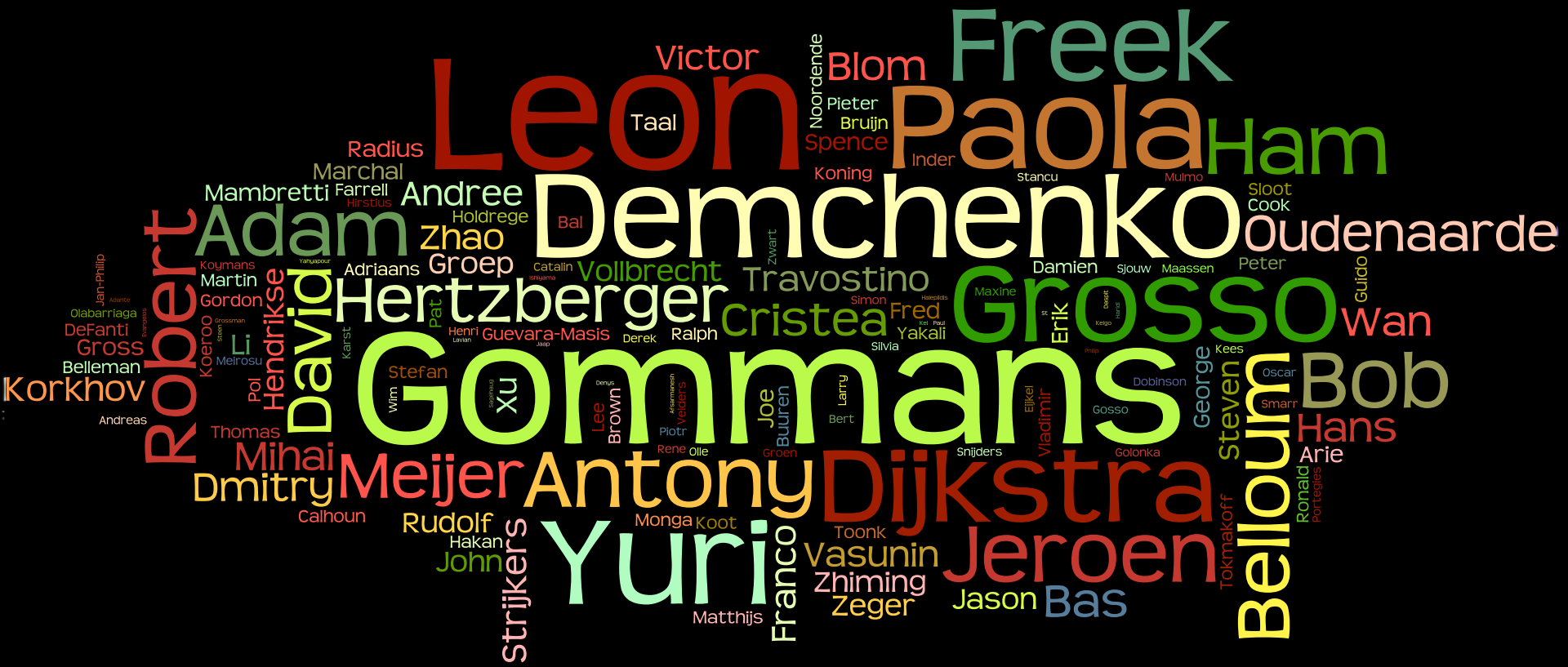
System and Network Engineering Research for Big Data Sciences

Cees de Laat

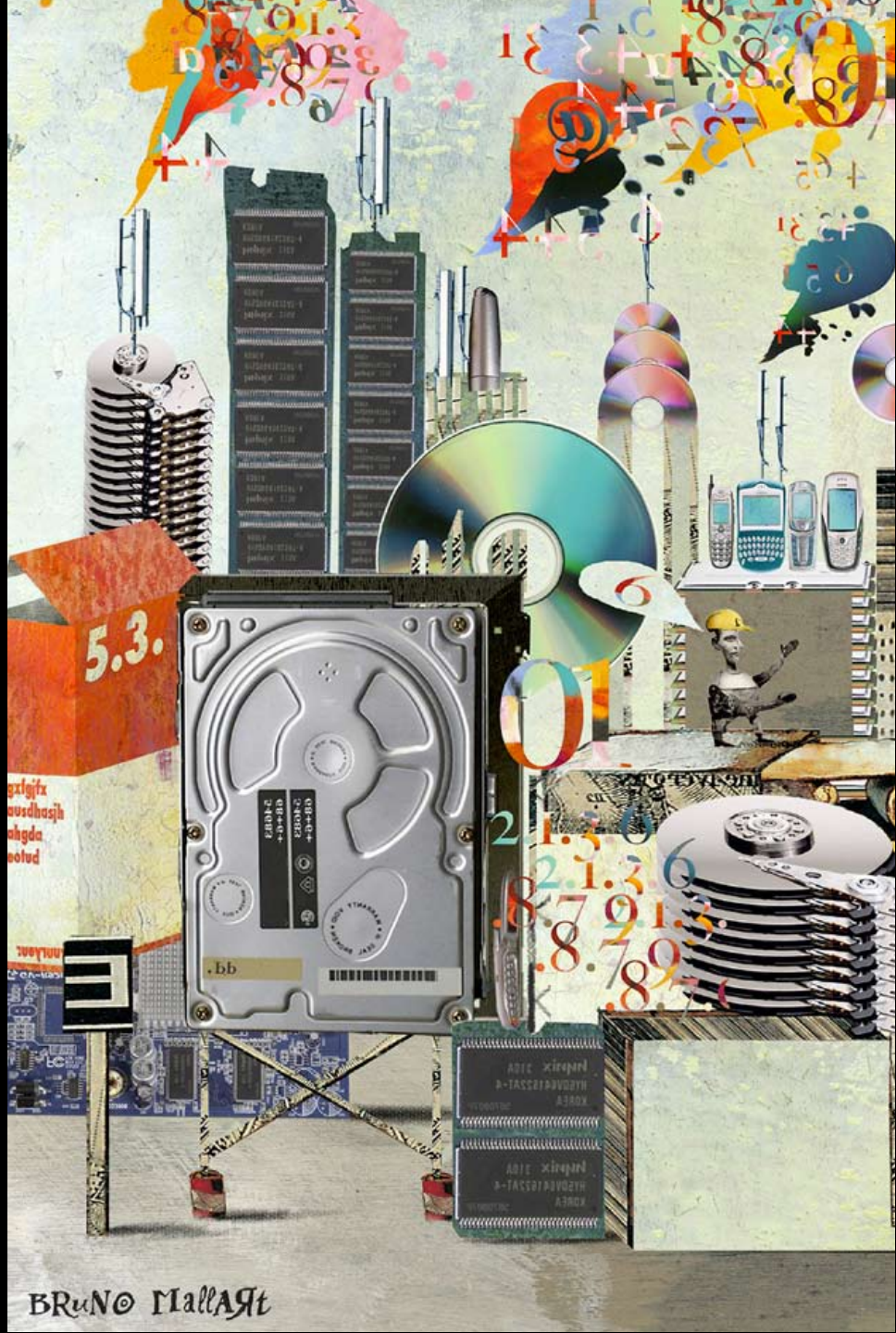


From King's Dutch Academy of Sciences

The Dutch Research Agenda

•"Information technology (IT) now permeates all aspects of public, commercial, social, and personal life. bank cards, satnav, and weather radar... IT has become completely indispensable."

•"But to guarantee the reliability and quality of constantly bigger and more complicated IT, we will need to find **answers to some** fundamental questions!"



•https://www.knaw.nl/nl/actueel/publicaties/the-dutch-research-agenda/@@download/pdf_file/20111029.pdf

Reduction of Complexity by Integration

•By combining services such as telephony, television, data, and computing capacity within a single network, we can cut down on complexity, energy consumption and maintenance.

- How can we describe and analyze complex information systems effectively?
- How can we specify and measure the quality and reliability of a system?
- How can we combine various different systems?
- How can we design systems in which separate processors can co-operate efficiently via mutual network connections within a much larger whole?
- Can we design information systems that can diagnose their own malfunctions and perhaps even repair them?
- How can we specify, predict, and measure system performance as effectively as possible?

•SNE addresses a.o. the **highlighted** questions!

•http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20111029.pdf

Internet developments

... more data!

Speed

Volume

Deterministic

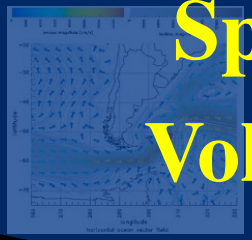
Real-time

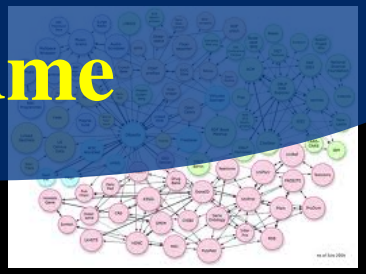
more realtime!

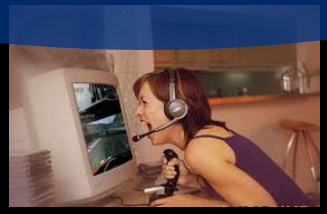
Scalable

Secure

... more users!







Mission SNE

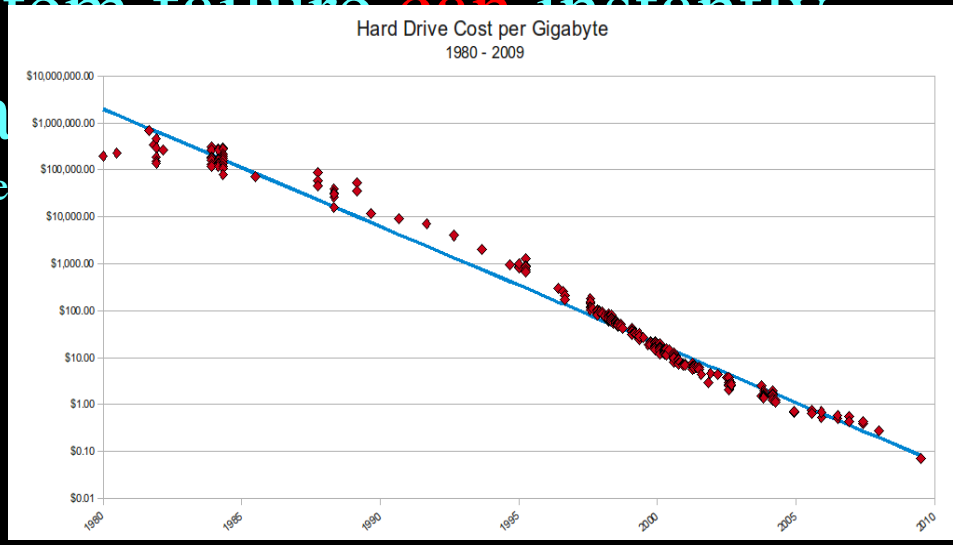
Can we create smart and safe data processing infrastructures that can be tailored to diverse application needs?

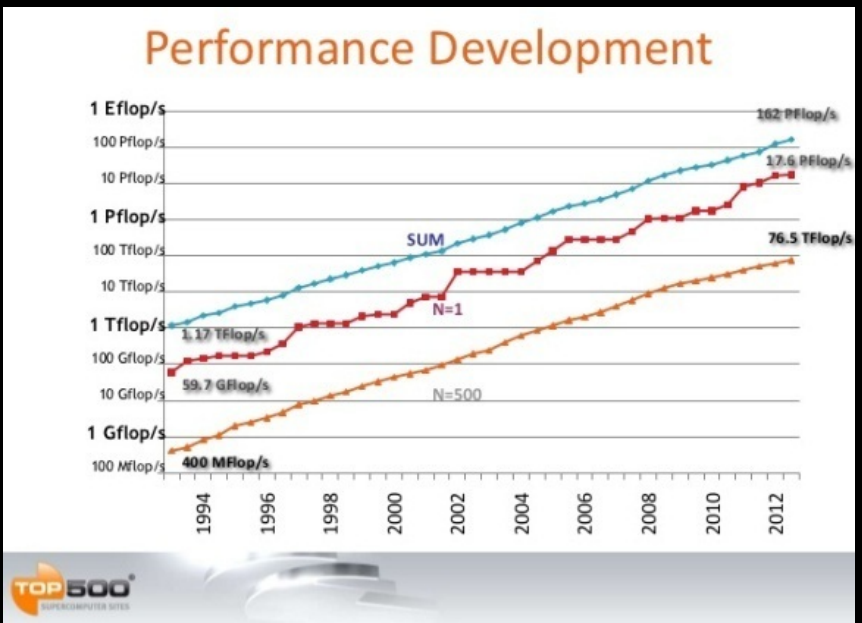
- *Capacity*
  - *Bandwidth on demand, QoS, architectures, photonics, performance*
- *Capability*
  - *Programmability, virtualization, complexity, semantics, workflows*
- *Security*
  - *Authorization, Anonymity, integrity of data in distributed data processing*
- *Sustainability*
  - *Greening infrastructure, awareness*
- *Resilience*
  - *Systems under attack, failures, disasters*

Reliable and Safe!

- This omnipresence of IT makes us not only strong but also vulnerable.
- A virus, a hacker, or a system failure can instantly send digital shockwaves a

- The hardware and software that allow all our systems to operate is becoming bigger and more complex all the time, and the capacity of networks and data storage is increasing by leaps and bounds.

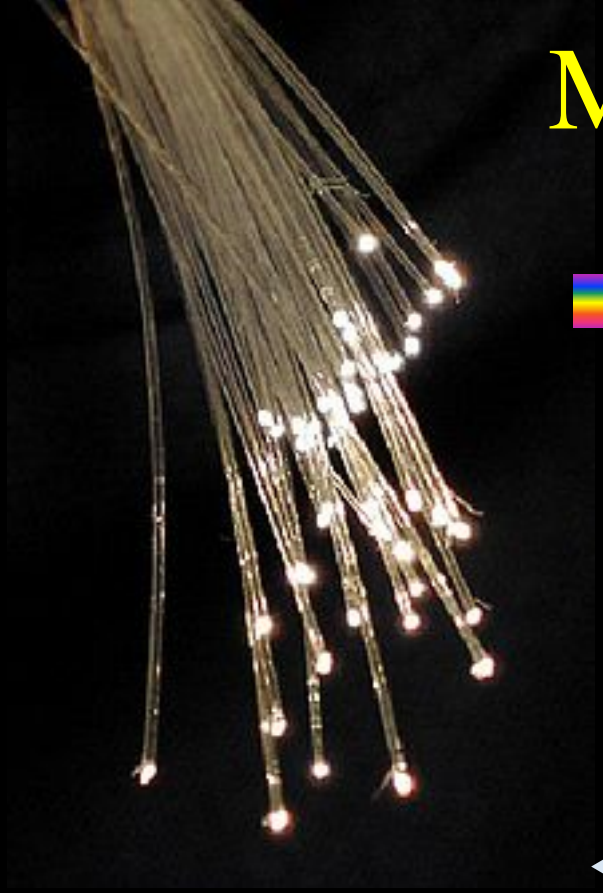


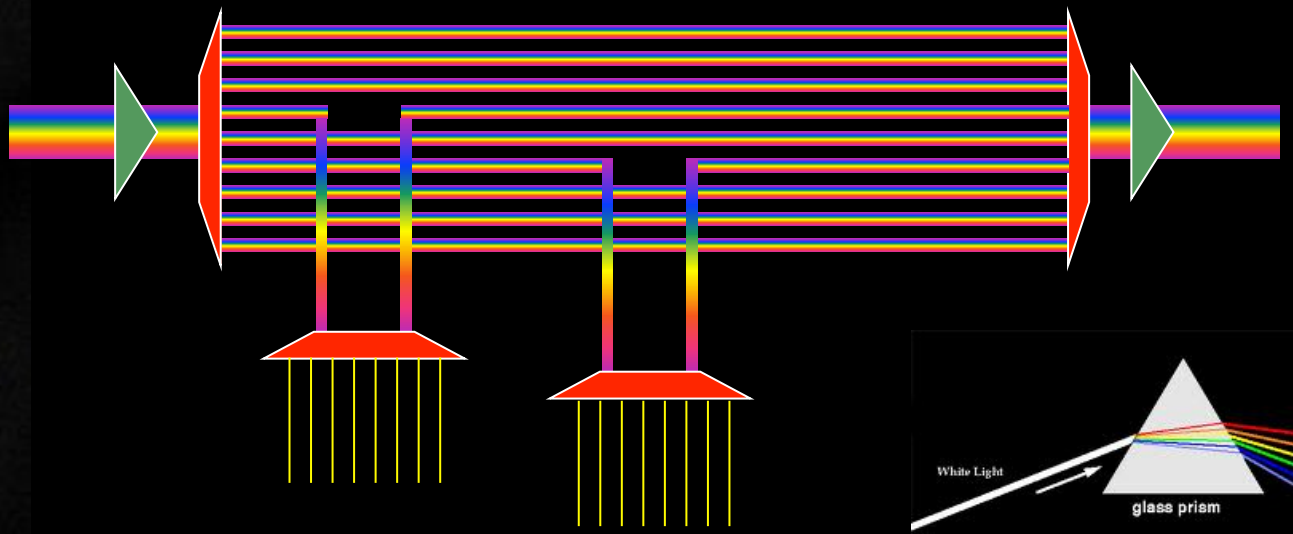


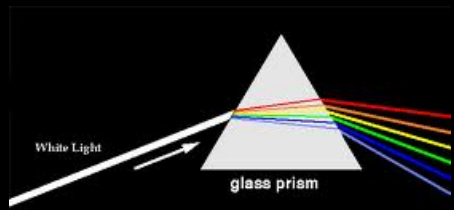
- We will soon reach the limits of what is currently feasible and controllable.

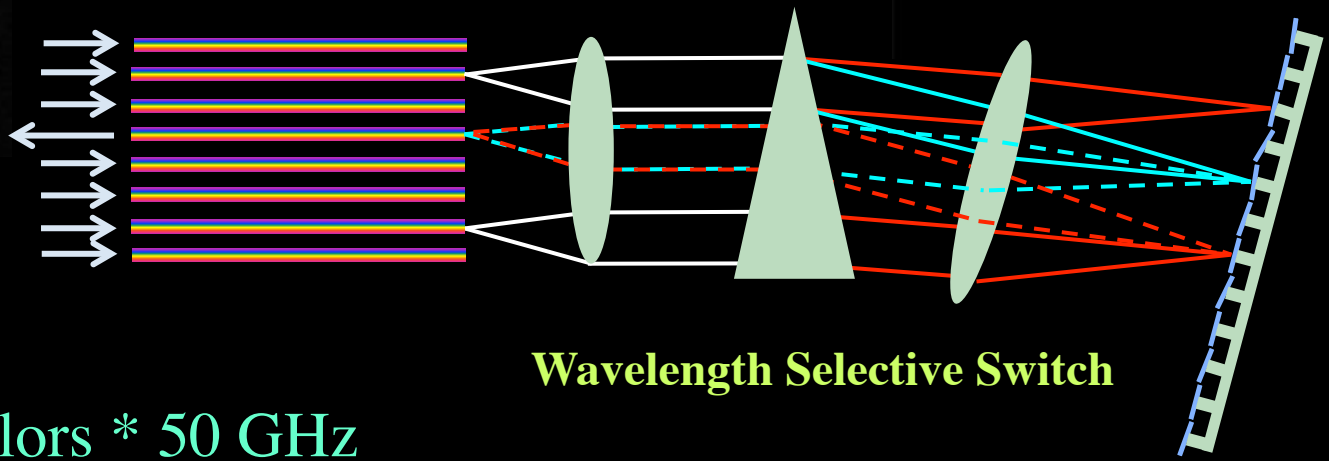
- http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20111029.pdf

Multiple colors / Fiber









Wavelength Selective Switch

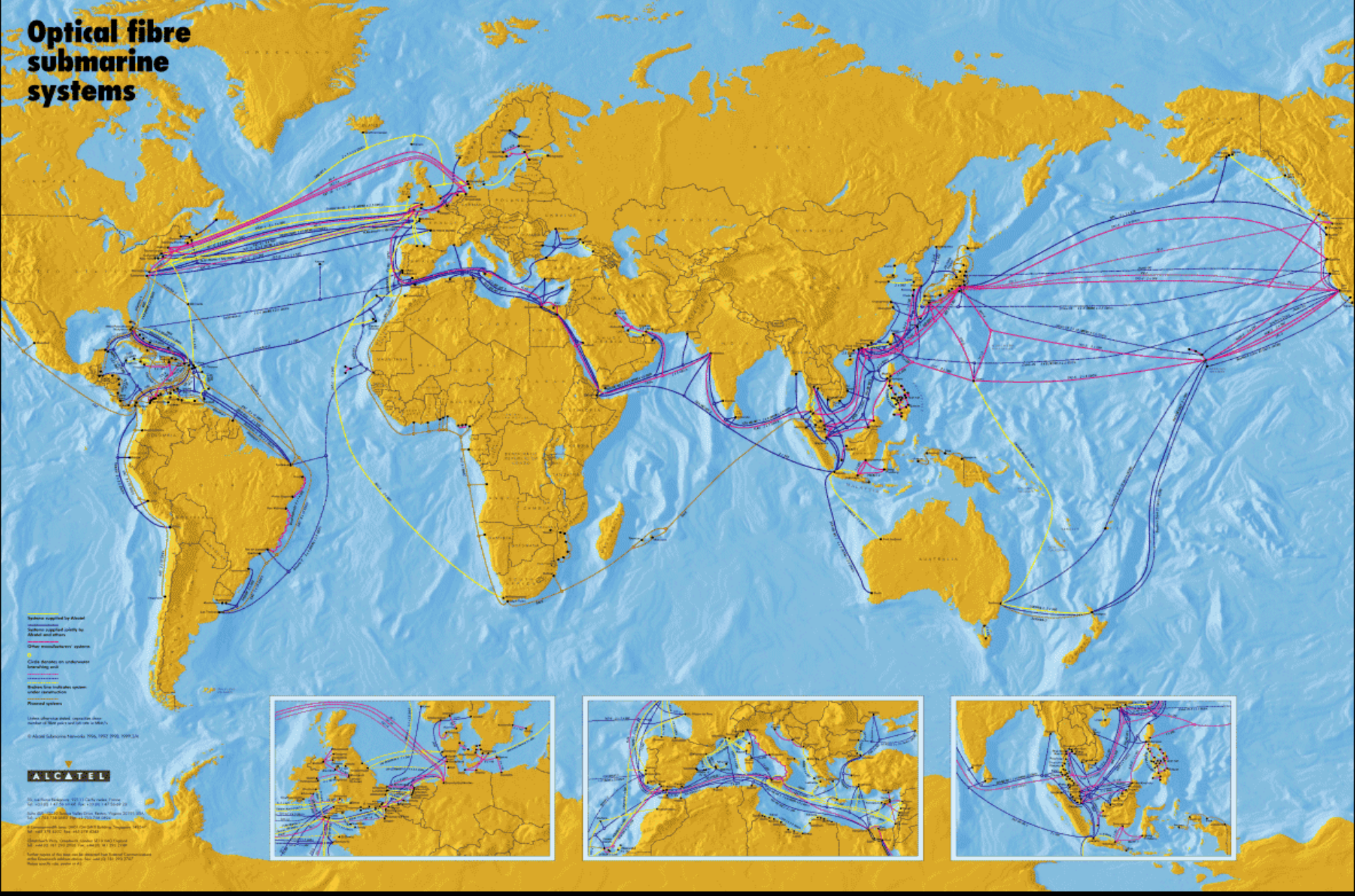
Per fiber: ~ 80-100 colors * 50 GHz

Per color: 10 – 40 – 100 Gbit/s

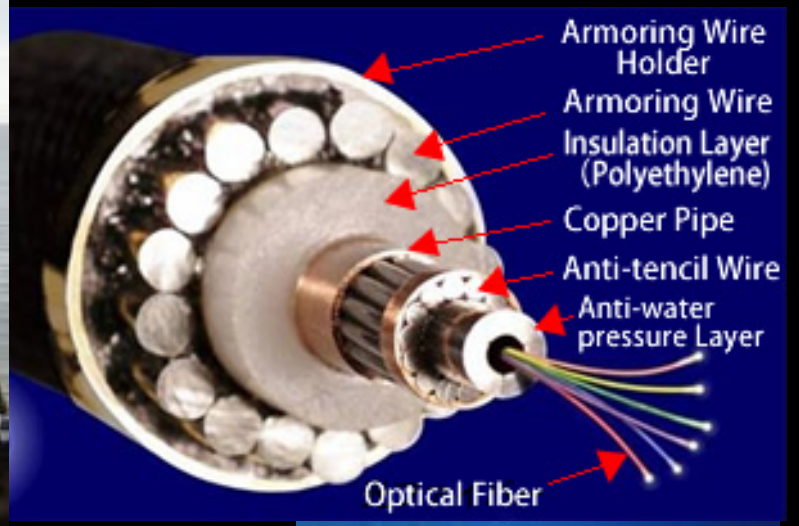
BW * Distance ~ $2*10^{17}$ bm/s

New: Hollow Fiber!

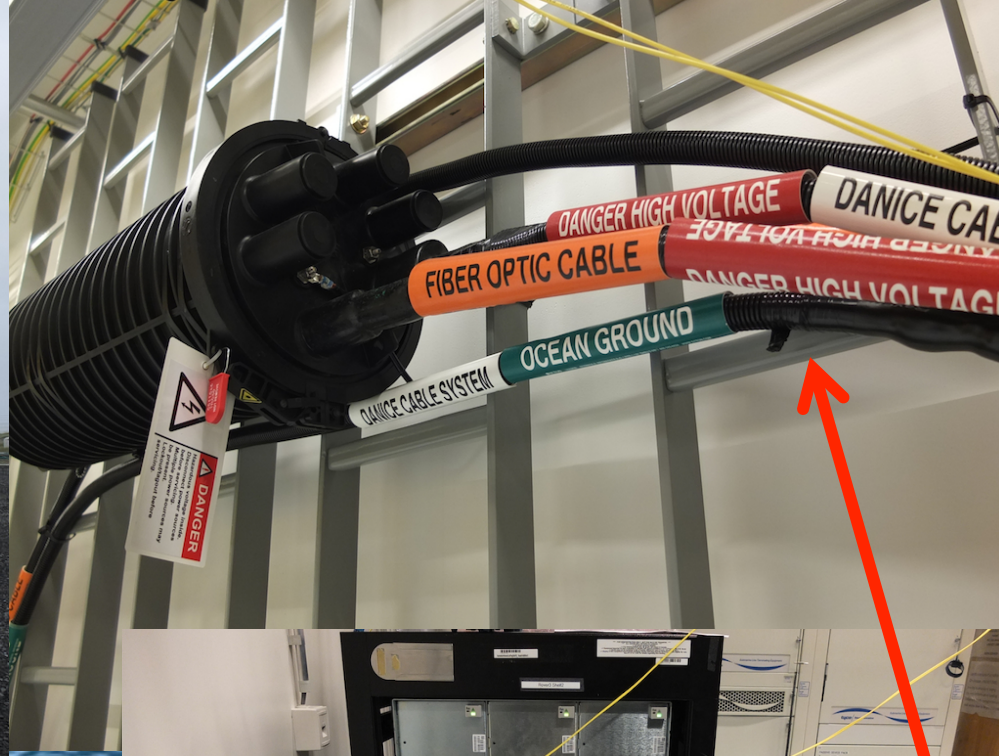
➔ less RTT!



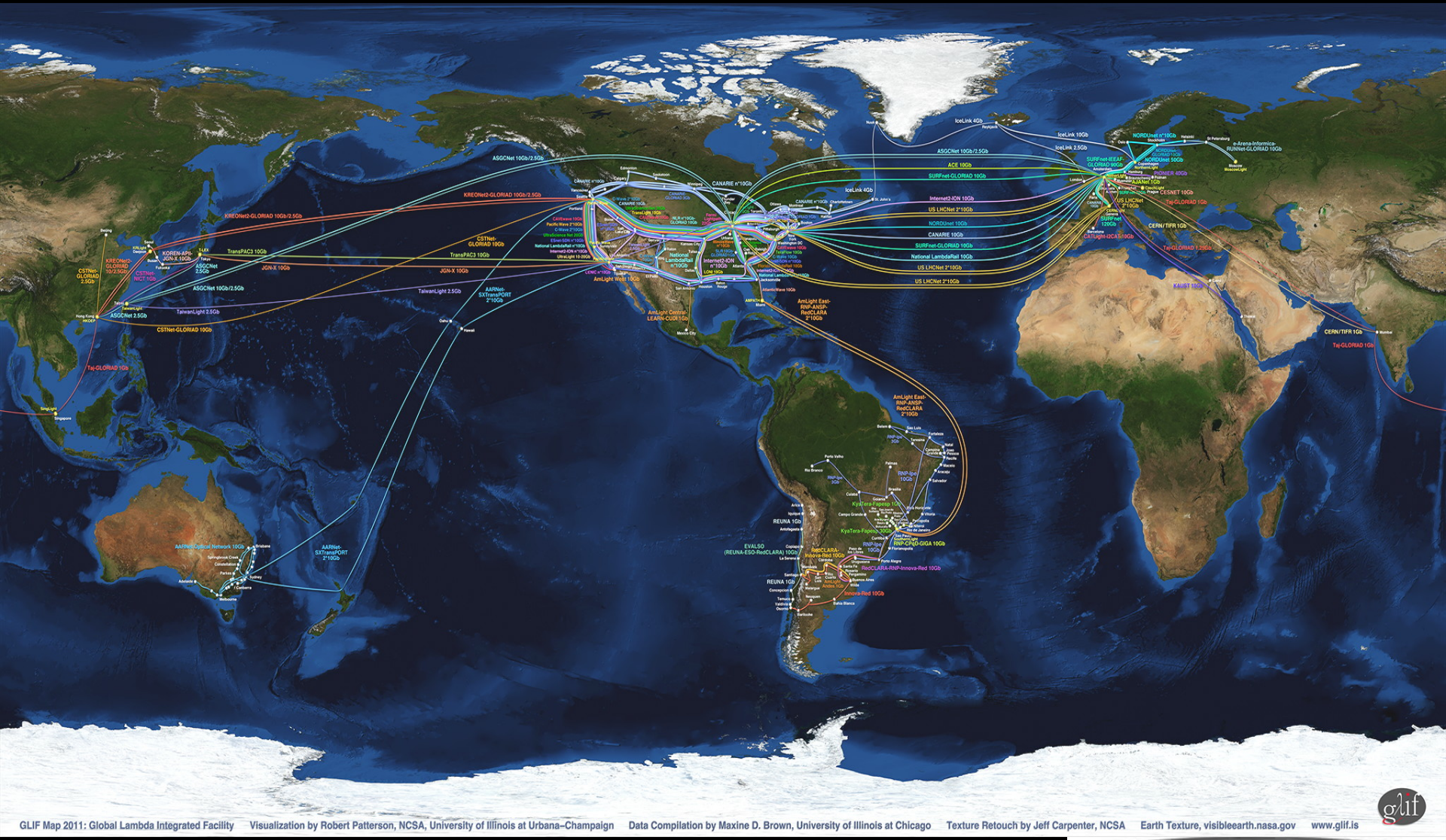
Undersea Cable System



•A **cable landing station** may or may not be required, depending on whether, for example, the submarine cable requires power to power submarine repeaters or amplifiers. The voltages applied to the cables can be high **3,000 to 4,000 volts** for a typical trans-Atlantic telecommunications cable system, and 1,000 volts for a cross-channel telecommunications cable system. Submarine power cables can operate at many kilovolts: for example, the Fenno-Skan power cable operates at 400 kV DC.

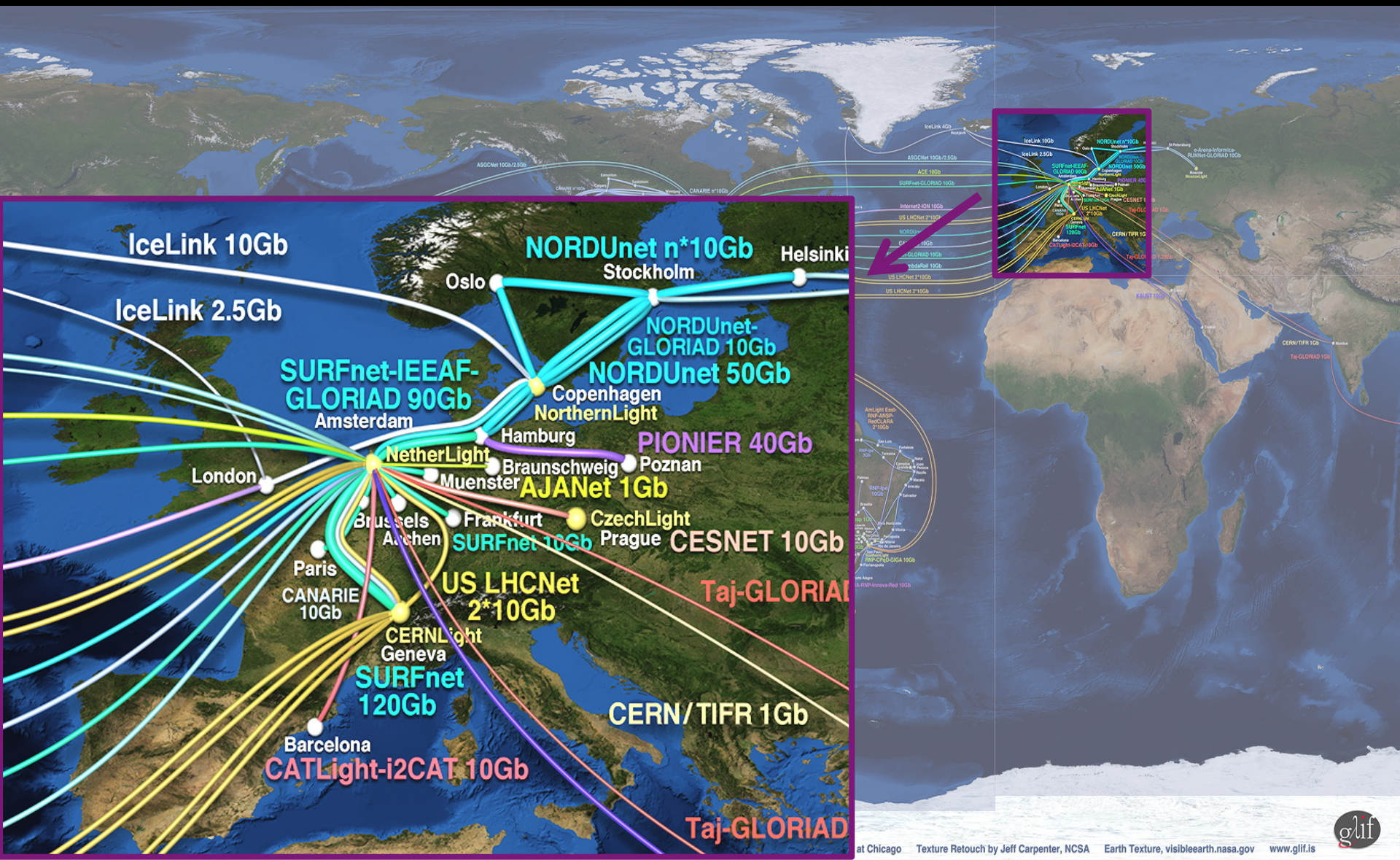


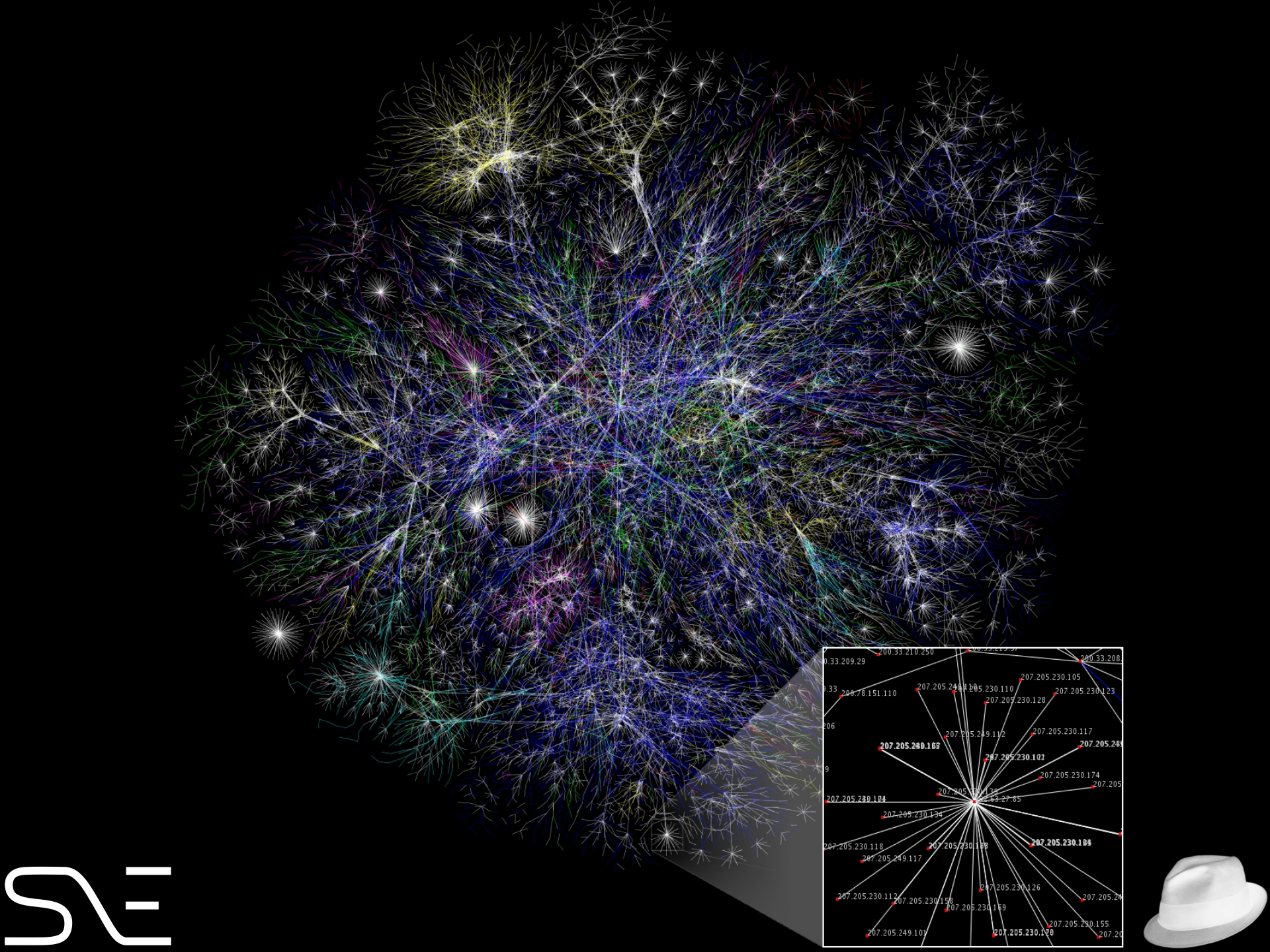
Undersea Cable HV



We investigate: TOMTOM for complex networks!

- F Dijkstra, J van der Ham, P Grosso, C de Laat, “A path finding implementation for multi-layer networks”, Future Generation Computer Systems 25 (2), 142-146.




200.33.210.250
207.205.230.105
207.205.230.110
207.205.230.123
200.78.151.110
207.205.230.128
207.205.249.112
207.205.230.117
207.205.230.174
207.205.230.134
207.205.230.118
207.205.249.117
207.205.230.126
207.205.230.156
207.205.230.159
207.205.230.155
207.205.249.101

Alien light From idea to realisation!

40Gb/s alien wavelength transmission via a multi-vendor 10Gb/s DWDM infrastructure

NCF

Alien wavelength advantages

- Direct connection of customer equipment[1] → cost savings
- Avoid OEO regeneration → power savings
- Faster time to service[2] → time savings
- Support of different modulation formats[3] → extend network lifetime

Alien wavelength challenges

- Complex end-to-end optical path engineering in terms of linear (i.e. OSNR, dispersion) and non-linear (FWM, SPM, XPM, Raman) transmission effects for different modulation formats.
- Complex interoperability testing.
- End-to-end monitoring, fault isolation and resolution.
- End-to-end service activation.

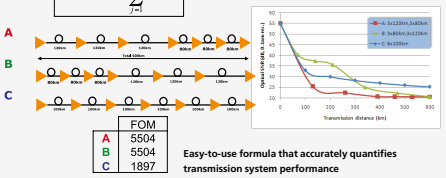
In this demonstration we will investigate the performance of a 40Gb/s PM-QPSK alien wavelength installed on a 10Gb/s DWDM infrastructure.

New method to present fiber link quality, FoM (Figure of Merit)

In order to quantify optical link grade, we propose a new method of representing system quality: the FOM (Figure of Merit) for concatenated fiber spans.

$$FOM = \sum_{j=1}^{N} 10^{\left(\frac{L_j}{10}\right)}$$

Lj, span losses in dB
N, number of spans

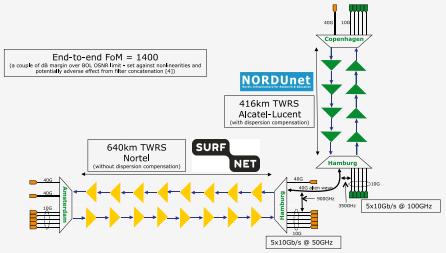


	FOM
A	5504
B	5504
C	1897

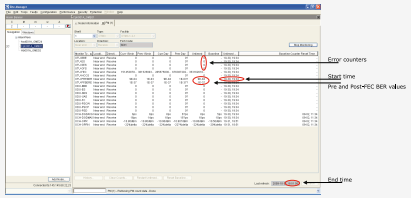
Easy-to-use formula that accurately quantifies transmission system performance

Transmission system setup

JOINT SURFnet/NORDUnet 40Gb/s PM-QPSK alien wavelength DEMONSTRATION.



Test results



Error-free transmission for 23 hours, 17 minutes → BER < 3.0 10^{-16}

Conclusions

- We have investigated experimentally the all-optical transmission of a 40Gb/s PM-QPSK alien wavelength via a concatenated native and third party DWDM system that both were carrying live 10Gb/s wavelengths.
- The end-to-end transmission system consisted of 1056 km of TWRS (TrueWave Reduced Slope) transmission fiber.
- We demonstrated error-free transmission (i.e. BER below 10-15) during a 23 hour period.
- More detailed system performance analysis will be presented in an upcoming paper.

NORTEL NORDUnet telindus together with ISTY SURF NET

REFERENCES [1] "OPERATIONAL SOLUTIONS FOR AN OPEN DWDM LAYER", O. GERSTEL ET AL, OFC'2009 | [2] "AT&T OPTICAL TRANSPORT SERVICES", BARBARA E. SMITH, OFC'09 [3] "OPEX SAVINGS OF ALL-OPTICAL CORE NETWORKS", ANDREW LORD AND CARL ENGINEER, ECOC2009 | [4] NORTEL/SURFNET INTERNAL COMMUNICATION

ACKNOWLEDGEMENTS WE ARE GRATEFUL TO NORDUNET FOR PROVIDING US WITH BANDWIDTH ON THEIR DWDM LINK FOR THIS EXPERIMENT AND ALSO FOR THEIR SUPPORT AND ASSISTANCE DURING THE EXPERIMENTS. WE ALSO ACKNOWLEDGE TELINDUS AND NORTEL FOR THEIR INTEGRATION WORK AND SIMULATION SUPPORT

What Happens in an **Internet Minute?**

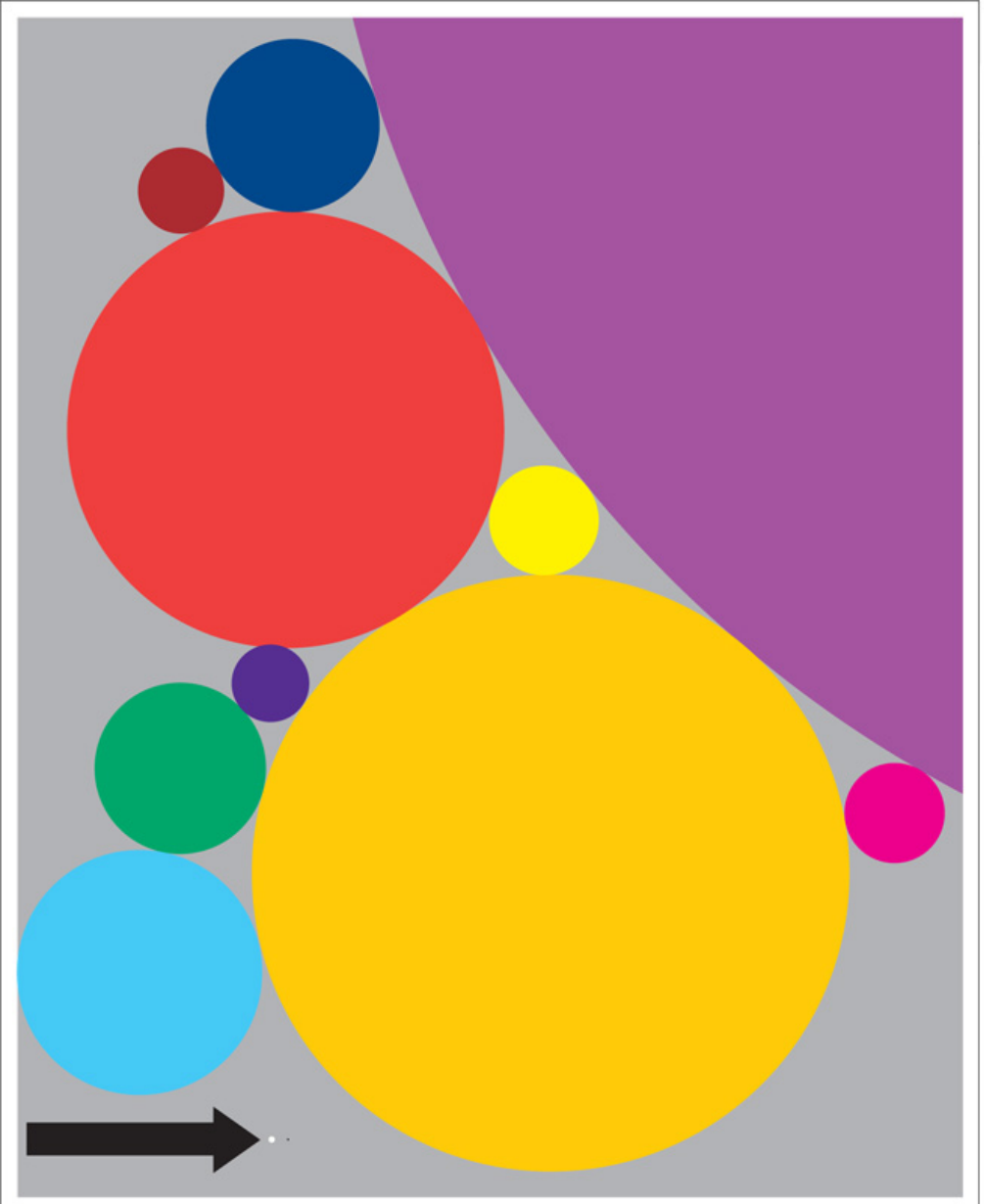
And **Future Growth** is **Staggering**

Today, the number of networked devices = the global population

By **2015**, the number of networked devices = **2x** the global population

In **2015**, it would take you **5 years** to view all video crossing IP networks each **second**

intel



Size of data sets in terabytes

Business email sent per year	2,986,100	National Climactic Data Center database	6,144
Content uploaded to Facebook each year	182,500	Library of Congress' digital collection	5,120
Google's search index	97,656	US Census Bureau data	3,789
Kaiser Permanente's digital health records	30,720	Nasdaq stock market database	3,072
Large Hadron Collider's annual data output	15,360	Tweets sent in 2012	19
Videos uploaded to YouTube per year	15,000	Contents of every print issue of WIRED	1.26

There
is
always
a
bigger
fish

MERKEL ON THE PHONE, SHE'S VERY ANGRY
NSA
CHAPPATTE
International New York Times

Fact & Fiction

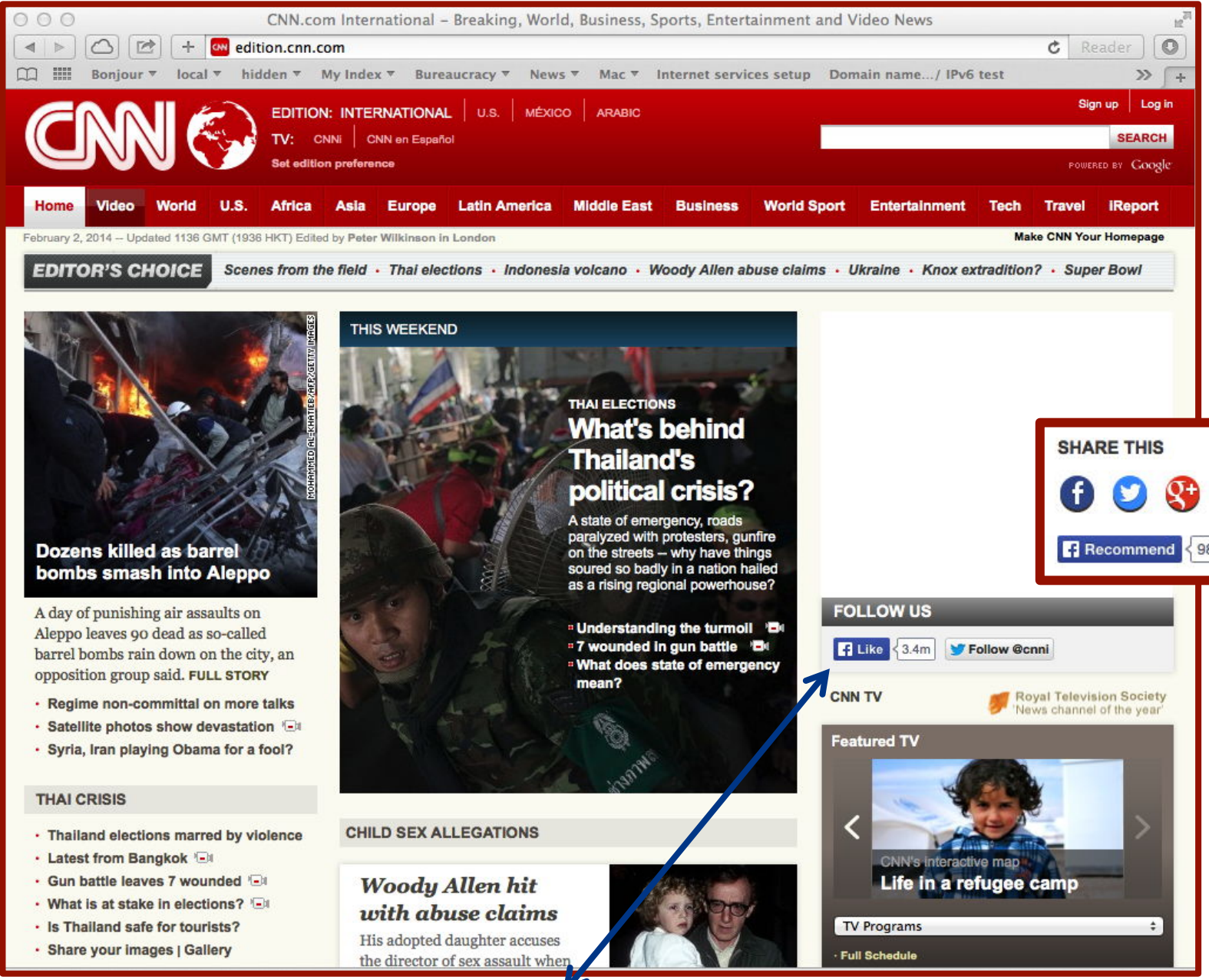
- Claims that BT pre-cooks adsl modems to send information from home networks to NSA and British Intelligence
  - http://cryptome.org/2013/12/Full-Disclosure.pdf
  - Modem connects to specific IP addresses at boot time
- Critical responses:
  - DOD uses lots of address space that is not publicly routed
    - http://blog.erratasec.com/2013/12/dod-address-space-its-not-conspiracy.html
  - See also the comment:
    "lucent uses 152.148.0.0/16 for 'management' on lots of their old big telco iron as if it was RFC-1918 space. (...)"
  - Also BT-competitor AAISP claims this is FUD:
    - https://s.aa.net.uk/1871
  - Claims: "They use DOD space because it's not internet-routable, and it's for the TR-069 (http://en.wikipedia.org/wiki/TR-069) service. This is *NOT* news."
    - http://www.bit-tech.net/news/hardware/2013/12/17/bt-back-door/1

Wat heeft dit van doen met de 2013 nationale wetenschapsquiz?

- Vraag 13: Voor een ziekte waar 1 op de 1000 mensen aan lijdt, is een 99% betrouwbare test ontwikkeld. Wat is de kans dat je ook echt ziek bent als de test dat uitwijst?

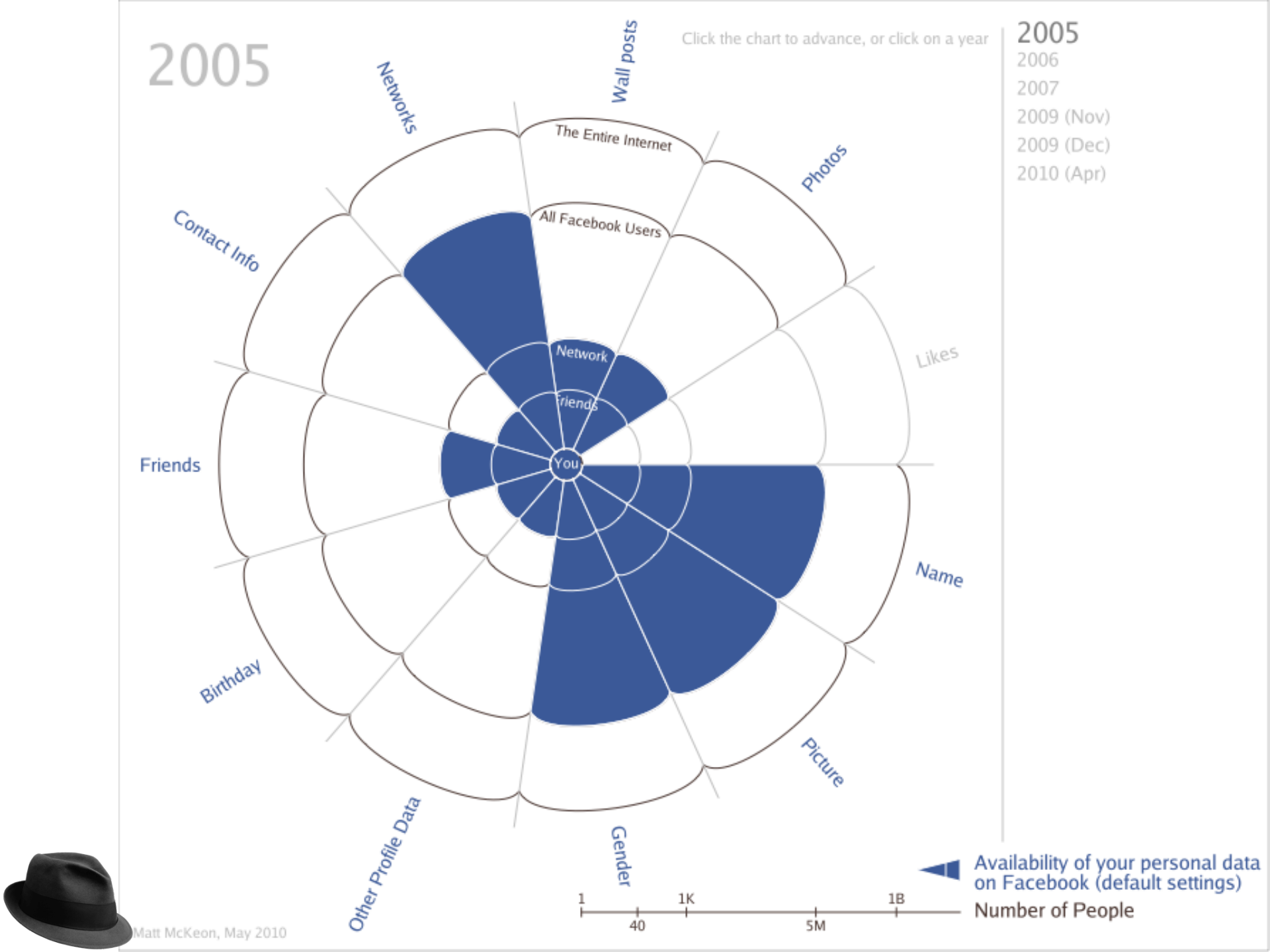
- Stel dat PRISM, Tempora, Xkeyscore, etc. in 99% van de gevallen betrouwbaar zijn en dat 1 op de 100000 subjects echt terroristen zijn…

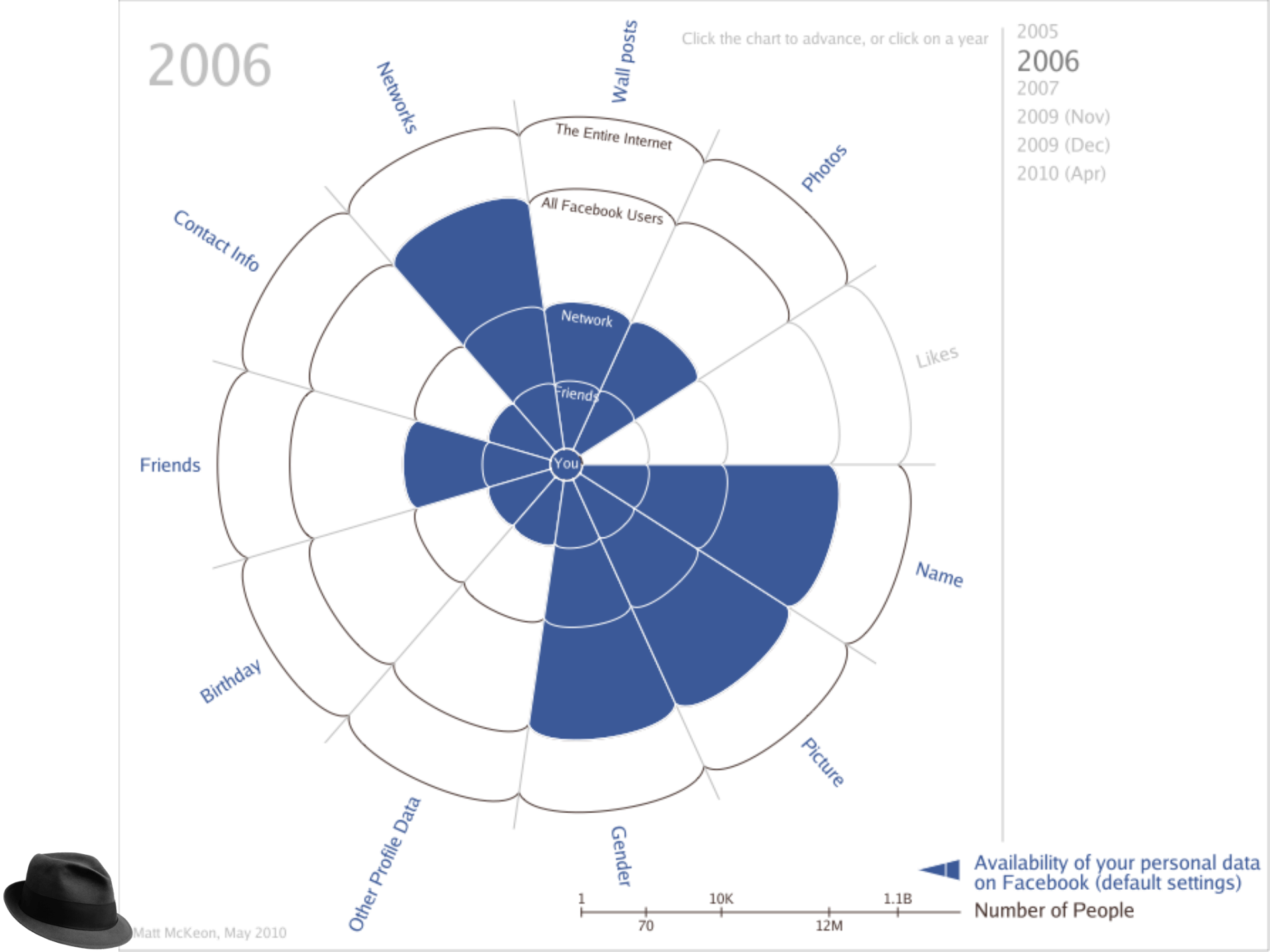
- False positives… ~1000 !

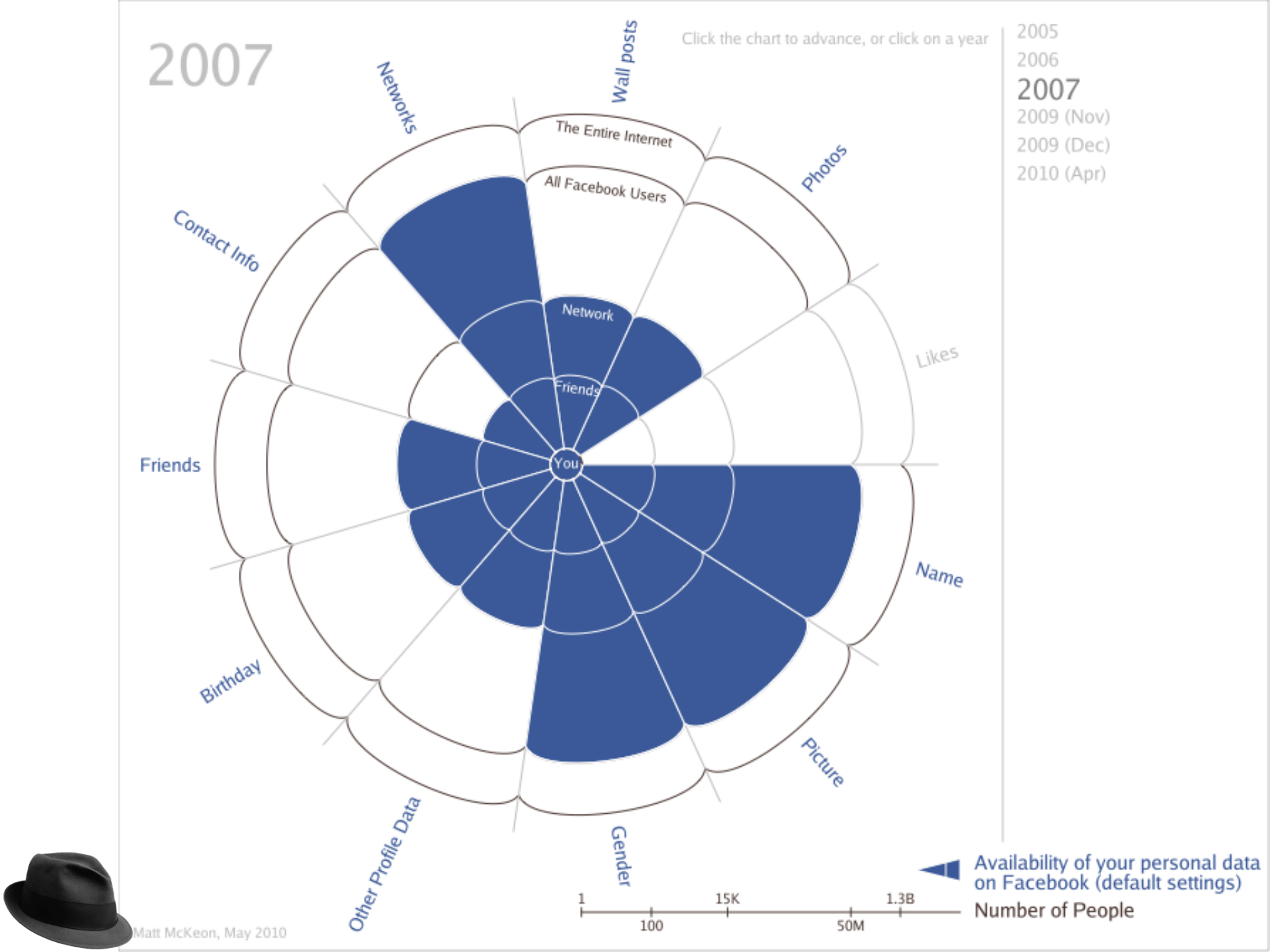


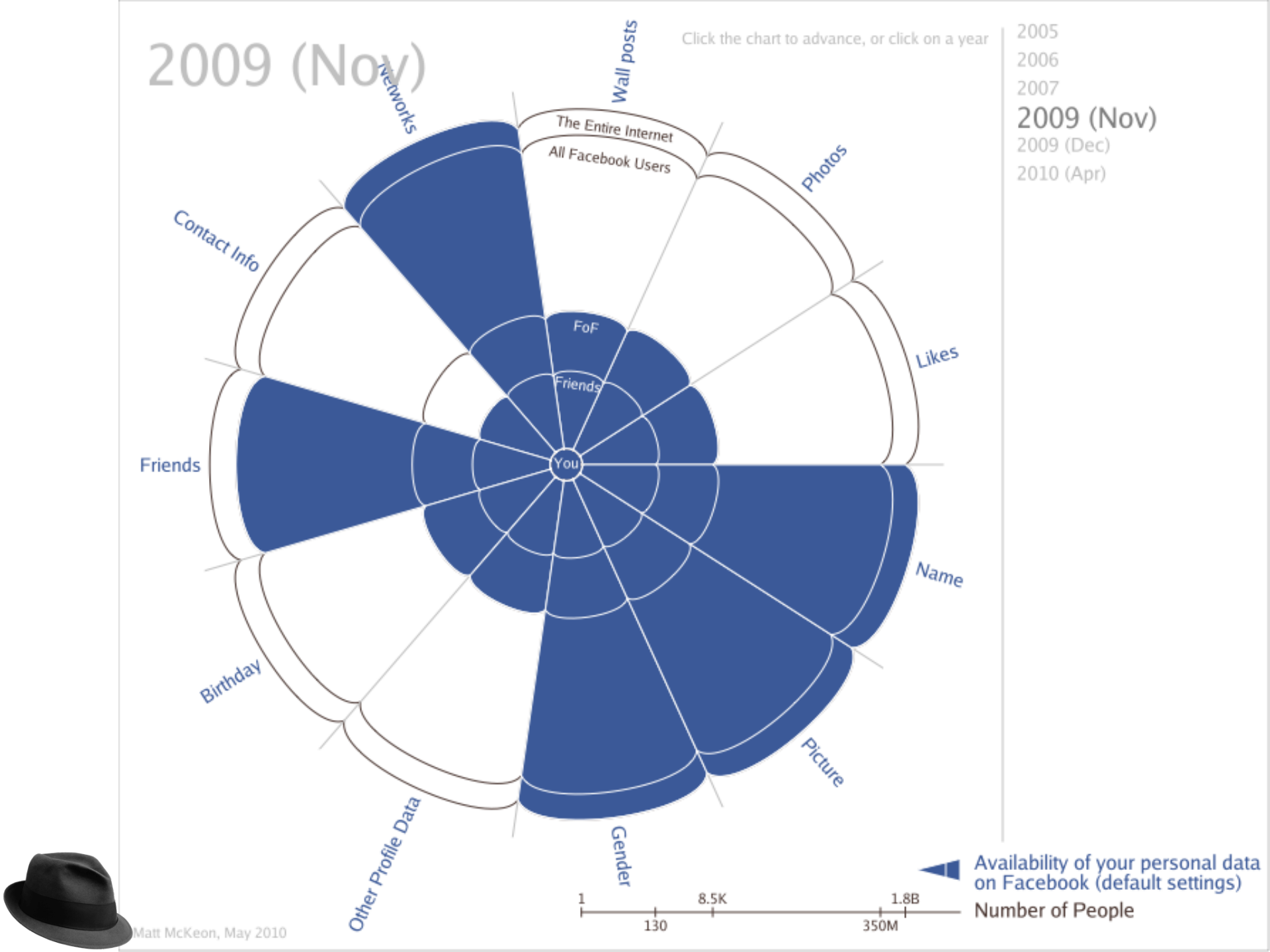
I will follow you!

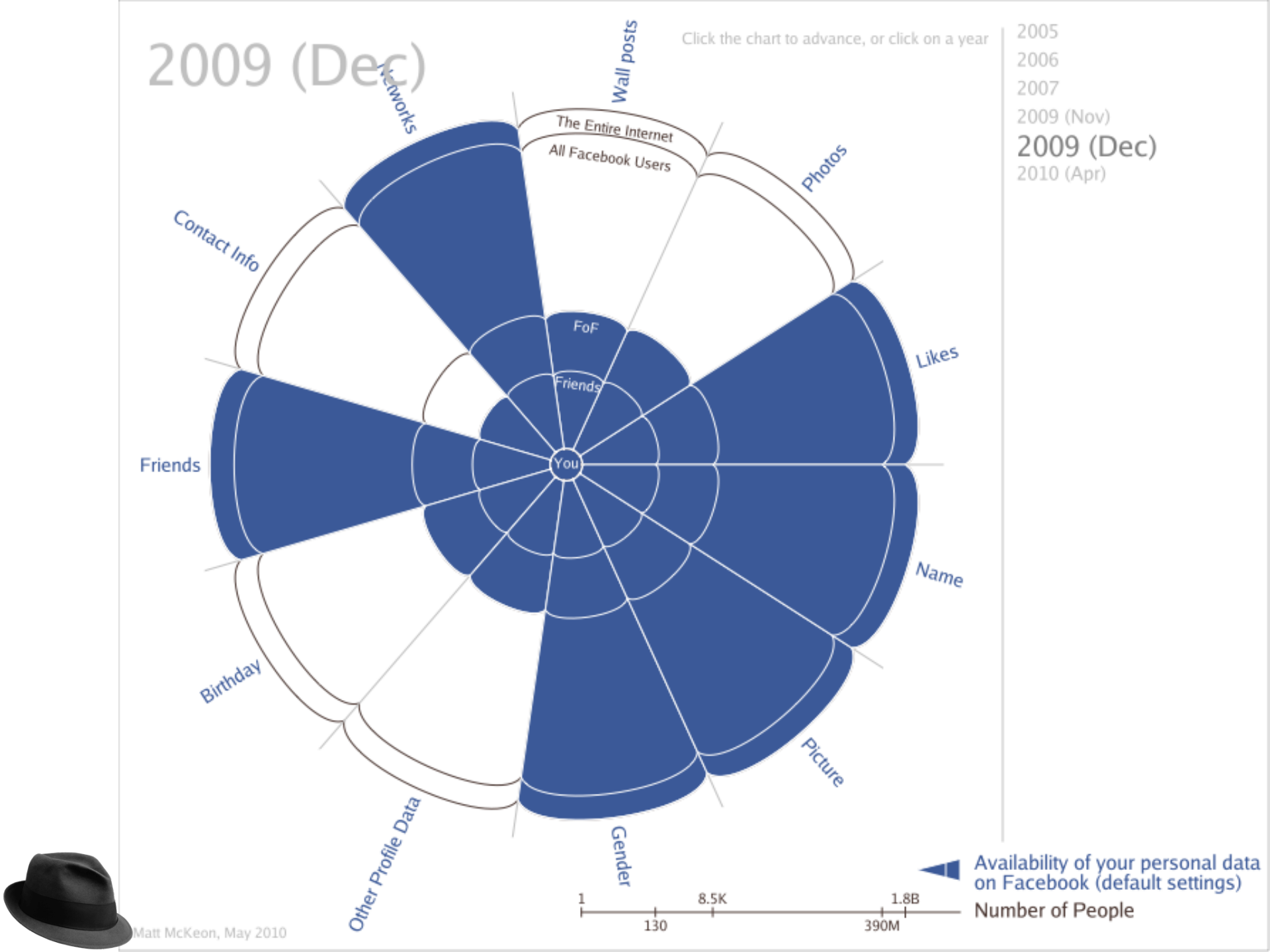
<iframe src="//www.facebook.com/plugins/like.php?href=http%3A%2F%2Fwww.facebook.com%2Fcnninternational&send=false&layout=button_count&width=450&show_faces=false&action=like&colorscheme=light&font=arial&height=21" …></iframe>

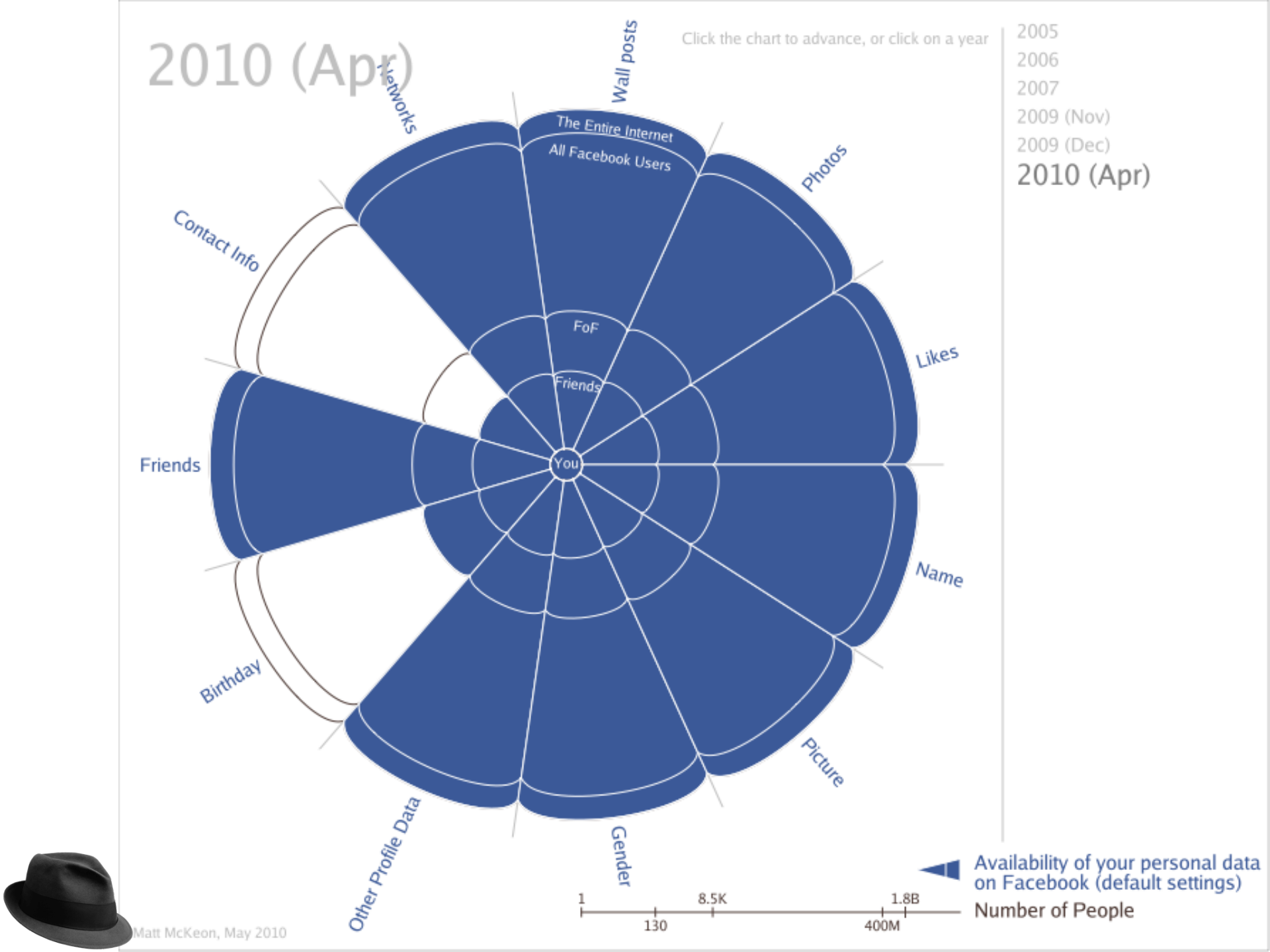

2005
Click the chart to advance, or click on a year
2005
2006
2007
2009 (Nov)
2009 (Dec)
2010 (Apr)
Wall posts
Photos
Likes
Name
Picture
Gender
Other Profile Data
Birthday
Friends
Contact Info
Networks
The Entire Internet
All Facebook Users
Network
Friends
You
Availability of your personal data on Facebook (default settings)
1
40
1K
5M
1B
Number of People
Matt McKeon, May 2010


2006
Click the chart to advance, or click on a year
2005
2006
2007
2009 (Nov)
2009 (Dec)
2010 (Apr)
Wall posts
Photos
Likes
Name
Picture
Gender
Other Profile Data
Birthday
Friends
Contact Info
Networks
The Entire Internet
All Facebook Users
Network
Friends
You
Availability of your personal data on Facebook (default settings)
1
70
10K
12M
1.1B
Number of People
Matt McKeon, May 2010


2007
Click the chart to advance, or click on a year
2005
2006
2007
2009 (Nov)
2009 (Dec)
2010 (Apr)
Networks
Wall posts
Photos
Likes
Name
Picture
Gender
Other Profile Data
Birthday
Friends
Contact Info
The Entire Internet
All Facebook Users
Network
Friends
You
Availability of your personal data on Facebook (default settings)
1
100
15K
50M
1.3B
Number of People
Matt McKeon, May 2010


2009 (Nov)
Click the chart to advance, or click on a year
2005
2006
2007
2009 (Nov)
2009 (Dec)
2010 (Apr)
Networks
Wall posts
Photos
Likes
Name
Picture
Gender
Other Profile Data
Birthday
Friends
Contact Info
The Entire Internet
All Facebook Users
FoF
Friends
You
Availability of your personal data on Facebook (default settings)
1
8.5K
1.8B
130
350M
Number of People
Matt McKeon, May 2010

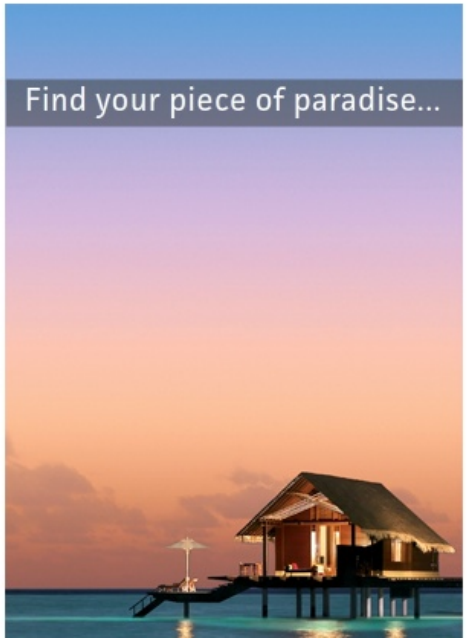

2009 (Dec)
Click the chart to advance, or click on a year
2005
2006
2007
2009 (Nov)
2009 (Dec)
2010 (Apr)
Networks
Wall posts
Photos
Likes
Name
Picture
Gender
Other Profile Data
Birthday
Friends
Contact Info
The Entire Internet
All Facebook Users
FoF
Friends
You
Availability of your personal data on Facebook (default settings)
1
8.5K
1.8B
130
390M
Number of People
Matt McKeon, May 2010


2010 (Apr)
Click the chart to advance, or click on a year
2005
2006
2007
2009 (Nov)
2009 (Dec)
2010 (Apr)
Networks
Wall posts
Photos
Likes
Name
Picture
Gender
Other Profile Data
Birthday
Friends
Contact Info
The Entire Internet
All Facebook Users
FoF
Friends
You
Availability of your personal data on Facebook (default settings)
1
8.5K
1.8B
130
400M
Number of People
Matt McKeon, May 2010

WIRED

You are Facebook's product, not customer

TECHNOLOGY / 21 SEPTEMBER 11 / by OLIVIA SOLON



People need to understand that they are the product of Facebook and not the customer, according to media theorist and writer Douglas Rushkoff.

Speaking at the inaugural Hello Etsy conference in Berlin, the author of *Program or Be Programmed* said: "Ask a kid what Facebook is for and they'll answer 'it's there to help me make friends'. Facebook's boardroom isn't talking about how to make Johnny more friends. It's talking about how to monetise Johnnny's social graph."

Flickr.com/designbyfront

Losse opmerkingen

- Lang niet iedereen is interessant ;-)
- False positives desastreus
- Het Internet vergeet niet
- Asymtotisch verlies aan privacy
- Proberen te verdwijnen valt net zo goed op!
- Regeringen kunnen eng zijn, vergeet industrie niet!
- NSA candy store:
  - http://en.wikipedia.org/wiki/NSA_ANT_catalog

The constant factor in our field is Change!

The 50 years it took Physicists to find one particle, the Higgs, we came from:

"Fortran goto", Unix, c, SmallTalk, DECnet, TCP/IP, c++, Internet, WWW, Semantic Web, Photonic networks, Google, grid, cloud, Data^3, App

to:

DDOS attacks destroying Banks and Bitcoins.

Conclusion:

Need for Safe, Smart, Resilient Sustainable Infrastructure.

CHANGE!