

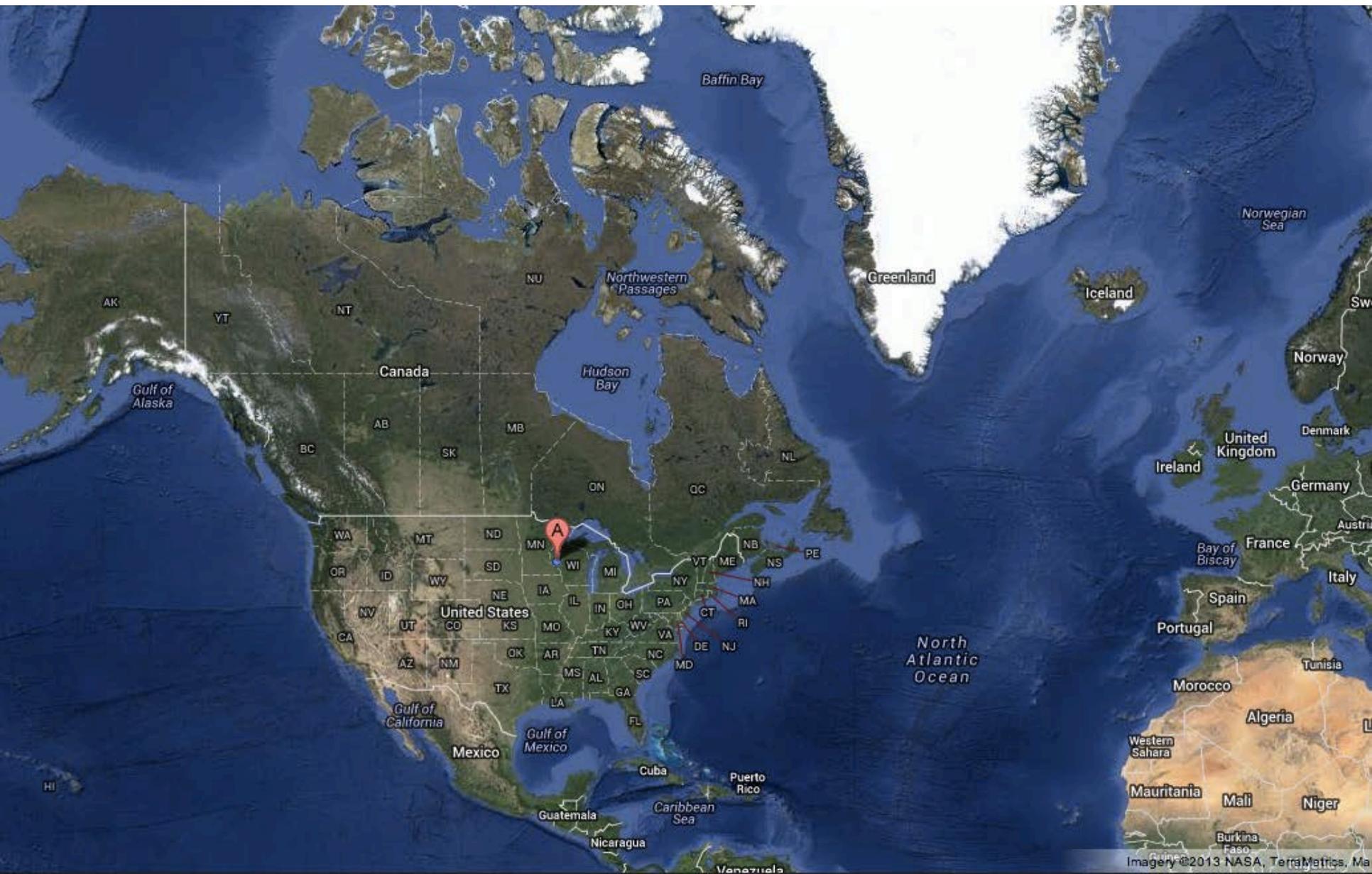
The New “Normal”: Revisiting Ethical Issues in Internet Research

Elizabeth A. Buchanan, Ph.D.

University of Wisconsin-Stout

buchanane@uwstout.edu

First....



Today's Goals

- Discuss the (rapidly) shifting infrastructure and modalities of “the Internet” and the impact on research
- Consider the ways in which “personal” identities are no longer individuated, but shared (with consent or not)
- Review current research regulatory models and the challenges they are facing vis-à-vis “the new normal”

Internet Research Defined

(SACHRP 2013)

- Research studying information that is already available on or via the Internet without direct interaction with human subjects (harvesting, mining, profiling, scraping—observation or recording of otherwise-existing data sets, chat room interactions, blogs, social media postings, etc.)
- Research that uses the Internet as a vehicle for recruiting or interacting, directly or indirectly, with subjects (Self-testing websites, survey tools, Amazon Mechanical Turk®, etc.)
- Research about the Internet itself and its effects (use patterns or effects of social media, search engines, email, etc.; evolution of privacy issues; information contagion; etc.)
- Research about Internet users—what they do, and how the Internet affects individuals and their behaviors
- Research that utilizes the Internet as an interventional tool, for example, interventions that influence subjects' behavior
- Others (emerging and cross-platform types of research and methods, including m-research (mobile))
- Recruitment in or through Internet locales or tools, for example social media, push technologies

What was the Old Normal?

World Wide Web

The WorldWideWeb (W3) is a wide-area [hypermedia](#) information retrieval initiative aiming to give documents.

Everything there is online about W3 is [linked](#) directly or indirectly to this document, including an [executive summary](#), [Policy](#), November's [W3 news](#), [Frequently Asked Questions](#).

[What's out there?](#)

Pointers to the world's online information, [subjects](#), [W3 servers](#), etc.

[Help](#)

on the browser you are using

[Software Products](#)

A list of W3 project components and their current state. (e.g. [Line Mode](#), [X11 Viola](#), [NeXTStep](#))

[Library](#)

[Technical](#)

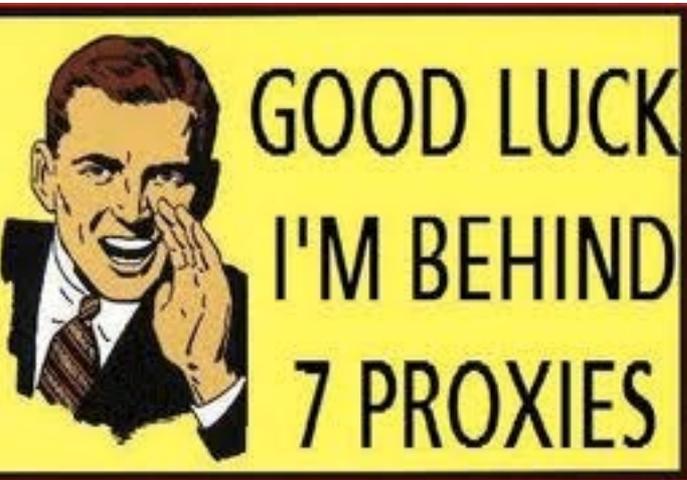
Details of protocols

The Good Old Days

- We went to “it”
- Oftentimes, we downloaded “it” (and maybe “owned” it) to our own device (we controlled what we got)
- Sometimes, we even downloaded it WITHOUT SIGNING A TOS/EULA! Does anyone remember anonymous FTP?
- We decided what we wanted to see, explore, do
- We could be “invisible” if desired
- We had to “connect” –connectivity was not ubiquitous
 - <http://www.youtube.com/watch?v=gsNaR6FRu00>

**I MISS
MY
PRE-INTERNET
BRAIN**

Moving From:



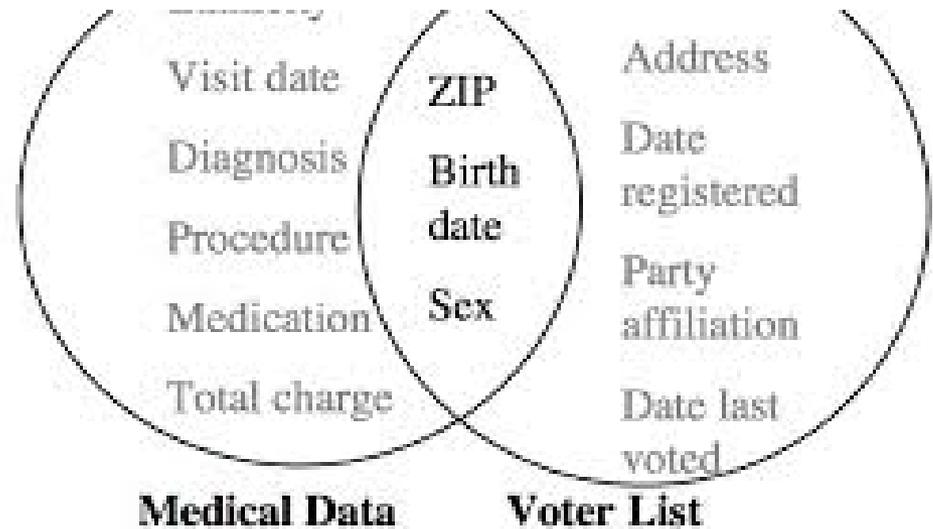
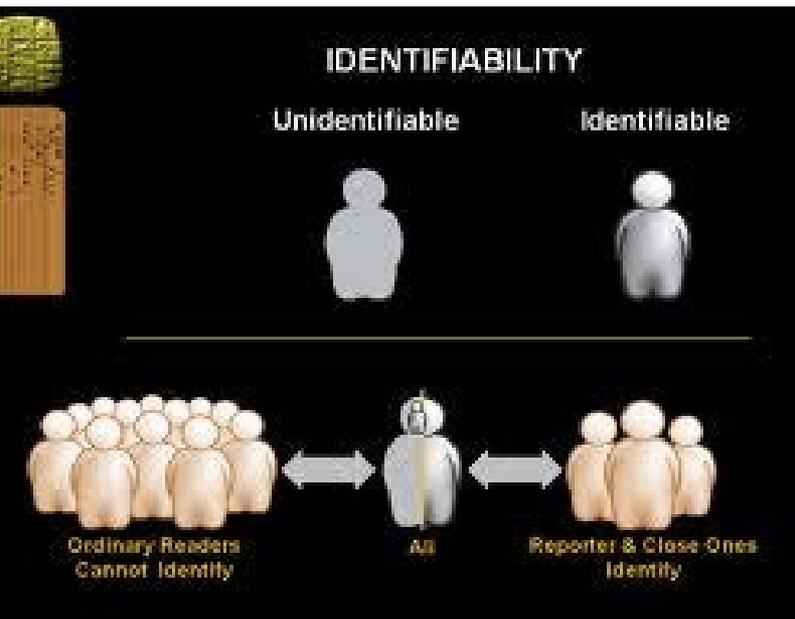
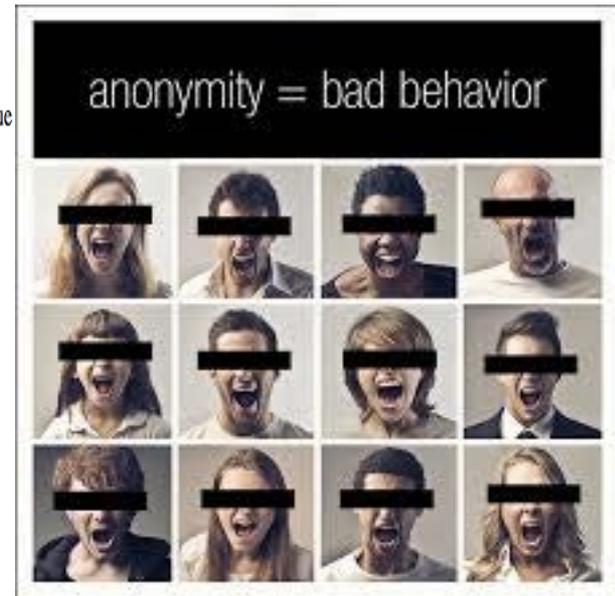
How Unique are You?

Enter your ZIP code, date of birth, and gender to see how unique you are (and therefore how easy it is to identify you from these values).

Date of Birth

Gender Male Female

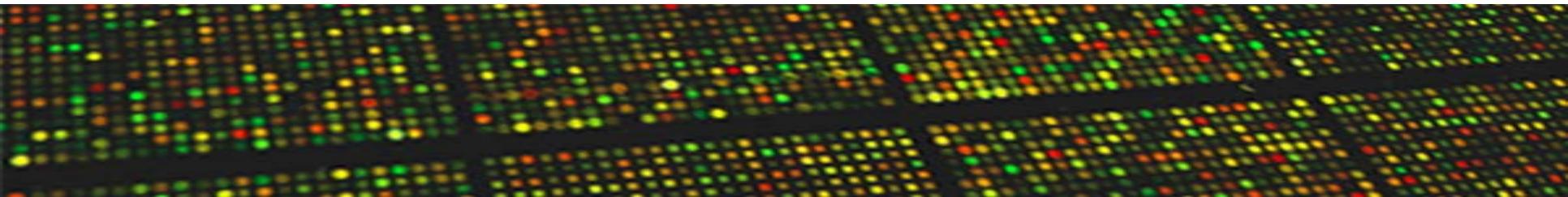
5-digit ZIP



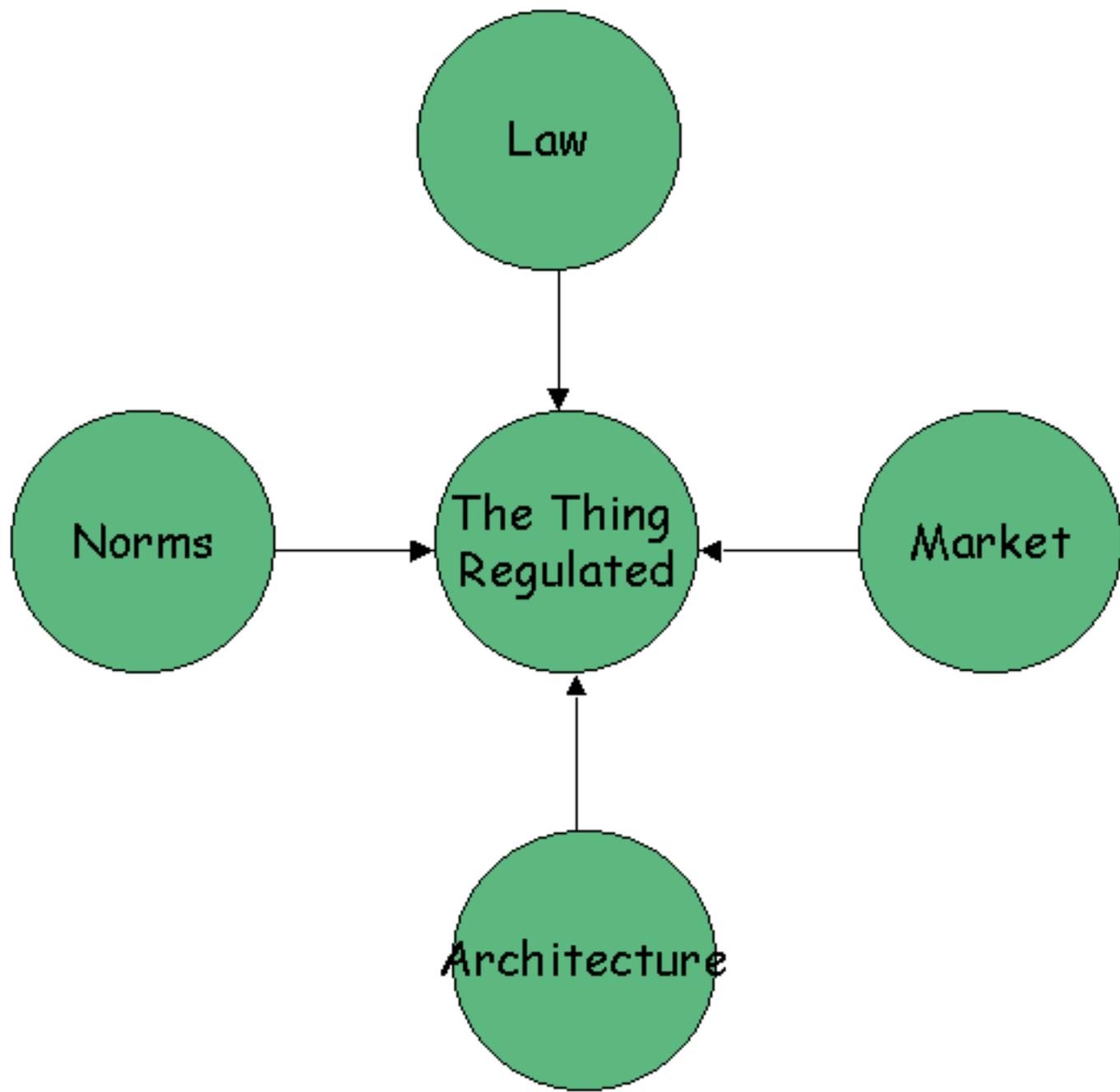
There is zero privacy anyway, get over it

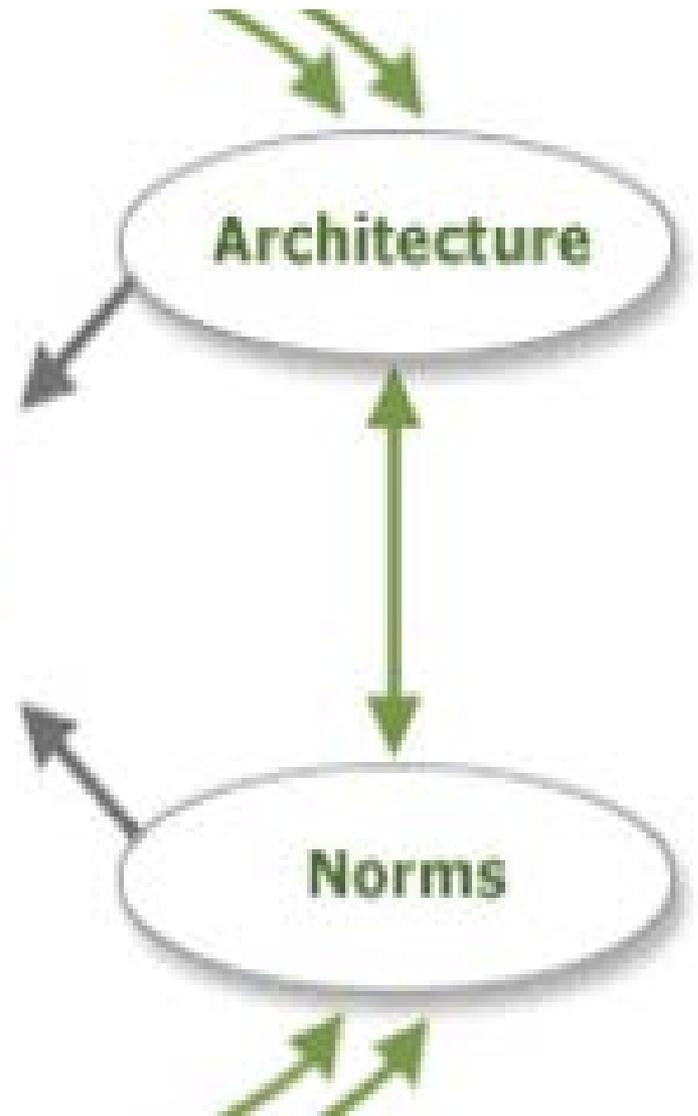
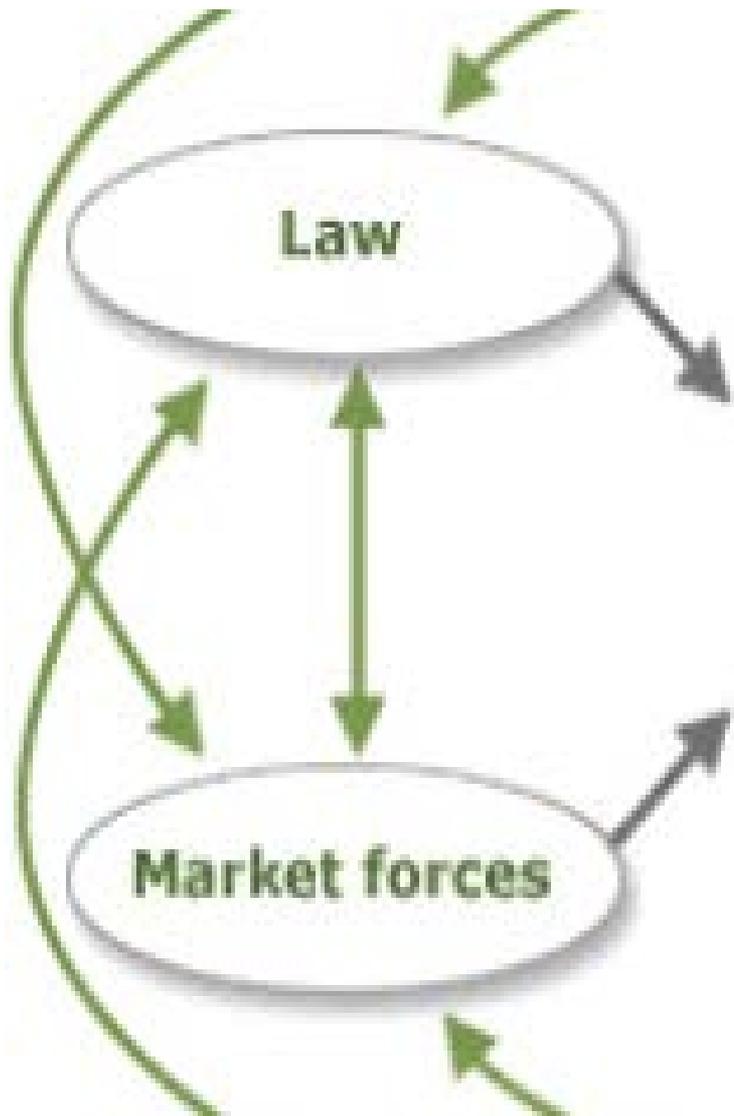


Although you can upload your data using a pseudonym, there is no way to anonymously submit data. Statistically speaking, it is really unlikely that your medical and genetic information matches that of someone else. By uploading you do not only disclose information about yourself, but also about your next kinship (parents and siblings) that shares half of a genome with you. Before uploading any genetical data you should make sure that those people approve of you doing so.” ([openSNP](#))



Why? And When did this Happen?





Social Media Landscape



FredCavazza.net

Big Data Landscape (Version 2.0)

Infrastructure

NoSQL Databases
 10gen, DATASTAX, basho, COUCHBASE, CLOUDANT, HYPERTABLE, Neo4j, SOCRATES, Oracle NoSQL Database

NewSQL Databases
 MarkLogic, paradigm4, memsql, SQLFire, DRAWNSCALE, VoltDB, NUODB

Hadoop Related
 cloudera, HADAPT, infochimps, Hortonworks, MAPR, HSTREAMING, Zettaset, MORTAR, IBM InfoSphere Business, Microsoft, GREENPLUM, A DIVISION OF EMC, amazon, Qubole, agril

MPP Databases
 VERTICA, An HP Company, Kognitio, PARACCEL, GREENPLUM, A DIVISION OF EMC, TERADATA, N, NETEZZA, InfiniDB, Microsoft SQL Server

Management / Monitoring
 OUTER THOUGHT, oceanSYNIC, StackIQ, bundy, DATADOG

Cluster Services
 LexisNexis, HPC Systems, Acunu, Security, Stormpath, IMPERA, TRACE VECTOR, codefortytwo software, DATAGUISE

Storage
 Cleversafe, panasas, nimblestorage, ANF DATA, Compuverde

Crowdsourcing
 CROWD COMPUTING SYSTEMS, CrowdFlower, amazon, mechanicalturk, Artificial Intelligence

Collection / Transport
 aspera, nodeable

Analytics

Analytics Solutions
 Palantir, platforma, PERSASIVE, Datameer, KARMA SPHERE, DataHive, DIGITAL REASONING, dataspora, PRECOG

Data Visualization
 Quid, visual.ly, ACTUATE, Kitenga, centrifuge, metaLayer, Ayasdi, ClearStory, +tableau, ISS, Quantum4D

Statistical Computing
 SKYTREE, p(k) Prior Knowledge, REVOLUTION ANALYTICS, MATLAB, SAS, SPSS, SAS IN-CLUSTER

Sentiment Analysis
 GENERAL SENTIMENT, crimson hexagon

Location / People / Events
 RapLeaf, Fliptop, Recorded Future, Place IQ, RADIUS

Real-Time
 CONTINUITY, ParStream, feedzai

Crowdsourced Analytics
 DataKind, kaggle

SMB Analytics
 sumall, RJMetrics, custora

Social Media
 bitly, bluefin, simple reach, Dataminr

Analytics Services
 THINK BIG, McKinsey & Company, accenture, OPERA, Mu Sigma

Big Data Search
 elasticsearch, Autonomy

IT Analytics
 splunk, sumologic

Applications

Ad Optimization
 DataXu, aggregate knowledge, m6d, MediaMath, bluekai, alMatch, rocketfuel, thetradedesk, TURN, 33 across

Publisher Tools
 VISUAL REVENUE, Yieldex, yieldbot

Marketing
 LATTICE ENGINES, Sailthru, bloomreach, CLICKFOX

Industry Applications
 NEXT BIG SOUND, KNEWTON, nestcash, wonga, numberFire, Mile Sense, Climate Solutions, Bloomberg, BILL GUARD

Application Service Providers
 collective

Data Marketplaces
 factual, DataMarket, Windows Azure Marketplace

Data Sources
 premise, DATA SIFT, knoema, Gnip, infochimps

Withings Personal Data
 JAWBONE, RunKeeper, BASIS, Nike, fitbit

Cross Infrastructure / Analytics

SAP, sas, IBM, Google, ORACLE, Microsoft, vmware, amazon, Hadoop, METAMARKETS, TERADATA, Autonomy, NetApp

Open Source Projects

Framework
 Hadoop, MapReduce, HDFS

Query / Data Flow
 Hive, Pig

Database Access
 Cassandra, SciDB, HBASE, CouchDB, Sqoop, mongoDB

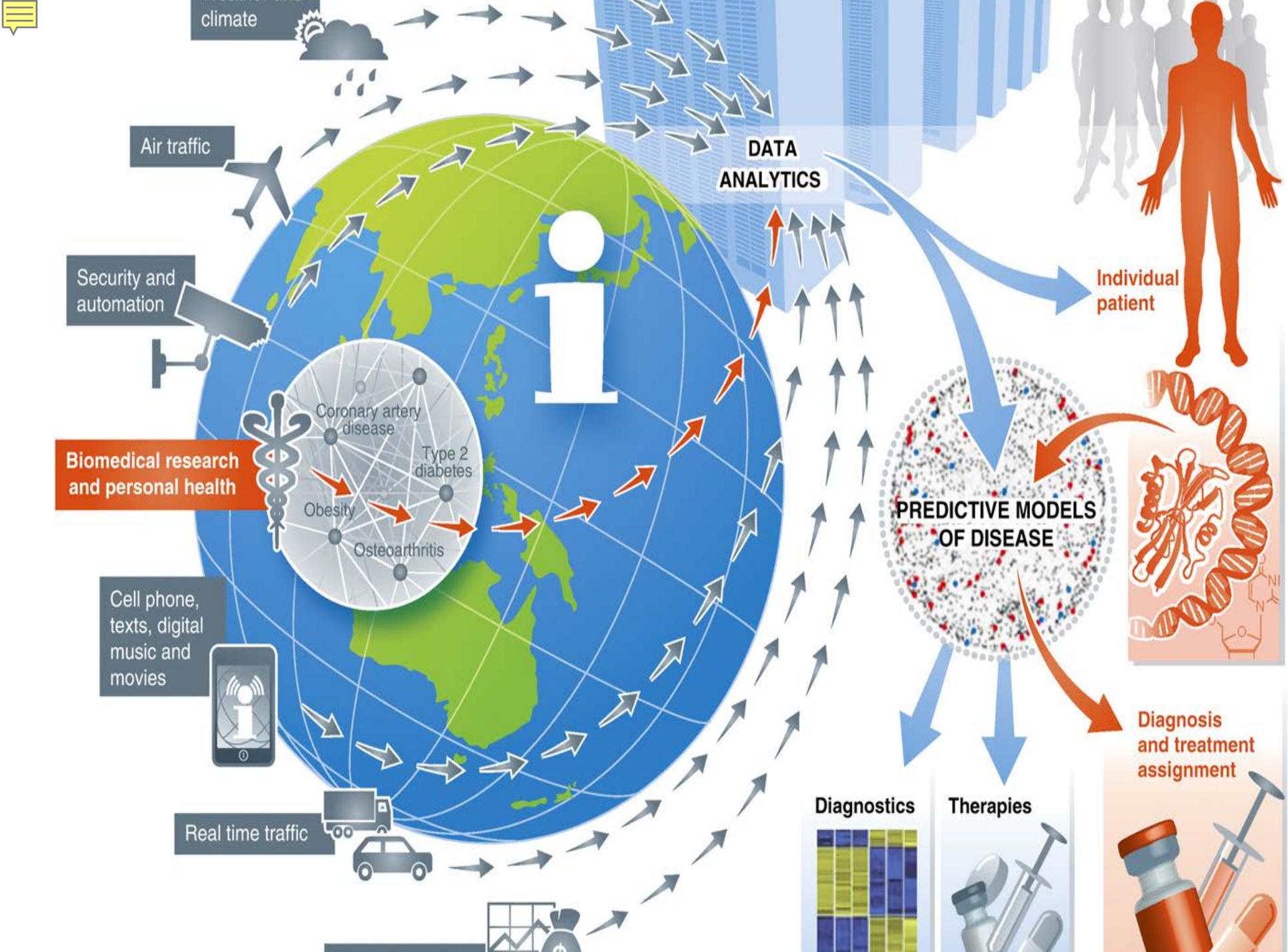
Coordination / Workflow
 ZooKeeper, talend, COOZEE

Real-Time
 Storm

Statistical Tools
 SciPy

Machine Learning
 Mahout

Cloud Deployment
 AWS



climate

Air traffic

Security and automation

Biomedical research and personal health

Cell phone, texts, digital music and movies

Real time traffic

DATA ANALYTICS

Individual patient

PREDICTIVE MODELS OF DISEASE

Diagnostics

Therapies

Diagnosis and treatment assignment



YOUR PII CHART™

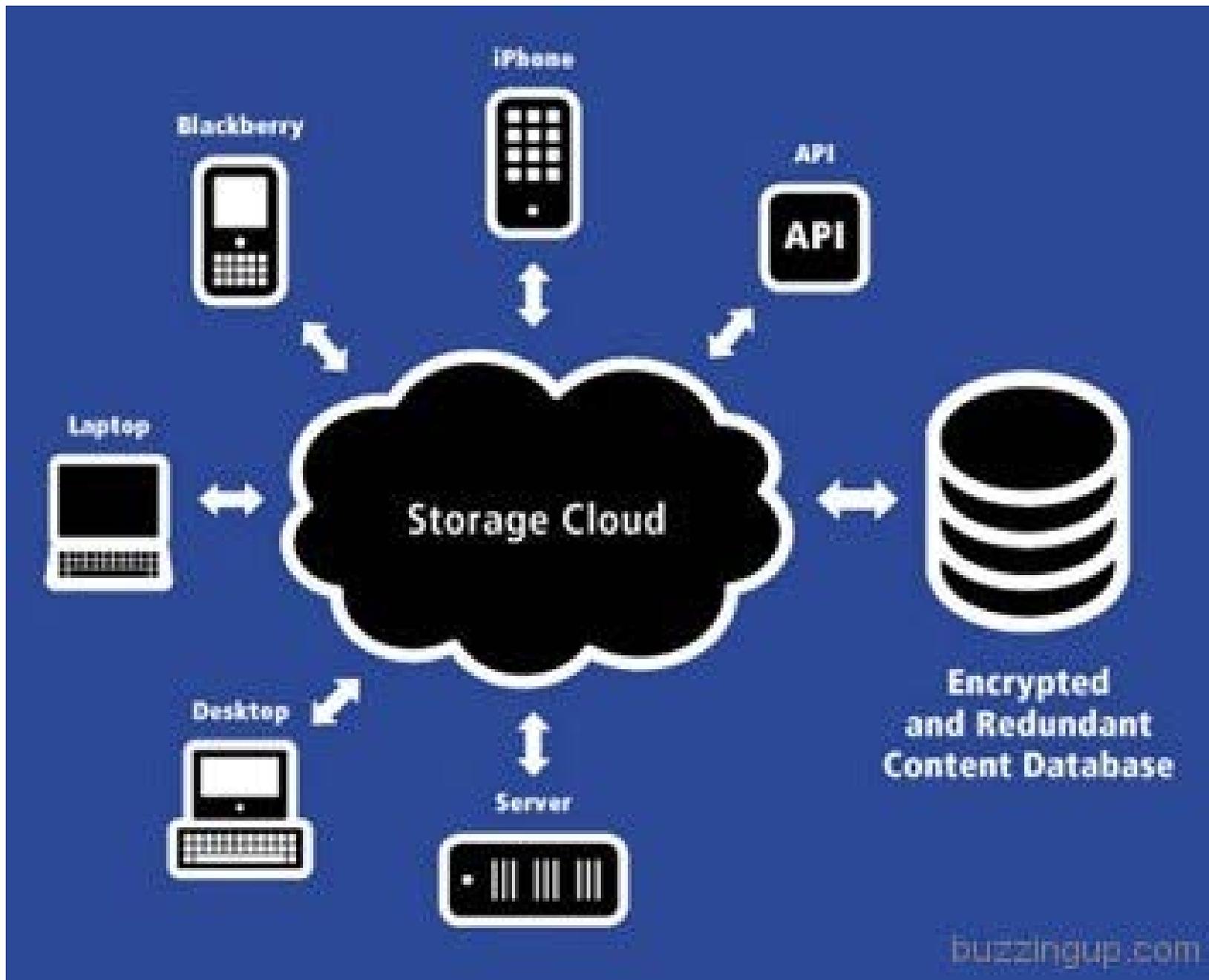
Take time to inventory the identity relationships you have with the companies, organizations, and individuals you entrust with your personally identifiable information or PII. See how your identity is a PII Chart™, a picture of relationships you've created. Once you visualize the slices of your PII, managing your identity assets becomes easier.

LEGEND

-  **SSN SOCIAL SECURITY NUMBER**
-  **CONTACT INFORMATION**
(email address, physical address, telephone and mobile numbers)
-  **GOVERNMENT-ISSUED IDENTIFICATION**
(driver's license, passport, birth certificate, library card)
-  **BIRTH DATE, BIRTH PLACE**
-  **ONLINE INFORMATION**
(Facebook, social media, passwords, PINs)
-  **GEOLOCATION**
(smartphone, GPS, camera)
-  **VERIFICATION DATA**
(mother's maiden name, pets' and kids' names, high school, passwords)
-  **MEDICAL RECORDS INFORMATION**
(prescriptions, medical records, exams, images)
-  **ACCOUNT NUMBERS**
(bank, insurance, investments, credit cards)



Recall our Big Data Graphic! Lots of flow and movement between and among spheres....



What Does all of this Mean for Research—and Research Ethics?

- Blurring of boundaries or venues
- Researchers are but one stakeholder in a larger framework
- Interoperability
- Synchronous sensing (eg, CenceMe©)
- Remote research—separation of researcher-subject
- Dissolving public v private sector data flows
- Non-consensual access to data
 - Pros and cons: DNA databases to identify family members in War Zones; potential of identifying families of war criminals?
 - Secondary subjects and third parties could benefit from rights to have their family members identified **or** could become more vulnerable and targeted
- Fundamental principles of autonomy, consent, privacy do not align
- Individuals are known by association or through a composite of data, true or untrue: your data persona
- How to isolate a research frame
 - Can CS help us social/behavioral researchers?
- More emphasis on DATA SECURITY versus individual privacy

Location, Location, Location

The screenshot shows a Google Chrome browser window with two tabs: "California Healthcare Atlas" and "Location sharing - Chrome". The address bar displays the URL "https://support.google.com/chrome/answer/142065?hl=en". The browser's toolbar includes various icons for apps, email, social media, and search engines. The page header features the Google logo, a search bar for "Search Chrome Help", and the user's email address "elizabeth.buchanan@gmail.com". The breadcrumb trail indicates the path: "Chrome > Help > How to use Chrome > Privacy and security settings". The main content area is titled "Location sharing" and includes a section "Why are sites asking for my location?" with explanatory text and a list of links for further information.

California Healthcare Atlas x Location sharing - Chrome x

https://support.google.com/chrome/answer/142065?hl=en

Apps Gmail Flickr Amazon Google+ JS School YMCA Wikipedia CNN Z CNN Editorial Manager@ Other Bookmarks

Google Search Chrome Help elizabeth.buchanan@gmail.com

Chrome > Help > How to use Chrome > Privacy and security settings Google Chrome Forum

Location sharing

Why are sites asking for my location?

You can allow some sites you visit to use your location information to make the information they show you more useful. For example, if you share your location with a site that specializes in restaurant reviews, the site can fine-tune its display and show reviews for restaurants near you. You have control over whether you share your location with sites.

Google Chrome never shares your location without your permission. By default, whenever you're on a site that wants to use your location information, Google Chrome alerts you by showing you a prompt at the top of the page. Your location is sent to the site only if you click **Allow** in the prompt.

If you agree to share your location with the site, the  location icon appears in the address bar to remind you that you've permitted the site -- or something embedded in the site, such as a map -- to access your location. To see more details or to clear location permissions for the site, just click the icon.

- + Enable or disable location sharing
- + How Google Chrome determines your location

Privacy and security settings

- Chrome's privacy settings
- Manage your browsing history
- Delete your browsing history
- Delete your cache and other browser data
- Manage your cookies and site data
- Adjust images, JavaScript, and other web content settings
- Phishing and malware detection
- Website settings
- Advanced security settings
- Location sharing
- Usage statistics and crash reports
- Chrome's privacy policy
- Notifications



Mobile Research

- Huge growth in this area
- FDA Guidance on Mobile Medical Apps now released
<http://www.fda.gov/downloads/MedicalDevices/DeviceeRegulationandGuidance/GuidanceDocuments/UCM263366.pdf>
- “Geo-intent:” “ Beyond location awareness, it is far more important to know where a person is headed and his needs and wants at the destination” (Klein, 2013)
- What is consent in this domain: “what you are consenting to when I can know where you are and what you're doing and you're with, and even how you're feeling” (Donna Spruijt-Metz)

GIS and Satellite Research

- Ongoing use and potential in international research
 - Surveillance of crowds, zones, locales, war zones, emergency/disasters, environment over time—where is the subject?
 - Movement towards societal well being over individual???
- Risk benefit considerations

Extant Regulatory Models

- In the US: Do not address the confounding issues around HUMAN SUBJECTS research vis-à-vis today's technological environment
 - Attempts (Buchanan and Gallant/SACHRP, 2013)
- Some better guidelines in Canada
- EU focus on data privacy and ownership
- But research is now multi-national, trans-border, globally conducted. Who's or what ethical models should apply?
- Context-dependent approaches seem most appropriate

As researchers...

- “Classic consents must therefore transition away from attempting to guarantee individuals’ privacy. Rather, new forms of consent should aim at educating research subjects on what the data collected on them can say and the degree to which it can or cannot be protected” (Schadt, 2012).

As Researchers...

- Research is now an ENTERPRISE—complex and multifaceted. Team approach is best: computer scientists, security, ethicists, methodologists.
- Good ethics = good science and good research
- Let ethics drive methods

Enquiries and Comments



For more information

Elizabeth Buchanan, Ph.D.
Endowed Chair
Center for Applied Ethics
University of Wisconsin-Stout
Menomonie, Wisconsin, USA
715 232 5184
buchanane@uwstout.edu
@stout_ethics