Data Sharing and Trading

Paul Heidhues

April 17, 2023

Düsseldorf Institute for Competition Economics, Heinrich-Heine Universität

Based largely on Seim, Bergemann, Crawford, Crémer, Dinielli, Fletcher, Groh, Heidhues, Schaefer, Schnitzer, Scott Morton, and Sullivan (Forthcoming, Yale Journal of Regulation):

Market Design for Personal Data.

Data is crucial for understanding the modern economy

"The world's most valuable resource is no longer oil, but data." (The Economist, 2017)

Data is crucial for understanding the modern economy

"The world's most valuable resource is no longer oil, but data." (The Economist, 2017)

- Big data influence the economy throughout (e.g., Farboodi and Veldkamp, 2022).
  - Market value of largest US firms mainly based on their data.
    - Huge and growing market for data.

Data is crucial for understanding the modern economy

"The world's most valuable resource is no longer oil, but data." (The Economist, 2017)

- Big data influence the economy throughout (e.g., Farboodi and Veldkamp, 2022).
  - Market value of largest US firms mainly based on their data.
    - Huge and growing market for data.
  - Can facilitate search and thereby improve matching.
    - Increases efficiency.

Data is crucial for understanding the modern economy

"The world's most valuable resource is no longer oil, but data." (The Economist, 2017)

- Big data influence the economy throughout (e.g., Farboodi and Veldkamp, 2022).
  - Market value of largest US firms mainly based on their data.
    - Huge and growing market for data.
  - Can facilitate search and thereby improve matching.
    - Increases efficiency.
  - Allows to customize or identify the need for new products.
    - Fosters innovation.
  - Allows to target users with (relevant) adds.

Data is crucial for understanding the modern economy

"The world's most valuable resource is no longer oil, but data." (The Economist, 2017)

- Big data influence the economy throughout (e.g., Farboodi and Veldkamp, 2022).
  - Market value of largest US firms mainly based on their data.
    - Huge and growing market for data.
  - Can facilitate search and thereby improve matching.
    - Increases efficiency.
  - Allows to customize or identify the need for new products.
    - Fosters innovation.
  - Allows to target users with (relevant) adds.
  - Can replace or add to decision making (as in medicine).
    - Improves decision making and reduces costs.

Data is crucial for understanding the modern economy

"The world's most valuable resource is no longer oil, but data." (The Economist, 2017)

- Big data influence the economy throughout (e.g., Farboodi and Veldkamp, 2022).
  - Market value of largest US firms mainly based on their data.
    - Huge and growing market for data.
  - Can facilitate search and thereby improve matching.
    - Increases efficiency.
  - Allows to customize or identify the need for new products.
    - Fosters innovation.
  - Allows to target users with (relevant) adds.
  - Can replace or add to decision making (as in medicine).
    - Improves decision making and reduces costs.
  - Tool for prediction.
    - Help to manage inventory, reduce investors' risk exposure.

Data is crucial for understanding the modern economy

"The world's most valuable resource is no longer oil, but data." (The Economist, 2017)

- Big data influence the economy throughout (e.g., Farboodi and Veldkamp, 2022).
  - Market value of largest US firms mainly based on their data.
    - Huge and growing market for data.
  - Can facilitate search and thereby improve matching.
    - Increases efficiency.
  - Allows to customize or identify the need for new products.
    - Fosters innovation.
  - Allows to target users with (relevant) adds.
  - Can replace or add to decision making (as in medicine).
    - Improves decision making and reduces costs.
  - Tool for prediction.
    - Help to manage inventory, reduce investors' risk exposure.

  Huge realized and potential benefits!

  - Disadvantage in popular discussion: privacy costs.

Data Requires New Economic Thinking: Competition Issues

Competition issues:

Data Requires New Economic Thinking: Competition Issues

Competition issues:

- Data is often generated in markets with strong network effects.
  - Weak competition in the market and wrong innovation incentives.

Data Requires New Economic Thinking: Competition Issues

Competition issues:

- Data is often generated in markets with strong network effects.
  - Weak competition in the market and wrong innovation incentives.
- Often increasing return to data use (feedback loop).
  - More data $\rightarrow$ higher productivity $\rightarrow$ higher market share $\rightarrow$ more data.
  - Risk of market tipping. Other markets?

Data Requires New Economic Thinking: Competition Issues

Competition issues:

- Data is often generated in markets with strong network effects.
  - Weak competition in the market and wrong innovation incentives.
- Often increasing return to data use (feedback loop).
  - More data $\rightarrow$ higher productivity $\rightarrow$ higher market share $\rightarrow$ more data.
  - Risk of market tipping. Other markets?

Keep in mind:

- Basic theory of value (general-equilibrium model) has no role for data.

Data Requires New Economic Thinking: Competition Issues

Competition issues:

- Data is often generated in markets with strong network effects.
  - Weak competition in the market and wrong innovation incentives.
- Often increasing return to data use (feedback loop).
  - More data $\rightarrow$ higher productivity $\rightarrow$ higher market share $\rightarrow$ more data.
  - Risk of market tipping. Other markets?

Keep in mind:

- Basic theory of value (general-equilibrium model) has no role for data.
  - No presumption that we should interfere only if "market failure".

Data Requires New Economic Thinking: Fundamental Issues

- Data can be reused. (No rivalry in usage.)
  - Firms do not want to share productivity-increasing data with rivals.

Data Requires New Economic Thinking: Fundamental Issues

- Data can be reused. (No rivalry in usage.)
  - Firms do not want to share productivity-increasing data with rivals.
- Data of individuals often used to predict behavior of others.

Data Requires New Economic Thinking: Fundamental Issues

- Data can be reused. (No rivalry in usage.)
  - Firms do not want to share productivity-increasing data with rivals.
- Data of individuals often used to predict behavior of others.
  - Data externality is important.

Data Requires New Economic Thinking: Fundamental Issues

- Data can be reused. (No rivalry in usage.)
  - Firms do not want to share productivity-increasing data with rivals.
- Data of individuals often used to predict behavior of others.
  - Data externality is important.
- People "pay" for service often with (personal) data. Barter exchange.

Data Requires New Economic Thinking: Fundamental Issues

- Data can be reused. (No rivalry in usage.)
  - Firms do not want to share productivity-increasing data with rivals.
- Data of individuals often used to predict behavior of others.
  - Data externality is important.
- People "pay" for service often with (personal) data. Barter exchange.
  - Almost all economic transactions create data as by-product. (Real price?)

Data Requires New Economic Thinking: Fundamental Issues

- Data can be reused. (No rivalry in usage.)
  - Firms do not want to share productivity-increasing data with rivals.
- Data of individuals often used to predict behavior of others.
  - Data externality is important.
- People "pay" for service often with (personal) data. Barter exchange.
  - Almost all economic transactions create data as by-product. (Real price?)
- Often zero-price lower bound for monetary price component.
  - Cannot enter with lower monetary price (in exchange for lower quality).

Data Requires New Economic Thinking: Fundamental Issues

- Data can be reused. (No rivalry in usage.)
  - Firms do not want to share productivity-increasing data with rivals.
- Data of individuals often used to predict behavior of others.
  - Data externality is important.
- People "pay" for service often with (personal) data. Barter exchange.
  - Almost all economic transactions create data as by-product. (Real price?)
- Often zero-price lower bound for monetary price component.
  - Cannot enter with lower monetary price (in exchange for lower quality).
- "Behavioral / default effects" crucial regarding consumer choices.
  - Google pays a large fraction of its revenue on Apple devices to be default. Why?

What is Personal Data?

GDPR: personal data is information "related to an identified or identifiable natural person".

Our definition: all data describing an individual's characteristics, transaction histories, browsing histories, etc generated by an individual (even if the actual person cannot be identified from the data generated).

- Advertiser may value information that I looked for blue shoes in Berlin even if they cannot identify me.

Why Refuse Sharing Personal Data?

Consumer may oppose the collection of personal data because

1. Some things are inherently "private" and do not want to share.
   - Spy camera in the shower.
   - Press reports on Tesla this week.
2. Knowledge can be used to my disadvantage.
   - Discrimination (e.g. on sexual orientation), blackmail (stolen data).
   - Price discrimination based on willingness to pay (IO literature).
   - Inability to insure (health data).
   - May enable firms to target me with bad products (scams / bad steering).
3. Fear negative consequence on societal level.
   - Firms treat non-sharing customers differently.
     - Insurance markets break down (Google / Fitbit).
     - Price discrimination.
   - Firms may exploit vulnerable consumers.
     - Gambling adds targeting gambling addicts, etc....
4. Feel they should be reimbursed for sharing their data.

Effective Privacy Control?

Currently, consumers can click on (confusing) pop-ups to try and make privacy choices.

- Real cost of learning about terms of service high - just fiction to think consumers know it.
- With switching costs, learning only so useful as these terms can be changed at any time.
- Impossible to predict what can be inferred from what data and how secure the data is.
- Often all-or-nothing nature.

Effective Privacy Control?

Currently, consumers can click on (confusing) pop-ups to try and make privacy choices.

- Real cost of learning about terms of service high - just fiction to think consumers know it.
- With switching costs, learning only so useful as these terms can be changed at any time.
- Impossible to predict what can be inferred from what data and how secure the data is.
- Often all-or-nothing nature.
- Decision strongly influenced by the frame in which consumers make decision (and default settings) — click fatigue.
- Profit-max. firms will find that frame that induced the from their perspective desirable decision, i.e. data sharing.

Effective Privacy Control?

Currently, consumers can click on (confusing) pop-ups to try and make privacy choices.

- Real cost of learning about terms of service high - just fiction to think consumers know it.
- With switching costs, learning only so useful as these terms can be changed at any time.
- Impossible to predict what can be inferred from what data and how secure the data is.
- Often all-or-nothing nature.
- Decision strongly influenced by the frame in which consumers make decision (and default settings) — click fatigue.
- Profit-max. firms will find that frame that induced the from their perspective desirable decision, i.e. data sharing.

Arguably, little—if any—effective privacy control.

Problems in Current Data Markets?

Status quo:

- A handful of firms control vast amount of personal data, yielding significant market power.
- Users' valuable resource — private data — is extracted without monetary payment.
  - And the barter exchange does not seem "fair" with largest firms persistently earning billions of dollars in profits.
- Data is shared at best selectively and non-optimally.
  - Not used for public benefits (as preventing digital addiction or transportation planing) and shielded from rivals or new entrants.

A well functioning data market design should address the

1. failure to provide users effective control over their data;
2. failure to provide users a way to benefit financially;
3. failure to ensure that data is put to highest use (non-profit, government, or other firms).

Towards a Solution: Data Intermediaries

Users need to "negotiate collectively" to receive fair share:

- Transaction costs.
- Typically: value of a marginal consumer's data is less than the average value.
  - Often use data of a group of (similar) users to predict an individual user's behavior (incremental prediction accuracy decreases).
- Large online firms have market power and can block attempts of users not to share their data.

Towards a Solution: Data Intermediaries

Users need to "negotiate collectively" to receive fair share:

- Transaction costs.
- Typically: value of a marginal consumer's data is less than the average value.
  - Often use data of a group of (similar) users to predict an individual user's behavior (incremental prediction accuracy decreases).
- Large online firms have market power and can block attempts of users not to share their data.

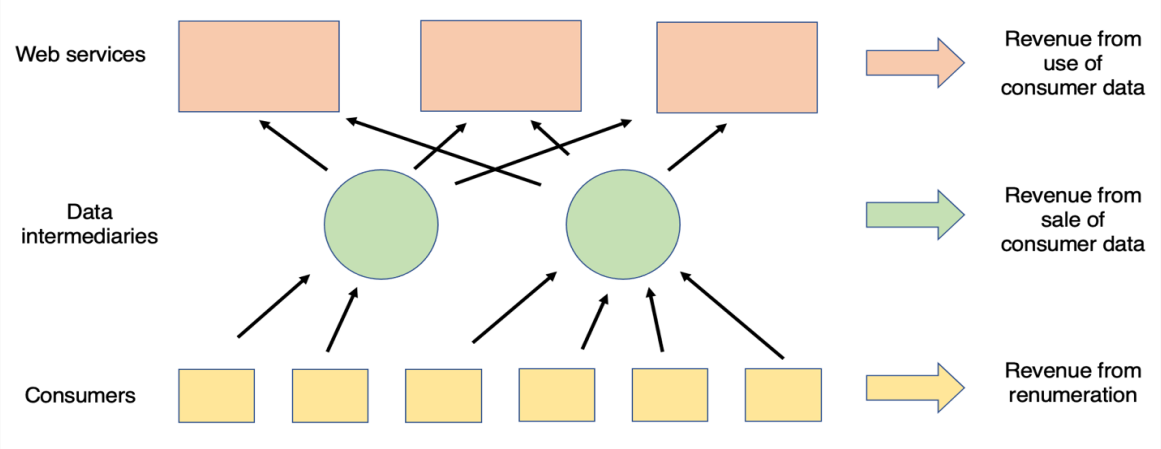
Propose governments intervention to create licensed intermediaries to collect and monetize personal data on behalf of user.

- Key: prohibit any firm other than licensed intermediary from monetizing personal data or control the use of personal data (exclusive control).
- Note: web service would only be able to use an individual's data to provide the web service in question, not for targeted advertising or other commercial purposes.

Proposal for Data Markets: Basic Structure

- Competing intermediaries offer renumeration contracts to consumers for selling their data.
  - Reimbursement amount depends on how much data consumer is willing to share.
- Web services who want to use personal data for ads need to purchase the data from the consumer's intermediary.

Schematic identifying principal categories of participants in data intermediary markets



Web Service: Data for Servicing Users

Web services can use data they collect to service users, such as use data to

- deliver ordered products (shipping and billing address);
- receive directions from a map service (real-time location data);
- send messages (directories of social connections).

Web Service: Data for Servicing Users

Web services can use data they collect to service users, such as use data to

- deliver ordered products (shipping and billing address);
- receive directions from a map service (real-time location data);
- send messages (directories of social connections).

Classification depends on purpose of use, not characteristic of information

Web Service: Data for Servicing Users

Web services can use data they collect to service users, such as use data to

- deliver ordered products (shipping and billing address);
- receive directions from a map service (real-time location data);
- send messages (directories of social connections).

Classification depends on purpose of use, not characteristic of information

- and privacy levels do not apply to above data use.

Web Service: Data for Servicing Users

Web services can use data they collect to service users, such as use data to

- deliver ordered products (shipping and billing address);
- receive directions from a map service (real-time location data);
- send messages (directories of social connections).

Classification depends on purpose of use, not characteristic of information

- and privacy levels do not apply to above data use.

Importantly, web services cannot simply use the (same) data to target users.

- Defn: targeting leverages private signals about intent to promote service that cost the user additional resources (money or data).

Web Service: Data for Servicing Users

Web services can use data they collect to service users, such as use data to

- deliver ordered products (shipping and billing address);
- receive directions from a map service (real-time location data);
- send messages (directories of social connections).

Classification depends on purpose of use, not characteristic of information

- and privacy levels do not apply to above data use.

Importantly, web services cannot simply use the (same) data to target users.

- Defn: targeting leverages private signals about intent to promote service that cost the user additional resources (money or data).
  - Eg. advertising based on a query (“best running shoes”) is targeting.

Web Service: Data for Servicing Users

Web services can use data they collect to service users, such as use data to

- deliver ordered products (shipping and billing address);
- receive directions from a map service (real-time location data);
- send messages (directories of social connections).

Classification depends on purpose of use, not characteristic of information

- and privacy levels do not apply to above data use.

Importantly, web services cannot simply use the (same) data to target users.

- Defn: targeting leverages private signals about intent to promote service that cost the user additional resources (money or data).
  - Eg. advertising based on a query ("best running shoes") is targeting.
  - Ads on a website about running don't use data generated by web user (contextual).

Web Service: Data for Servicing Users

Web services can use data they collect to service users, such as use data to

- deliver ordered products (shipping and billing address);
- receive directions from a map service (real-time location data);
- send messages (directories of social connections).

Classification depends on purpose of use, not characteristic of information

- and privacy levels do not apply to above data use.

Importantly, web services cannot simply use the (same) data to target users.

- Defn: targeting leverages private signals about intent to promote service that cost the user additional resources (money or data).
  - Eg. advertising based on a query ("best running shoes") is targeting.
  - Ads on a website about running don't use data generated by web user (contextual).
  - Recommendations on an e-commerce site is targeting, while recom. when users have fixed subscription (Spotify or Netflix) is not.

Web Service: Data for Analytics

Data generated by web users in the past is used for analytics (develop and improve an algorithm).

Web Service: Data for Analytics

Data generated by web users in the past is used for analytics (develop and improve an algorithm).

- Doesn't rely on identity of user who generated the data.

Web Service: Data for Analytics

Data generated by web users in the past is used for analytics (develop and improve an algorithm).

- Doesn't rely on identity of user who generated the data.
- Doesn't cost user additional resources.

Allowing usage of such data will improve web services products. But

Web Service: Data for Analytics

Data generated by web users in the past is used for analytics (develop and improve an algorithm).

- Doesn't rely on identity of user who generated the data.
- Doesn't cost user additional resources.

Allowing usage of such data will improve web services products. But

- reduces revenue that can be passed on to users; and

Web Service: Data for Analytics

Data generated by web users in the past is used for analytics (develop and improve an algorithm).

- Doesn't rely on identity of user who generated the data.
- Doesn't cost user additional resources.

Allowing usage of such data will improve web services products. But

- reduces revenue that can be passed on to users; and
- intermediaries could make the data available to entrants or rivals allowing them to overcome scale disadvantages.

Rules on Intermediaries

1. Intermediaries must be licensed.
   - Strict data protection, cybersecurity and resilence standards.
   - Data minimization.
2. Fiduciary duties and minimum payout rules (e.g. 70% of revenue).
3. Compete to be the exclusive agent of the individual once a year.
   - For personal data of individual's web devices and IoT.
   - Compete via renumeration and customer service, etc.
   - Must compete using a standardized interface (more later).
4. Web services and intermediaries cannot be vertically integrated.
   - Prohibitions of exclusive contracting.
   - Make sure the market for intermediaries does not monopolize.

Interface Design

To induce meaningfull choice, need to simplify and reduce the required choices (beyond just clear and plain language as in GDPR).

Interface Design

To induce meaningfull choice, need to simplify and reduce the required choices (beyond just clear and plain language as in GDPR). We propose standardized privacy levels such as:

1. Tier 1:
   - basic demographics (age, gender, location at postal code level);
   - apps installed in user device.
2. Tier 2 = Tier 1 + ...
   - browsing and app-usage data.
3. Tier 3 = Tier 2 + ...
   - approximate real-time location.
4. Tier 4 = Tier 3 + ... (things allowing precise attribution)
   - financial records;
   - online transactions;
   - email receipts.

Plus a compulsory Tier 0 for users that want to remain anonymous.

- Zero data monetization.

Desirable Data Usage Restrictions

The above design can be made consistent with data minimization principles.

Further data usage restrictions likely desirable:

Desirable Data Usage Restrictions

The above design can be made consistent with data minimization principles.

Further data usage restrictions likely desirable:

- The ability to target substance abuse users or gambling addicts with addiction-exploiting ads is likely undesirable.

Desirable Data Usage Restrictions

The above design can be made consistent with data minimization principles.

Further data usage restrictions likely desirable:

- The ability to target substance abuse users or gambling addicts with addiction-exploiting ads is likely undesirable.
- Google steered UK users searching for"high-return investment" to fraudulent ads.

Desirable Data Usage Restrictions

The above design can be made consistent with data minimization principles.

Further data usage restrictions likely desirable:

- The ability to target substance abuse users or gambling addicts with addiction-exploiting ads is likely undesirable.
- Google steered UK users searching for"high-return investment" to fraudulent ads.
- Uber develops an algorithm that predicts likelihood of a user being drunk.

Desirable Data Usage Restrictions

The above design can be made consistent with data minimization principles.

Further data usage restrictions likely desirable:

- The ability to target substance abuse users or gambling addicts with addiction-exploiting ads is likely undesirable.
- Google steered UK users searching for"high-return investment" to fraudulent ads.
- Uber develops an algorithm that predicts likelihood of a user being drunk.
- Companies like WeatherAds help firms to condition their ads on weather; problematic for durable good adds?

Steering in above examples based on consumer mistakes or perceived rather than "true" preferences.

Desirable Data Usage Restrictions

The above design can be made consistent with data minimization principles.

Further data usage restrictions likely desirable:

- The ability to target substance abuse users or gambling addicts with addiction-exploiting ads is likely undesirable.
- Google steered UK users searching for"high-return investment" to fraudulent ads.
- Uber develops an algorithm that predicts likelihood of a user being drunk.
- Companies like WeatherAds help firms to condition their ads on weather; problematic for durable good adds?

Steering in above examples based on consumer mistakes or perceived rather than "true" preferences.

- Far less likely to create value.

Desirable Data Usage Restrictions

The above design can be made consistent with data minimization principles.

Further data usage restrictions likely desirable:

- The ability to target substance abuse users or gambling addicts with addiction-exploiting ads is likely undesirable.
- Google steered UK users searching for"high-return investment" to fraudulent ads.
- Uber develops an algorithm that predicts likelihood of a user being drunk.
- Companies like WeatherAds help firms to condition their ads on weather; problematic for durable good adds?

Steering in above examples based on consumer mistakes or perceived rather than "true" preferences.

- Far less likely to create value.
- Perhaps allowing consumers to state interests helpful.

Conclusion

The market for personal data is huge, and extremely profitable...

Conclusion

The market for personal data is huge, and extremely profitable...

- but consumers do not seem to get a fair share or are able to make meaningful privacy choices.

Conclusion

The market for personal data is huge, and extremely profitable...

- but consumers do not seem to get a fair share or are able to make meaningful privacy choices.
- We used economic principles to suggest a possible market design in which consumers get a fair share of the surplus.

The DMA does not address this aspect of fairness but "just" fairness to business users.

Conclusion

The market for personal data is huge, and extremely profitable...

- but consumers do not seem to get a fair share or are able to make meaningful privacy choices.
- We used economic principles to suggest a possible market design in which consumers get a fair share of the surplus.

The DMA does not address this aspect of fairness but "just" fairness to business users.

- Likely helpful but we think more is desirable.