

Data Collection and Processing in Consumer Smart Devices

Hien Pham – April 2020

"Consumer Smart Devices"??

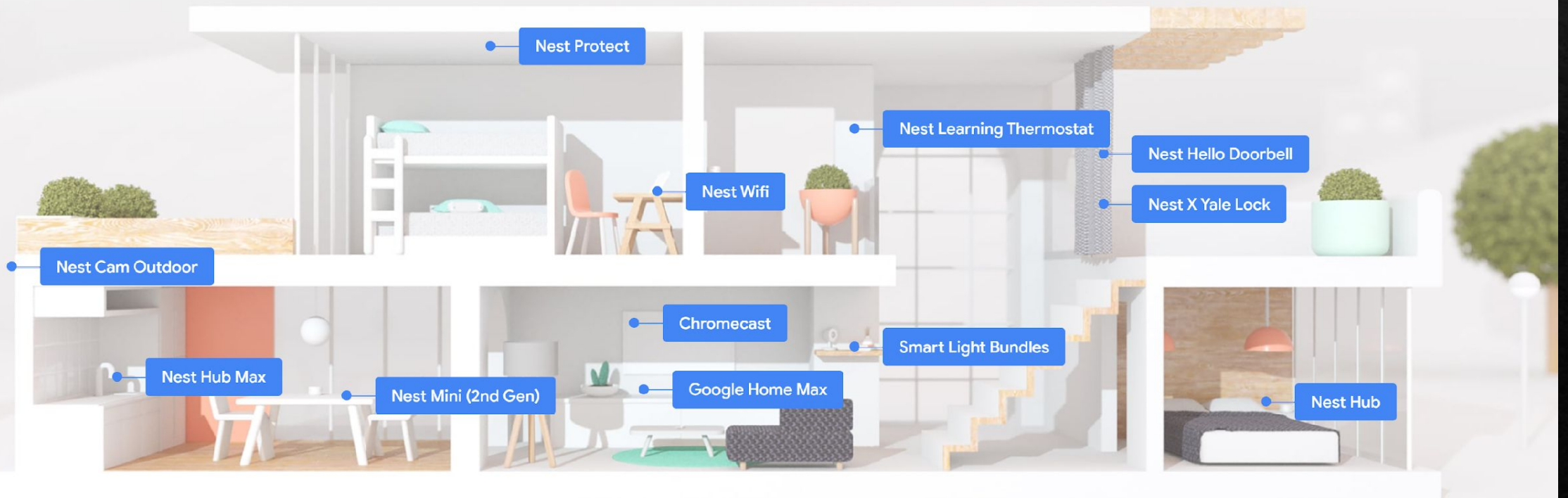
Home Assistance

Safety & Security

Entertainment

Energy & Lighting

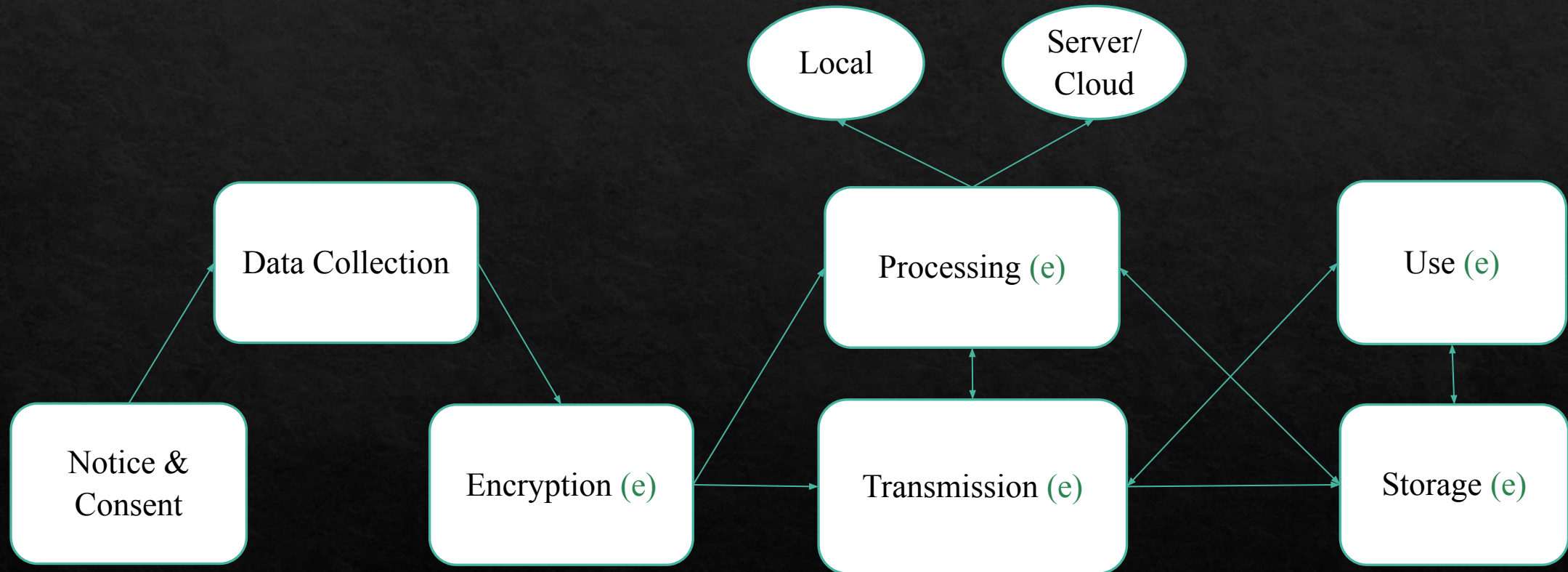
Connectivity



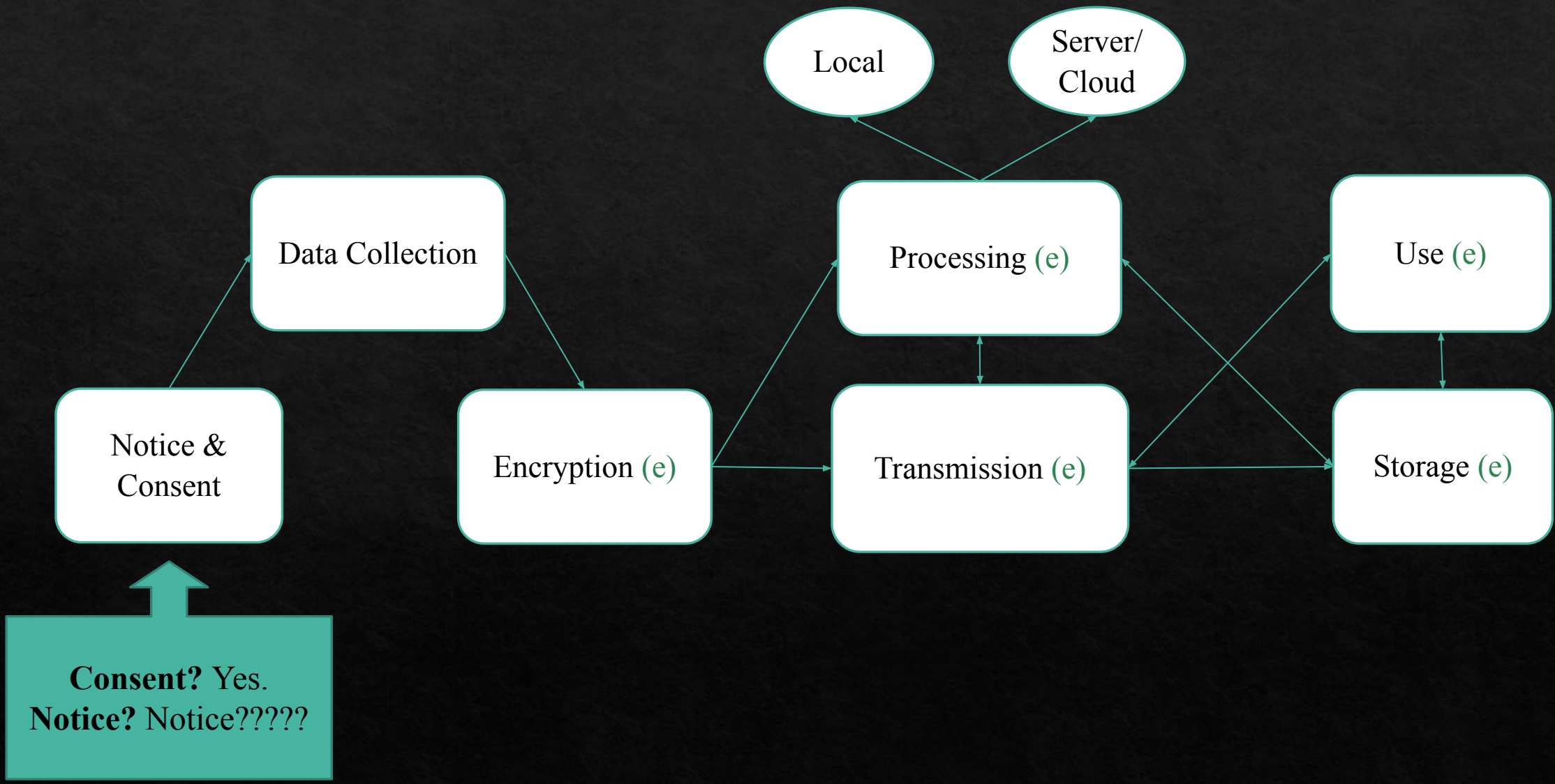
"Consumer Smart Devices"??

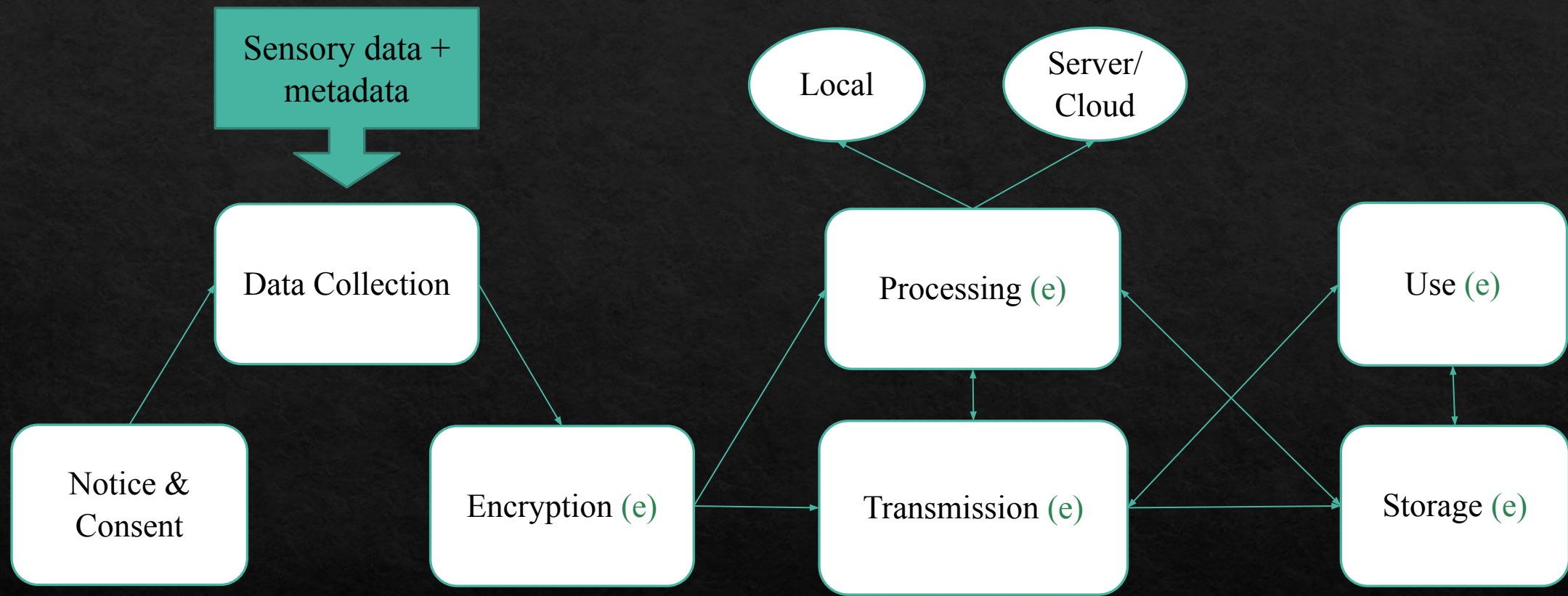


The Flow

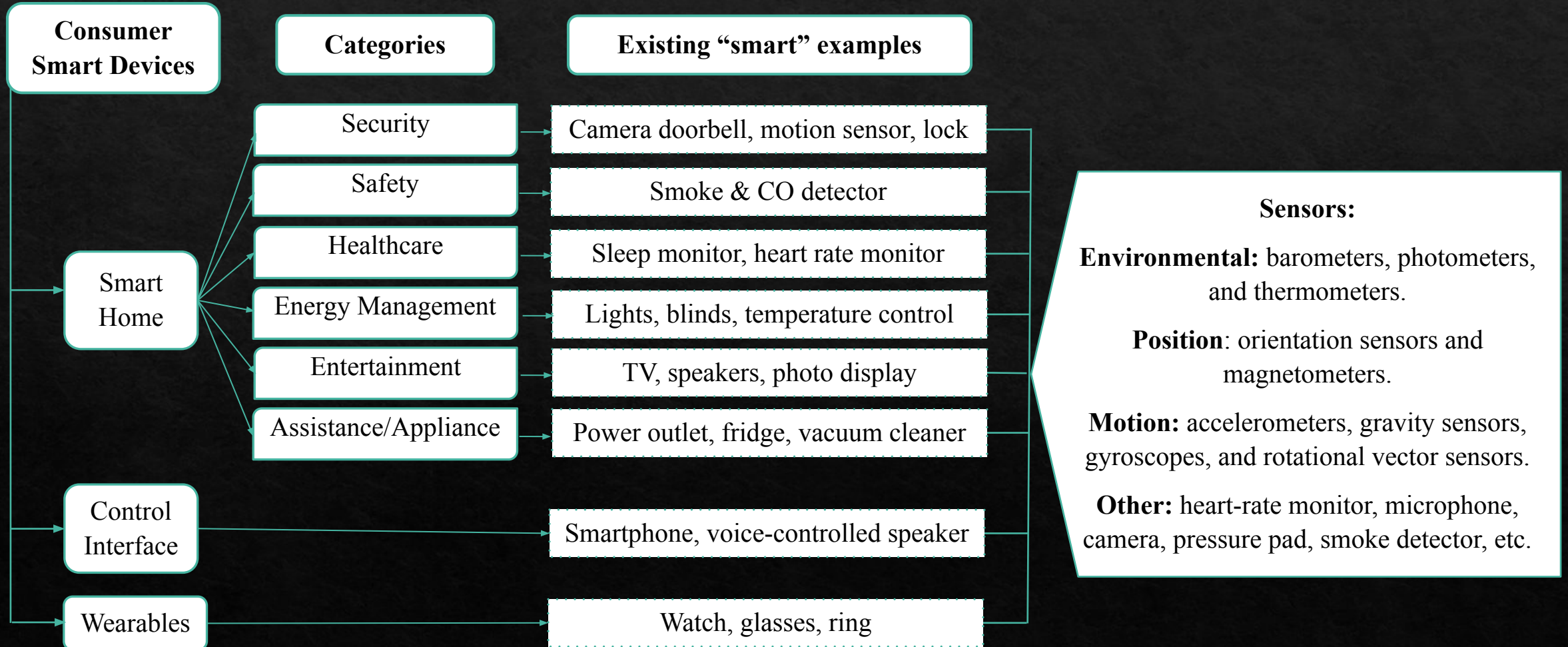


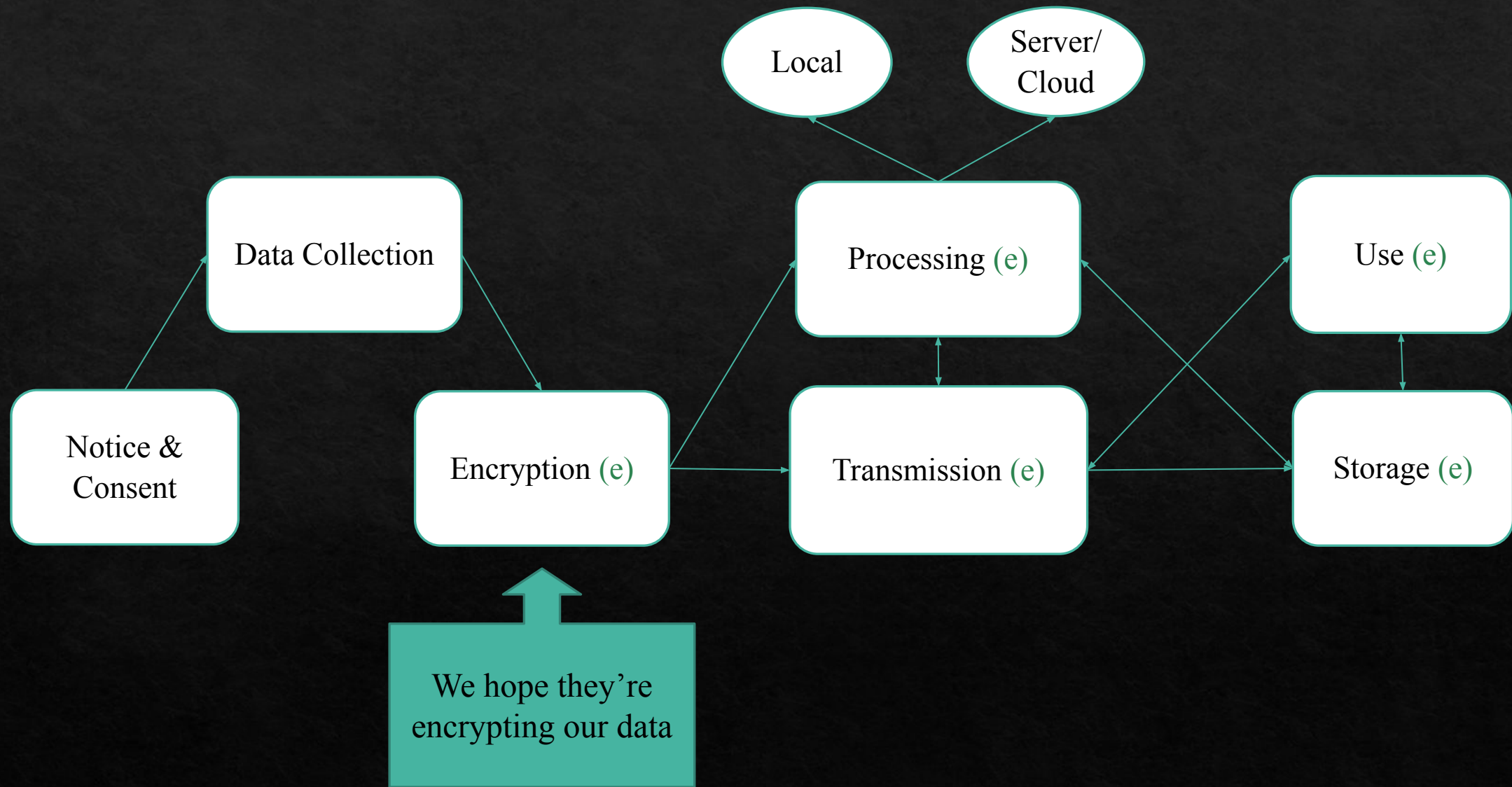
The (simplified and expected) process of collecting and processing data in consumer smart devices. All steps with active encryption are denoted with (e).

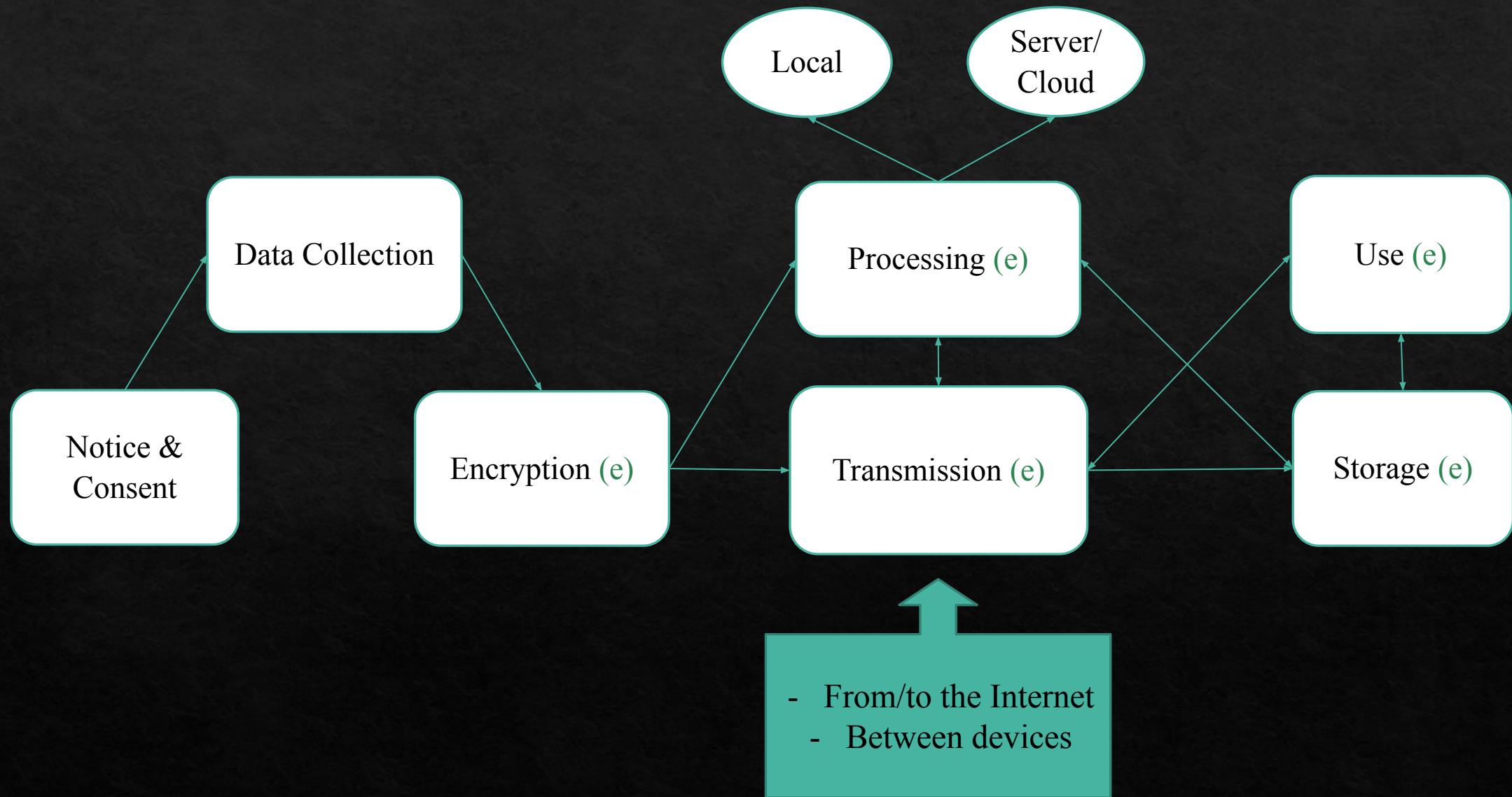




What's a smart device doing?








Security, Encryption & Transmission

- ◆ A 2019 research from Northeastern University & Imperial College on home smart devices found:
 - **72/81 devices shared data with third parties** – including IP addresses, location data, and user habit.
 - **“All devices expose information to eavesdroppers via at least one plaintext flow, and a passive eavesdropper can reliably infer user and device behavior from the traffic (encrypted or otherwise) of 30/81 devices.”**
- ◆ Wearables send a MAC address that every Bluetooth/Wi-Fi connected device has
- ◆ Anonymization: doesn't work if you're just removing the name.

Category	Cameras	Smart Hubs	Home Automation	TV	Audio	Appliances
Devices $N_{US}=46$ $N_{UK}=35$ $N_{US \cap UK}=26$ $N_{US \cup UK}=81$	<ul style="list-style-type: none"> Amazon Cloudcam Amcrest Cam Blink Cam Blink Hub Bosiwo Cam D-Link Cam Lefun Cam Luohe Cam Microseven Cam Ring Doorbell Wansview Cam WiMaker Spy Camera Xiaomi Cam Yi Cam ZModo Doorbell 	<ul style="list-style-type: none"> Insteon Lightify Philips Hue Sengled Smartthings Wink 2 Xiaomi 	<ul style="list-style-type: none"> D-Link Mov Sensor Flux Bulb Honeywell T-stat Magichome Strip Nest T-stat Philips Bulb TP-Link Bulb TP-Link Plug WeMo Plug Xiaomi Strip 	<ul style="list-style-type: none"> Apple TV Fire TV LG TV Roku TV Samsung TV 	<ul style="list-style-type: none"> Allure with Alexa Echo Dot Echo Spot Echo Plus Google Home Mini Google Home Invoke with Cortana 	<ul style="list-style-type: none"> Anova Sousvide Behmor Brewer GE Microwave Netatmo Weather Samsung Dryer Samsung Fridge Samsung Washer Smarter Brewer Smarter iKettle Xiaomi Cleaner Xiaomi Rice Cooker
Purpose	Devices offering or supporting a camera that can be accessed remotely such as smart cameras and doorbells.	Devices designed to integrate non-Wi-Fi wireless devices into an IP network.	Wi-Fi sensors and actuators such as switches, bulbs, movement sensors.	Smart TVs and devices designed to connect to TVs via HDMI.	Smart speakers offering a voice assistant.	Home appliances that offer remote control.
Interaction Experiments	Move in front of camera, watch remotely, record video, take picture, ring (doorbells).	Turn on/off, change brightness/color, move in front of movement sensor.	Turn on/off, change brightness/color, move in front of movement sensor, change temperature (thermostats).	Browse menu, voice command, change volume.	Voice command, change volume.	Start, stop, change temperature, view inside (fridge), voice/volume (fridge).

Table 1: IoT devices under test. From top to bottom: IoT devices by category, their common purpose within the category, and the interaction experiments we performed (if available) on all the devices within the category. Flags indicate the presence of the device in the US, UK, or both testbeds.



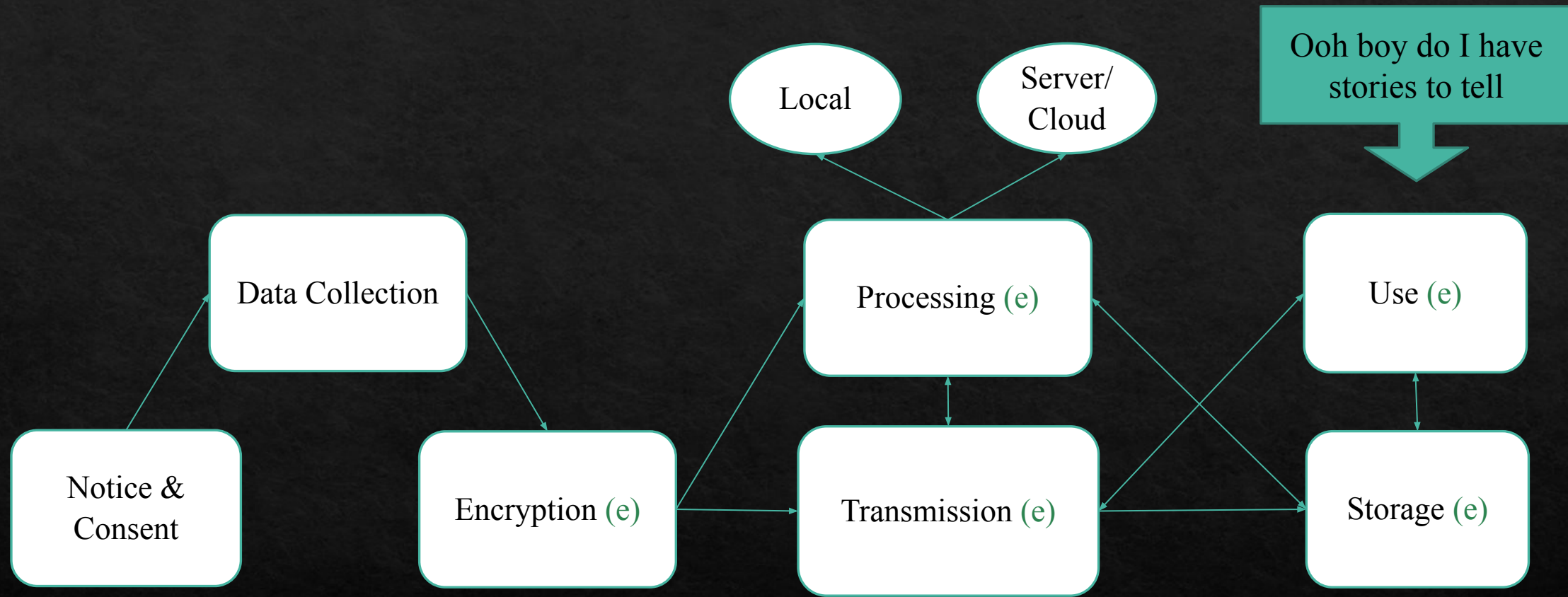
Connecting those pings reveals a diary of the person's life.

The New York Times were able to obtain location data on an individual's smart phone without knowing their name but were able to deduce who they were from where they go to in the morning (work) and where they return to at night (home).

Security & Encryption

Solutions (besides engineers):

- ◆ Open-sourced privacy tools? E.g. Google open-sourced differential privacy.
- ◆ Higher user standards?
 - No one likes 2FA & strong passwords but we need it
 - Customers unwilling to pay for devices with better processors that can handle end-to-end encryption vs. cheaper alternatives.
- ◆ Policies?
 - GDPR: rights to access, rectify, remove, transfer, or restrict data processing.
 - CCPA: protects Californians with rights to access, delete, opt-out of the sale of personal data.



What can you infer from the data?

In academia:

- ◆ **Activities:** Research in 2012 collected raw sensory data from smart home devices and wearables, labelled them with activities (running, sleeping, or eating), then trained models to do the same recognition with accuracy up to 100%(?)
- ◆ **User's biological traits:** Using sensory data from smartphone and wearables.

TABLE 3: Personal trait prediction.

Work	Sensor	Features	Algorithm	Traits	Results
[186]	Accelerometer	Time-domain features	ANN, J48 decision tree algorithms [187], and instance-based learning (IBk) [188]	Weight, height, and gender	71.2% for gender using IBk, 85.7% for height using ANN, and 78.9% for weight using IBk
[182]	Accelerometer and touchscreen	Time-domain features, touch pressure, and size	K -mean nearest neighbor	User identification	More than 96% for identification
[77]	Touchscreen	Delay between pressing two different keys	ANN, nearest neighbor, SVM, gradient descent bp, Euclidean distance, linear discriminant analysis, and another 5 algorithms	Classifying children from adults	More than 92% for SVM and 89% for linear discriminant analysis
[80]	Touchscreen	Delay and duration of pressing	SVM	Gender classification	Accuracy of 91%
[81]	Touchscreen, accelerometer, and gyroscope	29 features including: special keys, total keys pressed, number of backspaces used, edit distance, total completion time, average time between keys	Decision tree (number of keys), SVC linear kernel (age), SVC linear kernel (gender), logistic regression, K -nearest, and Gaussian NB	Number of fingers used, gender, and age	80% for the number of fingers, 75% for age, and 60% for gender
[189]	Touchscreen gestures, gyroscope, accelerometer	14 gesture features, total length, total time, width, height, area, pressure, speed, acceleration, arc distance, and angle start to end	SVM, logistic regression, naive Bayes, J48	Gender classification	71% accuracy for logistic regression
[190]	Fingerprint	Wavelet features and singular value decomposition	K -nearest	Gender classification	Accuracy exceeded 88%
[87]	Touchscreen	Swipe gesture speed in four directions and other features from [189]	Statistical	Thumb length and users' height	Accuracy of 72% of the relation between thumb length and height

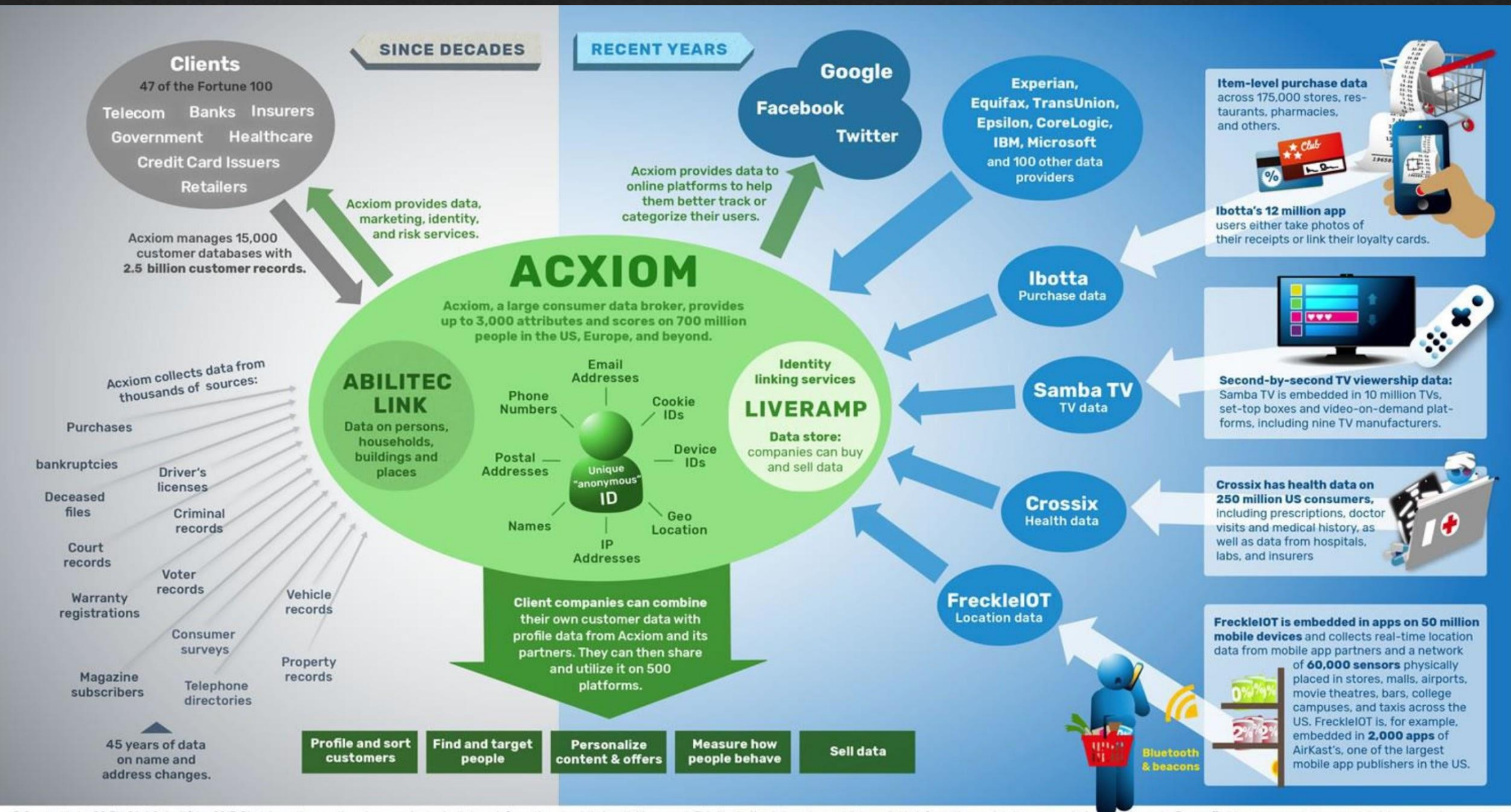
The [Utopian] Future

- ◆ **Ambient Intelligence** – a whole environment that “can extrapolate behavioral characteristics and generate pro-active responses.”
 - Context awareness, personalization, adaptivity and anticipatory behavior
 - “People-empowering smartness” instead of a “system-centric, importunate and automated smartness” system
 - Empower people with control and freedom, especially for semi-autonomous people such as elderly or disabled people.
- ◆ **Choice:** Automated (pre)processing & aggregation of data vs. human intervention & decision making.
 - Need transparency in data processing & trust.

Reality check!

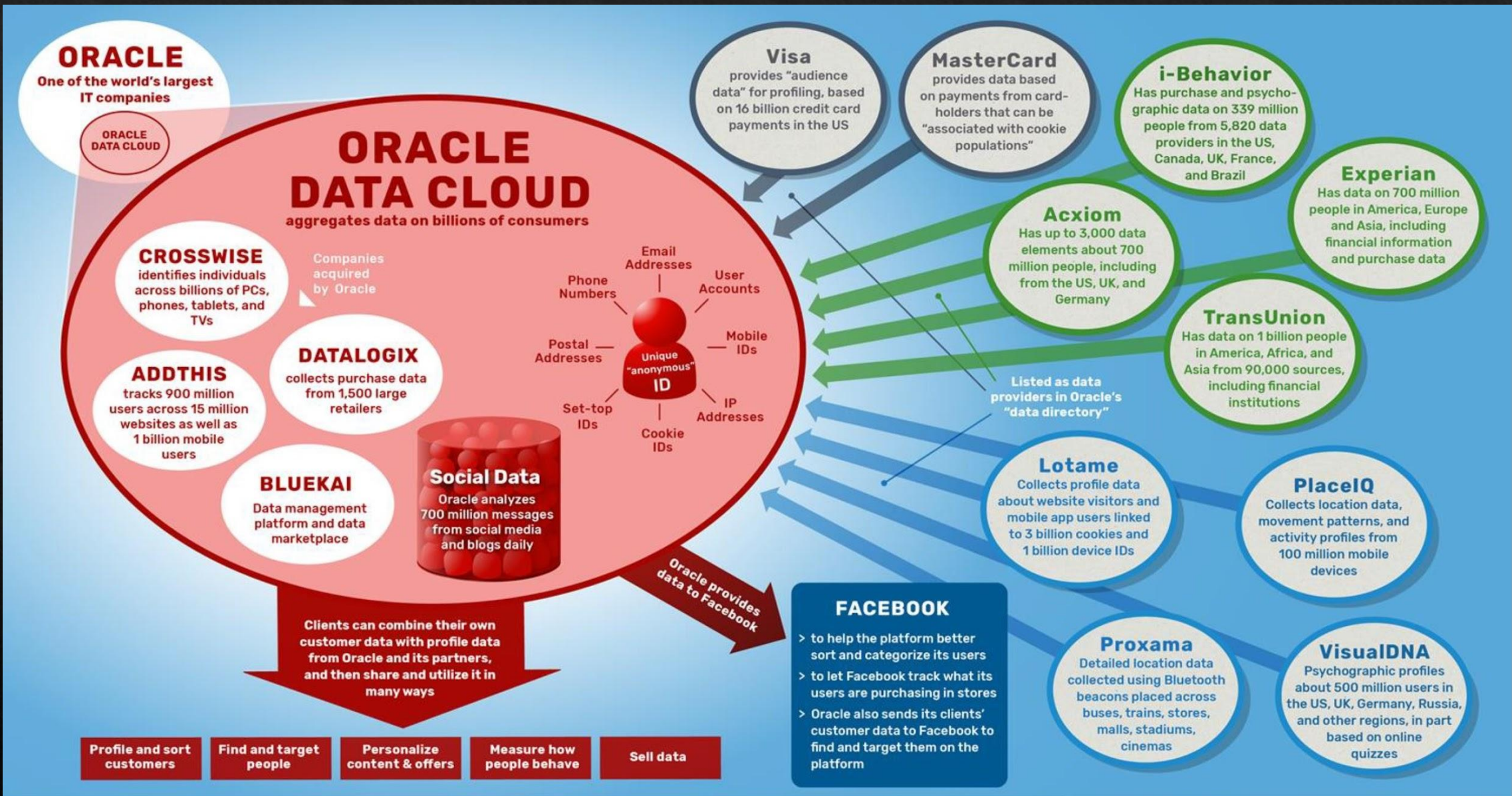
Data aggregation & mining

(because data from your smart devices alone isn't enough)



© Cracked Labs CC BY-SA 4.0, April/May 2017. Disclaimer: the mentioned companies typically keep information about their activities secret. This illustration is based on publicly available information, mainly the companies' own statements. Every effort has been made to accurately interpret and represent the companies' activities, but we cannot accept any liability in the case of eventual errors. Sources: Acxiom website, press releases, brochures, annual reports, and response to US congress inquiry. LiveRamp website, brochures, press releases, presentations. Ibotta website. Crossix website, press releases. FreckleIoT press releases. For details about the sources see the report "Corporate Surveillance in Everyday Life".

Figure 2: Acxiom and some of its data providers, partners and services.

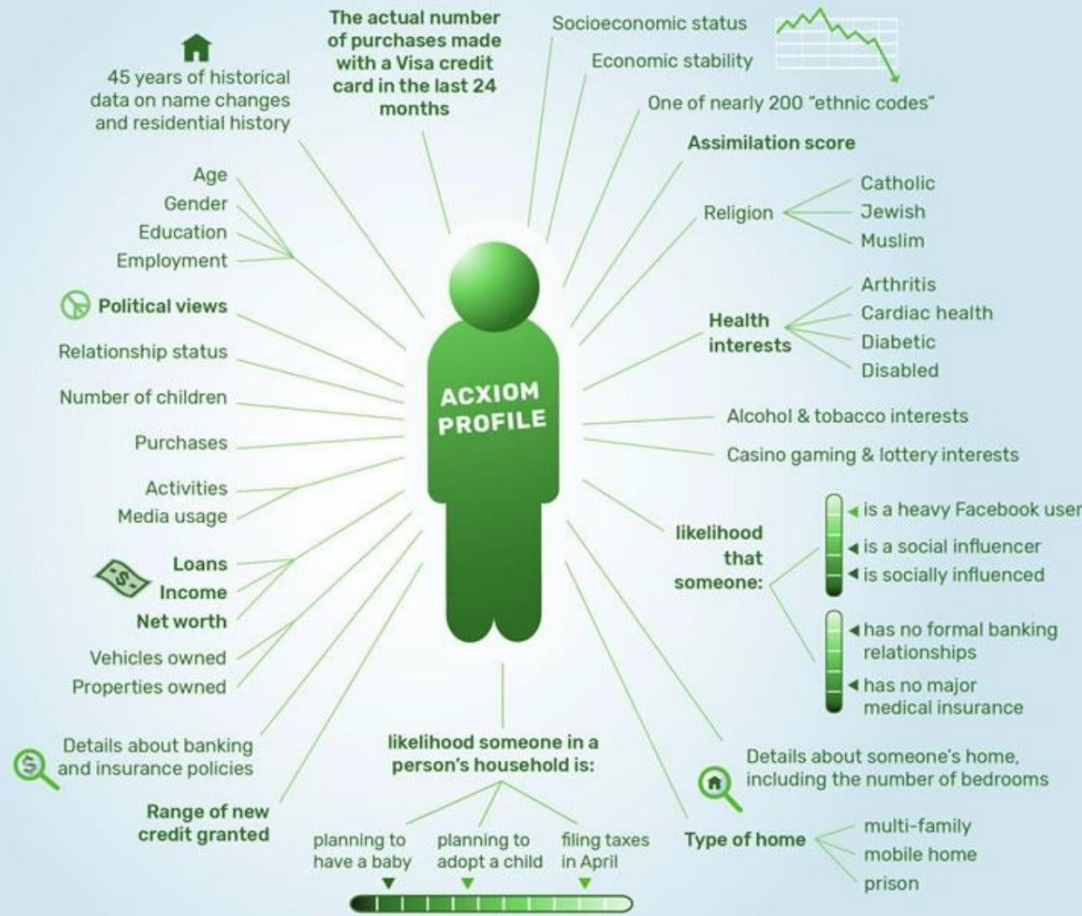


© Cracked Labs CC BY-SA 4.0, April/May 2017. Disclaimer: the mentioned companies typically keep information about their activities secret. This illustration is based on publicly available information, mainly the companies' own statements. Every effort has been made to accurately interpret and represent the companies' activities, but we cannot accept any liability in the case of eventual errors. Sources: Oracle website, press releases, data directory, brochures, presentations, MasterCard website, Acxiom annual report, TransUnion annual report, Lotame website, VisualDNA brochure, Facebook website, ProPublica article. For details about the sources see the report "Corporate Surveillance in Everyday Life".

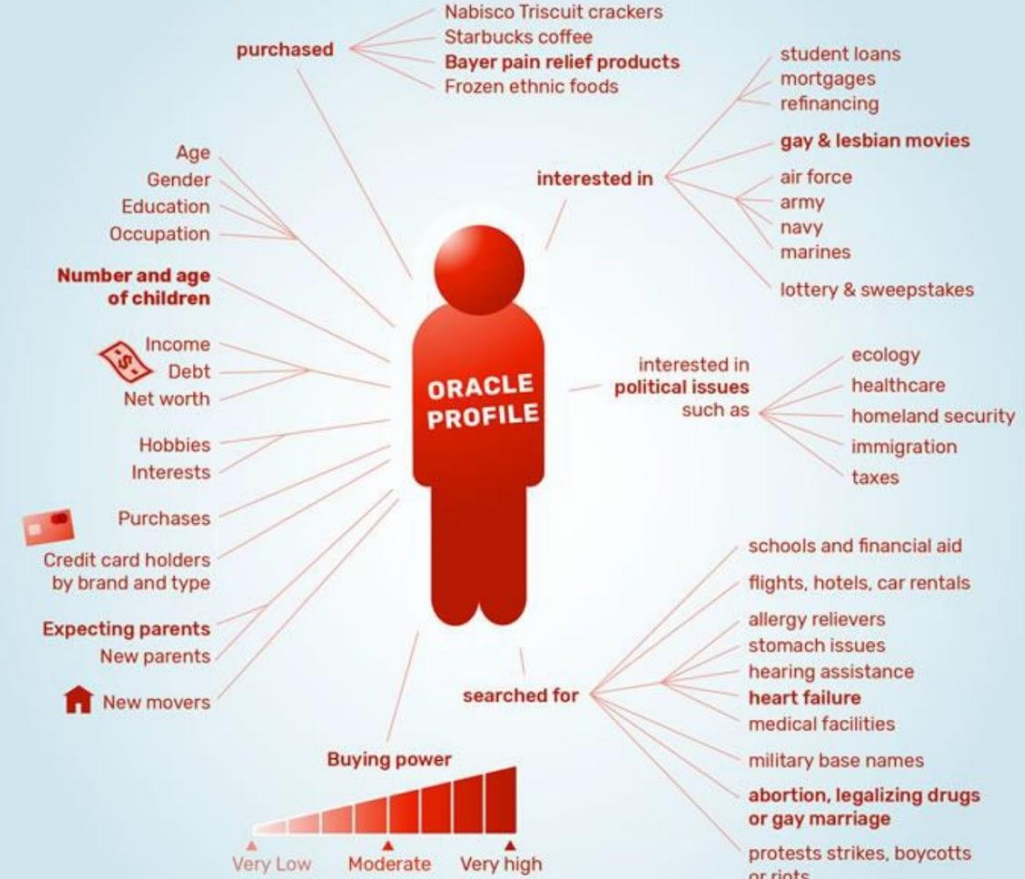
Figure 3: Oracle and some of its data providers, partners and services.

DATA BROKERS HAVE EXTENSIVE PROFILE INFORMATION ON ENTIRE POPULATIONS

Examples of data on consumers provided by Acxiom and Oracle



Acxiom provides up to 3,000 attributes and scores on 700 million people in the US, Europe, and other regions.



Oracle sorts people into thousands of categories and provides > 30,000 attributes on 2 billion consumer profiles

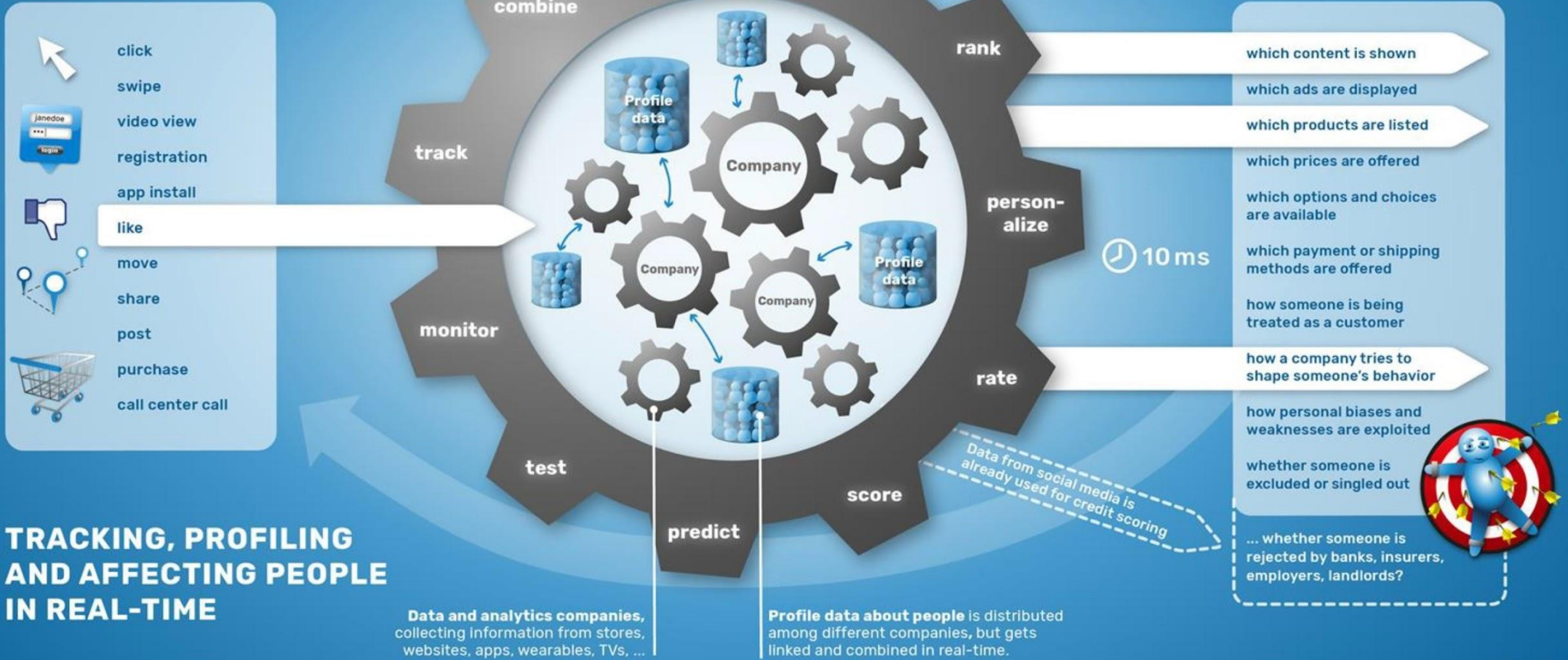
© Cracked Labs CC BY-SA 4.0, April/May 2017. Disclaimer: the mentioned companies typically keep information about their activities secret. This illustration is based on publicly available information by Acxiom and Oracle. Every effort has been made to accurately interpret and represent the companies' activities, but we cannot accept any liability in the case of eventual errors. Sources: Acxiom annual reports, developer website (API docs), Oracle press release, help center website, audience playbook, taxonomy updates for January, 2017 (Excel document). For details about the sources see the report "Corporate Surveillance in Everyday Life".

Figure 4: Examples of data on consumers provided by Acxiom and Oracle.

EVERY INTERACTION...

...TRIGGERS A WIDE RANGE OF DATA FLOWS BETWEEN MULTIPLE COMPANIES

...AND CAN AFFECT:



TRACKING, PROFILING AND AFFECTING PEOPLE IN REAL-TIME

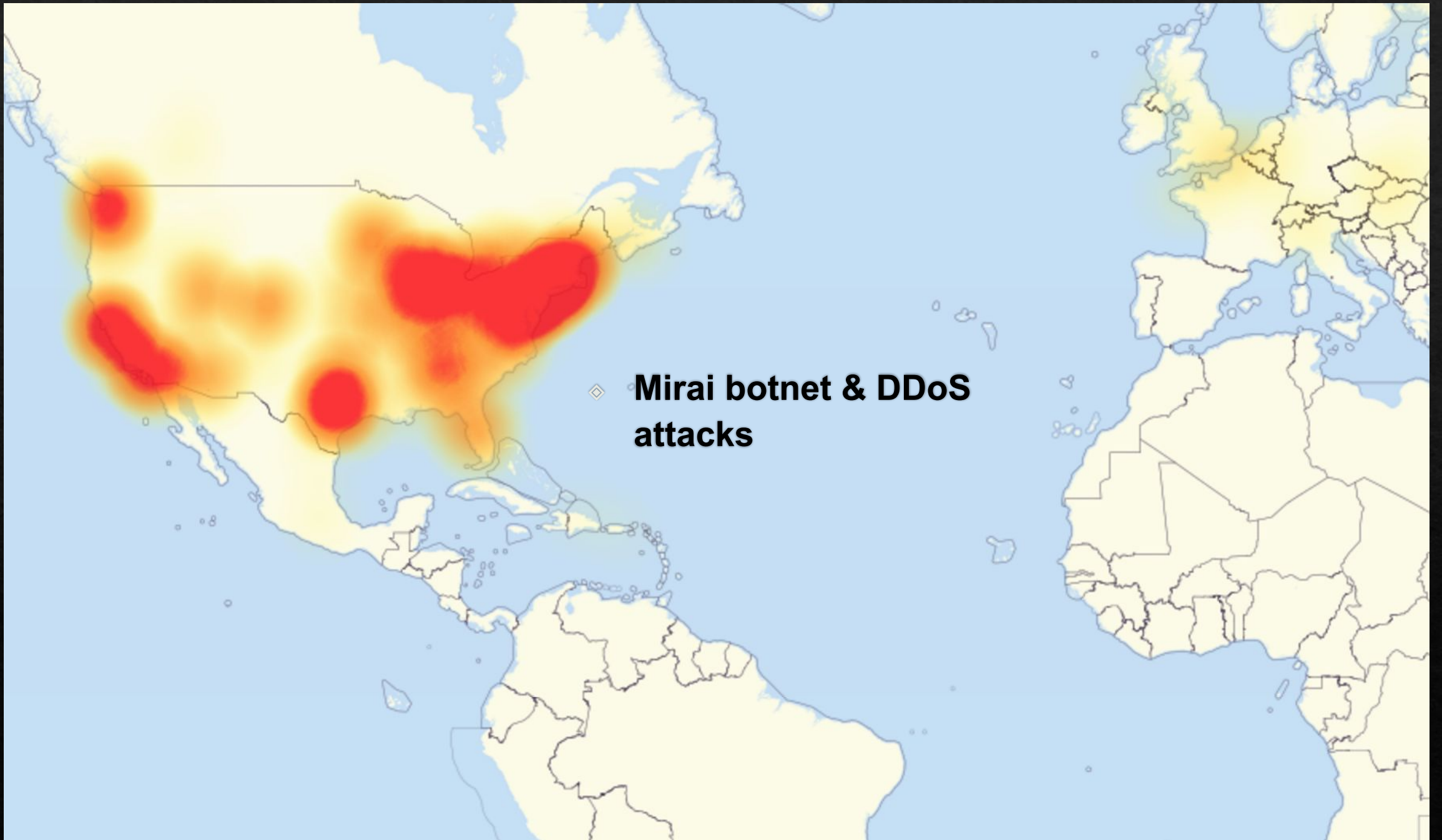
Data and analytics companies, collecting information from stores, websites, apps, wearables, TVs, ...

Profile data about people is distributed among different companies, but gets linked and combined in real-time.

Figure 6: Tracking, profiling and affecting people in real-time.

What's your data powering?

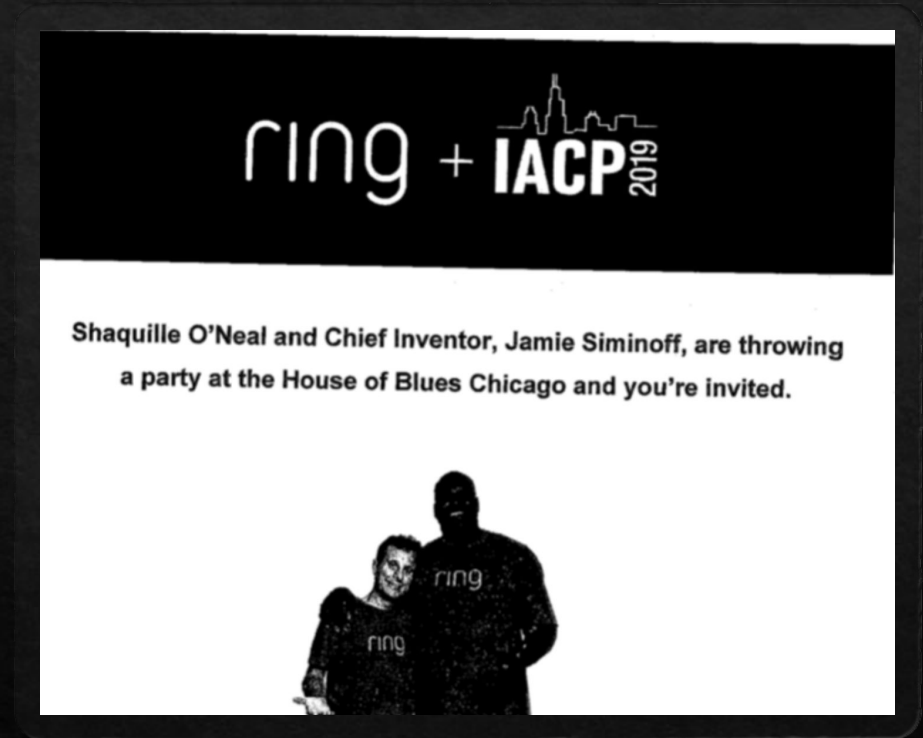
- ◆ **“Trustev, an online identity verification and fraud prevention company evaluates digital transactions for clients in financial services, government, healthcare, and insurance in real-time by analyzing digital behaviors, identities, and devices such as phones, tablets, laptops, game consoles, TVs, and even refrigerators.”**
- ◆ Insurance assessment
- ◆ Credit score calculation
- ◆ Risk assessment for inmates
- ◆ Hiring decisions
- ◆ Home and auto loans
- ◆ “Custom audiences” for targeted & behavioral advertising



◇ **Mirai botnet & DDoS attacks**

Ring & surveillance

- ◆ Smart (read: camera & internet connected) doorbell company owned by Amazon
- ◆ Has over 600 partnerships with law enforcement agencies (by Dec 2019)
- ◆ Law enforcement gets access to an interactive map with approximate location of all Ring devices in an area.
- ◆ Ring coaches police on how to convince people to give their camera footage to police without a warrant.
- ◆ According to a memo obtained from the police department of Pomona, CA, when camera owners are "uncooperative or unavailable," officers are instructed to contact Ring and request that the captured video be preserved.



Back to Data Collection

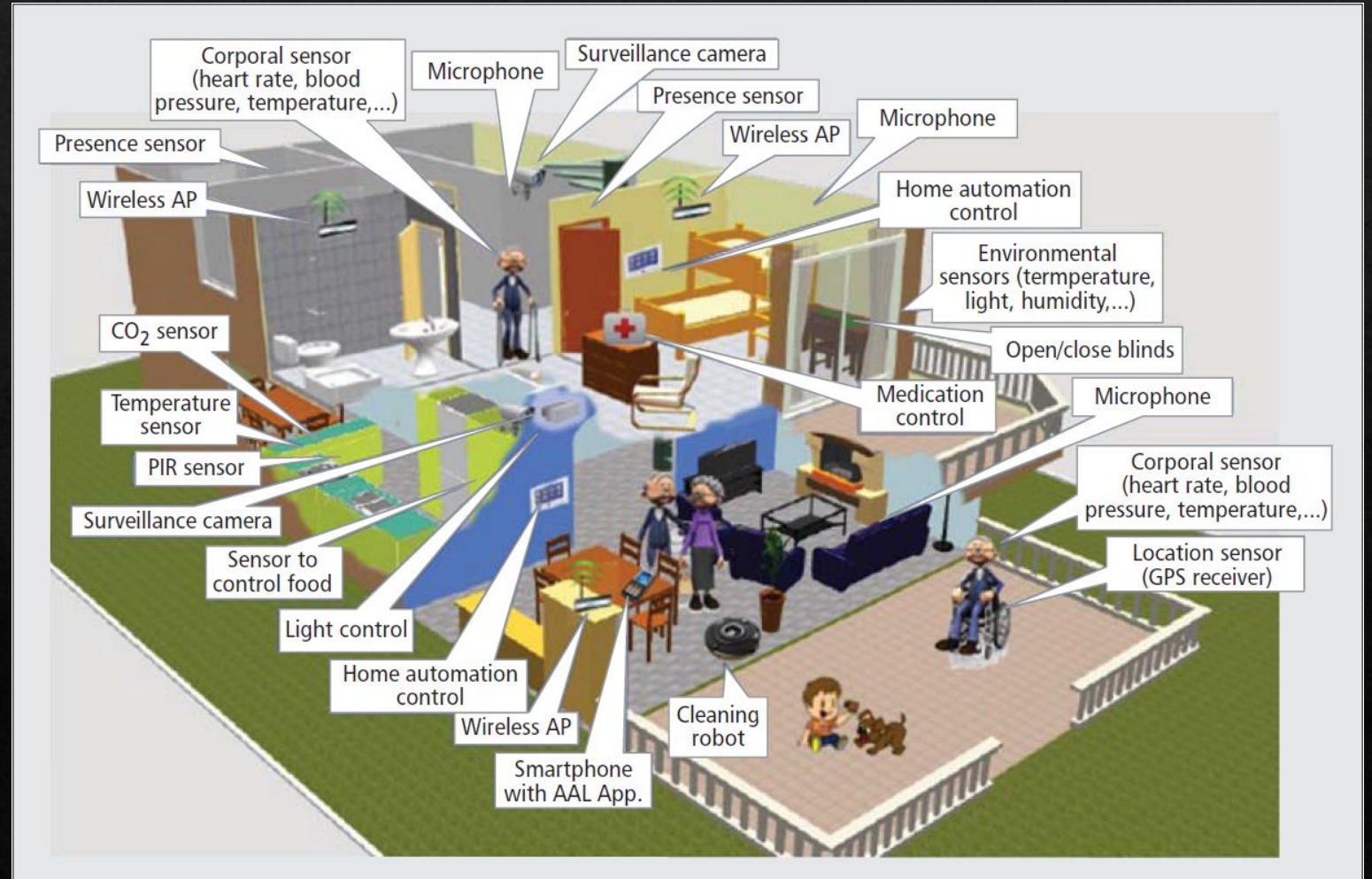
- ◆ **Data minimization** – collect only data for disclosed purpose
 - GDPR requires for sensitive data, but no US laws does
 - Doesn't account for data that can be mined
 - Doesn't help with machine learning's need for large amounts of data.
- ◆ **On-device processing**
- ◆ Alternatives?

Thoughts?



Internet-connected everything?

(Picture: smart toilet with built-in speakers, ambient lighting, and Amazon Alexa



Sources

- ◇ Christl, Wolfie. “Corporate Surveillance In Everyday Life. How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions.” Cracked Labs, June 8, 2017. <http://crackedlabs.org/en/corporate-surveillance>.
- ◇ CNN, Elizabeth Wolfe and Brian Ries. “A Hacker Accessed a Family’s Ring Security Camera and Told Their 8-Year-Old Daughter He Was Santa Claus.” *CNN*. Accessed January 4, 2020. <https://www.cnn.com/2019/12/12/tech/ring-security-camera-hacker-harassed-girl-trnd/index.html>.
- ◇ Fruhlinger, Josh. “The Mirai Botnet Explained: How IoT Devices Almost Brought down the Internet.” *CSO Online* (blog), March 9, 2018. <https://www.csoonline.com/article/3258748/the-mirai-botnet-explained-how-teen-scammers-and-cctv-cameras-almost-brought-down-the-internet.html>.
- ◇ Google Store. “Google Connected Home Devices.” Accessed January 10, 2020. https://store.google.com/us/category/connected_home
- ◇ Haskins, Lia Kantrowitz, Caroline. “Inside Ring’s Quest to Become Law Enforcement’s Best Friend.” *Vice*, December 4, 2019. https://www.vice.com/en_us/article/bjw9e8/inside-rings-quest-to-become-law-enforcements-best-friend.
- ◇ Masoud, Mohammad, Yousef Jaradat, Ahmad Manasrah, and Ismael Jannoud. “Sensors of Smart Devices in the Internet of Everything (IoE) Era: Big Opportunities and Massive Doubts.” Research article. *Journal of Sensors*, 2019. <https://doi.org/10.1155/2019/6514520>.
- ◇ Richard, Lauren. “Cyber Privacy: Notice-and-Consent vs. Responsible Use.” *Paymetric* (blog), August 6, 2014. </uncategorized/cyber-privacy-notice-consent-vs-responsible-use/>.
- ◇ Wikipedia. “2016 Dyn Cyberattack.” In *Wikipedia*, October 31, 2019. https://en.wikipedia.org/w/index.php?title=2016_Dyn_cyberattack&oldid=923850977.
- ◇ Android Developers. “Sensors Overview.” Accessed April 26, 2020. <https://developer.android.com/reference/android/sensors>

More Sources

- ◆ Angwin, Julia, and Jeff Larson. "Machine Bias." *ProPublica*, May 23, 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- ◆ Cassagnol, Danielle. "U.S. Consumer Tech Sales To Surpass \$400 Billion Milestone in 2019, Says CTA." Blog. Consumer Technology Association, January 6, 2020. [https://www.cta.tech/News/Press-Releases/2019/July/U-S-Consumer-Tech-Sales-To-Surpass-\\$400-Billion-M.aspx](https://www.cta.tech/News/Press-Releases/2019/July/U-S-Consumer-Tech-Sales-To-Surpass-$400-Billion-M.aspx).
- ◆ Los Angeles Times. "CES 2020: Surveillance Tech Is All the Rage at Annual Gadget Expo," January 6, 2020. <https://www.latimes.com/business/technology/story/2020-01-06/surveillance-tech-is-all-the-rage-at-ces-2020>.
- ◆ Christl, Wolfie. "Corporate Surveillance In Everyday Life. How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions." Cracked Labs, June 8, 2017. <http://crackedlabs.org/en/corporate-surveillance>.
- ◆ CNN, Elizabeth Wolfe and Brian Ries. "A Hacker Accessed a Family's Ring Security Camera and Told Their 8-Year-Old Daughter He Was Santa Claus." *CNN*. Accessed January 4, 2020. <https://www.cnn.com/2019/12/12/tech/ring-security-camera-hacker-harassed-girl-trnd/index.html>.
- ◆ Combs, Veronica. "Predictive Tech Needs More of Everything: Education, Data, and Regulation." *TechRepublic*. Accessed January 9, 2020. <https://www.techrepublic.com/article/predictive-tech-needs-more-of-everything-education-data-and-regulation/>.
- ◆ Cook, Diane. "Learning Setting-Generalized Activity Models for Smart Spaces." *IEEE Intelligent Systems* 27, no. 1 (January 2012): 32–38. <https://doi.org/10.1109/MIS.2010.112>.
- ◆ Corrigan, Jack. "CBP Plans to Use Facial Recognition For 'All Passenger Applications.'" Nextgov.com, August 9, 2019. <https://www.nextgov.com/emerging-tech/2019/08/cbp-plans-use-facial-recognition-all-passenger-applications/159086/>.
- ◆ Cowan, Jill, and Natasha Singer. "How California's New Privacy Law Affects You." *The New York Times*, January 3, 2020, sec. U.S. <https://www.nytimes.com/2020/01/03/us/ccpa-california-privacy-law.html>.
- ◆ Cox, Joseph. "We Tested Ring's Security. It's Awful." *Vice*, December 17, 2019. https://www.vice.com/en_us/article/epg4xm/amazon-ring-camera-security.
- ◆ Cyphers, Bennett, and Gennie Gebhart. "Behind the One-Way Mirror: A Deep Dive Into the Technology of Corporate Surveillance." Electronic Frontier Foundation, December 11, 2019. <https://www.eff.org/document/behind-one-way-mirror-deep-dive-technology-corporate-surveillance>.
- ◆ EU. Art. 12 GDPR – Transparent information, communication and modalities for the exercise of the rights of the data subject. Accessed January 10, 2020. <https://gdpr-info.eu/art-12-gdpr/>.
- ◆ EU. Chapter 4 GDPR – Controller and processor. Accessed January 10, 2020. <https://gdpr-info.eu/chapter-4/>. 20

Even More Sources

- ◇ EU. Recital 51 - Protecting Sensitive Personal Data. Accessed January 10, 2020. <https://gdpr-info.eu/recitals/no-51/>.
- ◇ Facebook. "About Custom Audiences from Customer Lists." Facebook Ads Help Center. Accessed January 10, 2020. <https://www.facebook.com/business/help/341425252616329>.
- ◇ Facebook. "Facebook Advertising Targeting Options." Facebook for Business. Accessed January 10, 2020. <https://www.facebook.com/business/ads/ad-targeting>.
- ◇ Gizmodo. "Facebook Is Giving Advertisers Access to Your Shadow Contact Information." Accessed January 4, 2020. <https://gizmodo.com/facebook-is-giving-advertisers-access-to-your-shadow-co-1828476051>.
- ◇ Federal Trade Commission. "Privacy Online: A Report To Congress." Federal Trade Commission, June 1998.
- ◇ Forrest, Connor. "Google Assistant: A Cheat Sheet." *TechRepublic*, October 15, 2019. <https://www.techrepublic.com/article/google-assistant-the-smart-persons-guide/>.
- ◇ Fruhlinger, Josh. "The Mirai Botnet Explained: How IoT Devices Almost Brought down the Internet." *CSO Online* (blog), March 9, 2018. <https://www.csoonline.com/article/3258748/the-mirai-botnet-explained-how-teen-scammers-and-cctv-cameras-almost-brought-down-the-internet.html>.
- ◇ Google. "Google Nest Commitment to Privacy in the Home," January 6, 2020. https://store.google.com/magazine/google_nest_privacy.
- ◇ Google. "Manage Google Voice and Audio Recordings - Computer." Google Search Help, January 9, 2020. <https://support.google.com/websearch/answer/6030020?co=GENIE.Platform%3DDesktop&hl=en>.
- ◇ Google. "Sensors in Google Nest Devices - Google Nest Help," January 6, 2020. <https://support.google.com/googlenest/answer/9330256>.
- ◇ Google. "Targeting Your Ads - Google Ads Help." Google Ads Help. Accessed January 10, 2020. <https://support.google.com/google-ads/answer/1704368?hl=en>.
- ◇ Graham, Jefferson. "Hey Siri, Google and Alexa — Enough with the Snooping." *USA TODAY*. Accessed January 10, 2020. <https://www.usatoday.com/story/tech/2019/08/24/hey-siri-google-alex-and-cortana-stop-snooping-now/2083899001/>.
- ◇ Haskins, Caroline. "A Data Leak Exposed The Personal Information Of Over 3,000 Ring Users." News Site. BuzzFeed News, December 19, 2020. <https://www.buzzfeednews.com/article/carolinehaskins1/data-leak-exposes-personal-data-over-3000-ring-camera-users>.
- ◇ Haskins, Caroline. "How Ring Transmits Fear to American Suburbs." *Vice*, December 6, 2019. https://www.vice.com/en_us/article/ywaa57/how-ring-transmits-fear-to-american-suburbs.