

Data Ethics

Data ain't magic
Quinn Underriner

How do you create the largest amount of wealth ever geographically centralized in human history?

A: Arbitrage! Or as they say on Wall Street – buy low, sell high

People significantly misprice the value of their own data (not that many are even doing this calculation, or aware of the transaction they are participating in)



Largest 5 companies in 2007

Q: Why did Amazon get its start as a book seller?



Largest 5 companies in 2017

So what is your data worth?

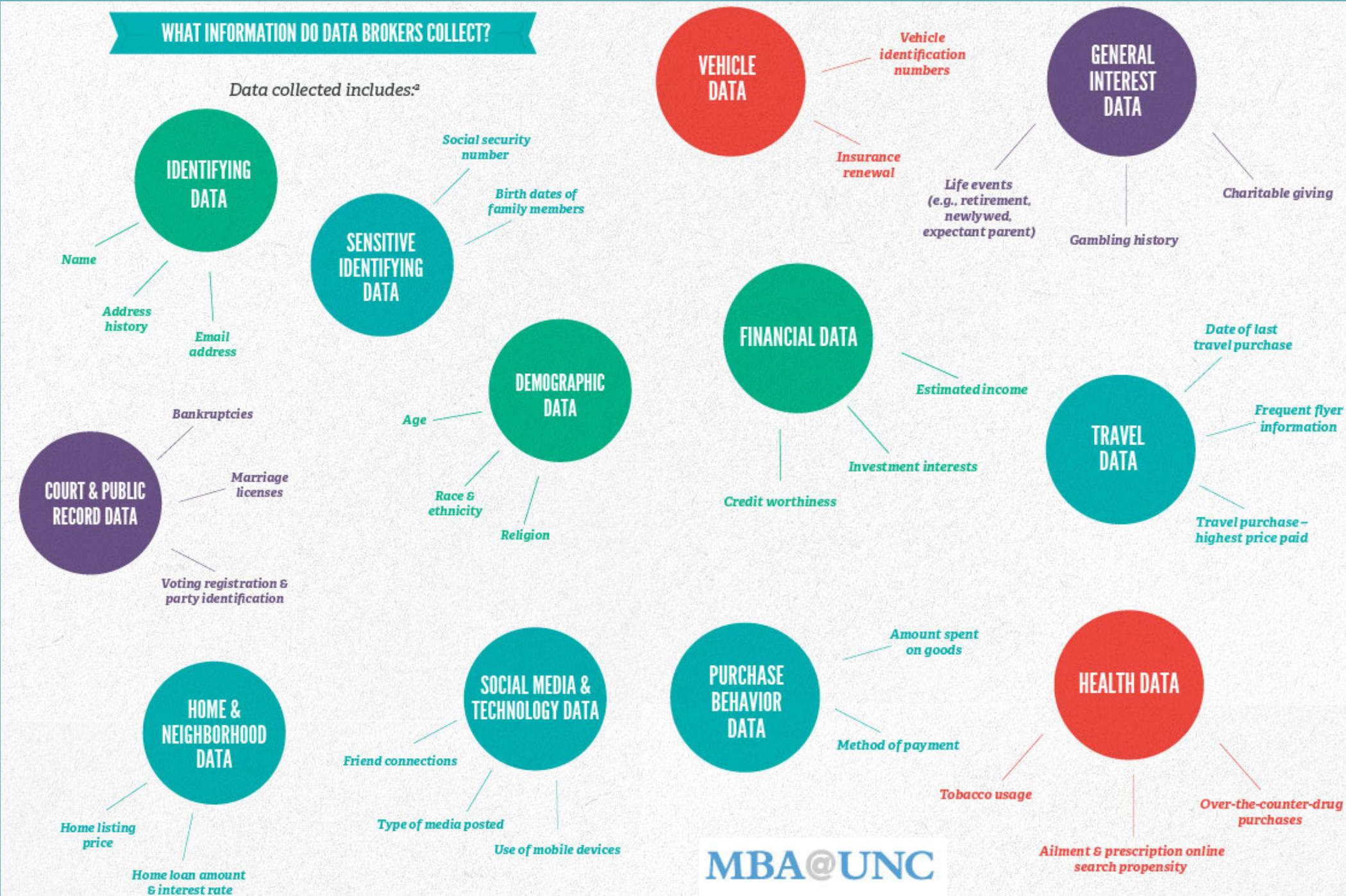
In Caesar's (the casino) chapter 11 bankruptcy filing some creditors valued their "Total Rewards" customer loyalty program data at \$1 billion, making it their largest asset (ahead of physical asset holdings!)

Why did Microsoft buy LinkedIn for \$26.2 billion? – Consumer data! While it's hard to breakdown specific costs (for reference, their revenue in 2015 was only \$2.9 billion).

Simple math shows us \$260 per monthly active user



What can data brokers figure out about you?



So uh who has my data?

	Company	Type of Company	Link to Privacy Policy	Opt-Out Possible?	Link to Opt-Out Page	Personal Info required for Opt-Out	Can you get your data?
1	411	people search	http://411.info/privacy	Yes	http://wpremove.411.info	phone number	Yes, email privacy@411.info
2	123 People	people search/aggregat	http://www.123people.com/page/people-manager	Yes			None
4	123 Verified	background	http://123verified.com	No			None
5	411Locate	people search	http://www.411locate.com	No			None
6	555-1212.com	data broker	http://www.555-1212.com	No			None
7	Accutellus.com	people search	http://www.accutellus.com	Yes	https://www.accutellus.com	address	None
8	Axiom	data broker	http://axiom.com/	Partial	https://isapps.axiom.com	phone number, address	Yes
9	American List Counsel (ALC)	data broker	http://www.alc.com	Yes	http://www.alc.com/privacy.htm		Yes
10	Ameridex	people search and vital records	http://ameridex.com	No, unless you are publicly listed	write to Ameridex PO Box 193061 San Francisco CA 94119-3055		none
11	Ancestry (associated with Ancestry.ca, Ancestry.co.uk)	genealogy and vital records	http://www.ancestry.com	Yes	Email customersolutions@DL		Yes, in some cases. http://ancestry.com/privacy.html#optout
12	Ancestry.com: Genealogy	genealogy and vital records	http://www.genealogy.com	No	http://www.genealogy.com/privacy.html#optout		None
13	Ancestry.com: My Family (The Generations Network)	genealogy and vital records	http://www.myfamilytree.com	Yes	email customersolutions@DL		None
14	Ancestry.com: Roots Web	genealogy and vital records	http://www.rootsweb.ancestry.com	Yes	Email customersolutions@DL		none
15	Any Who (part of AT&T. All info from Intelius)	people search	http://www.att.com	Yes	http://www.anywho.com/help/privacy		None
16	Aristotle: Integrity	background	http://integrity.aristotle.com	No			None
17	Aristotle: Voter Lists Online	voter list	http://www.voterlist.integrity.com	No			none
18	Arrests.org (varies from state to state)	mugshots		No	info@arrests.org		None
19	Background Check Gateway	background	http://www.backgroundchecksgateway.com	No			None
20	Background Check International (BCI)	background	http://www.bcint.com	No			None
21	BackgroundFinder.com	background		No			None
22	BCS Background Screening, LLC	background	http://www.ecriminology.com	No			None
23	BeenVerified	background	http://www.beenverified.com	Yes	http://www.beenverified.com	address	None
24	BlockShopper	real estate	http://blockshopper.com	No, unless you are publicly listed	http://www.blockshopper.com/faq.html#q4		None
25	Check Mate LLC	background	https://www.instantcheckmate.com	Yes	support@instantcheckmate.com	address, DOB	Possibly. It says you can email support@instantcheckmate.com
26	CIA Data (Consumer Intelligence Agency)	background	http://www.ciadata.com	No			Yes, legal@CIADATA.com
27	CIS Nationwide	background	http://cisenationwide.com	No, unless you are publicly listed	http://cisenationwide.com/optout.html		None
28	CIS Nationwide: Complete Background Checks	background	http://www.completebackgroundchecks360.com	No, unless you are publicly listed	http://www.completebackgroundchecks360.com/terms.html		None
29	CIS: Birth Records	background	http://www.birth-records.com	No, unless you are publicly listed	http://www.birth-records.com/index.php?xpath=lp_optout		None
30	CIS: Cell Phone Registry	people search	http://cellphoneregistry.com	No, unless you are publicly listed	http://cellphoneregistry.com/index.php?xpath=lp_optout		None

A very non-exhaustive list of shifty behavior

Bose wireless headphones noting your listening preferences to be sold to a third-party



Target predicted a teenage girl in Minnesota was pregnant before her parents knew and sent her targeted pregnancy advertisements



Facebook leak shows they create “ghost profiles” of people who are non-users



Vizio TVs tracking what television shows you watch to sell to 3rd parties



My personal favorite privacy violation:

SilverPush, Drawbridge, and Flurry and other data advertising companies who used inaudible noises to link your devices

Who even reads the privacy polices?

Unroll.me CEO Jojo Hedaya said that it was “heartbreaking to see that some of our users were upset to learn about how we monetize our free service.”

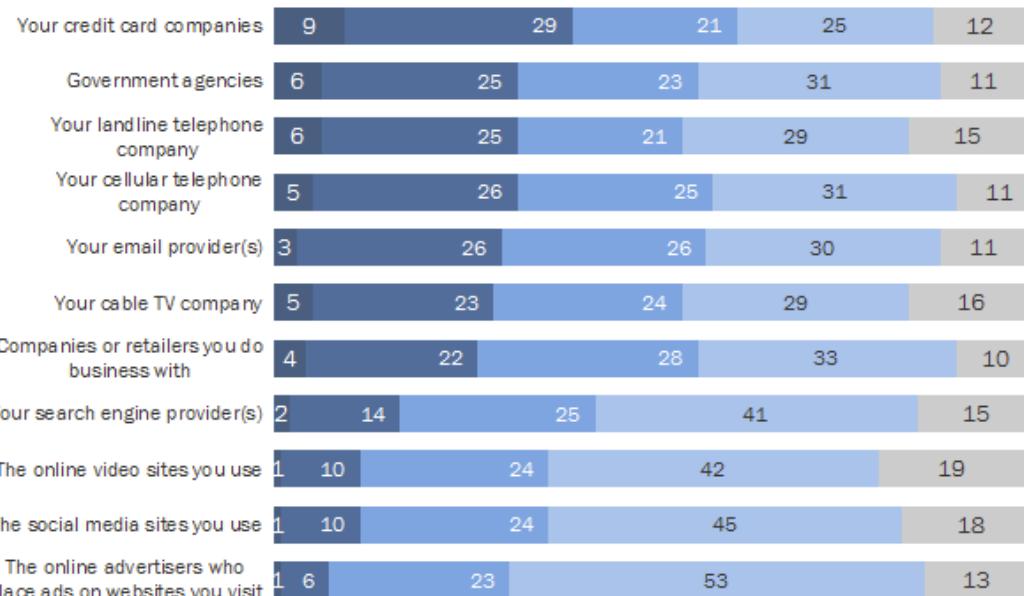


A [study](#) from Carnegie Mellon estimates that it would cost the U.S. economy **\$781 billion** if people actually read all the privacy polices they came across in a year (and this was in 2008!)

Few express confidence that their records will remain private and secure

% of adults who express varying levels of confidence that the records of their activity maintained by various companies and organizations will remain private and secure

■ Very confident ■ Somewhat confident ■ Not too confident ■ Not at all confident ■ Don't know



Do American's care about privacy?

Fully 91% of adults agree or strongly agree that consumers have lost control of how personal information is collected and used by companies

Some 74% say it is “very important” to them that they be in control of who can get information about them, and 65% say it is “very important” to them to control what information is collected about them.



U.S. vs. EU

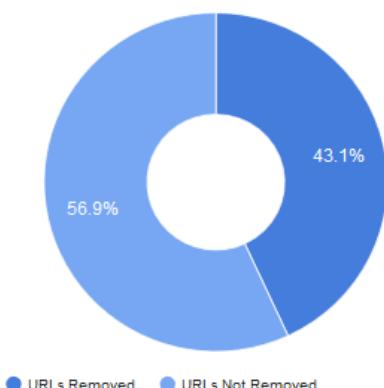
- Generally “pro-business”
- Regulations are a patchwork industry and/or state specific laws (e.g., HIPPA for Healthcare, COPPA for children)
- Opt-out consent
- Snowden revelation caused significant international anger and caused the European Court of Justice to invalidate the data sharing agreement (the Safe Harbor Agreement) between US and EU
 - This was replaced by the “Privacy Shield”, which is currently on shaky ground



- Privacy considered a fundamental human right in EU (helped by a historical fear of fascism) which allows, for example, for the “Right to be Forgotten”
- Strong Centralized Privacy Regulation
- Opt-in consent

URL removal request totals

The graph below shows data on the percentages of URLs we have reviewed and processed. The figures on the right are based on the total number of requests received. These data date back to the launch of our official request process on May 29, 2014.



Total URLs that Google has evaluated for removal:
2,010,298 URLs

Total requests Google has received: **715,729 requests**

The graph reflects URLs that have been fully processed, while the figures above reflect the total evaluated. URLs that require more information or are pending review are not included in the graph.

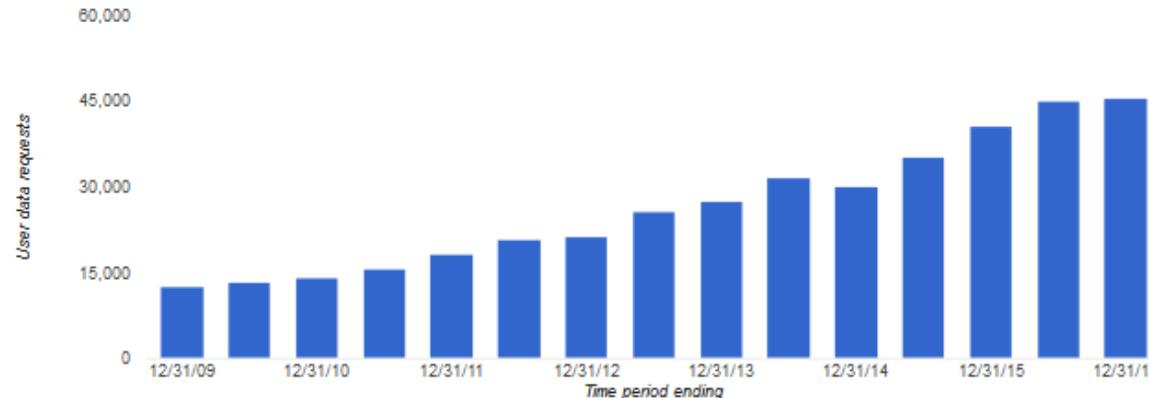
Brief history of EU- U.S. regulations

- EU negotiated the Safe Harbor Agreement of 2000 to allow U.S. companies and organizations to meet EU data protection requirements and permit the legal transfer of personal data between EU member countries and the United States
- Snowden revelation in June 2013 caused uproar, and eventually, in October 2015, the Court of Justice of the European Union invalidated the safe harbor agreement
- This scared the 4,500 U.S. companies who relied on this system
- In February 2016 U.S. & EU announced agreement “in principle” on a revised accord, called the Privacy Shield
 - detailed notice obligations, data retention limits, tightened conditions for onward transfers and liability regime, more stringent data integrity and purpose limitation principles, strengthened security requirements, increased enforcement from the FTC ability to dispute data beyond FTC with multiple redress opportunities

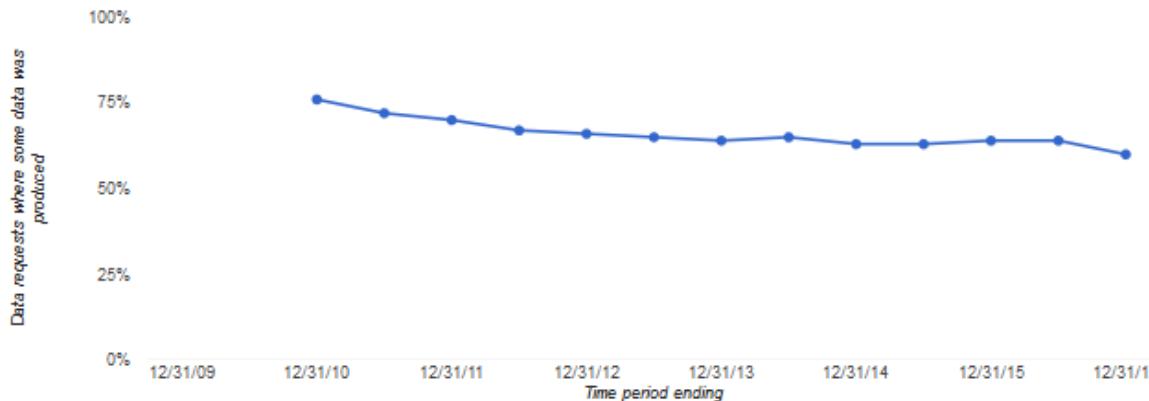
Post Snowden, companies have started releasing “Transparency Reports”

Requests by Reporting Period

User Data Requests | Users/Accounts



Percentage of Requests Where Some Data Produced



Sample from Google

Several other companies disclose data about government requests, including:

- AOL
- AT&T
- Apple
- Cheezburger
- Cloudflare
- Comcast
- Credo Mobile
- CyberGhost
- Daum Kakao
- Deutsche Telekom
- Dropbox
- Facebook
- GitHub
- Hong Kong Transparency Report
- Kickstarter
- Korea Internet Transparency Report
- LeaseWeb
- LinkedIn
- Lookout
- Microsoft
- Naver
- Nest
- Pinterest
- Rogers
- SaskTel
- Snapchat
- Sonic.net
- SpiderOak
- TELUS
- TekSavvy
- TeliaSonera
- Telstra
- Time Warner Cable
- TradeMe
- Tumblr
- Twitter
- Uber
- University of California, Berkeley
- Verizon
- Vodafone
- Wickr
- Wikimedia Foundation
- WordPress
- Yahoo!
- reddit

Other Organizations that produce Transparency Reports

How Cathy O'Neil characterizes “Weapons of Math Destruction”

- I. Algorithms that significantly impact peoples lives. She touches on systems such as:
 - I. loan rates
 - II. prison sentencing
 - III. teacher evaluations

- II. Black box systems :
 - I. Does the user understand how (and even if) they are being rated
 - II. As machine learning gets more sophisticated, this problem will be exacerbated

- III. Does it create a negative feedback loop?:
 - I. Is their a mechanism to test and change the system for biases and errors?

* If you don't read Cathy O'Neil's blog mathbabe, you're making a mistake

Credit Scores vs. “E-scores (data brokers)”

I. Credit scores:

- Governmental regulation
- Provide clear advice on how to raise score
- Legal (if inefficient) right to examine your score
- Legal (if inefficient) right to challenge and correct underlying data
- Models can see who actually defaults and then correct themselves

II. E-scores:

- No regulation
- No understanding on consumer name of bucket they are placed into, much less underlying data collected
- Many don't allow right of removal
- Unclear how they self-correct

The Hippocratic Oath for Data Scientists

- I solemnly pledge to practice my profession with conscience and dignity;
- To respect the privacy of the people whose data is confided in me;
- To maintain the utmost respect for the individuals whose data I am analyzing;
- To be transparent, open, and honest about the type of analysis I am applying to their data;
- To never use my knowledge to violate human rights and civil liberties, even under threat