

The Open World Framework

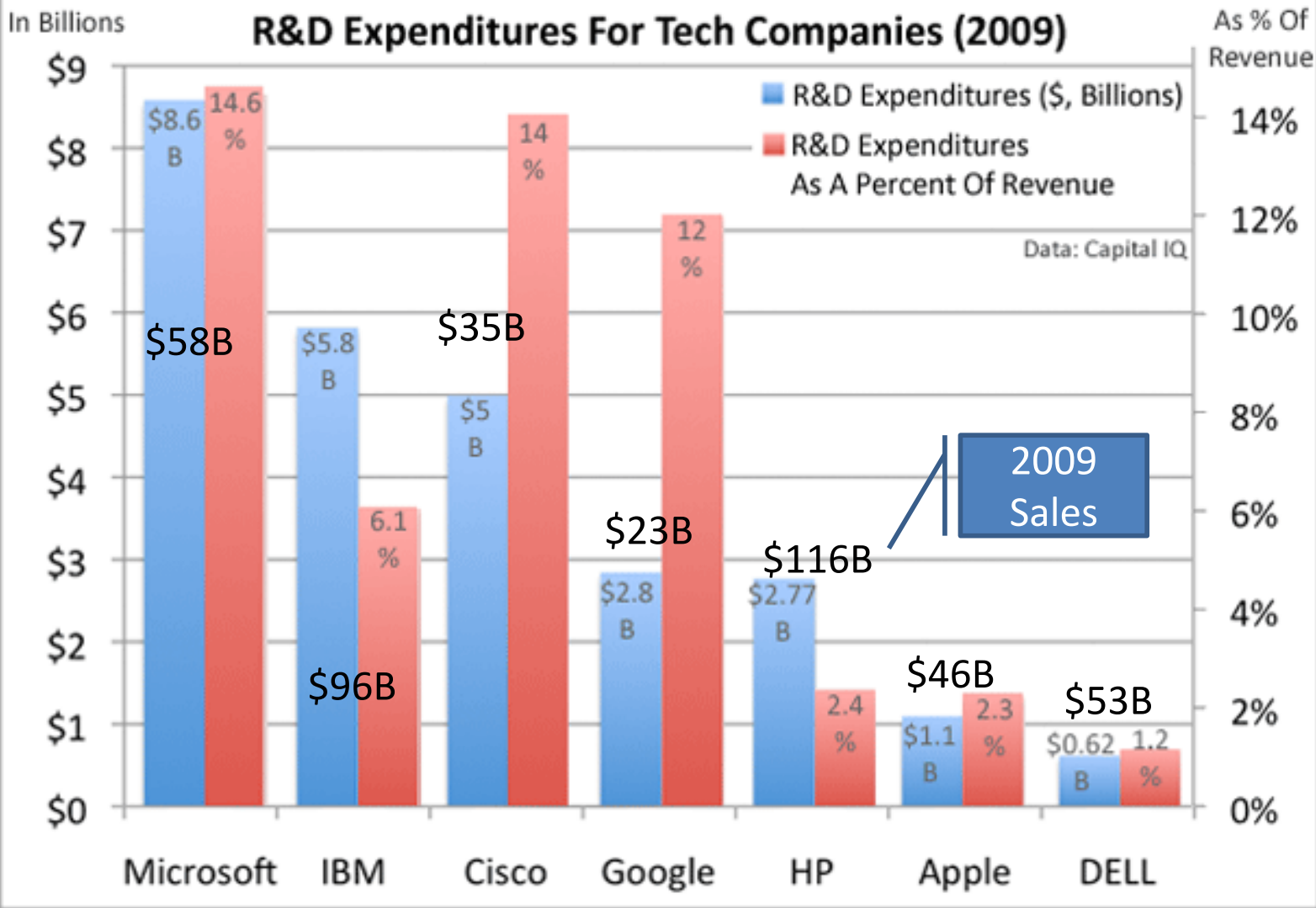
Sujata Millick
Visiting Researcher
HP Labs, Palo Alto

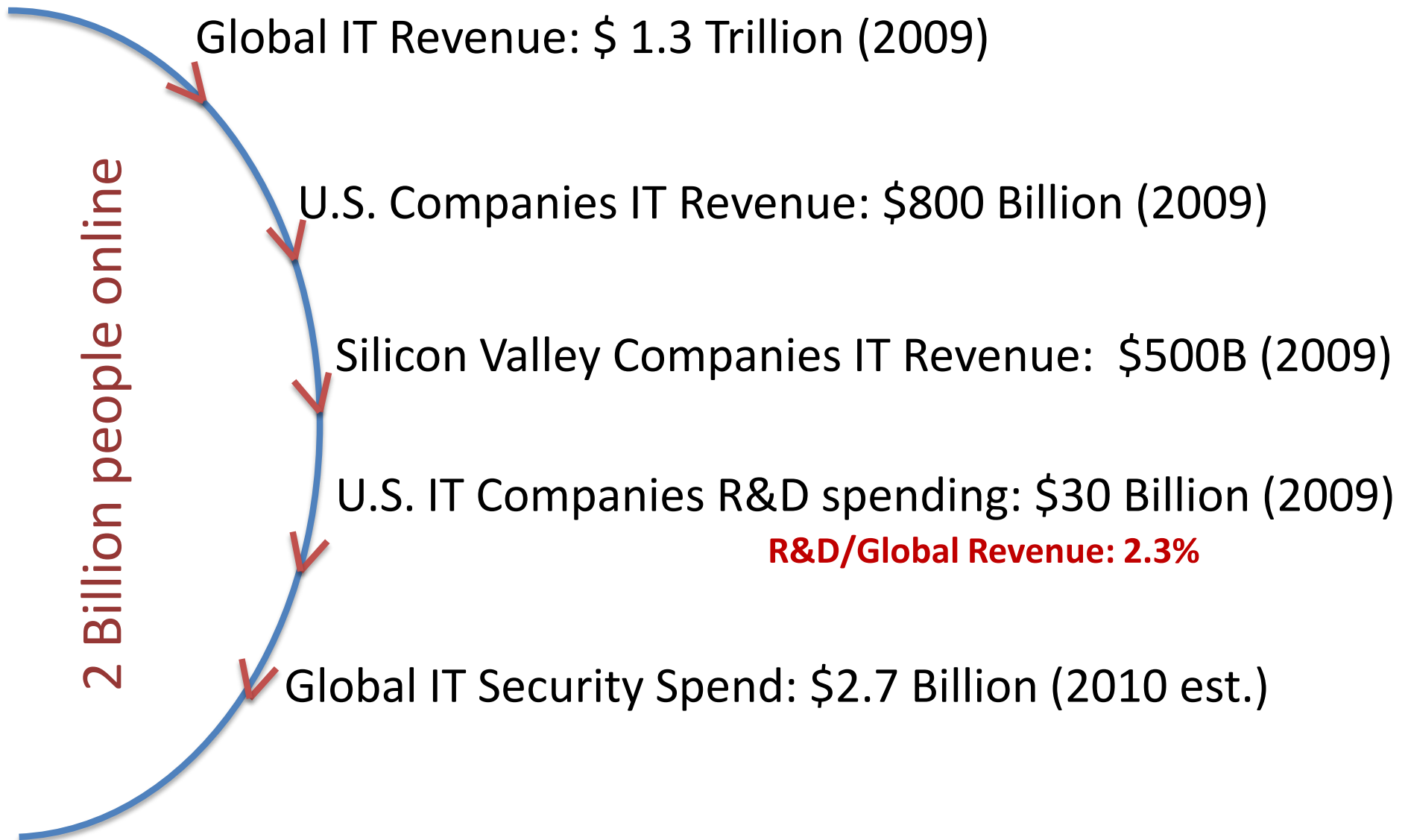
sujata.millick@hp.com
sujata.millick@navy.mil

Silicon Valley 150

RANK BY SALES				SALES (millions)			PROFIT (millions)		PROFIT MARGIN		MARKET CAP. (millions)	
2010 rank	2009 rank	Company	Business	End quarter	Last four quarters	% chg	Rank	Last four quarters	Rank	Profit/sales	Rank	Amount
1	1	Hewlett-Packard	Computers, peripherals	JAN	\$127,158	9%	3	\$9,116	85	7%	6	\$88,656
2	2	Apple	Computers	DEC	\$76,283	63%	1	\$16,639	23	22%	1	\$321,072
3	4	Intel	Semiconductors	DEC	\$43,623	24%	2	\$11,464	15	26%	4	\$110,748
4	3	Cisco Systems	Networking equipment	JAN	\$42,361	19%	6	\$7,578	36	18%	5	\$94,805
5	5	Oracle	Business software	FEB	\$34,479	43%	5	\$7,702	21	22%	3	\$169,186
6	6	Google	Internet search	DEC	\$29,321	24%	4	\$8,505	10	29%	2	\$189,076
7	13	Applied Materials	Semiconductor mfg. equip	JAN	\$10,386	88%	10	\$1,361	52	13%	13	\$20,609
8	9	Synnex	IT supply chain services	FEB	\$9,179	15%	42	\$123	111	1%	76	\$1,188
9	8	eBay	Internet-based auction services	DEC	\$9,156	5%	8	\$1,801	27	20%	7	\$40,277
10	10	Gilead Sciences	Therapeutic viral medicines	DEC	\$7,949	13%	7	\$2,901	6	36%	9	\$33,775
11	15	Sanmina-SCI	Electronic mfg. services	DEC	\$6,503	24%	48	\$91	109	1%	88	\$899
12	14	AMD	Semiconductors	DEC	\$6,494	20%	24	\$471	83	7%	35	\$5,893
13	11	Yahoo	Internet media content	DEC	\$6,325	-2%	12	\$1,232	29	19%	12	\$21,841
14	16	Franklin Resources	Investment services	DEC	\$6,309	33%	9	\$1,591	17	25%	10	\$27,901
15	12	Symantec	Network security software	DEC	\$6,048	2%	20	\$613	70	10%	18	\$14,151
16	17	Agilent Technologies	Measurement instruments	JAN	\$5,750	27%	14	\$798	47	14%	17	\$15,455
17	18	Con-way	Trucking, air freight	DEC	\$4,952	16%	113	\$4	117	0%	59	\$2,164
18	20	SanDisk	Flash-memory storage	DEC	\$4,827	35%	11	\$1,300	13	27%	23	\$10,910
19	19	NetApp	Network data storage devices	JAN	\$4,785	31%	18	\$624	54	13%	14	\$17,712
20	23	Juniper Networks	Internet infrastructure systems	DEC	\$4,093	23%	19	\$618	44	15%	11	\$22,510
21	27	Adobe Systems	Software for printing and display	FEB	\$3,969	32%	13	\$882	22	22%	15	\$16,737
22	24	Intuit	Personal financial software	JAN	\$3,554	10%	21	\$531	45	15%	16	\$16,070
23	22	Nvidia	3D-graphics processors	JAN	\$3,543	7%	32	\$253	86	7%	22	\$10,917
24	21	Electronic Arts	Entertainment software	DEC	\$3,478	-2%	150	(\$397)	135	-11%	31	\$6,529
25	25	Robert Half Int'l	Personnel services	DEC	\$3,175	5%	57	\$66	106	2%	40	\$4,477

\$387B





Outline

- **Snapshots**
 - Who's Online, Data Growth
- **Convergences**
 - Device, Relationship, E-Data
- **Changing World**
 - Information Age Model
 - Creativity, Data, Intellectual Property
- **Changing Paradigms**
 - Devices, Social Media Tools, Sentiment Analysis
- **Open World Framework**
 - IP, Relationships, Data
 - Content Creation, Collaborations, Data->Information, Security
- **Open World Experiment**

Statistic 1

WORLD INTERNET USAGE AND POPULATION STATISTICS

World Regions	Population (2010 Est.)	Internet Users Dec. 31, 2000	Internet Users Latest Data	Penetration (% Population)	Growth 2000-2010	Users % of Table
Africa	1,013,779,050	4,514,400	110,931,700	10.9 %	2,357.3 %	5.6 %
Asia	3,834,792,852	114,304,000	825,094,396	21.5 %	621.8 %	42.0 %
Europe	813,319,511	105,096,093	475,069,448	58.4 %	352.0 %	24.2 %
Middle East	212,336,924	3,284,800	63,240,946	29.8 %	1,825.3 %	3.2 %
North America	344,124,450	108,096,800	266,224,500	77.4 %	146.3 %	13.5 %
Latin America/Caribbean	592,556,972	18,068,919	204,689,836	34.5 %	1,032.8 %	10.4 %
Oceania / Australia	34,700,201	7,620,480	21,263,990	61.3 %	179.0 %	1.1 %
WORLD TOTAL	6,845,609,960	360,985,492	1,966,514,816	28.7 %	444.8 %	100.0 %

← Asia
 ← Europe
 ← N. America
79.7%

NOTES: (1) Internet Usage and World Population Statistics are for June 30, 2010.

Statistic 2

FACEBOOK USAGE AND INTERNET STATISTICS FOR AUGUST 31, 2010

Geographic World Regions	Population (2010 Est.)	Facebook Users	Facebook Penetration	Facebook Index(*)	Internet Users	Internet Penetration	
Asia	3,834,792,852	93,584,580	2.4 %	11.3 %	828,930,856	21.6 %	Asia
Africa	1,013,779,050	17,607,440	1.7 %	15.9 %	110,948,420	10.9 %	Africa
Europe	813,319,511	162,104,640	19.9 %	34.1 %	475,121,735	58.4 %	←
Latin America	550,924,250	68,189,920	12.4 %	35.0 %	195,042,230	30.4 %	
North America	344,124,450	149,054,040	43.3 %	56.0 %	266,224,500	77.4 %	←
Middle East	212,336,924	11,698,120	5.5 %	18.5 %	63,240,946	29.8 %	ME
The Caribbean	41,632,722	3,925,060	9.4 %	39.0 %	10,055,240	24.2 %	
Oceania / Australia	34,700,201	11,596,660	33.4 %	54.5 %	21,272,470	61.3 %	←
WORLD TOTAL	6,845,609,960	517,760,460	7.6 %	26.3 %	1,970,837,003	28.8 %	

NOTES: (1) Facebook World Users and Facebook World Penetration Statistics are consolidated for August 31, 2010

Statistic 3

E- Data...

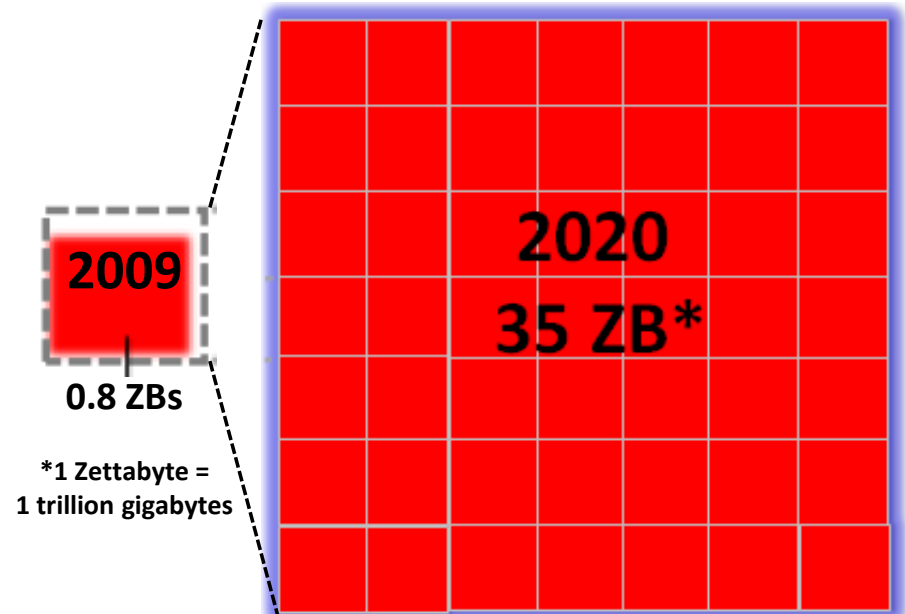
Digital data is increasing at a compounded growth rate of **60% per year** and this growth rate is expected to increase dramatically going forward

The amount of digital information increases **tenfold** every 5 years.

"Data, Data Everywhere: A special report on managing information" (The Economist, 27 Feb 2010).

"The Digital Universe Decade: Are You Ready?" (IDC – Sponsored by EMC, May 2010).

The Digital Universe: 2009-2020 Growing by a Factor of 44



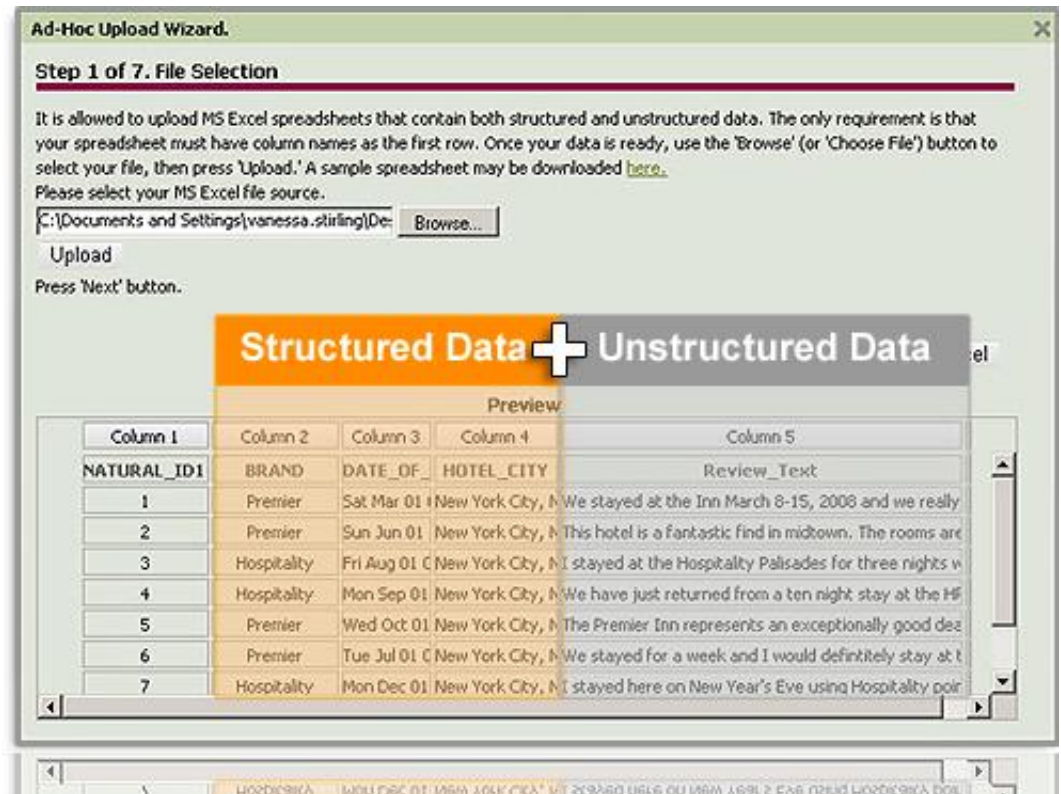
The Meaning of the Data Growth...

Structured Data

- Meta data
- Sensor data
- Sorted, Organized
- Analyzed using traditional statistical techniques
- Created by formal organizations

Unstructured Data

- Text
- Audio
- Video
- Analysis requires domain, context, sentiment info
- Created “primarily” by individual actors



Within a decade...

6 Billion Online

200% growth??

June 2010

2 Billion Online

444% growth

Dec 2000

360 Million Online

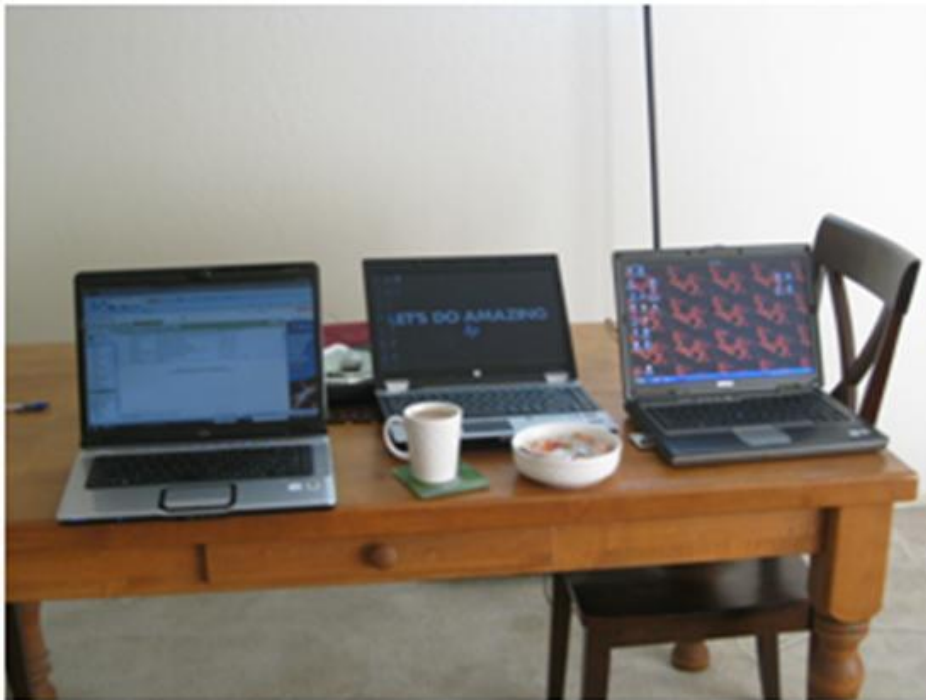
Drivers of Online Growth



Device Convergence



Relationship Convergence



Device Convergence

TV Remotes

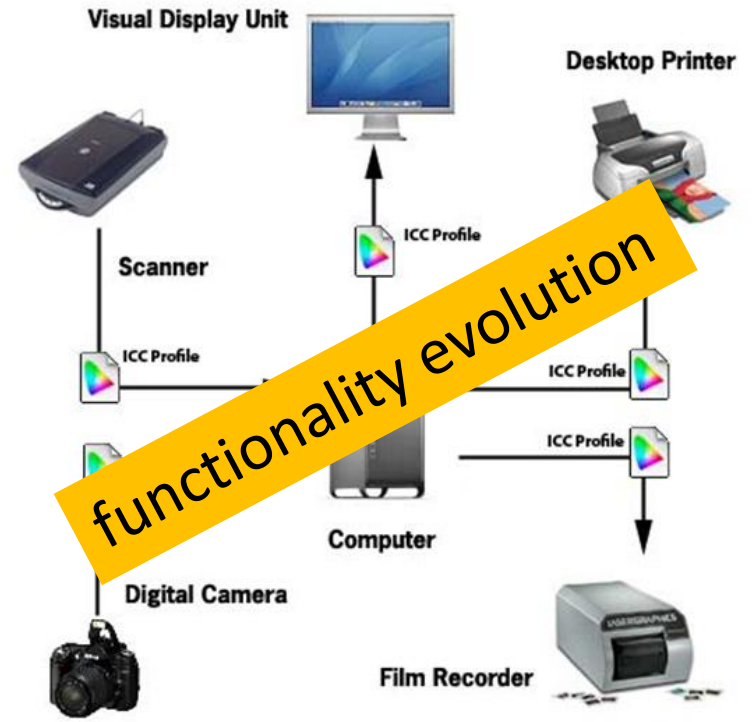


tool evolution



user evolution

Portable Entertainment



functionality evolution

Home Computer Systems

Relationship Convergence

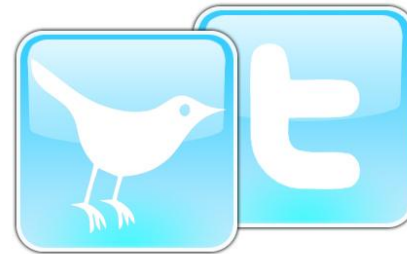
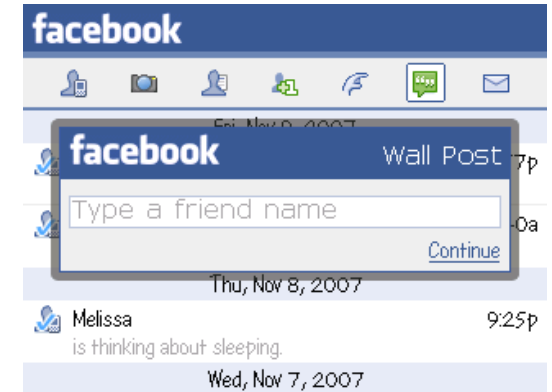
Information Technology is platform

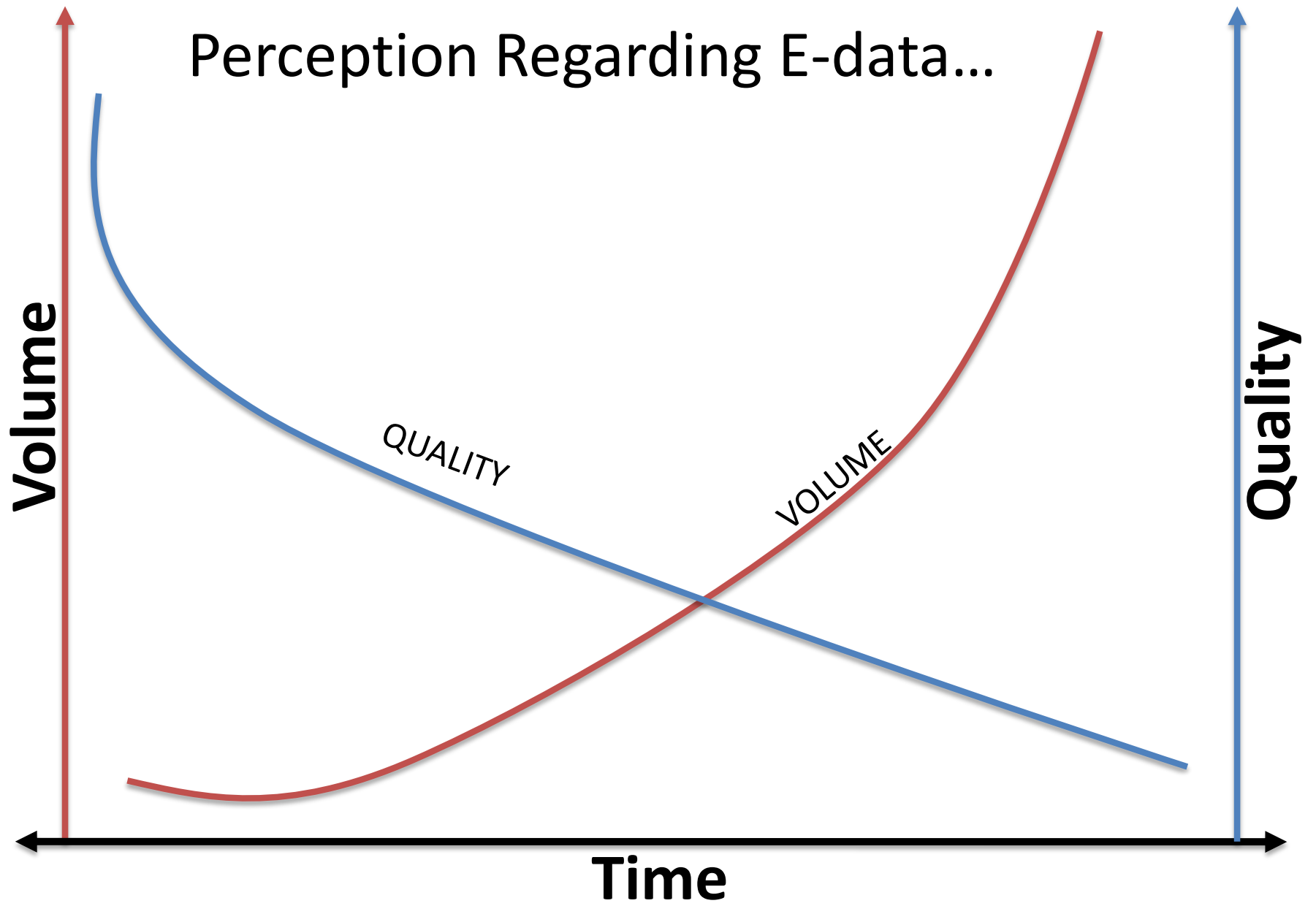
Social Networking is the medium

People (young adults) are the brains

Phones, pads, walls, blogs, tweets

are the tools





A photograph of a bright blue sky filled with scattered, fluffy white clouds. The clouds are of various sizes and shapes, some appearing as soft, rounded masses while others are more wispy and elongated. The overall scene is bright and clear, suggesting a sunny day.

Changing World

Industrial Age Model

Tangible Property

Formal Associations

Data generated and
controlled by entities

Information Age Model

Intangible Property

Informal Networks

Data volumes generated
by individuals

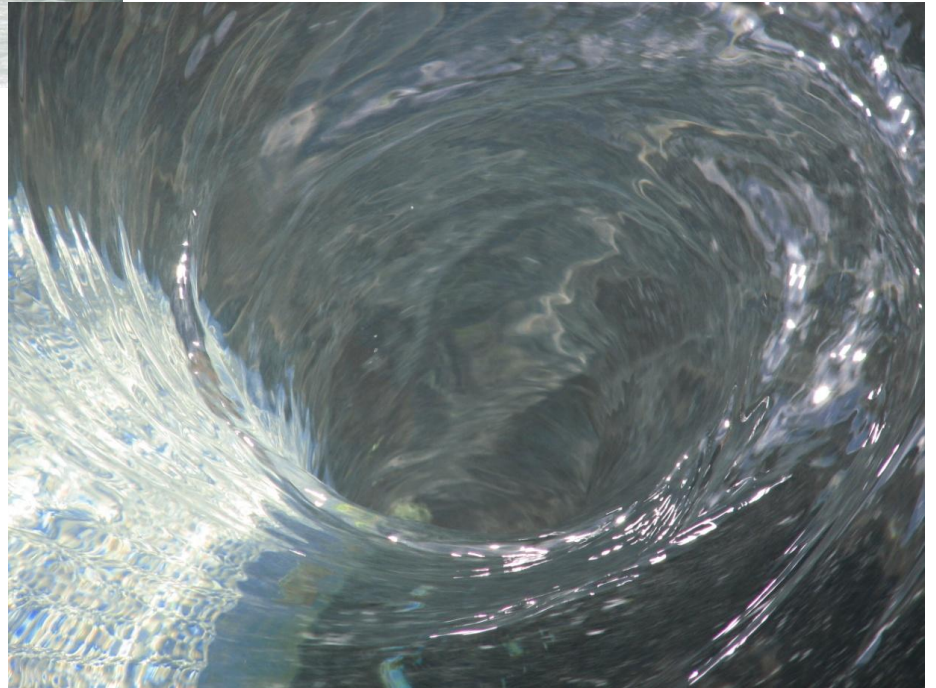


Watering Holes for Creativity?





Data Rivers for Security and Analytics?



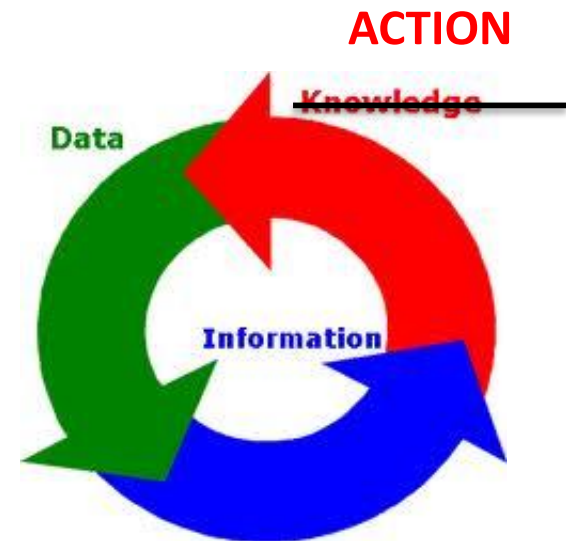
Data → Information

Traditional

- Statistical Analytics
- Correlations
- Select Population Surveys
- Experiments

New...

- Blogging
- Walls
- Tagging
- Tweeting
- Crowdsourcing
- Mashups
- Retweets



Data \leftrightarrow Information

- Tunisia, Egypt, and other
 - Information to Action predominated
 - Bypassed the “data” stage
 - Data analytics after the event
- Japan Tsunami...
 - Govt. transparency after Information

Intellectual Property

E- Content crosses....

Geography

Companies

Sovereign boundaries

Cultures

Languages

Internet Protocol Domains

What is E-Content...

Anything on the WWW

Content on formal websites

Personalized content

Early Examples of Common IP...

Open Source

Human Genome

World Wide Web

Guttenberg.org

Changing Paradigms

Devices

Social Media Tools

Sentiment Analysis

Topic: [Top](#) » Computer Science

View order

- Hot
- Popular
- Just published
- Recent
- Top Voted

Topic taxonomy

- No subtopics
- Feeling lucky

Type of content

- Event (120)
- Lecture (1830)
- Tutorial (160)
- Keynote (170)
- Interview (59)

Subtopics

Algorithmic Information Theory (3) Algorithms and Data Structures (12) **Artificial Intelligence** (76)
Bioinformatics (99) Chemoinformatics (1) Computational Biology (11) **Computational Linguistics** (30)
Computer Graphics (8) **Computer Vision** (159) Cryptography and Security (5) **Databases** (30)
Data Mining (252) Data Visualisation (21) Decision Support (4) Digital Signal Processing (5) Discrete Optimization (1)
Environmental Computation (11) Evolutionary Computation (3) Fuzzy Logic (8) Grid Computing (1) Human Computer Interaction (15)
Image Analysis (65) Information design (2) **Information Extraction** (46) **Information Retrieval** (100)
Information Theory (1) Intelligent Agents (8) **Interviews** (57) Logic (12) **Machine Learning** (971)
Multilingual Information Access (10) **Natural Language Processing** (61) **Network Analysis** (100)
Optimization Methods (8) **Pattern Recognition** (87) Programming languages (9) **Robotics** (40)
Search Engines (12) Semantic Search (14) **Semantic Web** (338) Software and Tools (15)
Spatial Data Structures (3) **Speech Analysis** (17) **Text Mining** (80) **Web Mining** (73) **Web Search** (22)

2479 items listed...

Page 1 of 65 > Next Last

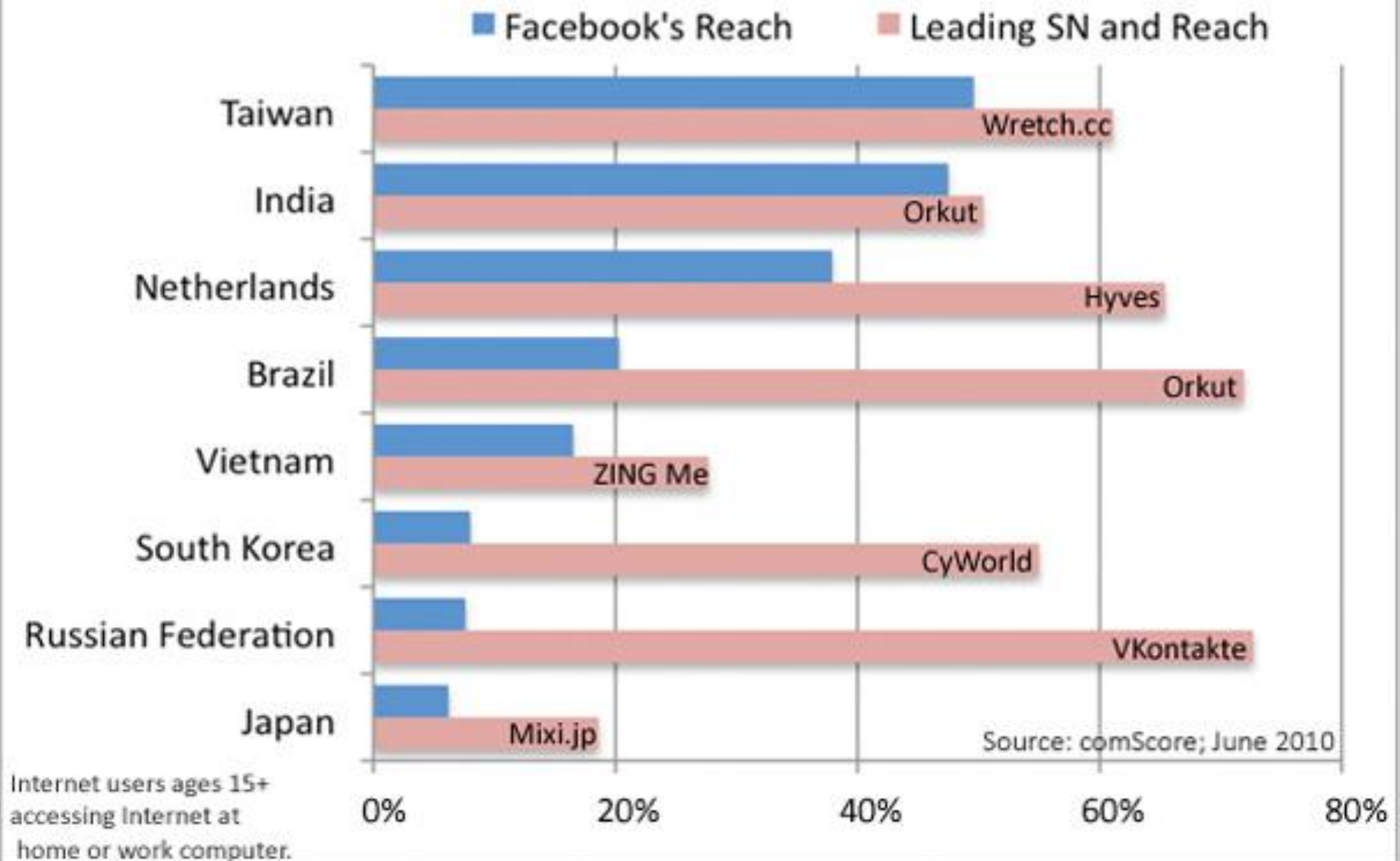
Cloud Tags

Context – Computer Science

Volume/Quantity, Domain Growth



Countries Where Facebook Is NOT The #1 Social Network By Reach



twitter Home Profile Mes

 **Sohaib Athar**
@ReallyVirtual Abbottabad Lahore Pakistan
An IT consultant taking a break from the rat-race by hiding in the mountains with his laptops.
<http://www.reallyvirtual.com>

+ Follow

Timeline Favorites Following Followers Lists

 **ReallyVirtual** Sohaib Athar
Here's a picture of OBL's hideout in Abbottabad, as shared by a friend @Rahat <http://yfrog.com/h7w4izmj>
7 minutes ago

 **ReallyVirtual** Sohaib Athar
I apologize for reporting the operation 'unwittingly/unknowingly' - had I known about it, I would have tweeted about it 'wittingly' I swear.
21 minutes ago

 **ReallyVirtual** Sohaib Athar
I just woke up and need to take care of some business before I (re)start blogging - please don't call me just yet. Thank you!
1 hour ago

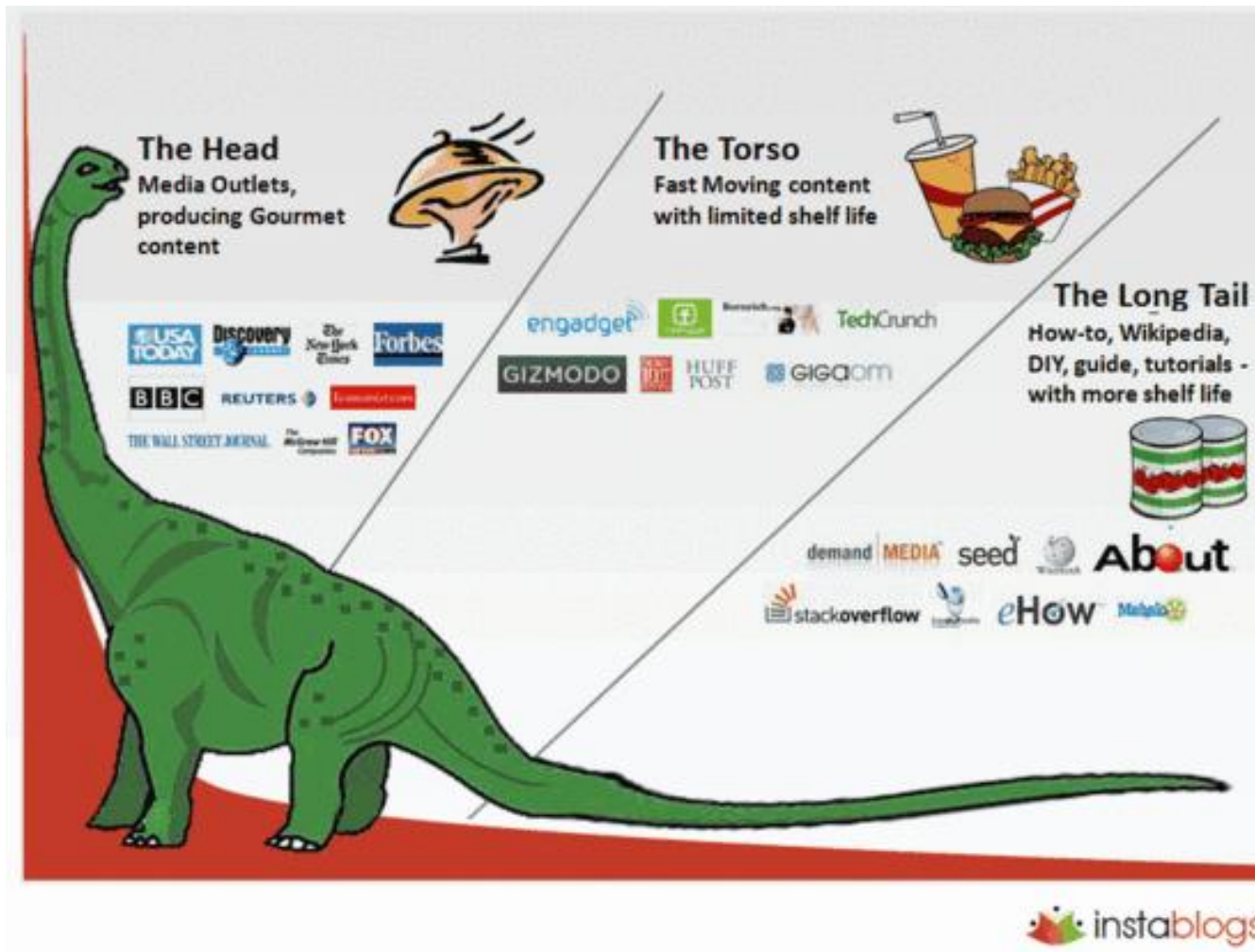
 **ReallyVirtual** Sohaib Athar

Abbottabad

Twitter User: #ReallyVirtual

* Did anyone monitor tweets coming out locally? (IP address, geography...)

Twitter Stats:
140+ Million registered users
60+ million tweets a day...



What's the value of "he knew before the world knew"?

Social analytics for long tails...

Devices of the Future...

- ...will be Mobile
- +
- ...work connect to a Cloud
- +
- ...will use low power/energy
- +
- ...will have composable Security

Form and fit of the device
Compact form
Personalized fit

Massive functionality
Seamless connectivity
Interchangeable
Mobile
Secure
Always on

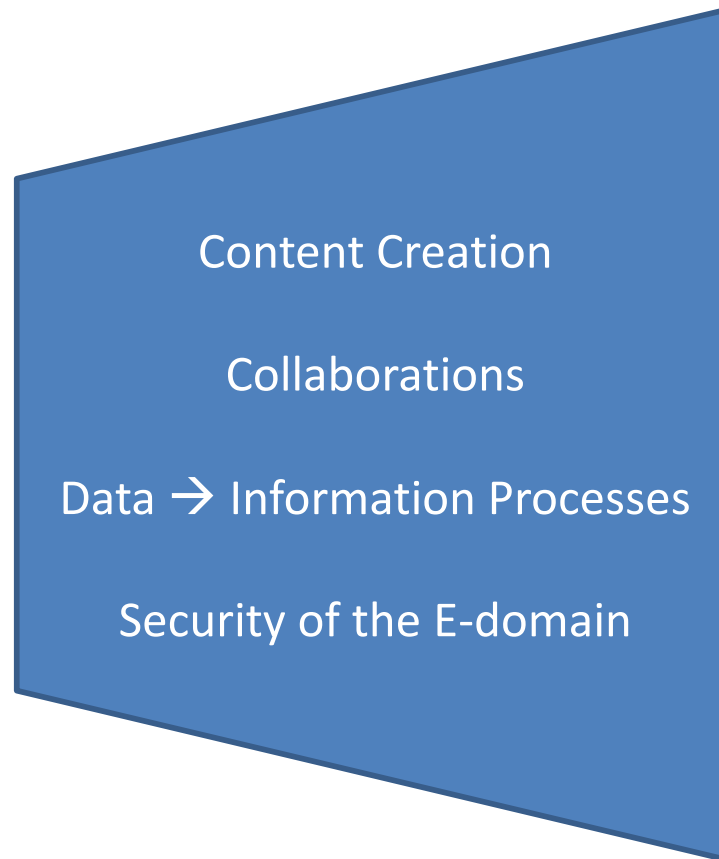


Open World

The human race is online....

Devices are always on...

Unstructured data is constantly created...



Watering Hole....
Collaborations
Content Creation



Raging River...
Data Analytics
E-Security



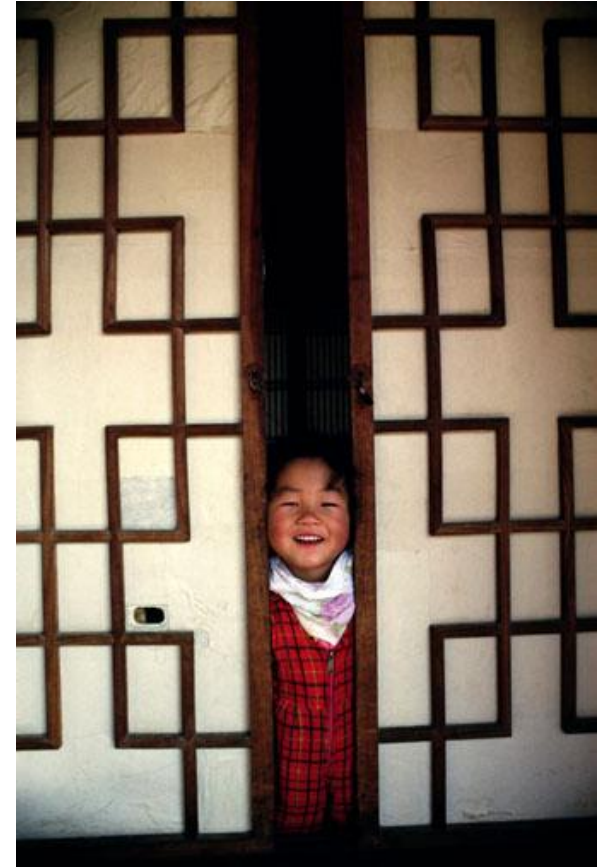
1. What is the Open World? What does it look like and how would we know it from anything else?
2. What are some new phenomena in the Open World that we may contend with?
3. What new technical problems and challenges will the Open World force us to face?
4. Are there security challenges in the Open World?
5. What are some examples of Open World phenomena that may already have made a presence and can we measure/manipulate them today?
6. Where will the Open World be manifested? Will it be in the cloud infrastructure that is coming up or in all parts of the Internet?
7. What does privacy look like in a world where everybody knows everything you say, everything you do, everywhere you go, and everyone you meet?
8. Would it be possible to establish the provenance of information in an open world?
9. What happens to intellectual property in an open world?
10. What incentives will there be to create and invest in intellectual property?
11. What would be a systematic way to gather information from multiple sources, combine it, to turn it into information that has more value than the sum of its parts?
14. Who controls information in an open world?
15. What are the implications of an open world on unstructured data analytics?
16. What constitutes an open data world, who are the players, at what rates and types of content will they play in/with?
17. Why is the Open World important? What are the consequences and changes that it will bring?
18. What is the opportunity for us now?
19. How can an open world be simulated and what analytics problems will best challenge this open world?
20. What could the consequences of ignoring the Open World be?
21. Suppose your content could seamlessly move not only between devices you own, but between billions of potential devices you encounter in your everyday life. How could you maintain any kind of confidentiality in such a world?

Open World Questions...



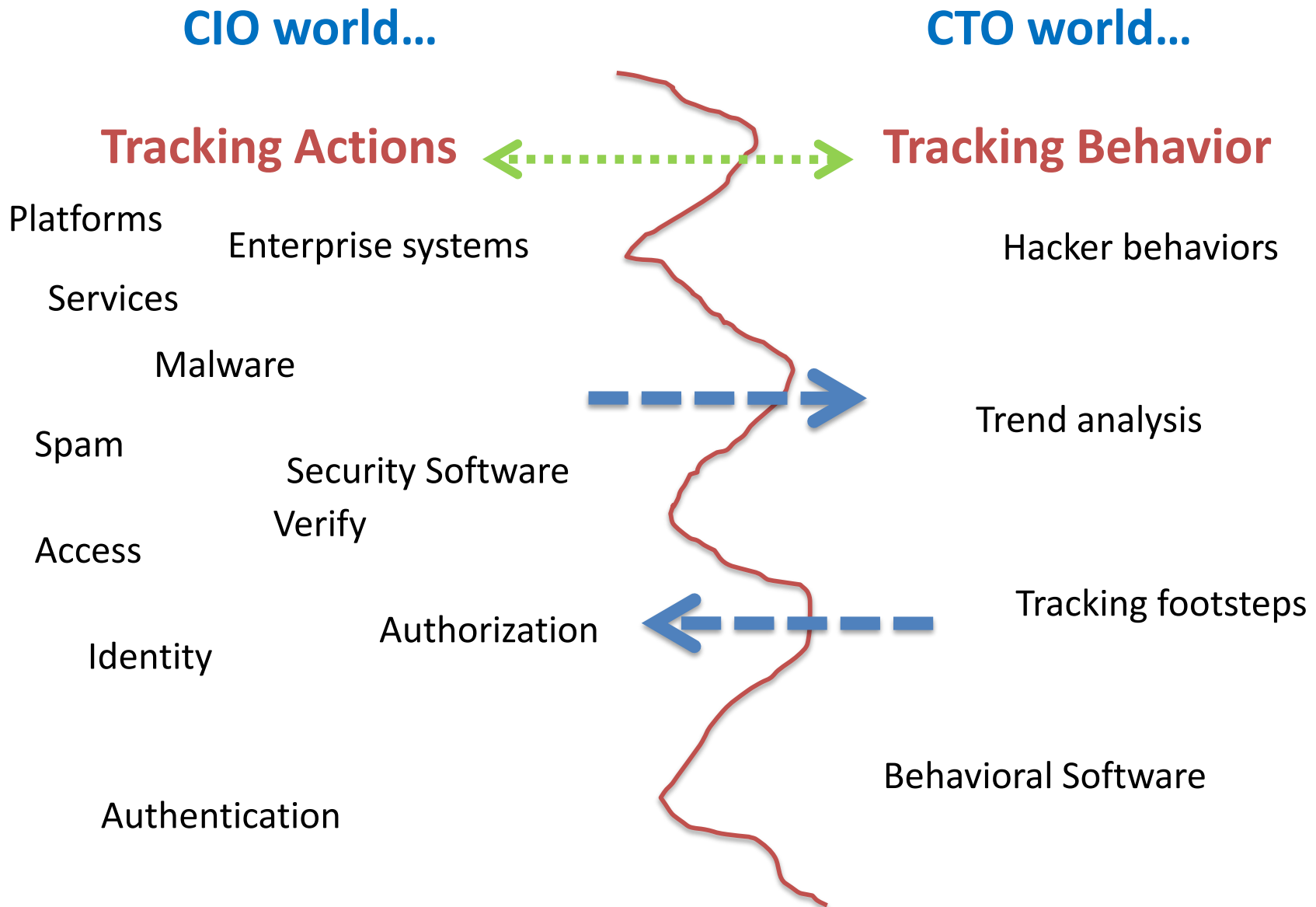
Why should you play?

How can you play?

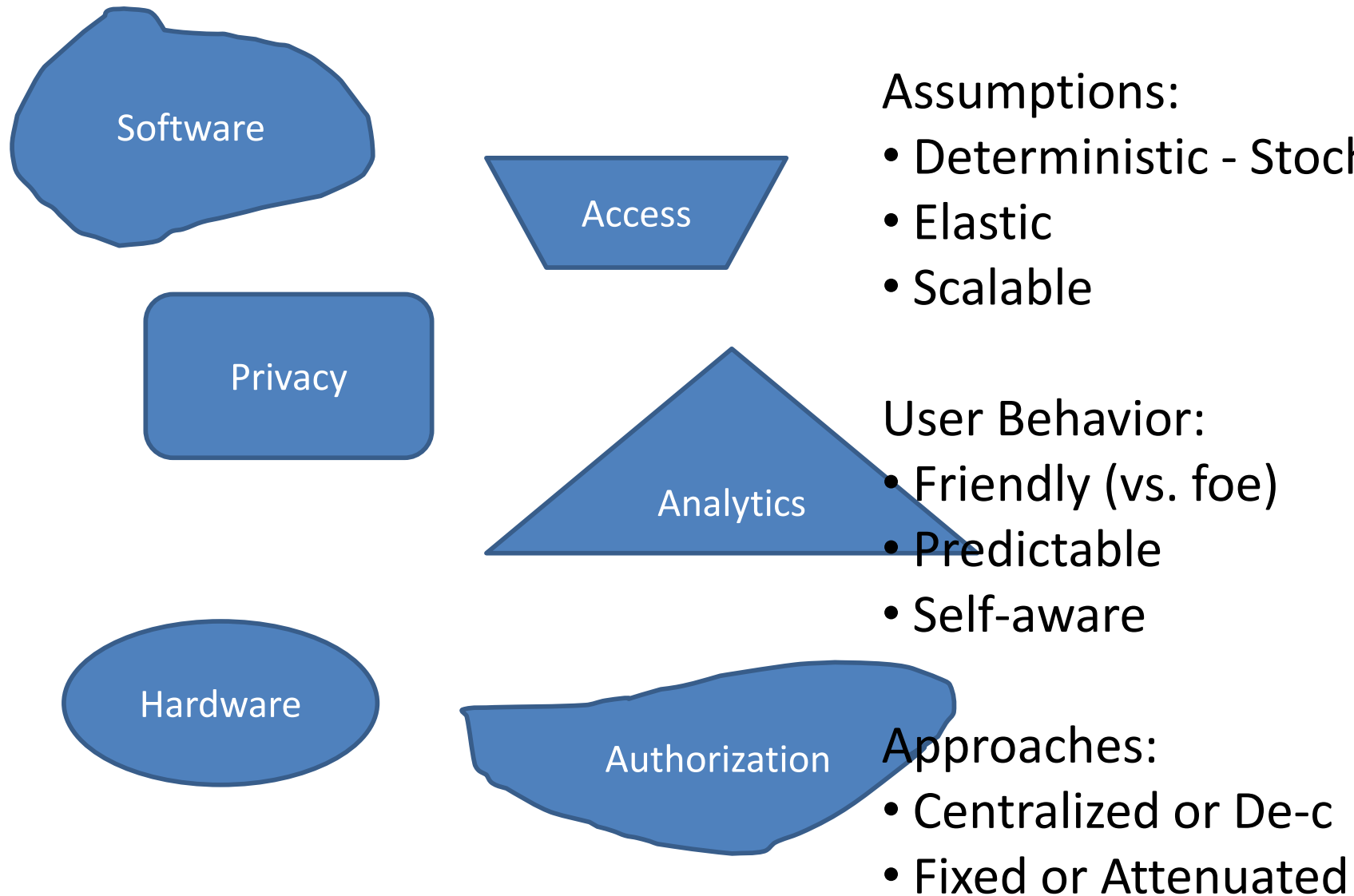


Actors/Agents managing the E-domain...

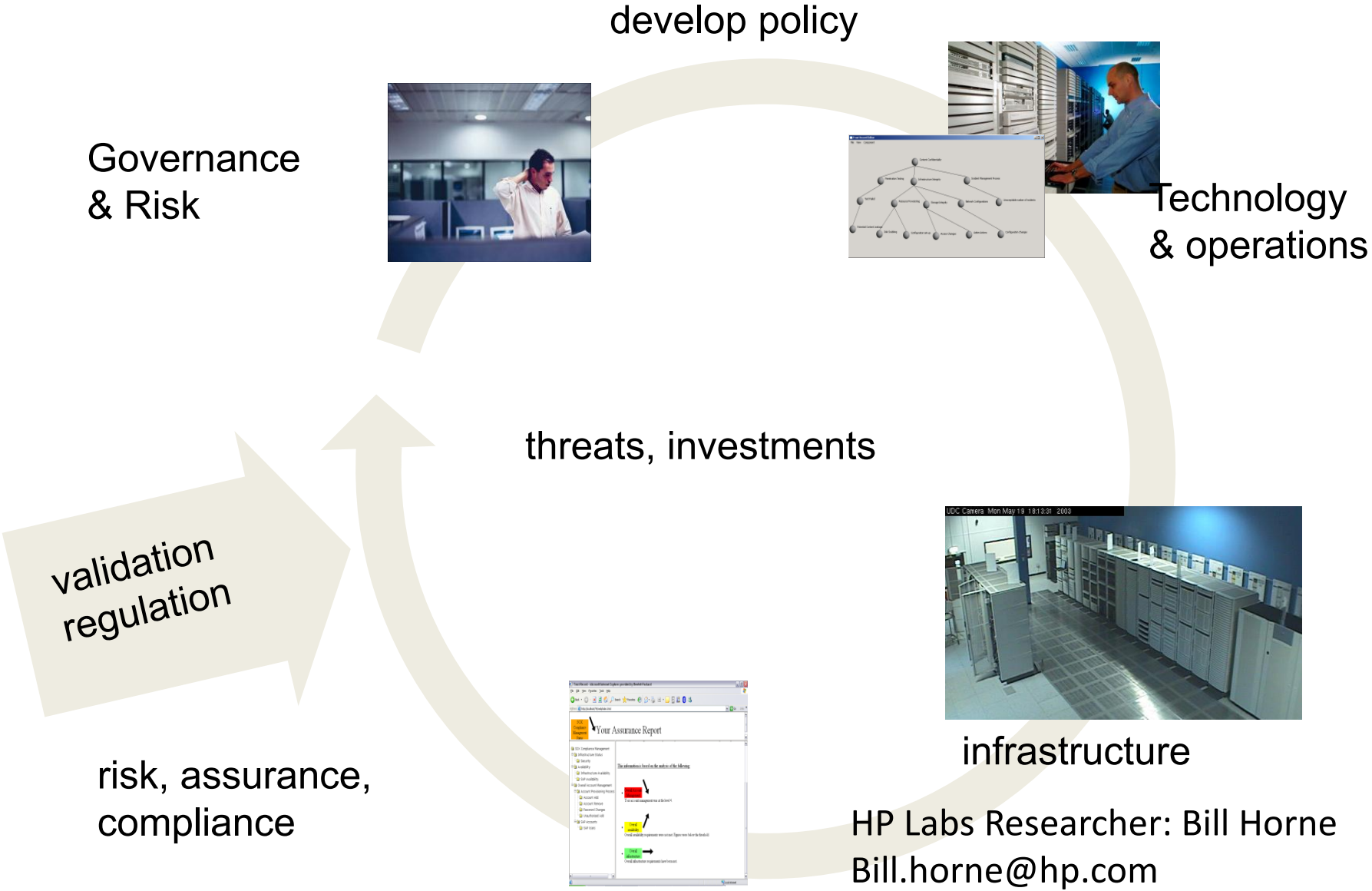
Realization 1



The molecules of E-security...



HPL Security Research



Governance & Risk

develop policy

Technology & operations

threats, investments

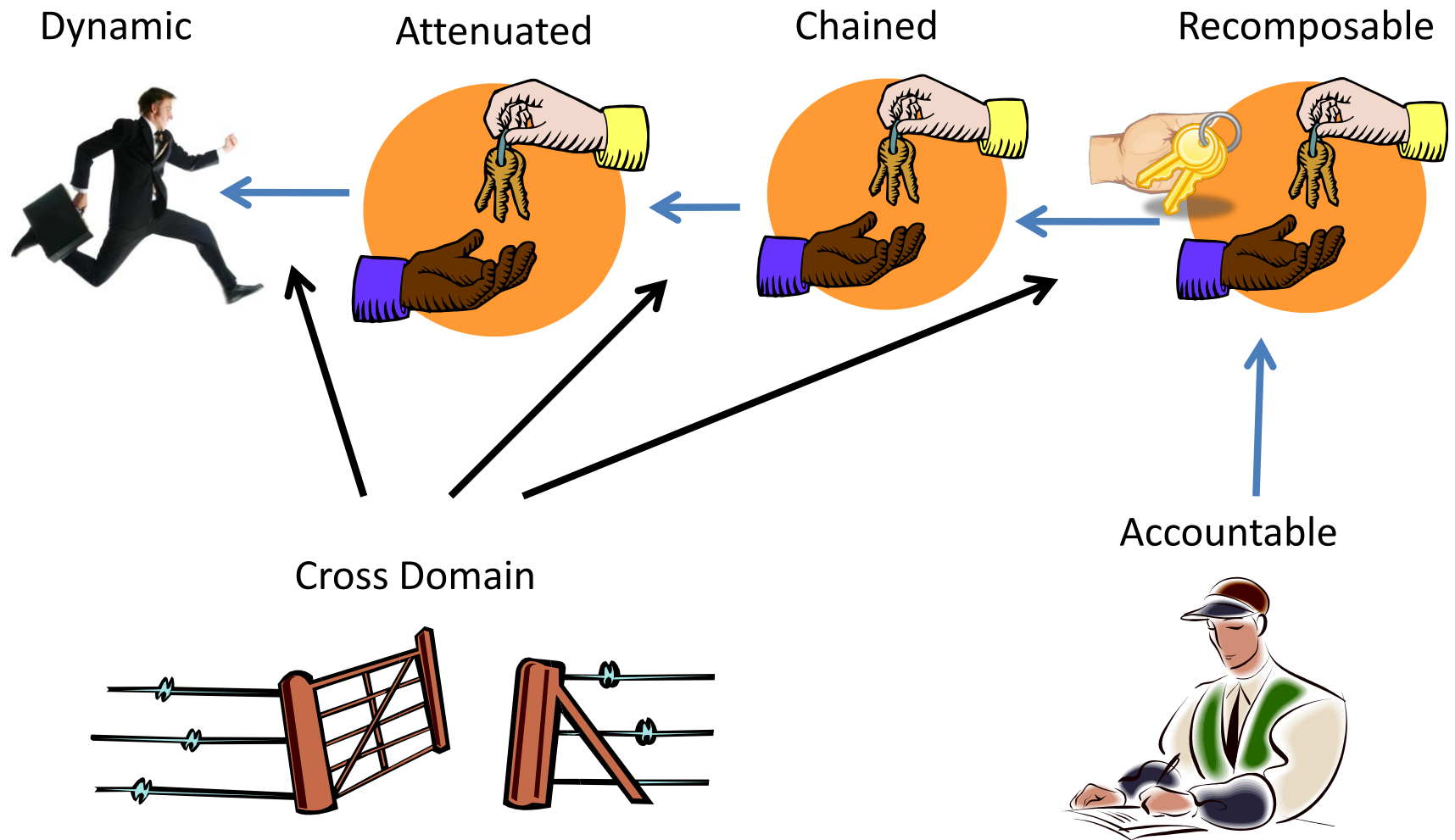
validation regulation

risk, assurance, compliance

infrastructure

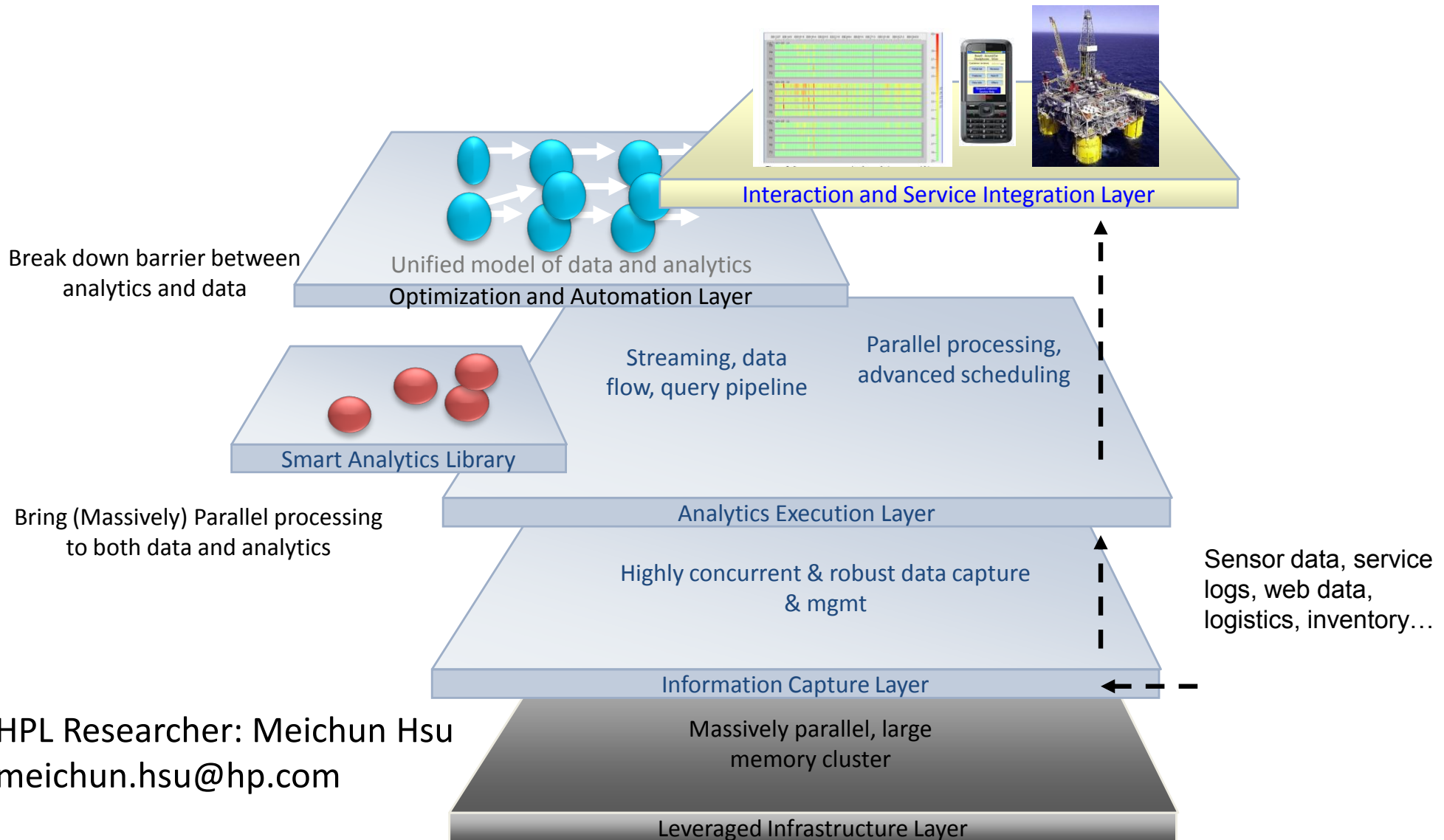
HP Labs Researcher: Bill Horne
Bill.horne@hp.com

People need to share to do their jobs



HP Labs Researcher: Marc Stiegler
marc.stiegler@hp.com

Live Analytics with Real Time Data Flow

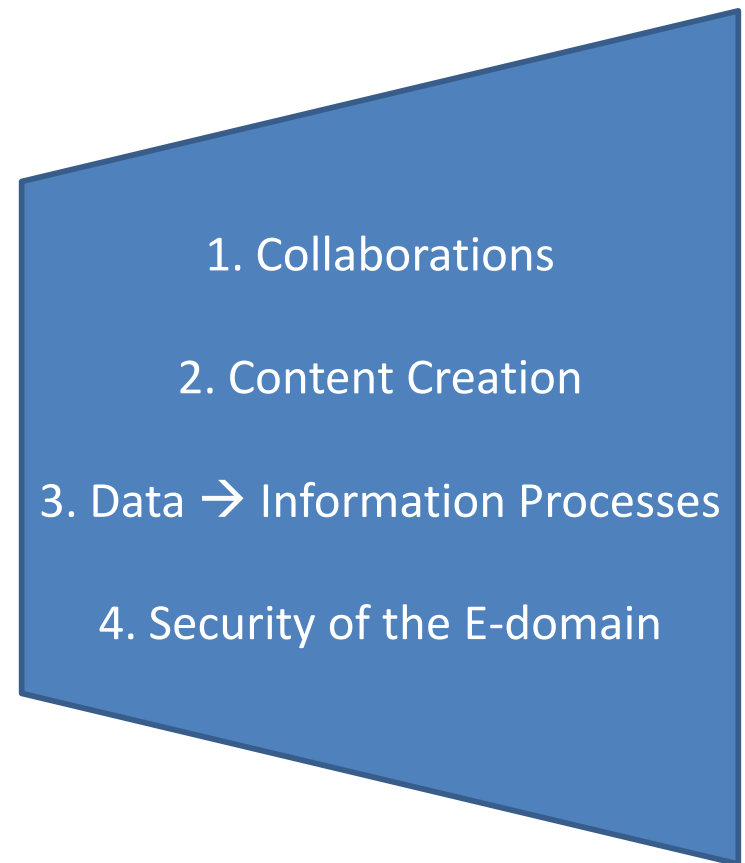


HPL Researcher: Meichun Hsu
 meichun.hsu@hp.com

Open World Principles

- 1. Democratic Intellectual Property**
- 2. Individualized Relationships**
- 3. Free flowing Data Rivers**

Open World Themes

- 
1. Collaborations
 2. Content Creation
 3. Data → Information Processes
 4. Security of the E-domain

1. Collaborations

- Informal Networks
- Communications (dynamic and invisible)
- Relationships (context and domain relevance)

2. Content Creation

- Visual and audio content (Youtube, Facebook)
- Creation via Suggestion (Netflix model..)
- Embedded content

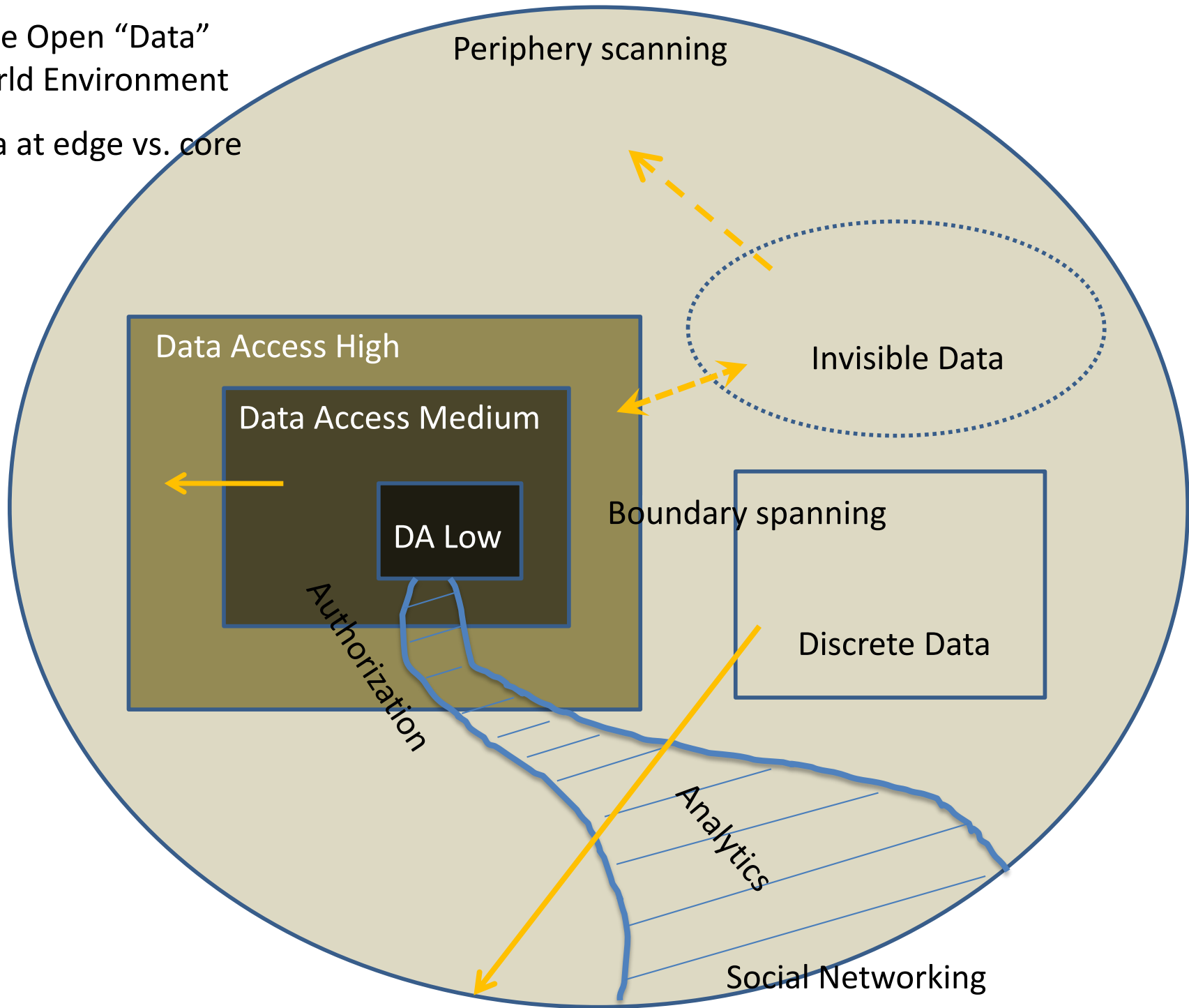
3. Data → Information

- Open data world
- Unstructured Content
- Data Warehousing
- It can enhance and exacerbate
 - Latency (timeliness)
 - False Reporting (accuracy)
 - Missed correlations (relevance)

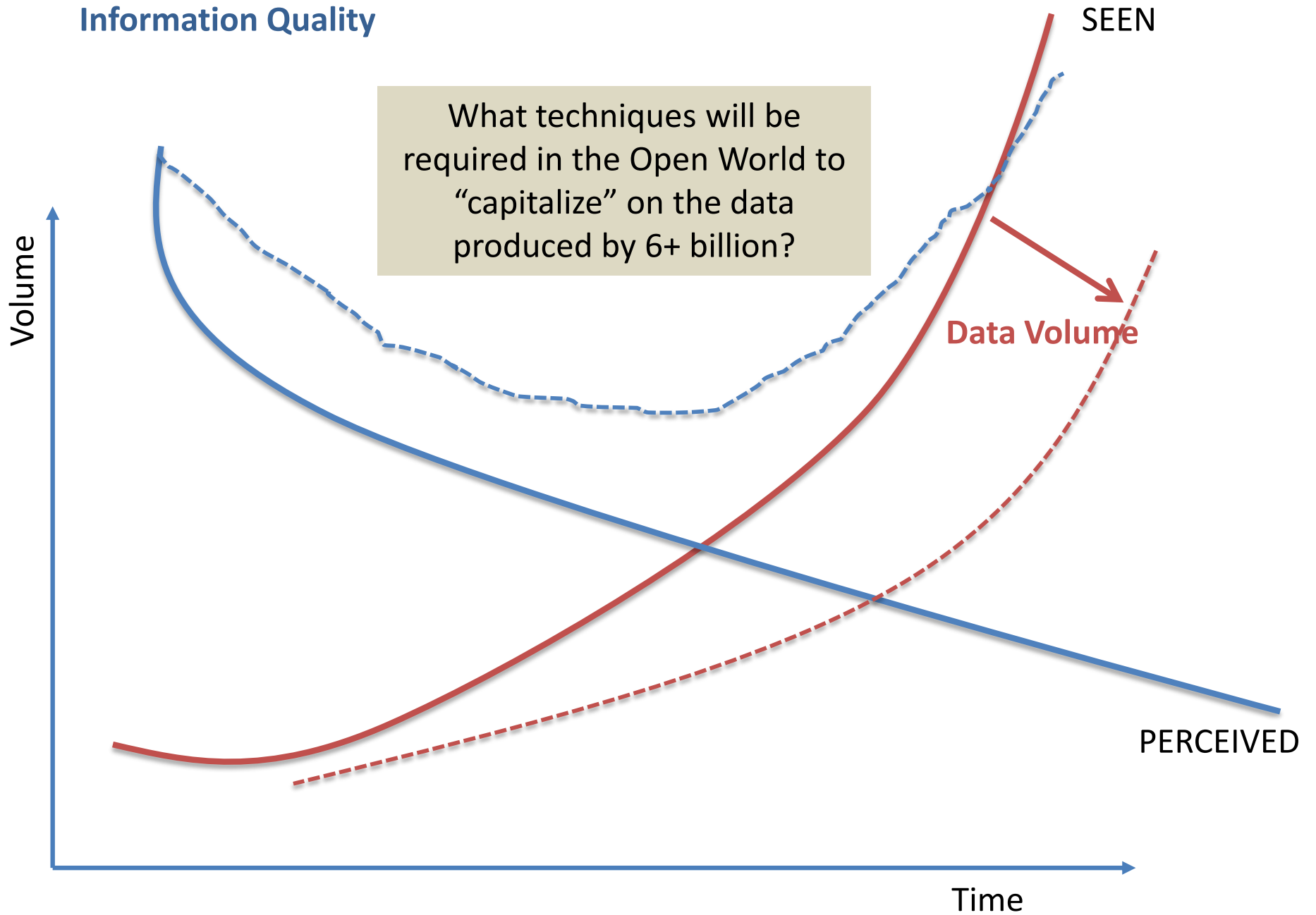
4. E-Governance/E-Security

- Governance
- Policies
- Architecture
- Data/Information Stewardship

The Open "Data"
World Environment
Data at edge vs. core



Information Quality

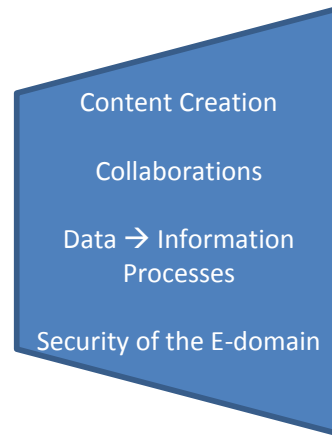


Open World Experiment

Democratic Intellectual Property

Individualized Relationships

Free flowing Data Rivers



Exercise 4 themes across 3 principles

Silicon Valley –communications
portal for the Open World

Experiment

Hosted via a virtual laboratory (cloud-based)
Researchers rotate out and nominate replacements
Learn by doing/play/participate
Multi-stage experimentation
Simulate context, domain, sentiment:
 Emergency Preparedness, Disaster Rescue
 News media, social media
 Healthcare
 Education

Incentives/Rewards

Ground-truth(ing) of social
predictions

Summary....

- Human race will be online, within a decade or so...
- This connectivity will impact: content creation and collaborations
- Changing paradigms of interaction require prototyping and experimentation in the “open world”