Priority Theme 2: Social Intelligence and e-Participation

Marko Grobelink, Jozef Stefan Institute

Contributors: Philipp Cimiano, Rob Gaizauskas, Marko Grobelink, Bernardo Magnini, Stephan Oepen, Rudi Studer, Ido Dagan

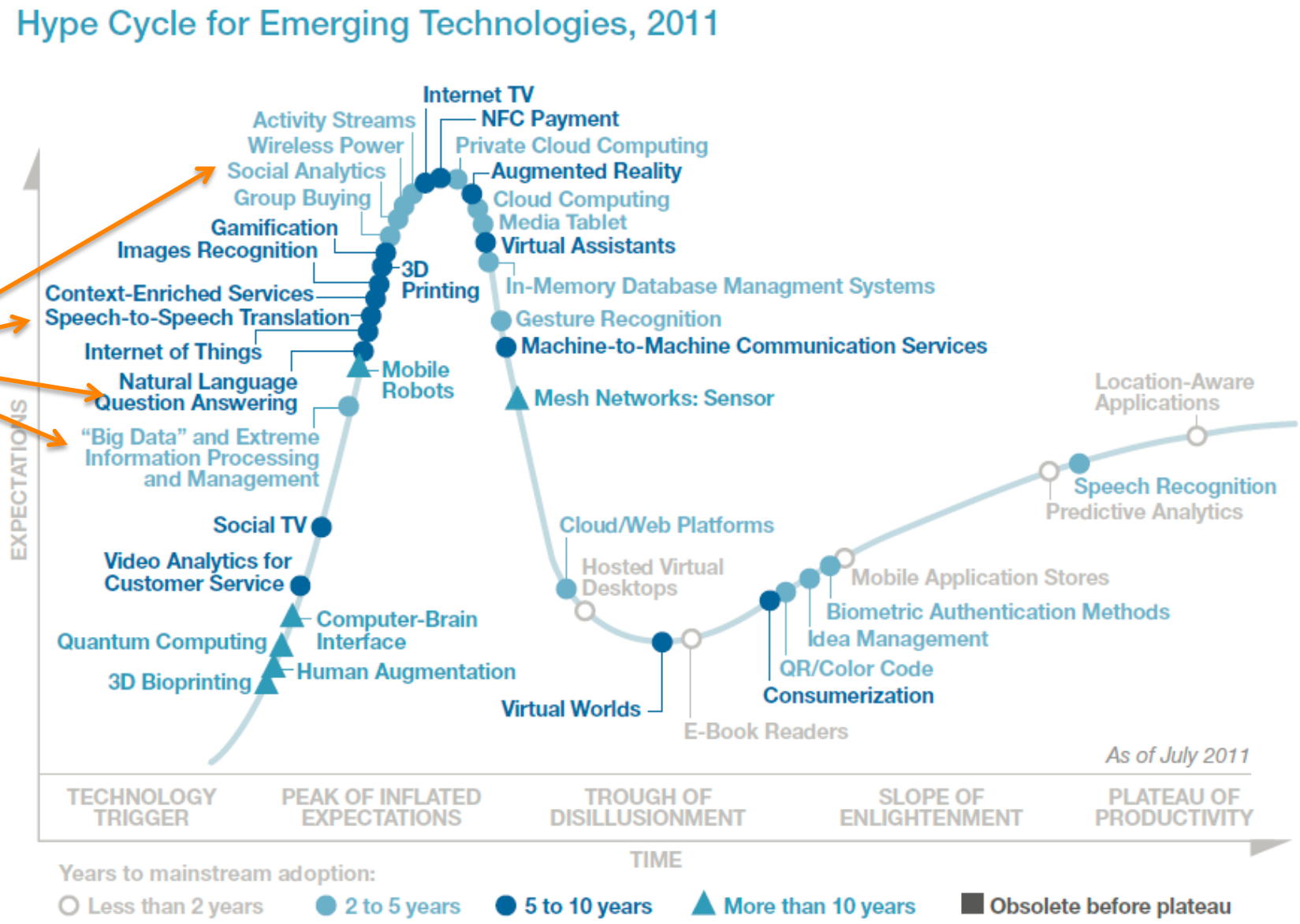
META FORUM 2012 A Strategy for Multilingual Europe
Brussels, Belgium, June 20/21

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

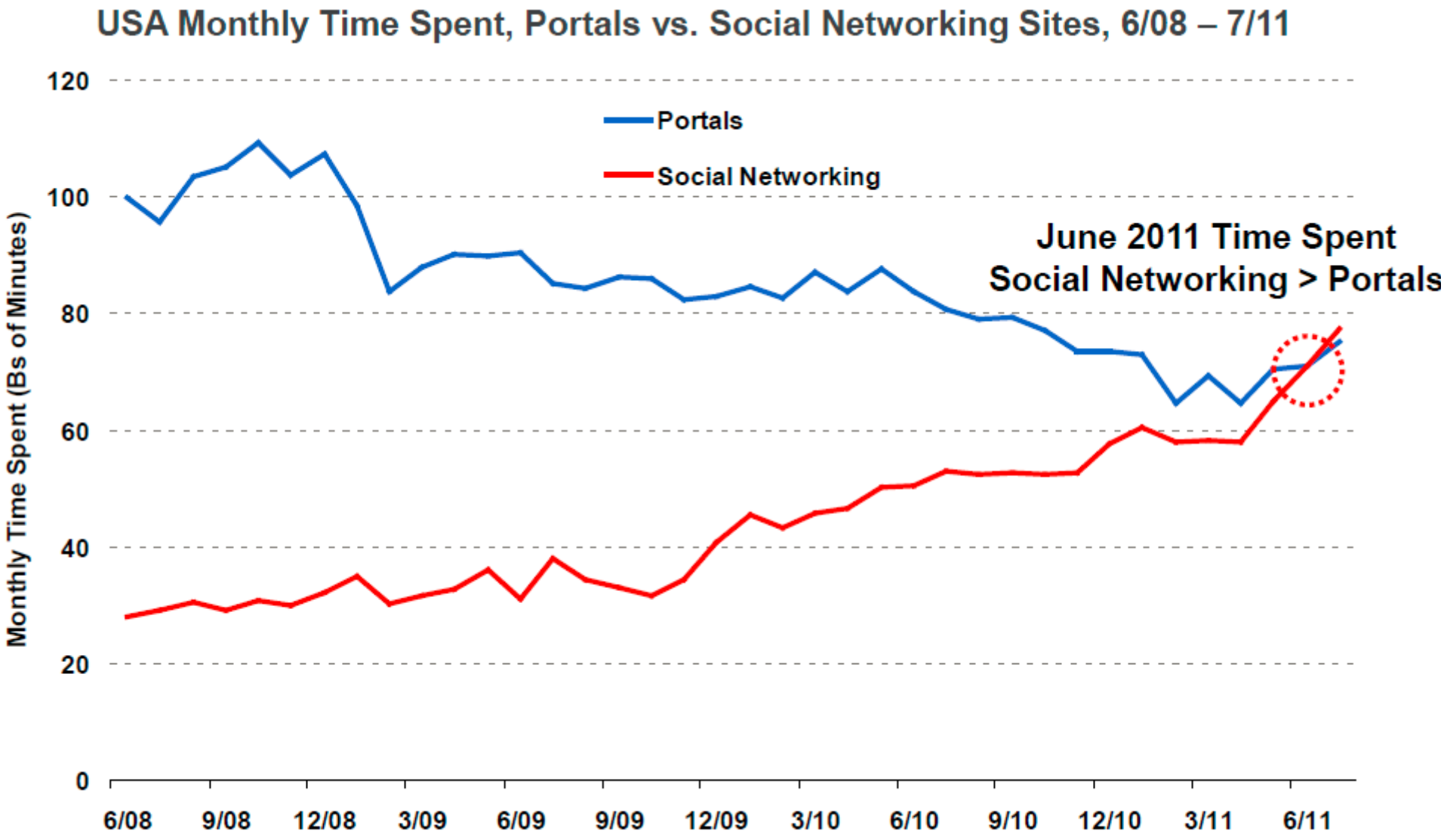
Setting-up the stage
trends and facts to be considered

Big-Data in Gartner Hype-Cycle 2011

Language Related Topics



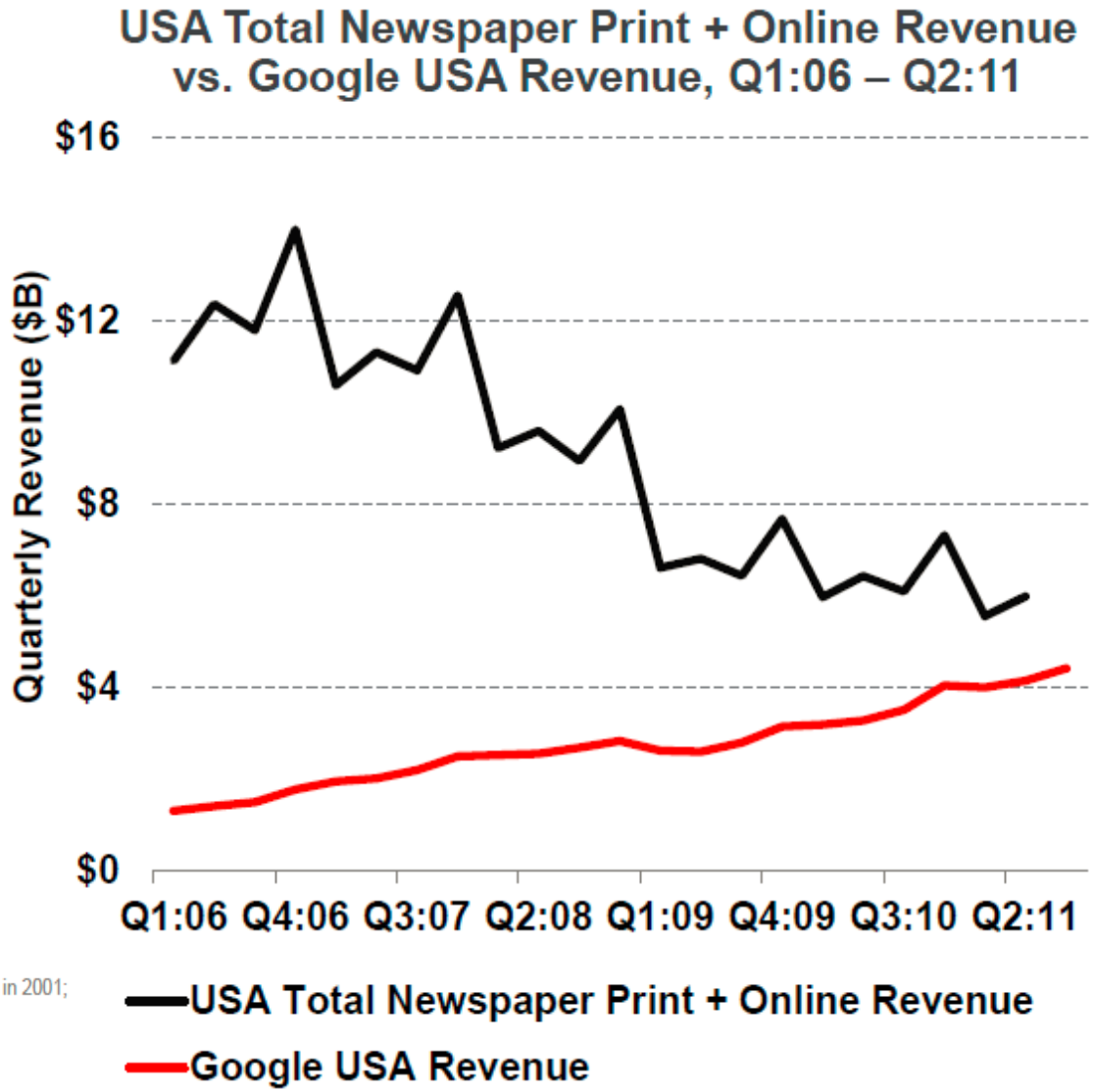
Social vs. Traditional Web



Source: comScore Media Metrix USA panel-only data.

Content consumption habits are changing rapidly

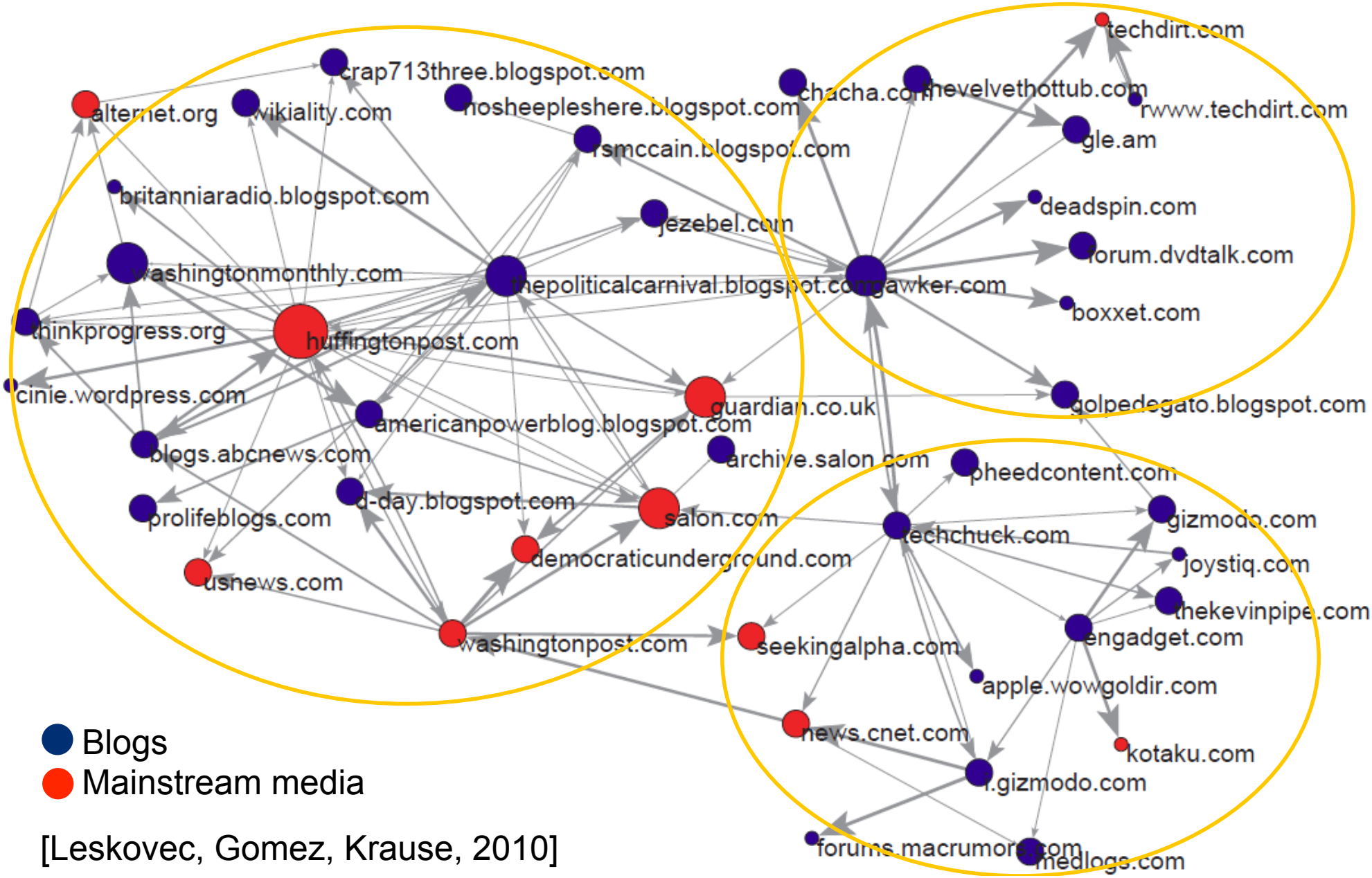
- Value is shifting from Creators to Aggregators
- Newspapers are experiencing 5th straight year of declining revenue.



Source: Demand Media.
Note that Search = ~50% of USA Online Ad Revenue ($15B based on Q2 run rate) in 2011 vs. <5% in 2001;
Display = 23% of Online Ad Revenue ($7B) in 2011 vs. 62% in 2001

Publishers ecosystem

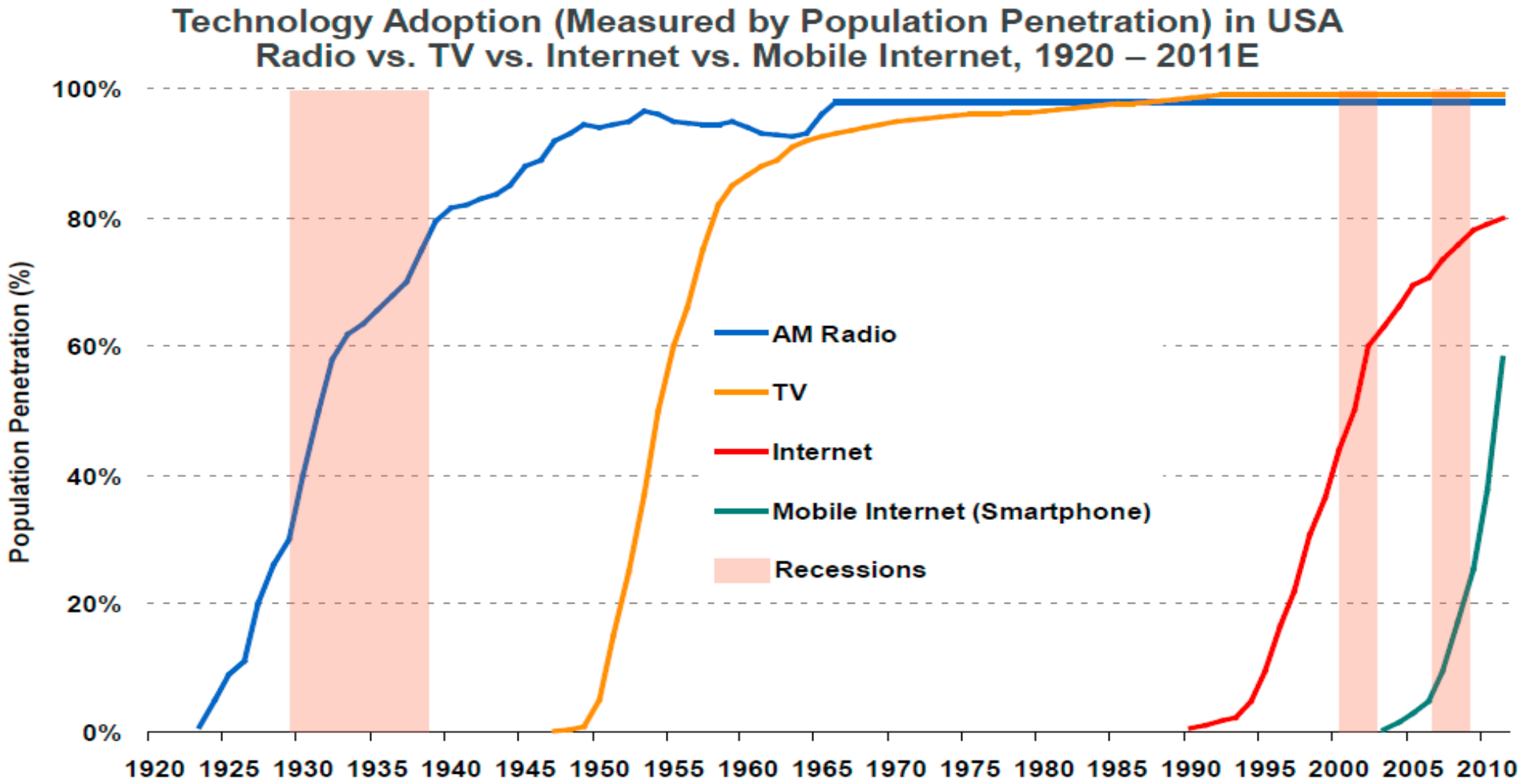
News diffusion across main-stream publishers and blogosphere



Blogs
Mainstream media

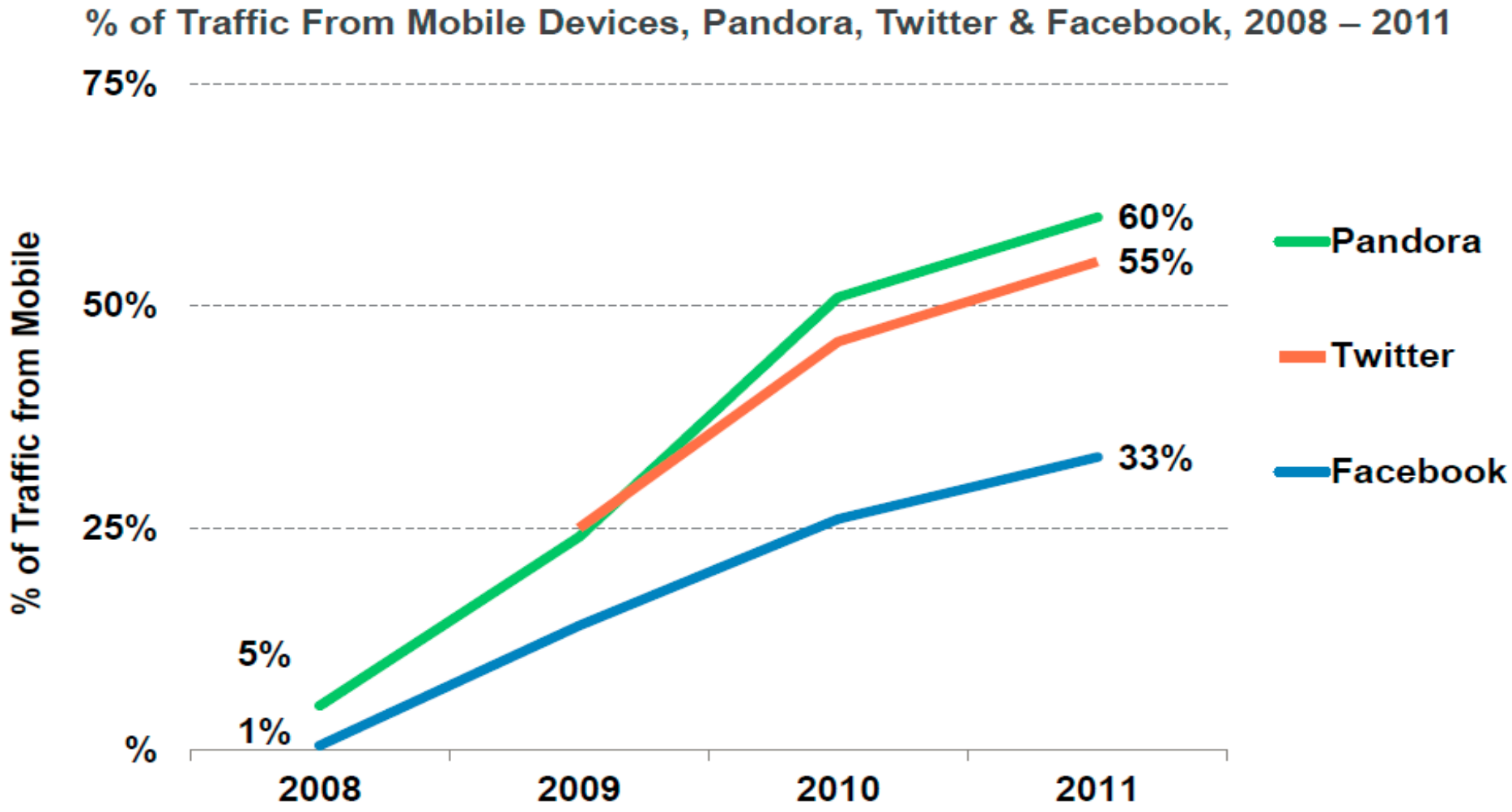
[Leskovec, Gomez, Krause, 2010]

Mobile internet is getting to the first tier



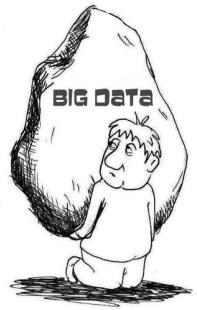
Source: Radio penetration data per Broadcasting & Cable Yearbook 1996, Internet penetration data per World Bank / ITU, Mobile Internet (smartphone) data per Morgan Stanley Research; 3G data per Informa.

Mobile devices are overtaking internet traffic



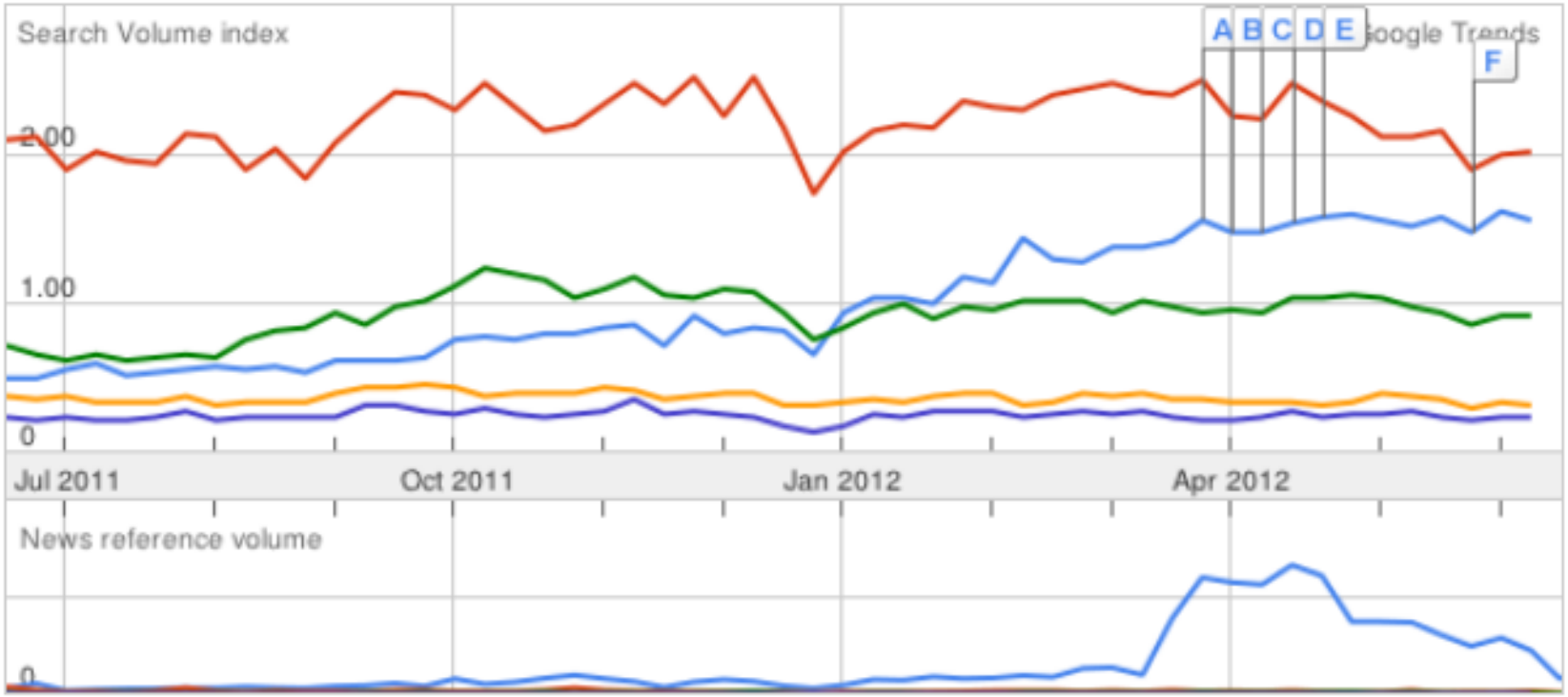
Source: Pandora S1, Twitter, Facebook.

Big-Data hype – how big is it?



big data 1.00 data mining 2.22 semantic web 0.36

machine learning 0.94 language technology 0.26



A Spectra Logic Delivers ExaScale Storage for 'Big Data'; Announces Series of Products and Advancements and Unveils World's Highest Capacity Storage System
MarketWatch - Nov 1 2011

B Webcast: Obama Goes Big on Big Data
Wired News - Mar 27 2012

C Cisco Joins Forces with EMC to Advance IT Skills in Cloud, Big Data and Data Center Technologies
Justmeans - Apr 3 2012

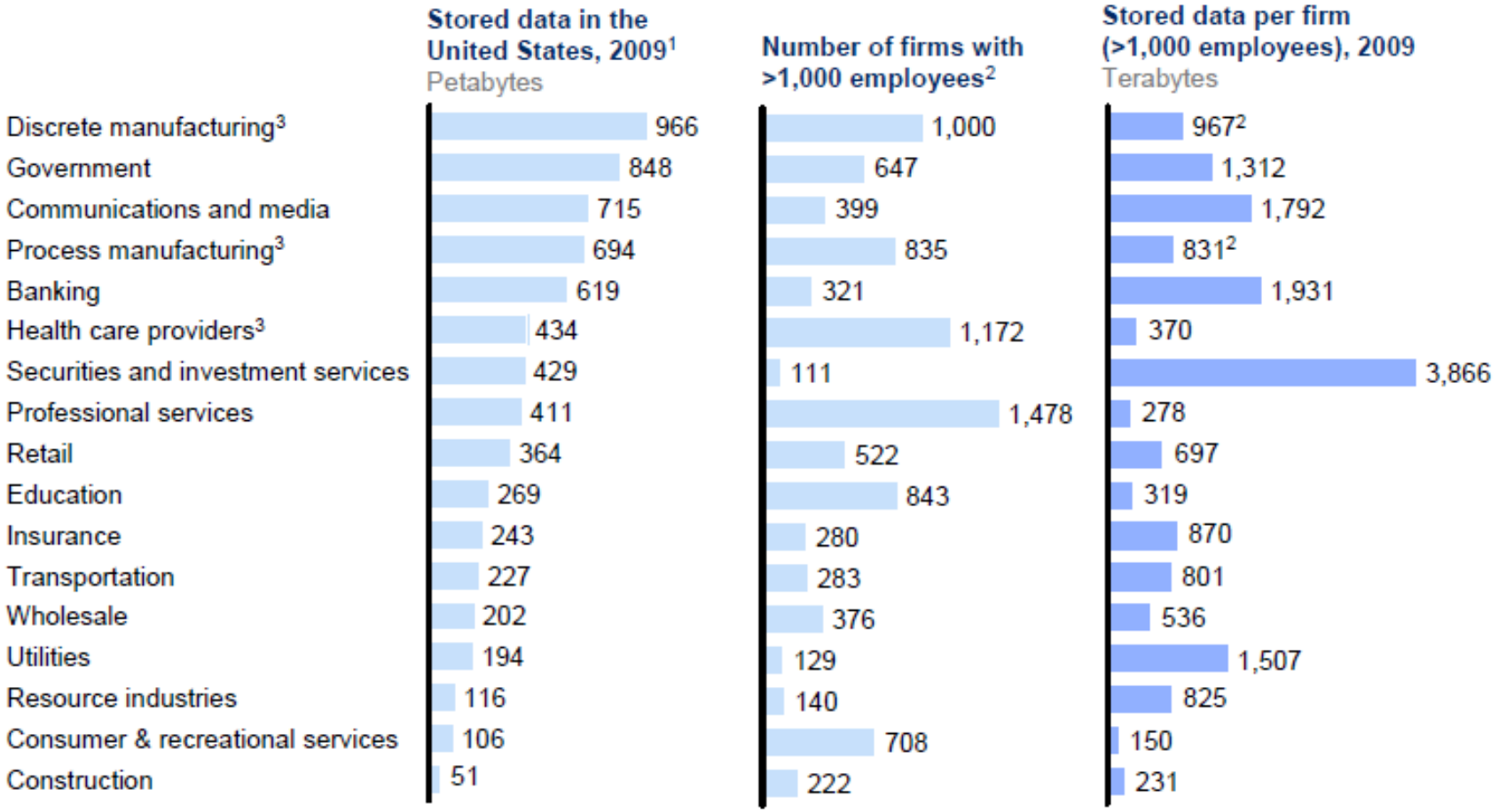
D Ferranti Unveils its MECOMS™ "Big Data" Strategy for Utility Meter Data Management and Real Time Billing
Victoria Times Colonist - Apr 10 2012

E Deconstructing Big Data - BuildZoom Launches an Article Series that Reveals the Hype and Substance Behind Big Data
Houston Chronicle - Apr 17 2012

F Harvard Releases Big Data for Books
New York Times - Apr 24 2012

Who has data?

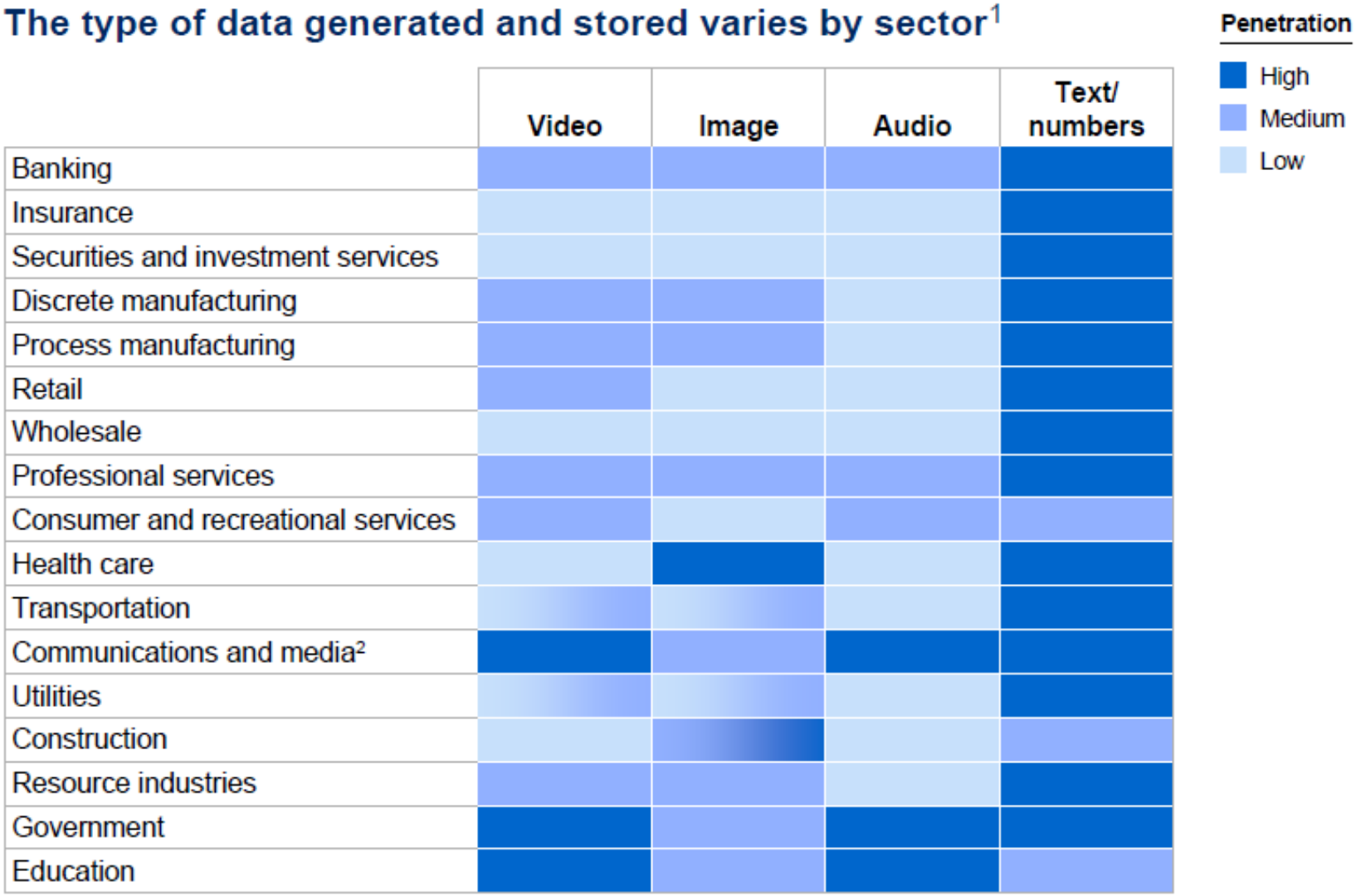
Companies in all sectors have at least 100 terabytes of stored data in the United States; many have more than 1 petabyte



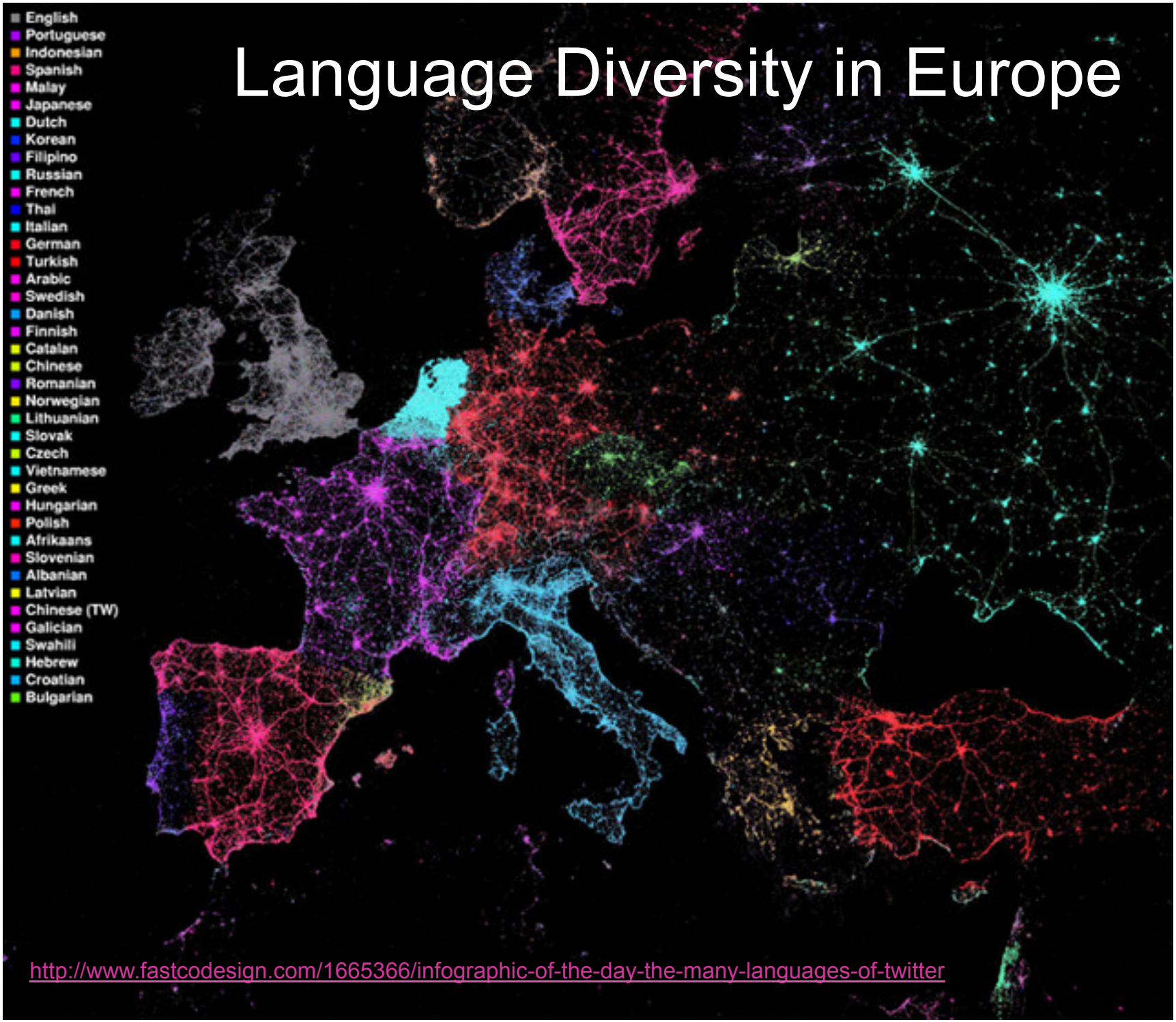
1 Storage data by sector derived from IDC.
2 Firm data split into sectors, when needed, using employment
3 The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

SOURCE: IDC; US Bureau of Labor Statistics; McKinsey Global Institute analysis

What kind od data are out there?



1 We compiled this heat map using units of data (in files or minutes of video) rather than bytes.
2 Video and audio are high in some subsectors.
SOURCE: McKinsey Global Institute analysis


English
Portuguese
Indonesian
Spanish
Malay
Japanese
Dutch
Korean
Filipino
Russian
French
Thai
Italian
German
Turkish
Arabic
Swedish
Danish
Finnish
Catalan
Chinese
Romanian
Norwegian
Lithuanian
Slovak
Czech
Vietnamese
Greek
Hungarian
Polish
Afrikaans
Slovenian
Albanian
Latvian
Chinese (TW)
Galician
Swahili
Hebrew
Croatian
Bulgarian
Language Diversity in Europe
http://www.fastcodesign.com/1665366/infographic-of-the-day-the-many-languages-of-twitter

Cultural diversity in New York

http://www.fastcodesign.com/1665366/infographic-of-the-day-the-many-languages-of-twitter

Novel ways of data acquisition

Recommendations for Technology & Research Roadmap

Summary Roadmap

Research Priority	Phase 1: 2013-2014	Phase 2: 2015-2017	Phase 3: 2018-2020
Social influence and incentives	Modelling social diversity of views across languages and cultures	Modelling social influence and incentives through game theoretic approaches using data from textual and social networking streams	Holistic modelling of society (or its segments) through observing variety of data sources
Information tracking	Tracking dynamics of information diffusion across languages, cultures and media	Transforming observed textual and social data streams into actionable deep knowledge representations	Prediction of future events and identification of causal relationships from textual and social streams
Multimodal data processing	Joining textual data and social networks, including spatial and temporal dimensions	Joining textual and social data with unstructured sources like sensor data (smart cities), video, images, audio	Detecting inconsistencies, gaps and completeness of collected knowledge from textual and social sources
Visualization and user interaction	Visualization of textual and social dynamics	Adaptive human-computer interfaces boosting specific aims in interaction	Adaptive interaction systems for communication with the whole or parts of society
Algorithmic fundamentals	Algorithms and toolkits for scalable processing of multi-modal data (Big-Data)	Real-time modelling and reasoning on massive textual and social streams	Algorithms and toolkits being able to deal with planetary scale analytics and reasoning with multimodal data

Five proposed research lines (1/2)

- **(1)** Social influence and incentives
  - Modeling social diversity of views across languages and cultures
  - Modeling social influence and incentives
  - Multipolar opinion mining (beyond usual sentiment analysis)

- **(2)** Information dynamics
  - Tracking dynamics of information diffusion across languages, cultures and media
  - Prediction of future events and identification of causal relationships from textual and social streams

Five proposed research lines (2/2)

- **(3)** Multimodal data processing
  - Joining textual data and social networks, including spatial and temporal dimensions
  - Joining textual and social data with unstructured sources like sensor data (smart cities), video, images, audio

- **(4)** Visualization and user interaction
  - Visualization of textual and social dynamics
  - Adaptive user interfaces

- **(5)** Algorithmic fundamentals
  - Algorithms and toolkits for scalable processing of multi-modal Big data
  - Real-time modeling and reasoning on massive textual and social streams

Q/A

Thank you.

office@meta-net.eu

http://www.meta-net.eu
http://www.facebook.com/META.Alliance