

Data Collection Practices of Large Online Platforms

Justin Brookman, Consumer Reports

Background about CR

6m+

Members

14m+

Unique
visitors/month

1m+

Survey responses

63

state-of-the-art
labs

327 acres

at Auto Test Center

2800+

products tested
annually

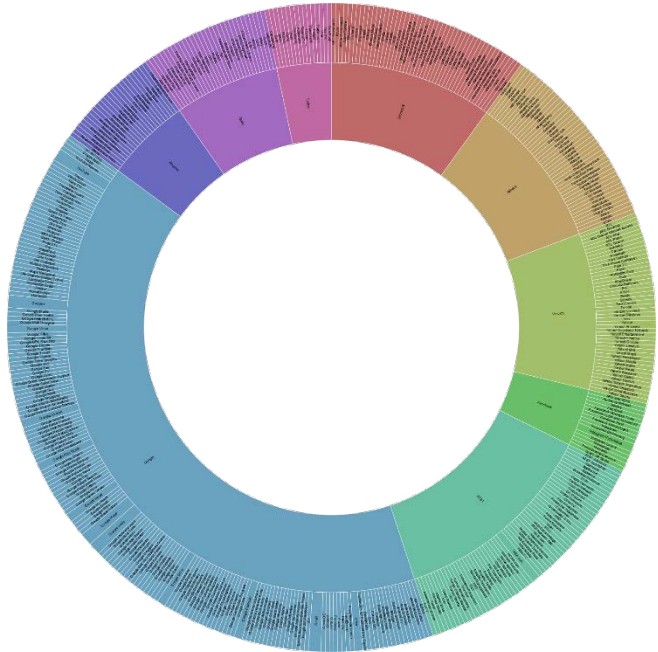
MOBILE PEER-TO-PEER PAYMENT SERVICES

SERVICE	OVERALL SCORE	PAYMENT AUTHENTICATION	DATA SECURITY	DATA PRIVACY	CUSTOMER SUPPORT	BROAD ACCESS
Apple Pay	76					
Venmo	69					
Cash App (Square)	64					
Facebook P2P Payments in Messenger	63					
Zelle (standalone app)	50					

Premise of the Study

- Look at the documentable data practices of 15 large internet platforms
 - Measurement
 - Policy analysis
 - External research and documentation

1. Alibaba
2. Amazon
3. Apple
4. AT&T
5. Charter
6. Comcast
7. Cox
8. Facebook
9. Google
10. Microsoft
11. Sprint
12. T-Mobile
13. Twitter
14. Verizon
15. Wikimedia



- **Data Collection:** What they collect
- **Data Use:** How they use data
- **Data Sharing:** With whom shared and why
- **Data Retention:** When deleted by default
- **Data Control:** Controls given to users re above

Methodology — Source Documents

- **136 Primary Sources:** Privacy policies, terms of use, conditions of use, account settings, EULAs, privacy control dashboards, etc.
- **86 Secondary Sources:** News articles, Congressional testimony, etc.
- **40+ Tertiary Sources:** Academic papers, environmental scans, etc.

Inferences

- Platforms use correlation and other techniques to infer more information than they explicitly gather
- Conducted a literature review to understand the forms of inference currently in use (or possible to use), and examples of their usage
- Sets up possibilities for future experiments and testing

Policy Constraints

- Public documentation about self-imposed limits
 - Section 5 liability
- Where everything technologically obtainable, policy constraints most meaningful limitations
- Testing may have to rely on affirmative forbearance

Collection Measurement

- Partnered with AppCensus to analyze data collection and sharing on 426 apps
 - Prolific use of (semi) permanent identifiers instead of Advertising ID
 - Encrypted data sent to measurement firms

Findings from Documentation

Activity on third-party apps
Activity on third-party sites
Ads clicked
Ads viewed
Apps used
Browser settings
Browser used
Browsing history
Calling-party phone number
Comments on videos made
Content created
Content received
Content uploaded
Crash reports
Date and time of call
Date and time of message
Date and time of search
Device location
Device operating system

Device settings
Device type
Device unique identifier
Documents saved on platform
Duration of call
Email content
Emails received
Emails written
Forwarding numbers of call
IP address
Locations history
Locations search history
Metadata user search history
Mobile network carrier name
Mobile network phone number
Mouse movements
Other users connected w/user
Other users sharing content w/user

Payment information
Phone number calling
Phone number messaging
Photos saved on platform
Publicly available information of user
Receiving-party phone number
Receiving-party phone number
Routing information of call
Search history
System activity
Type of call
User account information
User cookies information
User location
Videos saved on platform
Videos watched
Voice and audio information
Websites visited

Other Findings

- Only incomplete information available
- Many limitations on sharing, some limitations on use
- Few self-imposed restrictions on collection limitation/data minimization
 - First-party collection, presence on other services, data purchase or other matching
- Few limitations on sharing across products or contexts