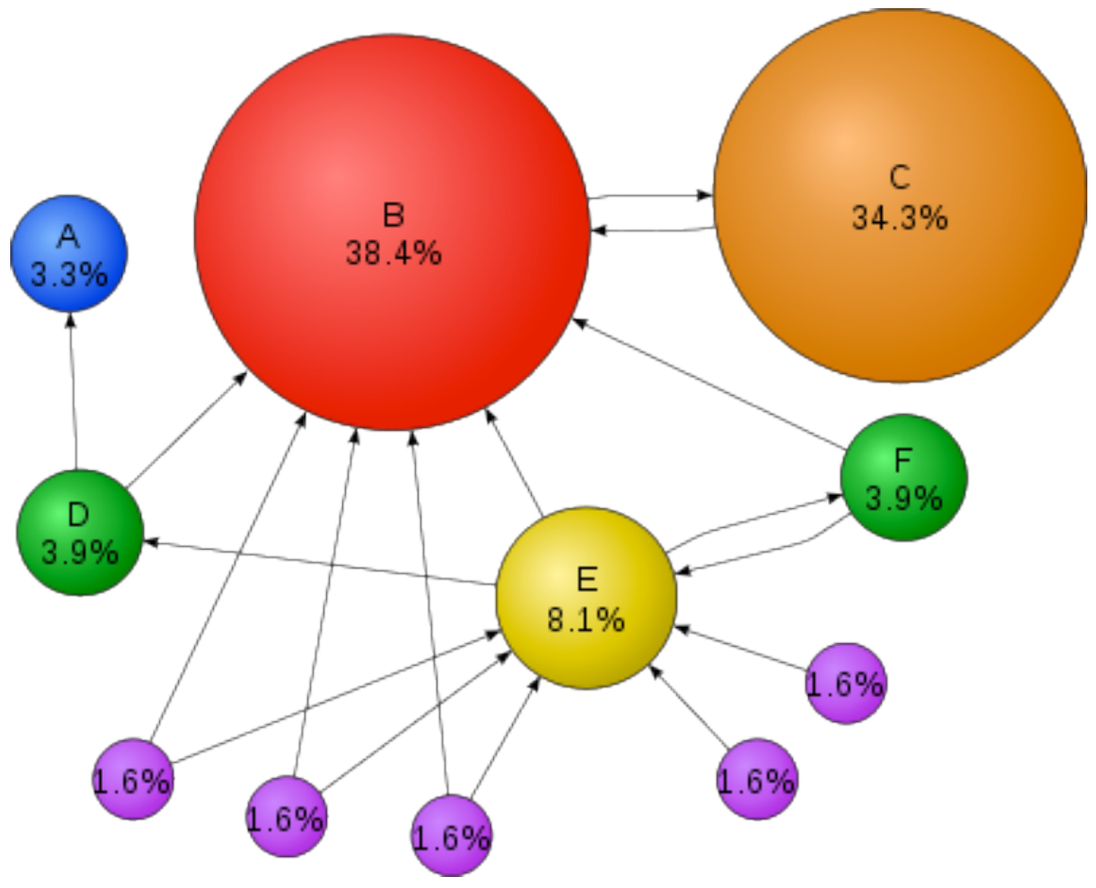
Google ranks ca 50 billion Web pages

in 150 languages

using several hundreds criteria

Answers a billion queries a day

The first service used online



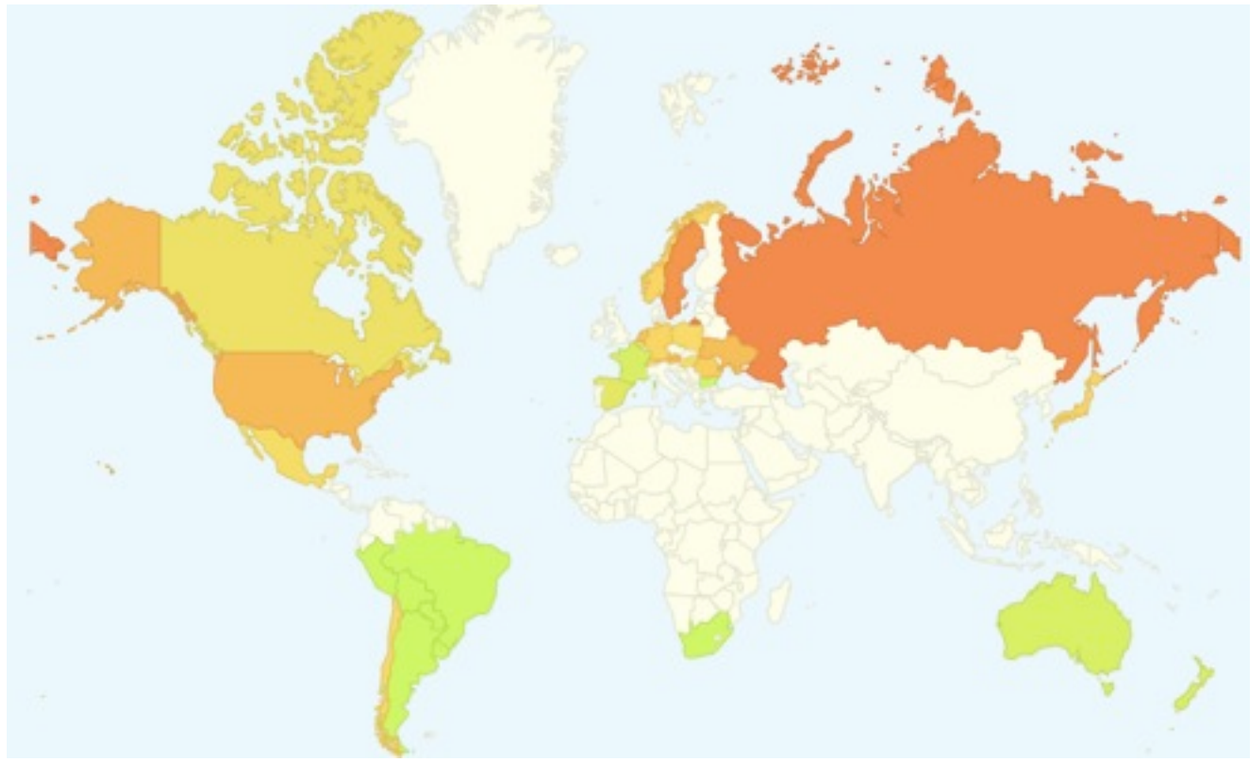
The search engine: knowledge creation

Business model of Google: ad technology

The list of queries of a user => profile of the user

Knowledge on populations

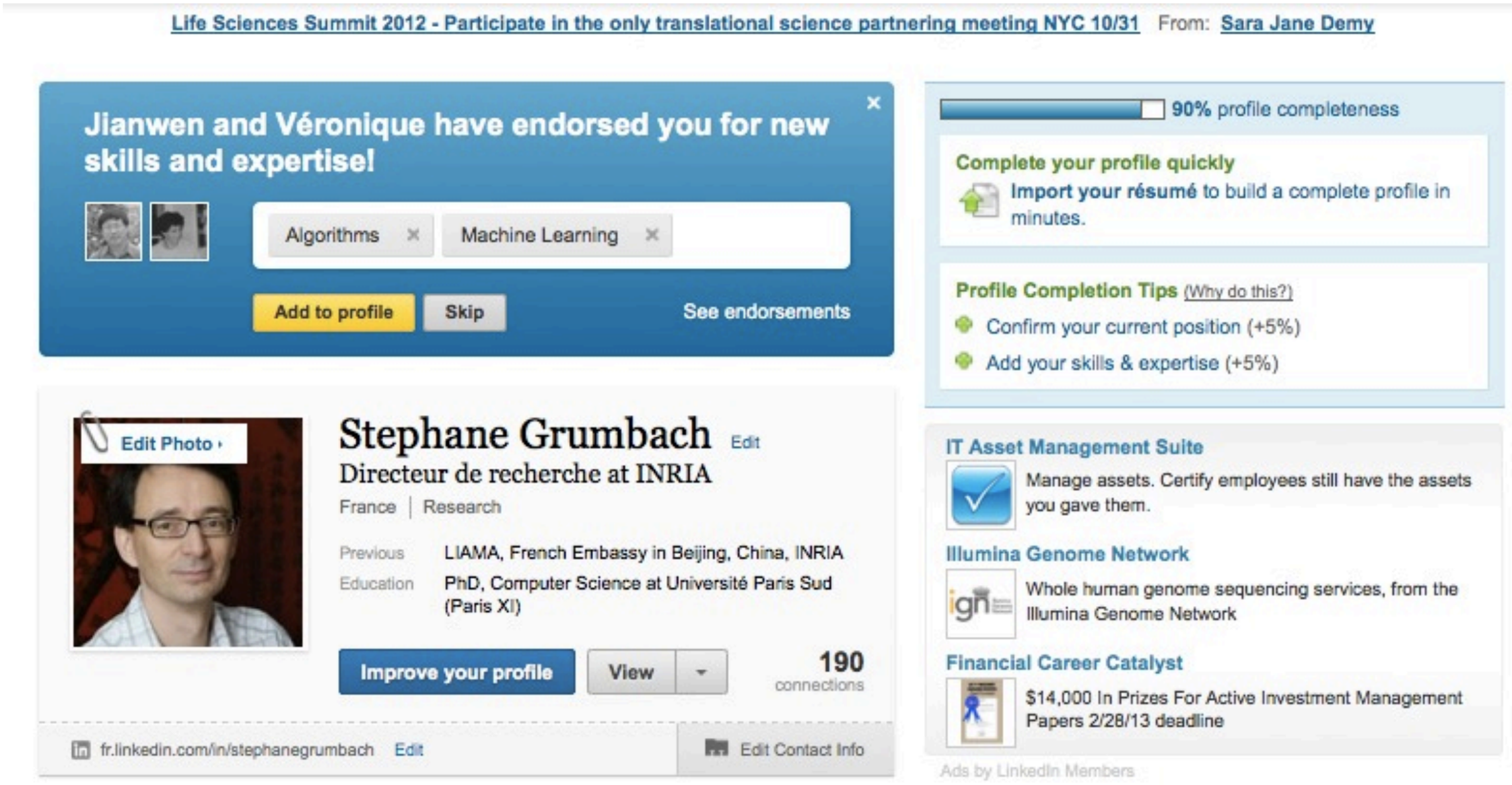
Google Flu
monitoring related queries



Crowdsourcing

collaborative mapping

Social networks



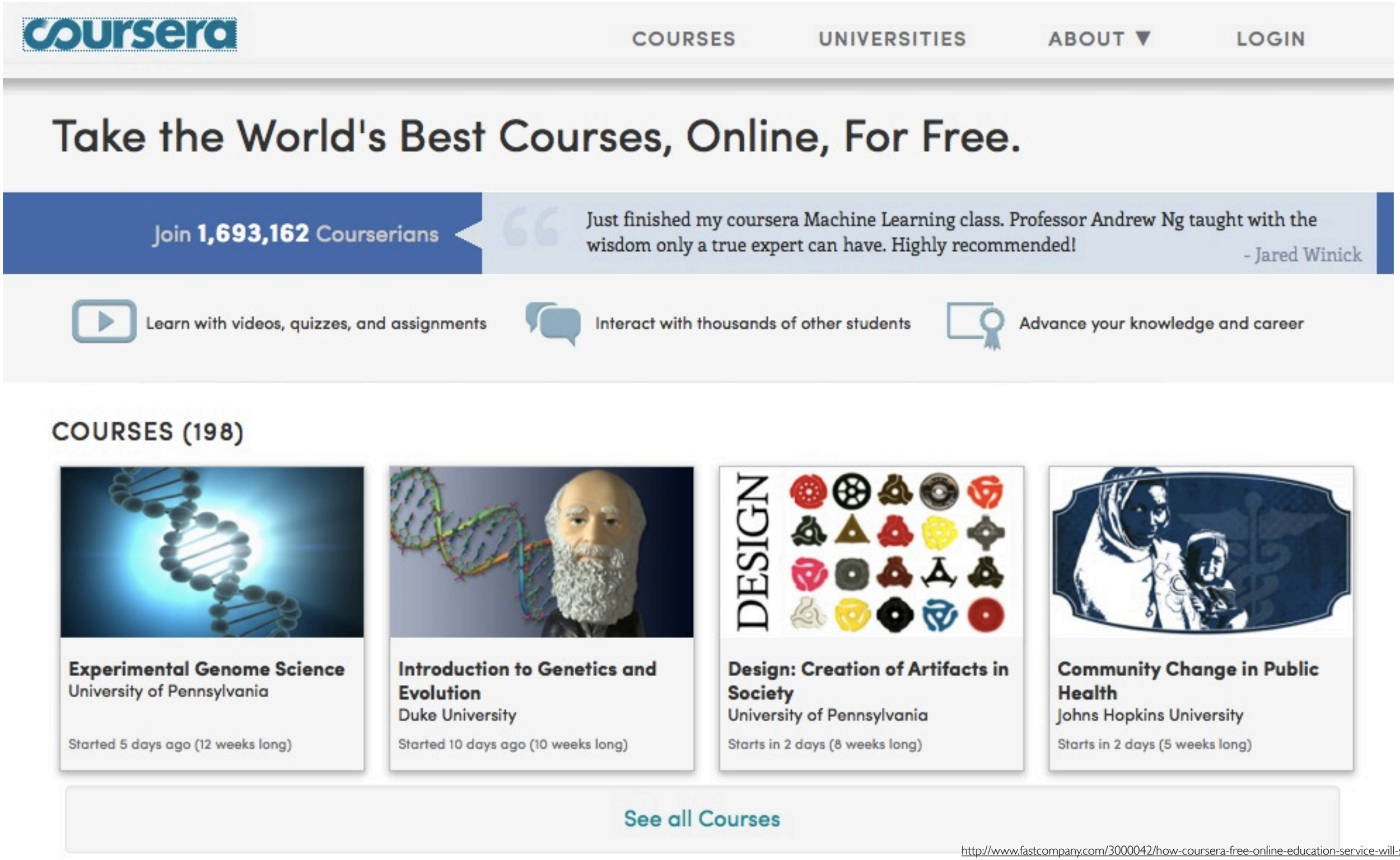
Premium services

Targeted Ads

Profiling for recruitment

http://www.fastcompany.com/3000042/how-coursera-free-online-education-service-will-scho

Massive Open Online Courses



http://www.fastcompany.com/3000042/how-coursera-free-online-education-service-will-scho

Mobile services

Relying on a few mobile operating systems

- iOS,
- Android

Library of Apps

- very knowledgeable
- personal assistant
- location based
- banking, ...

Google now

http://techcubeconsulting.com/5-great-free-mobile-apps-for-businesses/

Facebook? a new world!

The social platform

- 1 billion users
- 130 billions friend links
- 2.45 billion piece of content shared daily
- 350 millions pictures uploaded daily

Much more than a social network:

- storage, authentication, communication
- millions of Apps developers on the API
- 9 million Apps active

Towards “apperating systems”

From personal computers to dematerialized environment

Single interface
user / online environment

Complete data flow control

Facebook Home on top of Android

http://www.wired.com/gadgetlab/2013/04/rise-of-apperating-systems/

Chapters

I The age of information

II A Web of Services

III The Economy of Internet Data

IV Cartography of data flows

V Europe at the periphery

Data: building blocks of the information society

Data deluge in all sectors

U.S. Library of Congress:
235 Terabytes of data

Walmart: 2.5 petabytes of data,
1 million customer transactions / hour

Facebook: 30 Petabytes of user data

Google: processing 20 petabytes a day (2008)

World: 5 billion people calling,
tweeting, browsing on mobile phones

kilo 10^3
mega 10^6
giga 10^9
tera 10^{12}
peta 10^{15}
exa 10^{18}

The digital universe

2.7 zettabytes in 2012

Exponential increase

doubles every other year

with the capacity to store,
compute, and communicate

kilo 10^3
mega 10^6
giga 10^9
tera 10^{12}
peta 10^{15}
exa 10^{18}
zetta 10^{21}
yotta 10^{24}

35 zettabytes in 2020

The physical infrastructure

Data centers

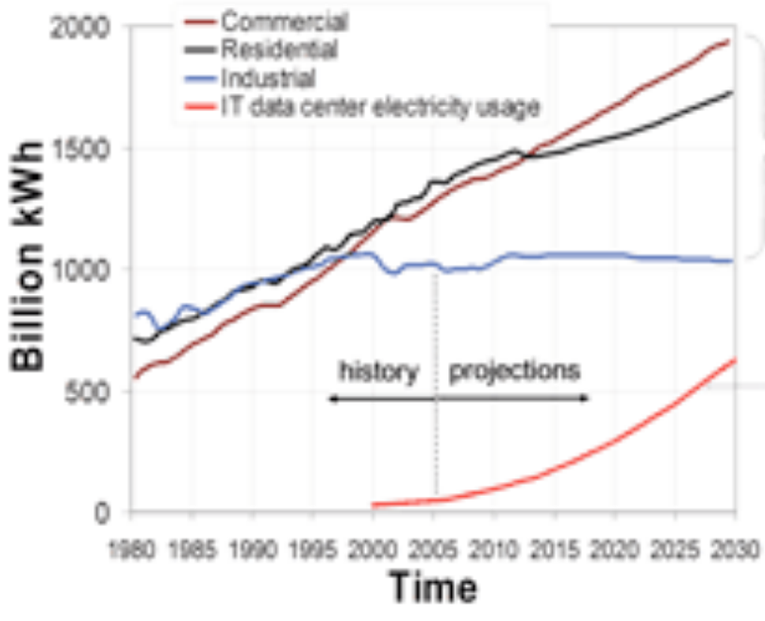
The physical infrastructure

Data centers

Energy consumption: 250 TWh in 2010

about half of electricity consumption in France

about 30 nuclear plants

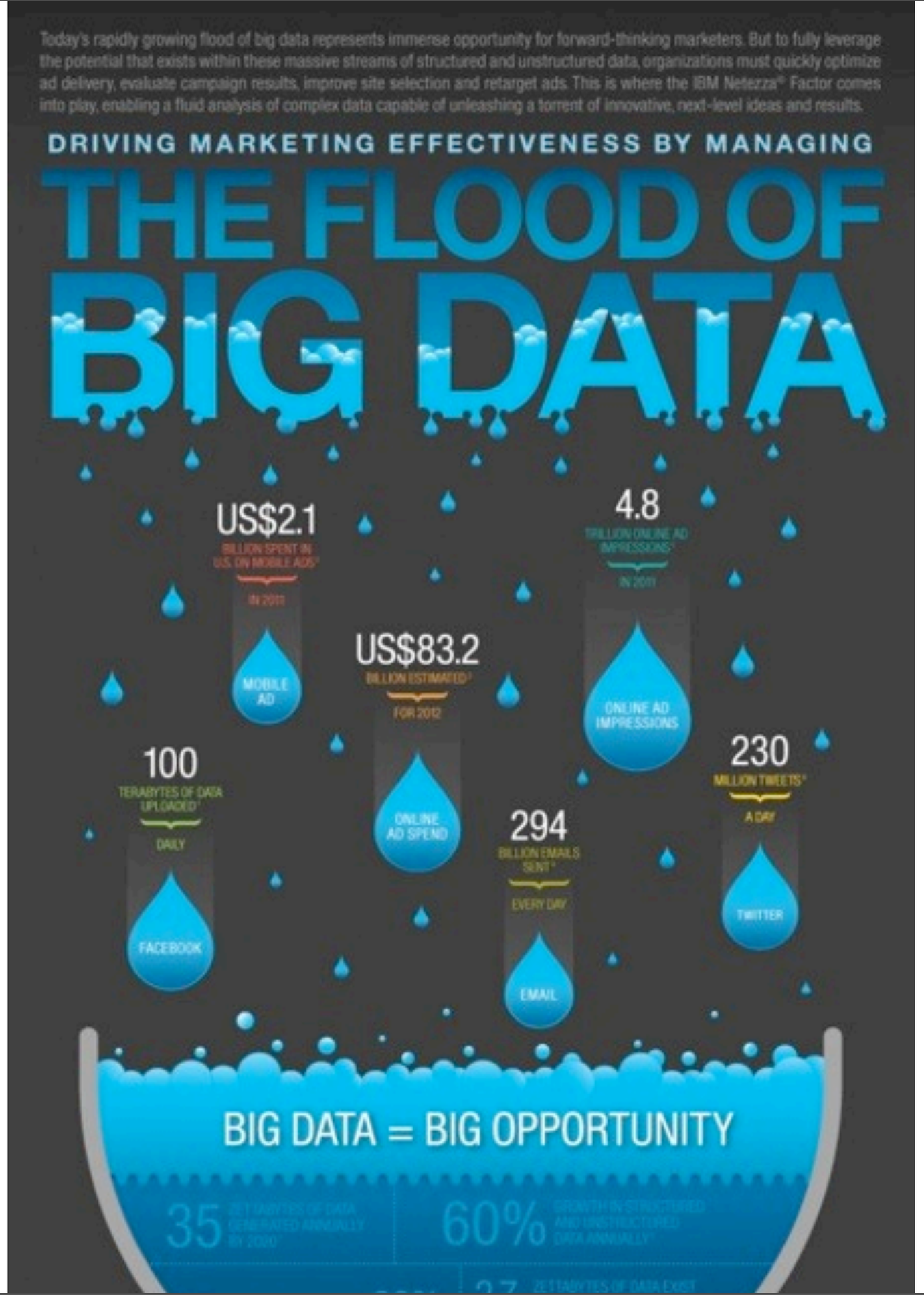


http://www.nytimes.com/2012/09/23/technology/data-centers-waste-vast-amounts-of-energy-belying-industry-image.html?_r=1&pagewanted=all
http://www.electronics-cooling.com/2010/12/energy-consumption-of-information-technology-data-centers/

Today's rapidly growing flood of big data represents immense opportunity for forward-thinking marketers. But to fully leverage the potential that exists within these massive streams of structured and unstructured data, organizations must quickly optimize ad delivery, evaluate campaign results, improve site selection and retarget ads. This is where the IBM Netezza® Factor comes into play, enabling a fluid analysis of complex data capable of unleashing a torrent of innovative, next-level ideas and results.

DRIVING MARKETING EFFECTIVENESS BY MANAGING

THE FLOOD OF BIG DATA



GROWTH
IN UNSTRUCTURED DATA*

5 EXABYTES
OF DATA GENERATED
EVERY TWO DAYS*

CAPITALIZING ON THIS OPPORTUNITY WILL REQUIRE:

RULES-DRIVEN INTEGRATION OF DISPARATE DATA

IMPROVED OPERATING INFRASTRUCTURES

NETWORK OF DATA-CENTRIC TECHNOLOGY AND PARTNERS

MARKETING DATA GOVERNANCE

MARKETERS CAN THEN OPTIMIZE IN FOUR PRIMARY WAYS:

AUDIENCE OPTIMIZATION

Identify high-potential audiences and accurately target them

CONTENT OPTIMIZATION

Enable the right message at the right time via the right content targeting

YIELD OPTIMIZATION

Maximize ad inventory by identifying high-value audiences across publisher properties

CHANNEL OPTIMIZATION

Optimize ad media purchase, understanding value of channels higher up in the funnel

NETEZZA FACTOR

NETEZZA FACTOR

NETEZZA FACTOR

NETEZZA FACTOR

20%
INCREASE IN REVENUE

48-hour
REDUCTION IN PROCESSING TIME

13 billion
IMPRESSIONS PER DAY

50%
DECREASE IN RUN TIME

Epsilon

Epsilon leveraged the IBM Netezza data warehouse appliance to simplify advanced analytics on massive data volumes, increasing retention and revenues by 20 percent.

Kelley Blue Book

KBB used the IBM Netezza data warehouse appliance to personalize content, improving conversion rates and reducing model processing time from 3 days to 1.

MediaMath

MediaMath used the IBM Netezza data warehouse appliance to measure 13 billion ad impressions a day, requiring half the manpower to deliver 10x the output.

Merkle

Merkle used IBM Netezza advanced analytics software to achieve a 50 percent decrease in end-to-end run time for marketing campaign execution.

NETEZZA

3 days
TO GET IT UP AND RUNNING

212%
ROI WITH FIVE-MONTH PAYBACK

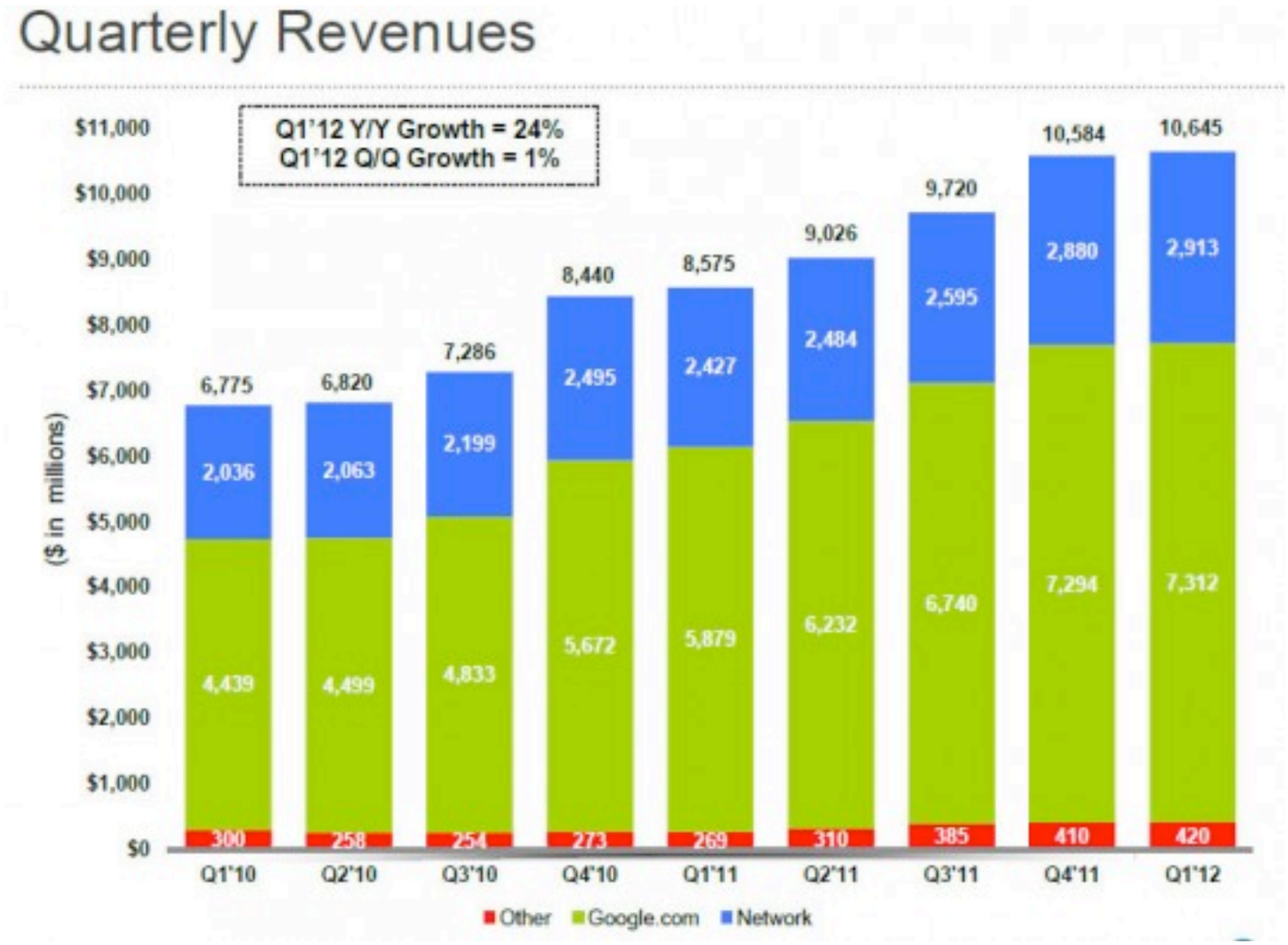
1
DATABASE ADMINISTRATOR REQUIRED

IBM

IBM, the IBM logo, and ibm.com are trademarks of IBM, registered in many jurisdictions worldwide. A current list of IBM trademarks is available at ibm.com/legal/copytrade.shtml. Netezza is a trademark or registered trademark of IBM International Group B.V., an IBM Company. Other company, product or service names may be trademarks or service marks of others. © Copyright IBM Corporation 2013.

“Google is a Vacuum cleaner for revenue”

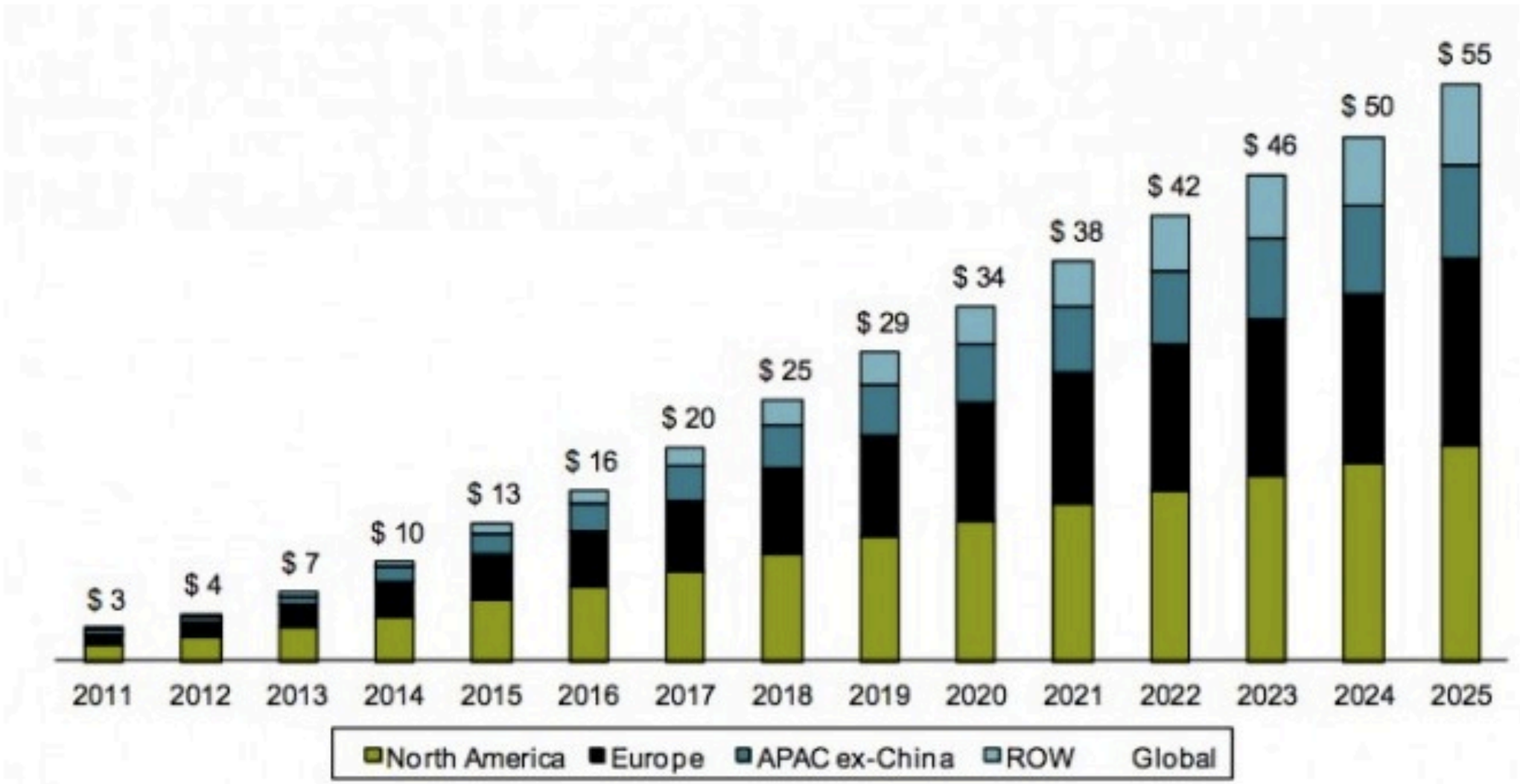
Barry Diller



$50 billion in 2012

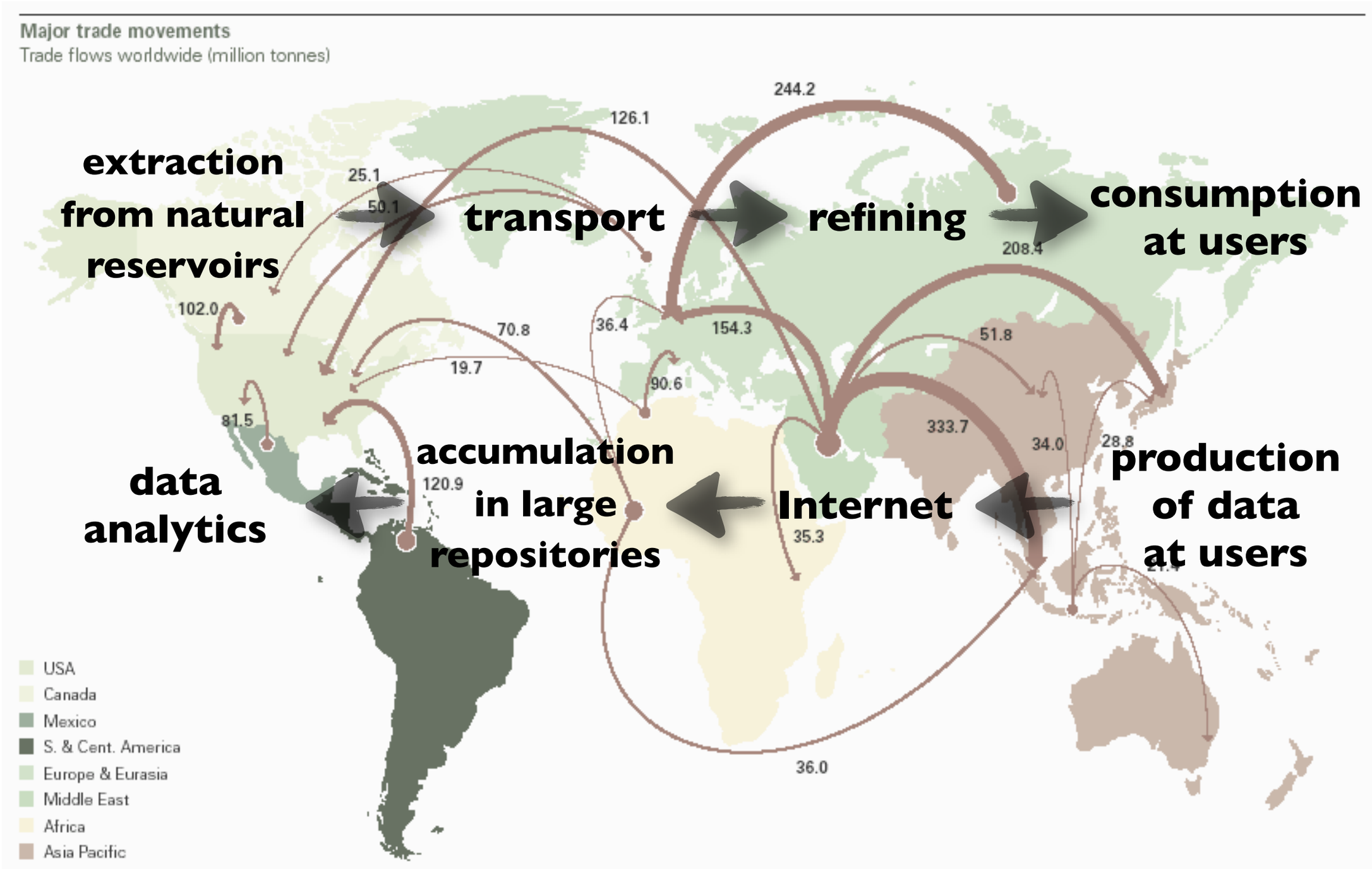
http://androidcommunity.com/google-posts-10-6-billion-in-revenue-for-q1-2012-20120412/
http://video.cnbc.com/gallery/?video=3000118210

Facebook revenue forecast



http://www.wired.com/business/2012/10/facebook-case-for-optimism/?utm_source=twitter&utm_medium=socialmedia&utm_campaign=twitterclickth

Data: raw material of the 21st century
much like crude oil



The challenges of the industry

First challenge:

capture users and data

scale up as much as possible

Second challenge:

extract value / knowledge

stay as open as possible

Third challenge:

protect data and privacy

Size matters exponentially

number of users of a search engine

=> traffic

=> interest of advertisers

=> word auctions

=> relevance (because of price)

=> probability of successful click

=> word covering

=> monetization covering

Thanks to François Bourdoncle

http://www.wordstream.com/blog/ws/2011/07/18/most-expensive-google-adwords-keywords

Data: new rules of the game

Data is a raw material,
to be transformed into value/information

Data is a money
"free" paradigm of the Web 2.0

Data can be duplicated at will
and is to ensure quality of service

Data can be transformed by people everywhere
Crowdsourcing

Internet giants as Extraterritorial powers

No real binding to the place of operation
Regulation, taxation: optimal use of national differences

Own raw material resources and industry
harvested without borders

Own legal systems
contracts users/corporations

Own monetary systems
emerging virtual currencies

http://www.nytimes.com/2013/04/08/business/media/bubble-or-no-virtual-bitcoins-show-real-worth.html?nl=technology&emc=edit_tu_20130408

Chapters

I The age of information

II A Web of Services

III The Economy of Internet Data

IV Cartography of data flows

V Europe at the periphery

Where are the data?

Huge concentration of data

85% of data handled by (large) corporations

Virtualization/dematerialization of infrastructures

Social networks, Cloud, ...

Most of the prominent corporations based in the USA

Google, Facebook, Amazon, Twitter, ...

Storage capacity of Europe = 70% USA [McKinsey 2011]

Data from the Web 2.0

produced by users everywhere in the world

but accumulated by corporations most often abroad

Percentage of national web corporations among top 25 by country

- USA: 100%
- China: 92% (only Google makes it in the top 25)
- France: 36% (but mostly marginal sites, not data intensive)

  leboncoin, Orange, Free, commentcamarche, lemonde, lequipe, lefigaro, pagesjaunes, sfr

Alexa.com

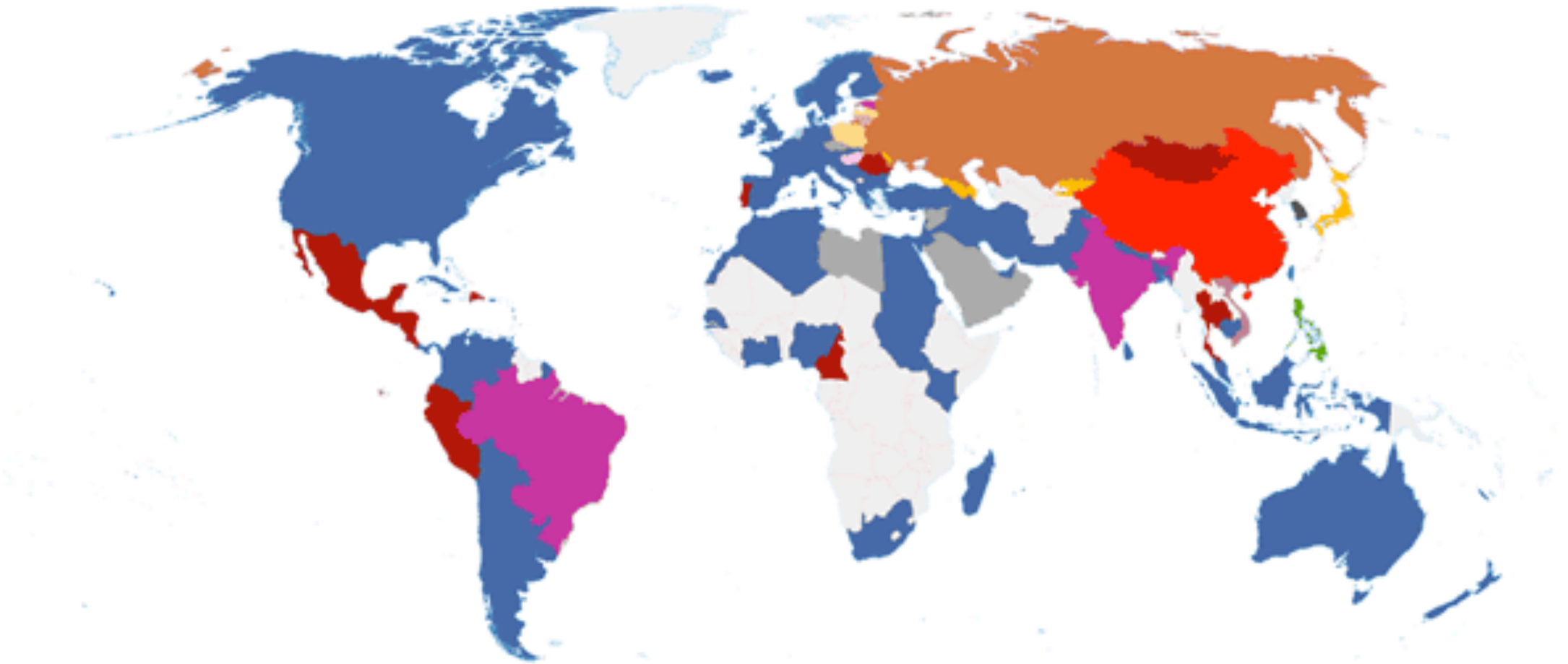
The Top 50 websites worldwide

- USA: 72 %
- China: 16 % (Baidu: 5; QQ: 8; Taobao: 13; Sina:17; 163: 28; Soso:29; Sina weibo:31; Sohu:43)
- Russia: 6 % (Yandex: 21; kontakte:30; Mail: 33;)
- Israel: 2 % (Babylon: 22)
- UK: 2 % (BBC: 46)
- Netherlands: 2 % (AVG: 47)

Facebook's territory

WORLD MAP OF SOCIAL NETWORKS

June 2009



credits: Vincenzo Cosenza www.vincos.it license: CC-BY-NC sources: Google Trends for Websites/Alexa

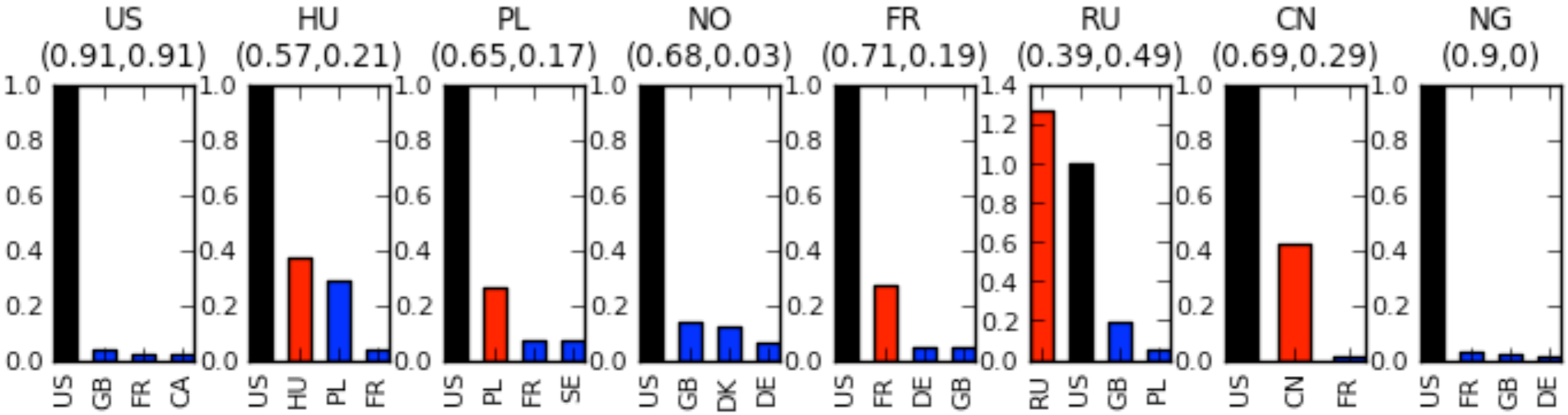
Diversity of search engines

- USA: Google: 65 % ; Bing: 15% ; Yahoo: 15%
- China: Baidu: 78% ; Google: 16%
- Russia: Yandex: 60% ; Google: 25%
- UK: Google: 91 % ; Bing: 5%
- France: Google: 92 % ; Bing: 3%

In France,

- Google has a de facto monopoly
- Google knows more about France than INSEE

Global tracking



Proportion of trackers in different countries

Joint work with Claude Castellucia & Lukasz Olejnik

Chapters

I The age of information

II A Web of Services

III The Economy of Internet Data

IV Cartography of data flows

V Europe at the periphery

What's at stake in Europe?

A general fear of new technologies

An inefficient industrial/innovation environment
no strong corporation emerging in IT

A strong suspicion with respect to personal data
very high concern for privacy protection

But carefree growing dependence on foreign systems

A disappointing path with IT

Failure to catch up with computing, networking, and digital technologies

in the 1970‘s with computers, chips,

in the 1990’s with the Web and its services

in the 2000’s with the information industry

in the 2010’s with the radical societal shift?

Although many concepts were invented in Europe

What personal data mean for Europeans?

http://www.germany.info/Vertretung/usa/en/__pr/P__Wash/Journalist__Trips/06__YJ__Germany__PR.html

What personal data mean for Europeans?

http://www.germany.info/Vertretung/usa/en/__pr/P__Wash/Journalist__Trips/06__YJ__Germany__PR.html

The digital precautionary principle

The European dream: allow systems with

a predefined service

using the minimal amount of data required for that service

The exact opposite of Facebook's approach

open to apps on private data with users consent

Can personal data be protected ? not clear to what extend

Shall systems be restrained ? not clear for which profit

IT is not yet another technology

IT will eventually controls everything in our economies / societies

commerce, transportation, production, energy, ...

education, public services, ...

IT forces to revisit our economic/monetary/fiscal systems

but technology evolves faster than regulation though

IT will eventually change political systems

online identity

political (democratic) processes

The risk of a strong dependency

Global imbalance

- information asymmetry (the knowledge is were the data is)

No impact on technological/societal choices

- Europe advocates “values”, but with no means to promote them
- Some regulations might not be implementable
- Technological changes are now much faster than societal changes

Underdevelopment

- Dependency on foreign powers for fundamental utilities

Loss of supremacy

- No full control of laws, currencies, etc.

Conclusion

Europe
is at
the periphery
of
the information society

Eventually
politics
will take place
on the net

Eventually
data
will surpass
crude oil
in importance