

# (What is Special about) Markets for Data

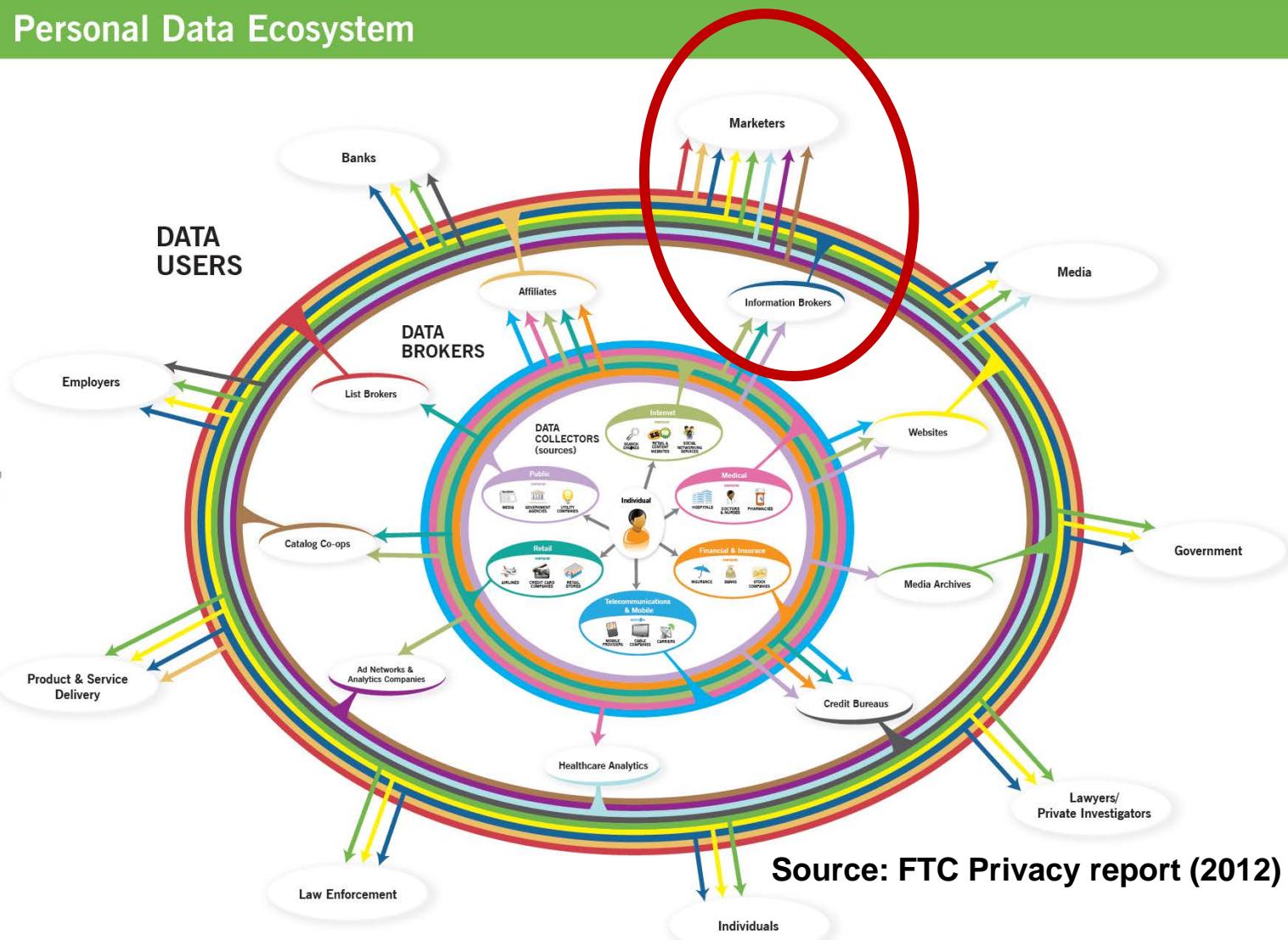
Alessandro Bonatti

MIT and TSE

The Economics of Intellectual Property,  
Software, and the Internet

Toulouse, January 12–13, 2017

# Personal Data Ecosystem



# “Original Lists”



products ▾ data packages perspectives partners about ▾ blog



Below are some of our most effective elements organized by channel, industry or area of interest.

## A Automotive

- Presence of Children
- Income Code -- Estimated Household
- NetWorth -- Gold
- In Market for a New Domestic Luxury Vehicle
- In Market for a New Domestic Regular Vehicle
- In Market for a New Japanese Luxury Vehicle
- In Market for a New Japanese Regular Vehicle
- In Market for a New Korean
- Pay Cash for a New Vehicle
- Finance a New Vehicle

## \$ Financial Services

- Income Code -- Estimated Household
- Home Owner / Renter
- Economic Stability Indicator
- UnderBanked
- NetWorth -- Gold
- Heavy Transactors
- Shopped for Banking Services Via the Internet
- Full Service Investor
- Self-Service Investor
- Have Online Trading Account

## Insurance

- Presence of Children
- Income Code -- Estimated Household
- Home Owner / Renter
- Networth Gold
- You and Other Household Members Have Medical Insurance
- You Alone Have Medical Insurance
- Obtain Medical Insurance from an Agent Representing One Company
- Obtain Medical Insurance from Agent or Broker Representing Many Companies
- Obtain Medical Insurance Through the Internet
- Have Long Term Care Insurance

## Heart Non-Profit

- Income Code -- Estimated Household
- NetWorth -- Gold
- Age in Two-Year Increments
- Childrens Age Ranges Present in Household
- Community Involvement -- Causes Supported Financially
- Green Living Lifestyle Propensity
- Written or Called any Politician at the State, Local, or National Level
- Written or Called any Politician at the State, Local, or National Level

## Checkmark Political

- Childrens Age Ranges Present in Household
- Home Owner / Renter
- Community Involvement -- Causes Supported Financially
- Green Living Lifestyle Propensity
- Written or Called any Politician at the State, Local, or National Level
- Congressional District
- State legislature assignment for the upper level (senate)

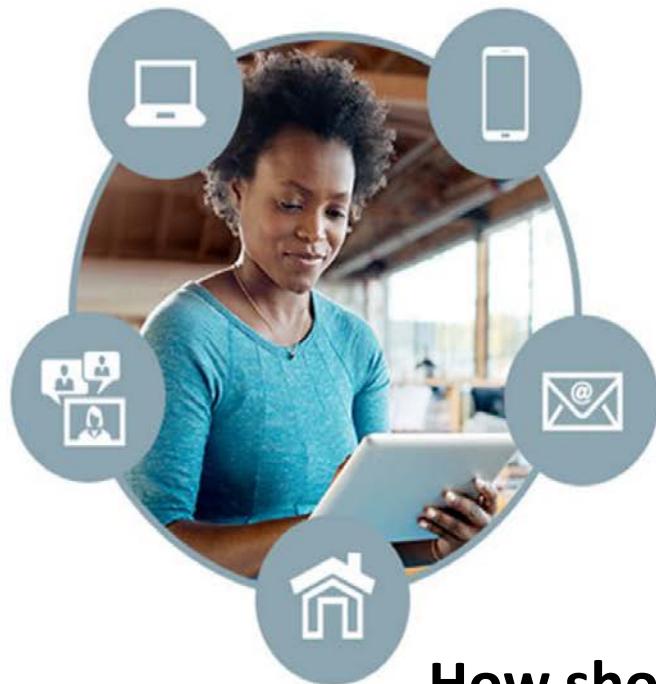
## Retail

- Childrens Age Ranges Present in Household
- Presence of Children
- Income Code -- Estimated Household
- Home Owner / Renter
- Networth Gold
- Technology Adoption
- Infobase--X Affordability
- Purchase Apparel from an Upscale Store

# “Data Appends”

## Win with Smart Data

Oracle ID Graph reaches more than 90 percent of US online consumers by connecting active cookies, mobile IDs, emails, registrations, and social IDs. It breaks down the silos, enabling you to build audiences you can prospect, connect, and personalize for every customer interaction.



- Leverage active IDs for recency and accuracy
- Activate marketing experiences across platforms for an individual
- Accurately measure media exposure, engagement, and sales

**How should these products be priced?**

# Selling Information: Theory

Information = input into a (strategic) decision problem

Inherently multidimensional (e.g., cookies, credit scores)

Initial information & preferences → optimal use of  
additional information (“production technology”)

Only valuable if it changes optimal actions (Blackwell)

# Selling Information: Implications

Profitable **versioning**: buyers value different dimensions

“Damaged goods” to low-value buyers

Peculiarity of information: buyers with lowest willingness to pay also have very specific preferences

E.g., very aggressive lenders are interested in very negative information only

# Equifax “Undisclosed Debt Monitoring”

Data Module	Description	Basic	Plus	Premium
Base Monitoring	<ul style="list-style-type: none"><li>• Debt to Income (DTI) <b>New!</b></li><li>• Trade Account Number</li><li>• Date to File</li><li>• Date Opened</li><li>• Forward and look back periods 120 days</li></ul>	X	X	X
New Collection	Any new collection found on borrower's file <b>New!</b>		X	X
New Public Record	Any new bankruptcies, foreclosures, judgments, or liens found on borrower's file <b>New!</b>		X	X
Balance Changes	Any increase in or zero out of balance on borrower's tradeline balance on borrower's file <b>New!</b>		X	X
Scheduled Payment	Any increase in scheduled payment amount on borrower's tradeline <b>New!</b>		X	X
New Late Payments	Any new late payments 30, 60, 90, 120 days past due <b>New!</b>		X	X
Tradeline Bankruptcy	Any new accounts found to be in bankruptcy on borrower's tradeline <b>New!</b>		X	X
Tradeline Collections	Any new accounts found to be in collections on borrower's tradeline <b>New!</b>		X	X
TransUnion® Inquiries	Inquiries sourced from TransUnion Pre/Post Closing Mortgage Notifications (PCMN) <b>New!</b>			X
Experian <sup>SM</sup> Inquiries	Inquiries sourced from Experian's Retention Triggers <sup>SM</sup> for Mortgage (MURS) <b>New!</b>			X

# Optimal Menu—Data Only

*The Design and Price of Information*  
(Bergemann, Bonatti and Smolin, 2017)

Seller offers a menu of information products

Damaged products = partially informative

Systematic distortions:

- Buyer incurs one type of statistical error only
- Lending example: no good borrowers are turned down

# Optimal Menu—Ads & Data

*The Price of Advice* (Eso and Szentes, 2007)

Suppose the buyer's advertising decision is contractible:

- Provide fully informative data
- Charge for the data only upon investment (CPM model)
- Price of data distorts advertising decisions downward
- As if data seller controls price of ads too (e.g., Google)

The nine data brokers the FTC studied collect information from many sources.\*



data brokers  
in the study



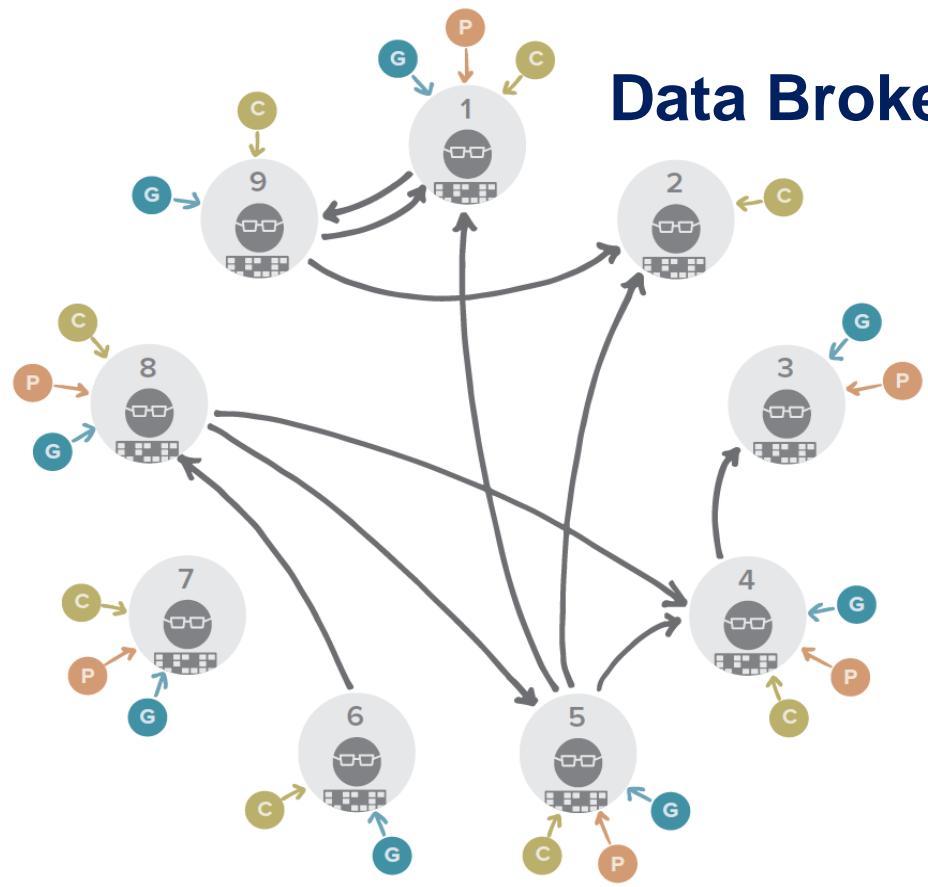
government  
sources<sup>†</sup>



publicly available  
sources<sup>‡</sup>



commercial  
sources



## Data Brokers' Sources (FTC 2014 report)

Krux Link Index

# Buying and Selling Data

Data acquired **repeatedly** from **multiple** sources

Equilibrium price of information:

- Incremental **continuation value** in a dynamic setting
- Depends on existing information, **correlation** structure

Requires:

- Flexible model for combining data sources
- Predictions for equilibrium market structure

# Welfare Implications

(In)efficient equilibrium allocation of information:

- market power (upstream or downstream)
- externalities (multiple uses of information)

Which uses of personal data harm consumers?

Which ones occur in equilibrium?

Privacy constrains equilibrium information revelation  
(Lots of recent work!)

# Takeaways

Information as a complex intermediate good

Demand heterogeneity → profitable versioning

No traditional supply chain

Dynamic multilateral trading of incremental information

No homogeneous asset market

Regulatory challenges with complex externalities

- e.g., parts suppliers' access to car data