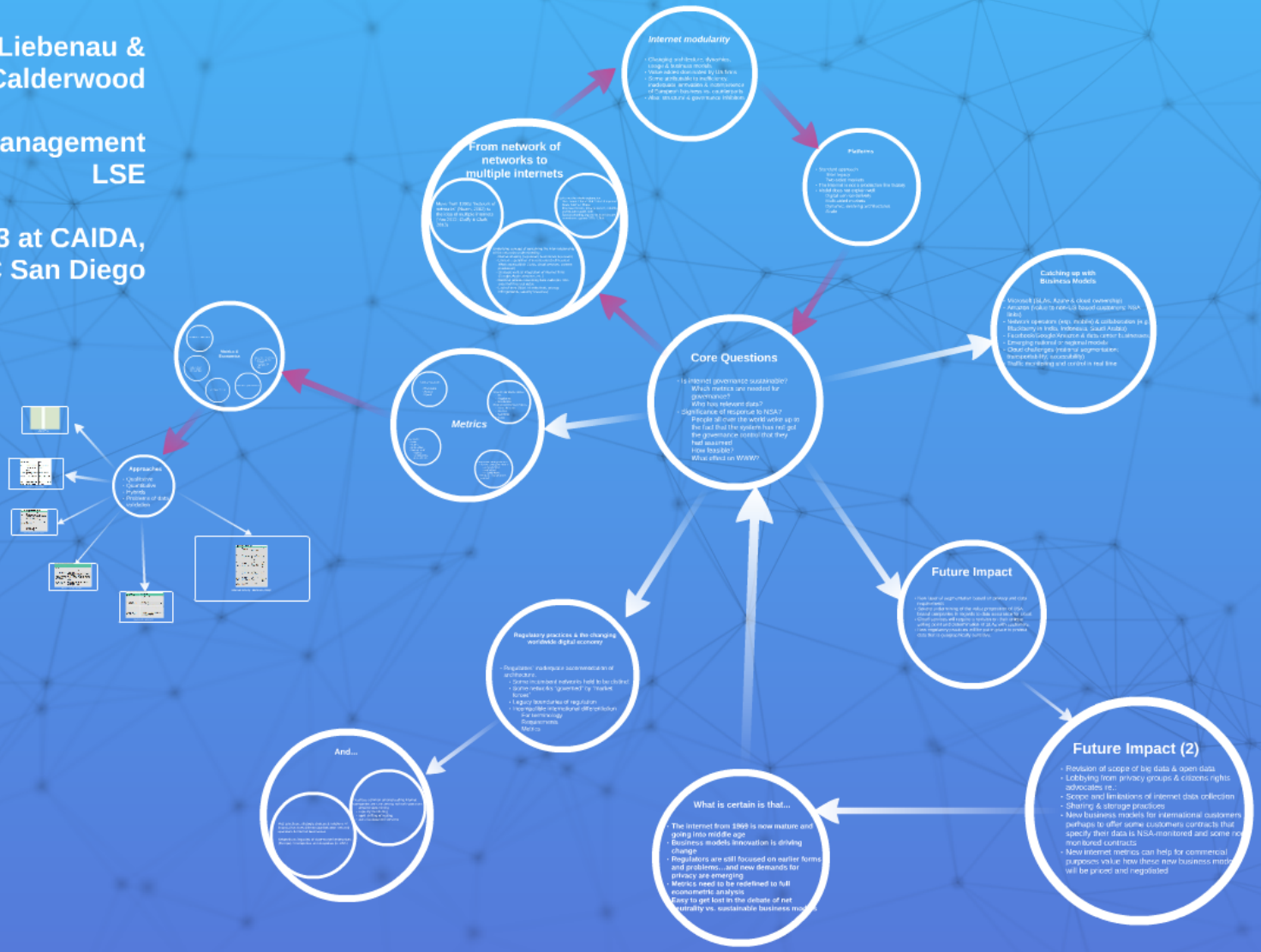
Measuring what might be regulated;
Regulating what might be measured in the internet

Dr. Jonathan Liebenau &
Dr. Silvia Elaluf-Calderwood

Department of Management
LSE

WIE 2013 at CAIDA,
UC San Diego



Measuring what might be regulated; Regulating what might be measured in the internet

Dr. Jonathan Liebenau &
Dr. Silvia Elaluf-Calderwood

Department of Management
LSE

WIE 2013 at CAIDA,
UC San Diego

Core Questions

- Is internet governance sustainable?
  - Which metrics are needed for governance?
  - Who has relevant data?
- Significance of response to NSA?
  - People all over the world woke up to the fact that the system has not got the governance control that they had assumed
  - How feasible?
  - What effect on WWW?

From network of networks to multiple internets

Move from 1990s “network of networks” (Noam, 2002) to the idea of multiple internets (Yoo 2012; Claffy & Clark, 2013)

NSA is not the whole explanation:

- NSA reveals forces that makes it apparent
- Starts 2007 w/ iPhone
- Business models, privacy, security, identity, architecture rapidly shift
- Special pleading arguments from telecom incumbents against OTTs; CDNs

Underlying concept of containing the inter-relationship of the networks undermined by:

- Market shaping (regulatory boundaries & powers)
- Limited capabilities of incumbents (half-hearted efforts to establish CDNs, cloud services, content production)
- Strategic vertical integration of internet firms (Google, Apple, Amazon, etc.)
- National policies restricting data trade (ex. non-export of financial data)
- Lack of trust (NSA; identity theft; privacy infringements, security breaches)

Move from 1990s “network of networks” (Noam, 2002) to the idea of multiple internets (Yoo 2012; Claffy & Clark, 2013)

Underlying conc
of the networks

affy & Clark,

Underlying concept of containing the inter-relationship of the networks undermined by:

- Market shaping (regulatory boundaries & powers)
- Limited capabilities of incumbents (half-hearted efforts to establish CDNs, cloud services, content production)
- Strategic vertical integration of internet firms (Google, Apple, Amazon, etc.)
- National policies restricting data trade (ex. non-export of financial data)
- Lack of trust (NSA; identity theft; privacy infringements, security breaches)

nets

NSA is not the whole explanation:

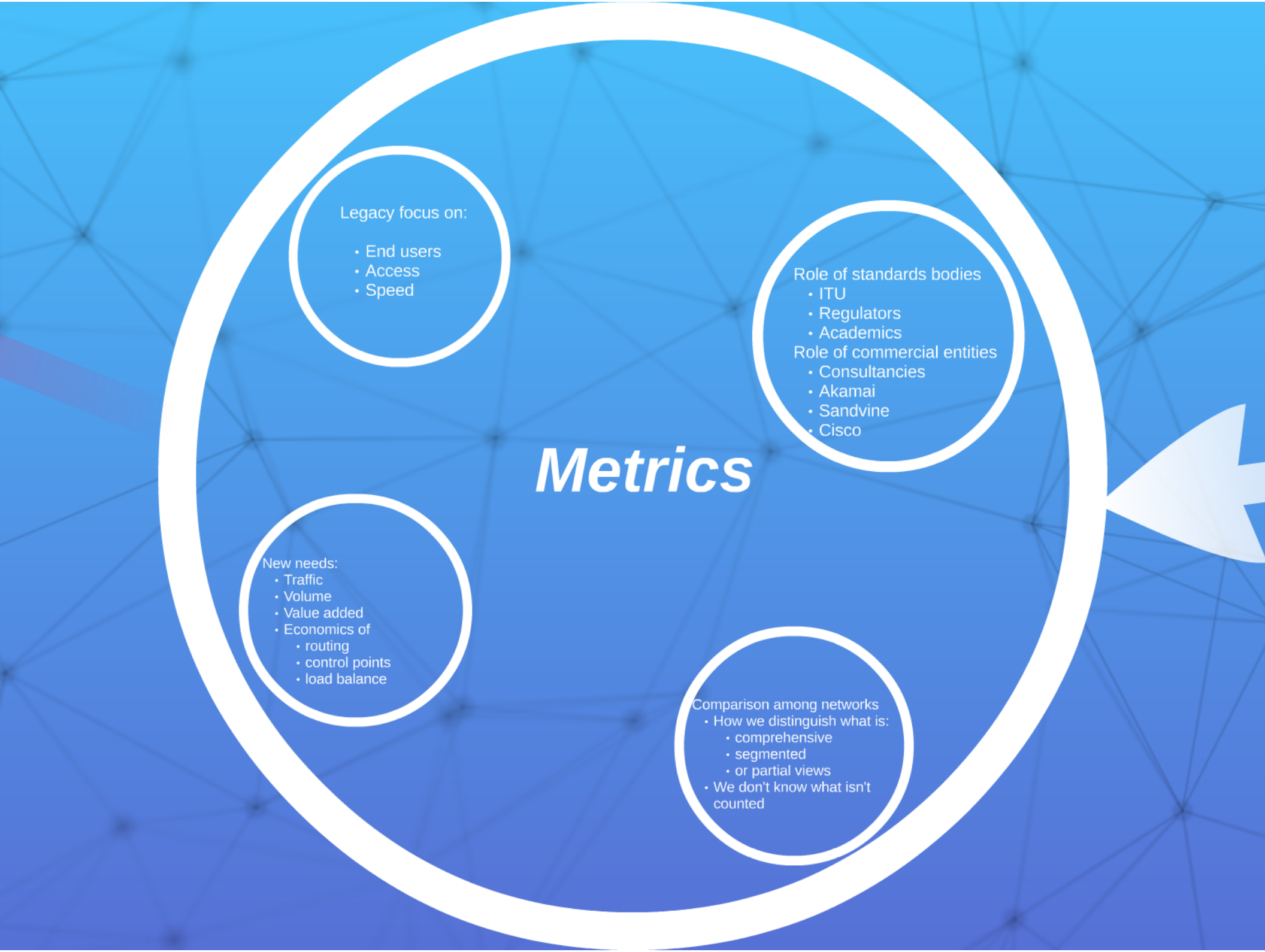
- NSA reveals forces that makes it apparent
- Starts 2007 w/ iPhone
- Business models, privacy, security, identity, architecture rapidly shift
- Special pleading arguments from telecom incumbents against OTTs; CDNs

Internet modularity

- Changing architecture, dynamics, usage & business models
- Value added dominated by US firms
- Some attributable to inefficiency, inadequate innovation & incompetence of European business vs. counterparts
- Also: structural & governance inhibitors

Platforms

- Standard approach
  - Intel legacy
  - Two-sided markets
- The internet is not a production line factory
- Model does not explain well:
  - Digital service delivery
  - Multi-sided markets
  - Dynamic, evolving architectures
  - Scale


Legacy focus on:
• End users
• Access
• Speed
Role of standards bodies
• ITU
• Regulators
• Academics
Role of commercial entities
• Consultancies
• Akamai
• Sandvine
• Cisco
Metrics
New needs:
• Traffic
• Volume
• Value added
• Economics of
• routing
• control points
• load balance
Comparison among networks
• How we distinguish what is:
• comprehensive
• segmented
• or partial views
• We don't know what isn't counted

Legacy focus on:

- End users
- Access
- Speed

New needs:

- Traffic
- Volume
- Value added
- Economics of
  - routing
  - control points
  - load balance

Comparison among networks

- How we distinguish what is:
  - comprehensive
  - segmented
  - or partial views
- We don't know what isn't counted

Role of standards bodies

- ITU
- Regulators
- Academics

Role of commercial entities

- Consultancies
- Akamai
- Sandvine
- Cisco


Metrics &
Economics
Economic variables of data
Cost of data vs.
cost of transport
(Lehr et al. 2011)
Asymmetric behavior
Market shaping and power issues
Differences in perspective
of traffic management
• IP transit
• private peering
• public peering
• CDN

Economic variables of data

Cost of data vs.
cost of transport
(Lehr et al. 2011)

Asymmetric behavior

Market shaping and power issues

Differences in perspective of traffic management

- IP transit
- private peering
- public peering
- CDN

Approaches

- Qualitative
- Quantitative
- Hybrids
- Problems of data validation

Report	Comments	Economic Analysis
Collectors of traffic counters connected to routers (e.g. CiscoVNI data)	Information based on Cisco's own collection. Considered industry benchmark. Lots of exceptions. No QoS. Traffic is analysed in categories that are not matched by other industry reports	No correlation between traffic and pricing., but regular reports comment on economic context.
Reports from firms generating traffic	Google, YouTube, Facebook http://www.google.com/transparencyreport/traffic/#expand=TJ%20or	No indicators for comparison
Classification of traffic: e.g. Sandvine	Focus sectors e.g. entertainment, home roaming, comm services. Sandvine Global Internet Phenomena Report	Some based primarily on billing, by type of traffic
Academic studies on the internet e.g. Economides	Theoretical and practical academic analysis of current status of the internet	Yes. Some of the pricing is indirectly estimated
Consultancy reports commissioned by interested parties	(incumbents, consumer groups, industry assoc., etc.) e.g. BCG, ATKearney (2013) Report on traffic and demand based on ETNO members data	Yes, based on own commissioned calculations
Regional, national, state regulatory frameworks	FCC http://www.gpo.gov/fdsys/pkg/FR-2011-09-23/html/2011-24259.htm	Yes. Based on specialized studies FCC commissioned

Explanation of Traffic Categories

The table below describes each of the traffic categories used in the Global Internet Phenomena Report: 2H 2013

Traffic Category	Description	Examples
Storage	Large data transfers using the File Transfer Protocol or its derivatives. Services that provide file-hosting, network back-up, and one-click downloads	FTP, Rapidshare, Mozy, zShare, Carbonite, Dropbox
Gaming	Console and PC gaming, console download traffic, game updates	Nintendo Wii, Xbox Live, Playstation 2, Playstation 3, PC games
Marketplaces	Marketplaces where subscribers can purchase and download media including applications, music, movies, books, and software updates	Google Android Marketplace, Apple iTunes, Windows Update
Administration	Applications and services used to administer the network	DNS, ICMP, NTP, SNMP
Filesharing	Filesharing applications that use a peer-to-peer or Newsgroups as a distribution models	BitTorrent, eDonkey, Gnutella, Ares, Newsgroups
Communications	Applications, services and protocols that allow email, chat, voice, and video communications; information sharing (photos, status, etc) between users	Skype, WhatsApp, iMessage, FaceTime
Real-Time Entertainment	Applications and protocols that allow "on-demand" entertainment that is consumed (viewed or heard) as it arrives	Streamed or buffered audio and video (RTSP, RTP, RTMP, Flash, MPEG), peercasting (PPStream, Octoshape), specific streaming sites and services (Netflix, Hulu, YouTube, Spotify,)
Social Networking	Websites and services focused on enabling interaction (chat, communication) and information sharing (photos, status, etc) between users	Facebook, Twitter, Linkedin, Instagram
Tunneling	Protocols and services that allow remote access to network resources or mask application identity.	Remote Desktop, VNC, PC Anywhere, SSL, SSH,
Web Browsing	Web protocols and specific websites	HTTP, WAP browsing

Sandvine traffic categories (2013)

TABLE 1: METRICS AND MEASURES FOR INTERNET INFRASTRUCTURE AND ACCESS

Metric	Data Sources
Penetration	• Business surveys: total number of subscriptions • Household and consumer surveys: proportion of houses connected to Internet/broadband
Speed	• Business surveys and market research: advertised speeds (e.g., OECD, FCC) • Content delivery networks and web services: download speeds (e.g., Akamai, Netflix) • Distributed client-side hardware: download and upload speeds (e.g., government partnerships with SamKnows) • Crowdsourcing: download and upload speeds (e.g., Ookla's Speedtest, M-Lab)
Price	• Market research: comparison of offers across different ISPs and countries (e.g., OECD, FCC) • Crowdsourcing: user-submitted information on prices (e.g., Ookla's Net Index)
Infrastructure: location, size, and routing	• IP address distribution • Allocation of domains • Number of Internet hosts • Number, size, and relationships of autonomous systems (AS) • Network bandwidth estimates • Internet exchange (IX) location and traffic • Route identification and analysis • National network status (e.g., Renesys, Arbor Networks) • International pipe location, traffic, and dependencies

Traditional metrics - Source Berkman Institute (2013)

TABLE 3: DATA SOURCES FOR MEASURING ONLINE ACTIVITY

General Data Type(s)/Origin(s)	Specific Source(s)
Reporting on individual behavior	• Client-side behavioral monitoring software (e.g., ComScore, Alexa) • Cookies and browsing history • Consumer surveys
Network monitoring: location, type, and quantity of traffic	• Monitoring by ISPs • Monitoring by network services (e.g., content distribution networks, Internet security companies)
Data collection by websites and services: visitors, contributors, content, links, comments, languages, locations	• Websites, including social media platforms (e.g., user-generated content sites, social network sites, blogging and micro-blogging sites) • Search data
Social media mapping: link- and/or content-based	• Landscape mapping: platform/service-based mapping (e.g., Twitter, Facebook, blogosphere) • Topical or issue-based mapping
Qualitative assessments	• Expert opinion surveys

Measuring directly and indirectly - Berkman (2013)

TABLE 3: DATA SOURCES FOR MEASURING ONLINE ACTIVITY

General Data Type(s)/Origin(s)	Specific Source(s)
Reporting on individual behavior	• Client-side behavioral monitoring software (e.g., ComScor Alexa) • Cookies and browsing history • Consumer surveys
Network monitoring: location, type, and quantity of traffic	• Monitoring by ISPs • Monitoring by network services (e.g., content distribution networks, Internet security companies)
Data collection by websites and services: visitors, contributors, content, links, comments, languages, locations	• Websites, including social media platforms (e.g., user-gene content sites, social network sites, blogging and micro-blo sites) • Search data
Social media mapping: link- and/or content-based	• Landscape mapping: platform/service-based mapping (e.g Twitter, Facebook, blogosphere) • Topical or issue-based mapping
Qualitative assessments	• Expert opinion surveys

Behaviour metrics - Berkman (2013)

TABLE 4: METHODS FOR MEASURING INTERNET ACTIVITY

Method	Strengths	Limitations	Examples
Individual surveys	• Robust sampling can offer data that is representative of general populations.	• Expensive to employ. • Response biases/rely on individuals to accurately recall and report information. • Opt-in surveys lose representativeness.	• Pew Internet • Eurostat
Business surveys	• Can provide near-comprehensive coverage of sector. • Reflect actual transactional data.	• Reporting biases/industry incentives to over or under report. • Typically exclude proprietary information from public release.	• Wireline and wireless subscription data
Expert surveys	• Able to address complex impact assessments.	• Often highly subjective. • Difficult to ensure cross-respondent consistency.	• World Economic Forum • Web Index
Crowdsourcing	• Often able to achieve broad coverage. • Inexpensive. • Can be highly responsive to changing contexts and events.	• Chronic representativeness issues.	• Herdict • Ookla
Web analytics	• Can provide comprehensive view of platforms/websites.	• Typically limited to specific platforms and sites. • Limited to transactional behavior. • Often proprietary.	• YouTube • Wikipedia • Facebook • Twitter • Search
Client-side monitoring	• Detailed individual online behavior.	• Sampling/representativeness issues.	• Comscore, Alexa, browsing trackers
Network monitoring	• Can offer broad view of users' Internet transactions and traffic.	• Proprietary.	• Sandvine • Akamai
Routing and hosting infrastructure	• View of infrastructure for conducting Internet business.	• Unreliable measure of user access and behavior.	• Internet hosts • IP addresses • International bandwidth

Internet Activity - Berkman (2013)

Catching up with Business Models

- Microsoft (SLAs, Azure & cloud ownership)
- Amazon (value to non-US based customers; NSA links)
- Network operators (esp. mobile) & collaboration (e.g. Blackberry in India, Indonesia, Saudi Arabia)
- Facebook/Google/Amazon & data center businesses
- Emerging national or regional models
- Cloud challenges (national segmentation; transportability; accessibility)
- Traffic monitoring and control in real time

Regulatory practices & the changing worldwide digital economy

- Regulators' inadequate accommodation of architecture.
  - Some incumbent networks held to be distinct
  - Some networks "governed" by "market forces"
  - Legacy boundaries of regulation
  - Incompatible international differentiation
    - For terminology
    - Requirements
    - Metrics

And...

Practices common among leading internet companies are rare among network operators:

- detailed data mining
- capacity monitoring
- rapid shifting of routing
- use of outsourced services

Mgt. practices, strategic choices & relations w/ financial services differentiate between network operators & internet businesses

Networks as legacies of state-owned enterprises (Europe) / monopolies and duopolies (in USA)

Mgt. practices, strategic choices & relations w/ financial services differentiate between network operators & internet businesses

Networks as legacies of state-owned enterprises (Europe) / monopolies and duopolies (in USA)

Practices co
companies a
- detailed
- capacity
- rapid shi
- use of ou

And...

Practices common among leading internet companies are rare among network operators:

- detailed data mining
- capacity monitoring
- rapid shifting of routing
- use of outsourced services

Mgt. practices, strategic choices & relations w/ financial services differentiate between network operators & internet businesses

Networks as legacies of state-owned enterprises (Europe) / monopolies and duopolies (in USA)

Future Impact

- New layer of segmentation based on privacy and data requirements
- Severe undermining of the value proposition of USA based companies in regards to data assurance for cloud.
- Cloud services will require a revision on their unique selling point and determination of SLAs with customers.
- New regulatory practices will be put in place to protect data that is geographically sensitive.

Future Impact (2)

- Revision of scope of big data & open data
- Lobbying from privacy groups & citizens rights advocates re.:
- Scope and limitations of internet data collection
- Sharing & storage practices
- New business models for international customers perhaps to offer some customers contracts that specify their data is NSA-monitored and some no monitored contracts
- New internet metrics can help for commercial purposes value how these new business mode will be priced and negotiated

What is certain is that...

- The internet from 1969 is now mature and going into middle age
- Business models innovation is driving change
- Regulators are still focused on earlier forms and problems…and new demands for privacy are emerging
- Metrics need to be redefined to full econometric analysis

Easy to get lost in the debate of net neutrality vs. sustainable business models