Thoughts on Adding "Privacy Considerations" to Internet Drafts

Alissa Cooper
John Morris
Center for Democracy & Technology

November 3, 2010

The Internet Privacy Workshop announcement [IAB] notes that while [RFC3552] provides guidance to authors of RFCs about what they should address in the "Security Considerations" sections of their documents, no similar guidance exists for privacy. The idea that RFC authors need guidance about documenting the privacy issues associated with the standardization of Internet protocols has been raised a number of times, including as part of a broader effort to provide guidance about policy issues of all kinds [I-D.morris-privacy-considerations][Morris] and at plenaries at IETF 56 [Blaze] and IETF 77 [Krishnamurthy]. Several recent drafts suggest approaches to providing such guidance and highlight some of the privacy threats to be addressed [I-D.brim-mobility-and-privacy][I-D.morris-policy-cons].

We believe that RFC authors do need guidance about documenting the privacy issues associated with the standards they create and that the idea of including a "Privacy Considerations" section in RFC text (and perhaps in W3C standards as well) holds promise. Just as every RFC author is not a security expert, not every author is a privacy expert. Thus privacy guidance for authors needs to be simple enough for those with little privacy expertise to use, but comprehensive enough to ensure that the less obvious privacy implications of the protocols being standardized are surfaced.

To help workshop participants conceptualize one way to provide privacy guidance, we have attached to this position paper the *Threshold Analysis for Online Advertising*, a document that CDT developed together with a range of Internet content and applications providers, ISPs, advertisers, and public interest groups. The term "threshold analysis" derives from the Privacy Threshold Analysis (PTA) program that US government agencies use to assess the privacy implications of agency systems [DHS].

The *Threshold Analysis for Online Advertising* provides a framework for describing the technical and business arrangements behind online advertising and for performing a privacy assessment of those arrangements. The framework is structured using a list of questions that advertising systems designers can ask about their designs (for example, "what data about individuals is used to target ads?") in order to evaluate the potential impact of those designs on privacy. The document also includes a list of definitions of relevant terms in Appendix A.

The *Threshold Analysis* is aimed at a different set of problems and a higher layer of the stack than most IETF protocols and many W3C standards. However, we believe it can serve as a helpful model for how to provide privacy guidance to RFC authors, both structurally and, to a lesser extent, substantively. The structure of providing questions for authors to answer has been used before in other contexts in the IETF [RFC3426].

The *Threshold Analysis* augments the question structure by suggesting ranges of answer values that map to levels of privacy risk; question 7, for example, asks about the length of data retention and provides a range of answer values, from shorter to longer, with examples of values within the range. The particular questions that are relevant to IETF protocols will be different from the questions in the *Threshold Analysis*, but the format of using questions with answer ranges may be just as valuable.

The other structural advantage of the *Threshold Analysis* is that it does not require an extensive discussion of the definition of privacy or the Fair Information Practices (FIPs) up front. While the "Evaluating Practices" section does make use of the FIPs, it is entirely separable from the "Describing Practices – Questions" section – the questions can be easily understood without an understanding of the FIPs. Given that RFC authors may have limited privacy experience, getting into the details of the different FIPs versions or different definitions of privacy (as [I-D.morris-privacy-considerations] does) is likely to be less effective than the simpler questions approach.

Substantively, the *Threshold Analysis* provides some content that could be usefully reused in RFC privacy guidance. The questions about which entities collect data (question 1), the relationship between those entities and individual users (2), what data is collected (3), the identifiability of the data (4), and data retention (7) could all be transposed to the IETF context with some adjustments (the notion of data "transmission" might replace data "collection," for example). More questions would certainly be needed, but these could provide useful starting points.

Some of the definitions in Appendix A -- the identifiability definitions in particular -- may also be reusable. While it is possible to describe the identifiability level of data with great nuance (as demonstrated in [I-D.hansen-privacy-terminology]), simpler definitions like the ones provided in the *Threshold Analysis* may be more easily understood by RFC authors and likely provide enough detail to describe most data implicated by IETF protocols.

The *Threshold Analysis* belongs to a different and less technical context than the standardization work that goes on in the IETF and the W3C. Nonetheless we hope that some of its concepts can catalyze a discussion about providing guidance to help authors of IETF and W3C specifications be more systematic in documenting the privacy implications of their work.

References

[Blaze] Blaze, M. and Morris, J., "Privacy Considerations for Internet Protocols," IETF 56 IESG plenary, March 2003, http://www.crypto.com/talks/ietf56-privacy.pdf.

[DHS] U.S. Department of Homeland Security, "Privacy Compliance," July 2010, http://www.dhs.gov/files/publications/gc_1209396374339.shtm.

[I-D.hansen-privacy-terminology]
Pfitzmann, A., Hansen, M, and Tschofenig, H., "Terminology for Talking

about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management,” draft-hansen-privacy-terminology-01 (work in progress), August 2010.

[I-D.brim-mobility-and-privacy]
Brim, S., Linsner, M., McLaughlin, B., and Wierenga, K., “Mobility and Privacy,” draft-brim-mobility-and-privacy-00 (work in progress), October 2010.

[I-D.morris-policy-cons]
Morris, J. and Davidson, A., “Public Policy Considerations for Internet Design Decisions,” draft-morris-policy-cons-00 (expired), June 2003.

[I-D.morris-privacy-considerations]
Morris, J., Tschofenig, H., Aboba, B., and Peterson, J., “Privacy Considerations for Internet Protocols,” draft-morris-privacy-considerations-00 (work in progress), October 2010.

[Krishnamurthy]
Krishnamurthy, B., “Privacy leakage on the Internet,” IETF 77 plenary, March 2010, http://www.ietf.org/proceedings/77/slides/plenaryt-5.pdf.

[Morris] Morris, J., and Davidson, A., “Policy Impact Assessments: Considering the Public Interest in Internet Standards Development,” *Proceedings of the 31st Research Conference on Communication, Information and Internet Policy (TPRC)*, August 2003, http://old.cdt.org/publications/pia.pdf.

[RFC3426] Floyd, S., ed., ”General Architectural and Policy Considerations,” RFC 3426, November 2002.

[RFC3552] Rescorla, E. and Korver, B., “Guidelines for Writing RFC Test on Security Considerations,” RFC 3552, July 2003.

CENTER FOR DEMOCRACY & TECHNOLOGY

Threshold Analysis for Online Advertising Practices

January 2009 – Version 1.0

The Threshold Analysis for Online Advertising Practices outlines a framework for describing and analyzing the landscape of practices involved in online advertising. Online advertising companies can use the framework as the first step in analyzing their own practices, as a precursor to a full privacy impact assessment or fair information practices analysis.

In December 2007, the Center for Democracy & Technology (CDT) hosted a meeting of its Internet Privacy Working Group (IPWG) about online advertising privacy issues. The group decided that it would be useful to create a way to describe and analyze online advertising practices. This document contains the analysis framework.

Participants at the original meeting agreed about the important role of online advertising in supporting a rich diversity of content, services, and applications provided without charge to Internet users. This framework aims to help IPWG members and others better understand and evaluate the landscape of online advertising practices. Online advertising companies can use the framework as the first step in analyzing their own practices, as a precursor to a full privacy impact assessment or fair information practices analysis. An explanation of how to apply the entire threshold analysis framework to a particular advertising practice appears in Section III.

DESCRIBING PRACTICES

Section II provides a framework for describing online advertising practices. It is based around the situation where an individual interacting with a first-party Web site or application receives a targeted ad. To fully describe the practice of targeting an ad within this framework, the following questions must be answered about the practice (detailed descriptions of the questions are in Section II):

1. What entities collect and use data to target the ad?
   - a. How many entities collect and use data to target the ad?
   - b. For each entity, what type of entity is it?

2. For each entity, describe the nexus between the entity and the individual.
   - a. What is the user's familiarity or degree of relationship with the entity in other contexts?
   - b. What is the user's reasonable expectation of the entity's involvement in ad targeting?

3. What data about the individual is collected and used to target the ad?
   - a. How many distinct data streams are collected and used to target the ad?
   - b. For each data stream, what types of data are collected and used to target the ad?

4. For each data stream, how identifiable is the data in the stream?

5. For each data stream, how specific is the data in the stream?

6. For each data stream, how long is data in the stream held before it is used for ad targeting?

7. For each data stream, how long is data in the stream retained?

EVALUATING PRACTICES

The analysis framework will also serve as a tool to assess practices from a privacy perspective. Such an assessment may be based on:

1. the impact of the practice on the individual viewing the ads; and
2. how well the entities engaged in the practice adhere to fair information practices (FIPs), taking the responses to the questions in Section II into account.

Section III explains these criteria in depth and provides a categorization scheme that can be applied to advertising practices based on these criteria.

The discussion of fair information practices in this document is consistent with the principles outlined in the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data: collection limitation, data quality, purpose specification, use limitation, security, openness, individual participation and accountability.[1] Other versions of FIPs exist, and this document takes no position on which one is the authoritative version.

EXAMPLES

Examples of how to use this framework appear in Appendix B.

[1] http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html

◪ Describing Practices – Questions

All of the questions focus on the situation where an individual interacting with a first-party site or application receives a targeted ad.

1. WHAT ENTITIES COLLECT AND USE DATA TO TARGET THE AD?

When an individual interacting with a first-party site or application receives a targeted ad, one or more entities may be involved in collecting and using data to target the ad. There are three types of entities:

- The first-party site or application
- Other sites or applications whose data collection and use for ad targeting is in some way controlled by the first party (for example, other Web properties owned by the first party)
- Third parties (ad networks, ISPs, or data aggregators, for example) that may use the data they collect or use for ad targeting on the first-party site or application for other purposes

These are general types that describe different kinds of entities involved in targeted advertising, but they do not fully describe the complexities of all potential online advertising business arrangements. Certain entities may not fall squarely into a single category.

These two questions should be answered to describe the entities that collect and use data to target an ad:

1. How many entities collect and use data to target the ad?
2. For each entity, what type of entity is it? Choose one:

- First party
- Other site/application whose data collection and use for ad targeting is in some way controlled by the first party
- Third party

2. FOR EACH ENTITY, DESCRIBE THE NEXUS BETWEEN THE ENTITY AND THE INDIVIDUAL.

The user's relationship with and awareness of an entity involved in ad targeting can impact how robust the entity's FIPs compliance must be in order for its advertising practice to be evaluated favorably from a privacy perspective. For example, if a user does not expect an entity to be involved in ad targeting or is not familiar with the entity, the principles of openness, individual participation, and accountability may be especially important.

For each entity listed in response to Question 1b, these two questions should be answered:

a. What is the user's familiarity or degree of relationship with the entity in other contexts?

The user's familiarity or degree of relationship with the entity in other contexts should be plotted along the following spectrum:

More familiarity or higher degree of relationship

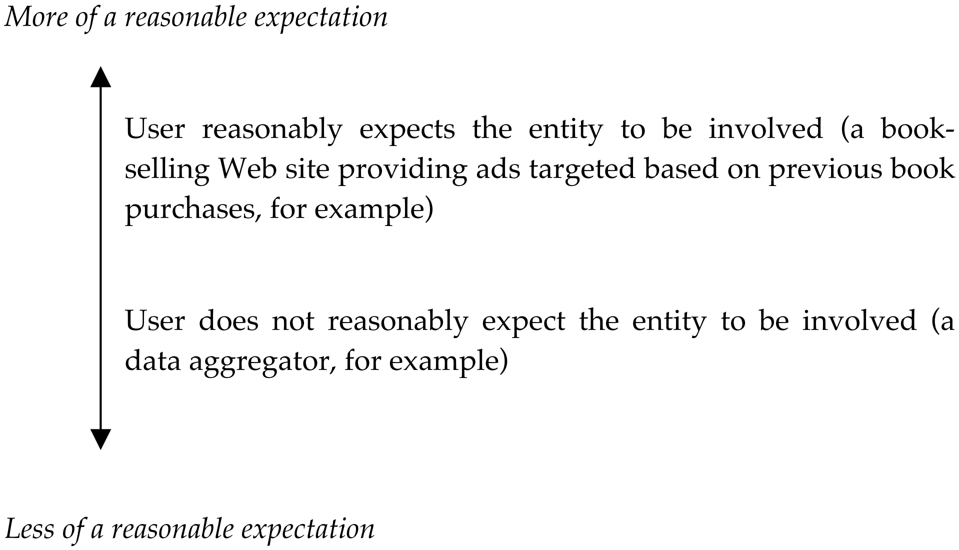
User has purchased goods or services from the entity

User has knowingly interacted with the entity (by visiting the entity's Web site, for example)

User has no familiarity or relationship with the entity

Less familiarity or lower degree of relationship *b. What is the user's reasonable expectation of the entity's involvement in ad targeting?*

The user's reasonable expectation of the entity's involvement in the ad targeting should be plotted along the following spectrum:



More of a reasonable expectation

User reasonably expects the entity to be involved (a book-selling Web site providing ads targeted based on previous book purchases, for example)

User does not reasonably expect the entity to be involved (a data aggregator, for example)

Less of a reasonable expectation

For the purposes of this question, a reasonable user does not need to be well educated about how online advertising works.

3. WHAT DATA ABOUT THE INDIVIDUAL IS COLLECTED AND USED TO TARGET THE AD?

When an individual interacting with a first-party site or application receives a targeted ad, data about the individual may be collected and used to target the ad. The definitions in Appendix A describe data of different types and qualities.

These questions should be answered to describe the data collected and used to target the ad:

a. *How many distinct data streams are collected and used to target the ad?* Two data streams are considered distinct from each other if their data identifiability levels, data specificity levels, data usage time spans, or data retention time spans are different (see Questions 4, 5, 6 and 7).

b. *For each data stream, what types of data are collected and used to target the ad?* Choose all that are applicable (the choices are not mutually exclusive):

- Clickstream data
- Communication content
- Derived data
- Non-resident data
- Offline data
- Public data
- Purchase data
- Resident data
- Sensitive data
- User-generated data
- Mobile location data
- Fixed location data
- Nomadic location data

The types of data collected and used for ad targeting can impact how robust the entity's FIPs compliance must be in order for its ad targeting practice to be evaluated favorably from a privacy perspective. For example, practices that involve the collection of more sensitive kinds of data may require stricter adherence to the principles of collection limitation and use limitation in order to be considered favorably than practices that involve less sensitive kinds of data.

4. FOR EACH DATA STREAM, HOW IDENTIFIABLE IS THE DATA IN THE STREAM?

The data identifiability spectrum describes the extent to which data collected and used for ad targeting can be used to identify an individual. For each data stream listed in response to Question 3b, the identifiability level of the data in that stream should be plotted on the spectrum below.

Less identifiable

↑ Aggregate or otherwise anonymous data

Inferably identifiable data

↓ Directly identifiable data

More identifiable

Data identifiability can impact how robust an entity's FIPs compliance must be in order for its ad targeting practice to be evaluated favorably from a privacy perspective. For example, an ad targeting practice that uses directly identifiable data may require stricter use limitations and higher levels of security than practices that use less identifiable data.

5. FOR EACH DATA STREAM, HOW SPECIFIC IS THE DATA IN THE STREAM?

The data specificity dimension describes the level of detail of the data collected and used for ad targeting. When an individual interacting with a site or application receives a targeted ad, that targeting may use data about the individual that is based on:

More generic data

Generic categorizations (“sports fan” or “frequent traveler,” for example)

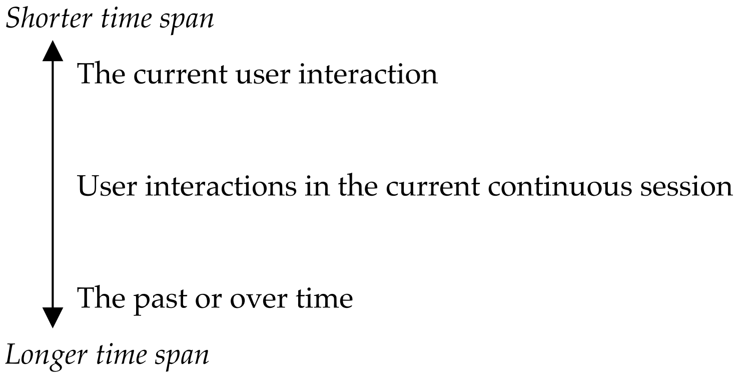
Specific data culled from clickstream data, communications content, or other sources (lists of URLs visited, for example)

More specific data

For each data stream listed in response to Question 3b, the specificity of the data in the stream should be plotted on the spectrum above.

6. FOR EACH DATA STREAM, HOW LONG IS DATA IN THE STREAM HELD BEFORE IT IS USED FOR AD TARGETING?

The data usage time span of a data stream describes the length of time between the collection of data and its use for ad targeting. When an individual interacting with a site or application receives a targeted ad, that targeting may be based on data collected during:


Shorter time span
The current user interaction
User interactions in the current continuous session
The past or over time
Longer time span

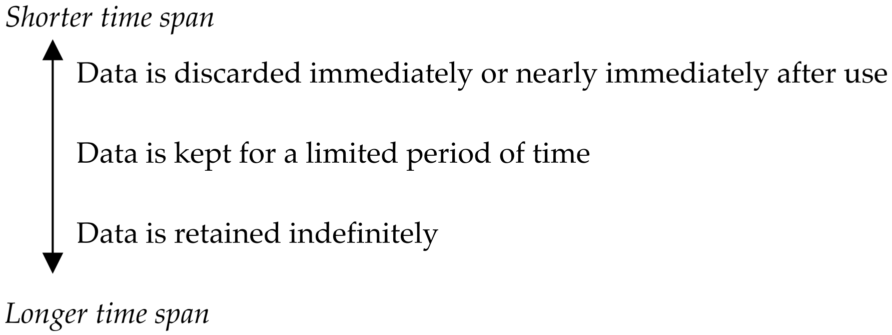
For each data stream listed in response to Question 3b, the data usage time span of the data in that stream should be plotted on the spectrum above.

7. FOR EACH DATA STREAM, HOW LONG IS DATA IN THE STREAM RETAINED?

The data retention time span spectrum describes how long data collected for ad targeting is retained by the entities involved. This is distinct from the data usage time span, which measures the time between collection and use. Data retention measures the time between collection and deletion.

An entity's data retention time can impact how robust the entity's FIPs compliance must be in order for its ad targeting practice to be evaluated favorably from a privacy perspective. For example, an entity retaining data for longer periods of time may need stronger security, use limitations, and individual participation capabilities than entities retaining data for shorter periods.

For each data stream listed in response to Question 3b, the data retention time for the stream should be plotted on the spectrum below.



Shorter time span

Data is discarded immediately or nearly immediately after use

Data is kept for a limited period of time

Data is retained indefinitely

Longer time span

Evaluating Practices

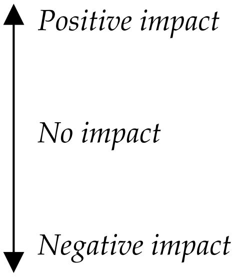
Describing an ad targeting practice using the questions in Section II provides a foundation for evaluating the practice from both a consumer protection perspective and a privacy perspective. This section outlines criteria that can be used to evaluate practices from both perspectives and a categorization scheme that can be applied based on these criteria.

Two criteria are useful in evaluating an advertising practice:

1. Consumer protection: the impact of the practice on the individual viewing the ads; and
2. Privacy: how well the entities engaged in the practice adhere to Fair Information Practices (FIPs), taking the responses to the questions in Section II into account.

CONSUMER PROTECTION: IMPACT

Targeted advertising practices can have a range of effects on the individuals viewing the ads. From a consumer protection standpoint, some of these effects may be positive, others may be negative, and some advertising practices will have little or no impact on individuals. The impact of an advertising practice thus exists along a spectrum:



A single ad may produce a positive impact; for example, when an individual is about to make a purchase online, he or she may appreciate seeing an ad for the same item at a lower price elsewhere. A targeted advertising practice as a whole may also produce a positive impact by providing benefits to individuals that untargeted or less-well-targeted advertising do not provide. For example, if a Web site is able to offer particular content or services for free or at a reduced price because the revenue it earns from targeted advertising is greater than from untargeted advertising, individuals may consider the ad targeting practice to have a positive impact. Whether an advertising practice has a positive impact is determined by the individual who views the ads and perceives the effects of the practice.

Conversely, a targeted advertising practice may have actual or potential negative effects on an individual. Negative impact describes the effects of a targeted advertising practice broadly, as opposed to any effects from a single ad. Some negative impacts may be technical or financial in nature, or they may depend on the invasiveness of the ad targeting practice. Negative impact is one way of characterizing what traditional consumer protection efforts have aimed to minimize.

Delivering targeted ads that purposefully damage an individual's computer is an example of a practice with negative technical impact. Redlining (in an online advertising context, where certain ads are not shown or the goods being advertised are marked up based on an individual's race, ethnicity, or other demographic information) is an example of a practice of potential negative impact. Because negative impact may not always be evident to the individual receiving the ad, it may be determined either by the individual or by the entities doing the ad targeting, or both.

PRIVACY: ADHERENCE TO FIPS

A privacy analysis of an entity's ad targeting practice can be conducted by evaluating how well the entity adheres to Fair Information Practices (FIPs). An entity's adherence to FIPs is a measure of how well the entity's ad targeting practice comports with FIPs given the features of the practice as described in response to the questions in Section II. In evaluating a particular practice, the features of the practice should be assessed against each FIP to answer the following question: given how this practice is described, is the entity's implementation of each FIP principle robust enough? Evaluation against some FIPs will require all the features of the practice to be taken into account, while others may only be based on a subset of features or a single feature.

For example, as noted in Question 4, practices that use directly identifiable data may require stricter use limitations than practices that uses less identifiable data. Practices that use directly identifiable data should be assessed to determine whether their use limitations are adequate considering the identifiability of the data involved. The evaluation of how well an entity adheres to the use limitation principle may also depend on other features of its practice aside from data identifiability.

In evaluating a practice it may be useful to consider which of the following three categories it falls into, based on the criteria above.

- **Unacceptable**: Practices are considered unacceptable only on the basis of consumer protection failings; a privacy analysis is not relevant to the determination of whether a practice is unacceptable. Practices in this category have a strictly negative impact. They may be illegal, fraudulent, unfair, and/or deceptive. While these practices may also be deficient from a privacy perspective (i.e., they may not adhere to some or all FIPs), a practice is determined to be unacceptable only on the basis of its impact. Conversely, even with robust FIPs compliance, these practices would still be considered unacceptable because of their negative impact.

- **High-risk**: These practices may not raise any consumer protection issues: they may have either positive or no impact. However, these practices may raise privacy issues in that their adherence to one or more FIPs is not sufficient given the features of the practice as described by answers to the questions in Section II.

- **Low-risk**: These practices may raise neither consumer protection issues nor privacy issues. They may have positive or no impact and their adherence to FIPs may be sufficiently robust given the features of the practice as described by answers to the questions in Section II.

HOW TO USE THIS FRAMEWORK

To use this framework to evaluate an advertising practice, first describe the practice by answering the questions in Section II. Then conduct an analysis of the practice from a consumer protection perspective to determine the impact of the practice on individuals. If the practice has a negative impact, the analysis is complete and the practice can be considered unacceptable. If the advertising practice has no impact or a positive impact on individuals, then conduct a privacy analysis by determining whether the entities involved adhere sufficiently to FIPs given the way the practice is described. If the FIPs adherence is sufficiently robust, the practice can be considered low-risk. Otherwise it can be considered high-risk.

FOR MORE INFORMATION

Please contact: Brock Meeks, Director of Communications
202-637-9800

Appendix A: Data Definitions

The data types in the list below should be regarded as descriptors or tags, and not as mutually exclusive categories. It may well be possible for a data element or data stream to belong to more than one category. For example, clickstream data can also be non-resident.

GENERAL DEFINITIONS

Clickstream data – Data collected about a individual's Web site visits or other online activities, with or without the individual's awareness. Some current examples are the individual's IP address and cookies, the date and time of the activity, the URL of a requested site, the individual's browser and operating system types, the links the individual clicks on, and the referring URL.

Communication content – The substance of a transmission destined for one or more specified individuals (as opposed to a site, service, or application). Current examples include the subject and body of an email and the content of a voice call or text chat.

Device – A computer, cell phone, PDA, or other machine capable of accessing the Internet.

Derived data – Data relating to an analysis about an individual derived from the individual's clickstream data, purchase data, user-generated data, communications content, or other data.

Non-resident data – Data related to a particular individual or device that is not stored on the device.

Offline data – Data that is not related to an individual's online activities. Current examples include driving records and voting records.

Public data – Data related to an individual that is obtainable by anyone (payment of a fee may be required).

Purchase data – Data relating to an individual's purchase or acquisition of goods or services, such as the items purchased, date and time of purchase, payment information, and shipping information.

Resident data – Data stored on an individual's device.

Sensitive data – Data related to an individual that is granted some measure of special treatment. Examples may include data related to health or medical conditions, finances, sexual behavior or orientation, race or ethnicity, or political opinions.

User-generated data – Data generated knowingly by an individual. Some current examples are search terms, input into online form fields, and posts on public forums.

User interaction – The transfer of data between an individual (or his or her device) and a site, service, or application.

LOCATION DEFINITIONS

The location definitions are based on terminology used by technical standards bodies focused on location information and privacy, most notably the Internet Engineering Task Force (IETF) Geographic Location/Privacy Working Group.

Civic location data – Data that describes the geographic location of an individual in terms of a postal address or civic landmark. Examples of such data are room number, street number, street name, city, ZIP+4, ZIP, county, state, and country. The precision of this data can be reduced by removing elements (for example, the precision of the combination of city, state and ZIP can be reduced by only using state).

Geodetic location data – Data that describes the geographic location of an individual in a particular coordinate system (for example, a latitude-longitude pair). The precision of this data can be reduced by specifying a geographic area of particular spectrums rather than a point (for example, a circle with a 300 meter radius centered at 40° North, 105° West). However, the limits of such a precision specification can be circumvented by repeatedly sampling an individual's geodetic location.

Mobile location data – Civic or geodetic location data that identifies the whereabouts of an individual or his or her device in real or near-real time.

Fixed location data – Civic or geodetic location data that describes a fixed location associated with an individual. Examples include a home or office location.

Nomadic location data – Civic or geodetic location data that identifies the whereabouts of an individual using a device that may be moved occasionally from its fixed location. For example, if an individual occasionally uses his or her laptop at an Internet cafe, the location of the laptop would be considered nomadic.

IDENTIFIABILITY DEFINITIONS

These definitions measure data identifiability from the perspective of the entity collecting and using data for online advertising (as opposed to an outside observer or statistician, for example). How easy or hard it may be for such an entity to use data to identify an individual depends on the other data sources available to the entity, the capabilities of the entity, and the time, effort, and cost required to identify individuals. Note that all inferably identifiable data is pseudonymous, but all pseudonymous data is not necessary inferably identifiable.

Aggregate data – Data about multiple individuals that cannot reasonably be used to directly or inferably identify any single individual.

Directly identifiable data – Data that directly and overtly identifies an individual, such as name, address, email address, phone number, government identifier, or financial identifier.

Inferably identifiable data – Data from which an individual's identity can be reasonably inferred, including combinations of data elements or data sets that would not, on their own, identify an individual.

Pseudonymous data – Data associated with a unique identifier that does not directly identify an individual.

Appendix B: Example Applications

The three example practices described and evaluated below are Single Site Personalization, an Ad Network, and Social Ads on a Social Network.

These examples are hypothetical practices, although they are based on real-world advertising practices. The FIPs implementation section for each practice provides examples of how the practice could be implemented in both a high-risk and a low-risk way (per the categories outlined in Section III of the Threshold Analysis). These FIPs descriptions are meant to be clear-cut: for each practice, one description provides a bare-bones FIPs implementation that could land the practice in the high-risk category, and the other description provides a highly robust FIPs implementation that could land the practice in the low-risk category. Because the examples are hypothetical, such FIPs implementations may not exist today.

PRACTICE 1: SINGLE SITE PERSONALIZATION

Description

An individual visits an e-commerce site that targets ads to him on the site based on his purchases on the site, specific items in which he expresses interest on the site, and aggregate data about other users with similar interests.

1. **What entities are involved in collecting and using data to target the ad?** E-commerce site

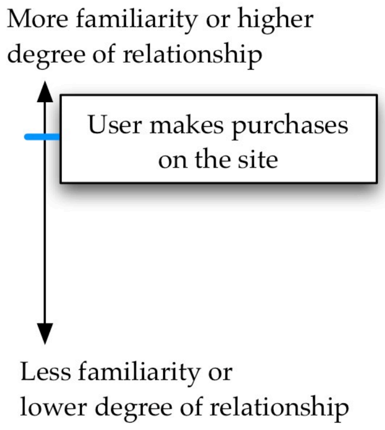
   a. How many entities are involved in collecting and using data to target the ad? 1

   b. For each entity, what type of entity is it? E-commerce site: First party

2. **For each entity involved, describe the nexus between the entity and the individual.**

   a. What is the user's familiarity or degree of relationship with the entity in other contexts?

E-commerce site:



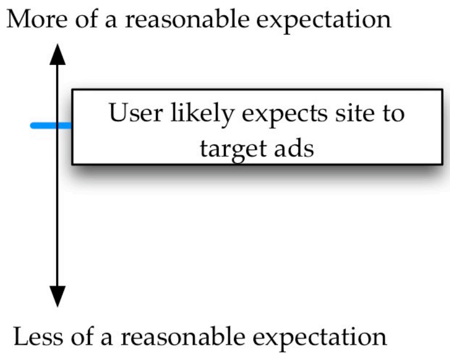
More familiarity or higher degree of relationship

User makes purchases on the site

Less familiarity or lower degree of relationship

   b. What is the user's reasonable expectation of the entity's involvement in ad targeting?

E-commerce site:



More of a reasonable expectation

User likely expects site to target ads

Less of a reasonable expectation

3. What data about the individual is collected and used to target the ad?

a. How many distinct data streams are collected and used to target the ad? 2

b. For each data stream, what types of data are collected and used to target the ad?

Data stream 1:

☒ Clickstream data
□ Communications content
□ Public data
☒ User-generated data
☒ Purchase data
□ Sensitive data
□ Derived data

□ Nonresident data
□ Resident data
□ Offline data
□ Mobile location data
□ Nomadic location data
□ Fixed location data

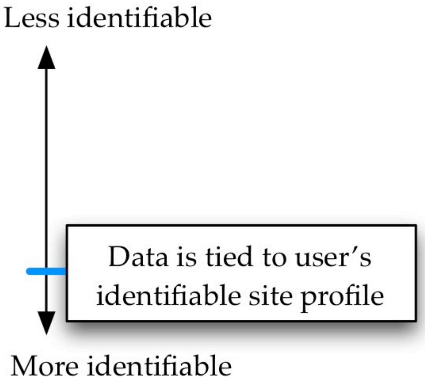
Data stream 2:

□ Clickstream data
□ Communications content
□ Public data
□ User-generated data
□ Purchase data
□ Sensitive data
☒ Derived data

□ Nonresident data
□ Resident data
□ Offline data
□ Mobile location data
□ Nomadic location data
□ Fixed location data

4. For each data stream, how identifiable is data in the stream?

Data stream 1:

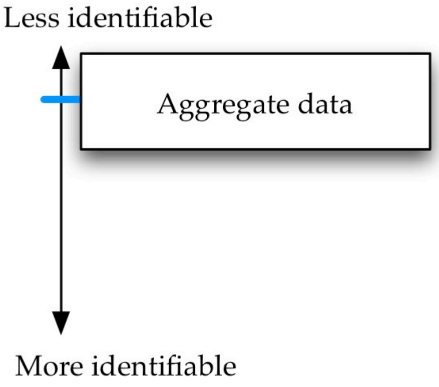


Less identifiable

Data is tied to user's identifiable site profile

More identifiable

Data stream 2:



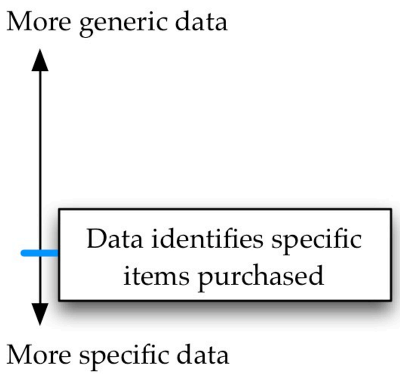
Less identifiable

Aggregate data

More identifiable

5. For each data stream, how specific is data in the stream?

Data stream 1:

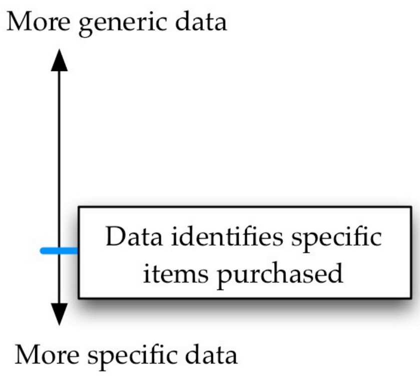


More generic data

Data identifies specific items purchased

More specific data

Data stream 2:



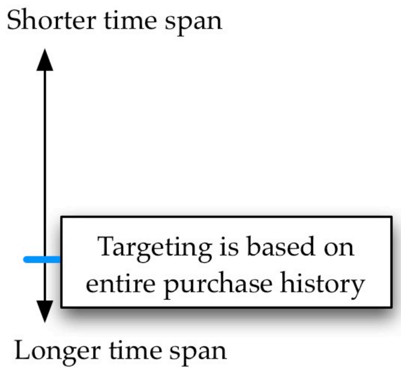
More generic data

Data identifies specific items purchased

More specific data

6. For each data stream, how long is data in the stream held before it is used for ad targeting?

Data stream 1:

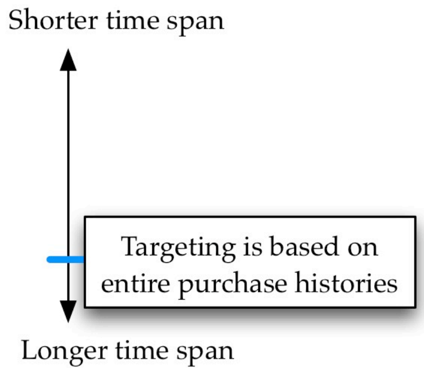


Shorter time span

Targeting is based on entire purchase history

Longer time span

Data stream 2:



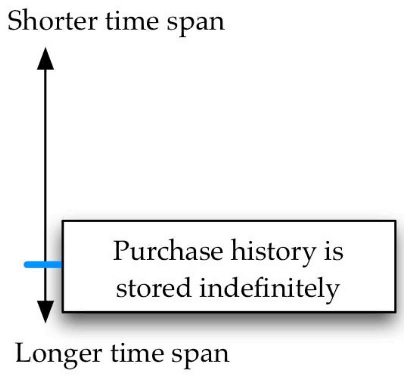
Shorter time span

Targeting is based on entire purchase histories

Longer time span

7. For each data stream, how long is data in the stream retained?

Data stream 1:

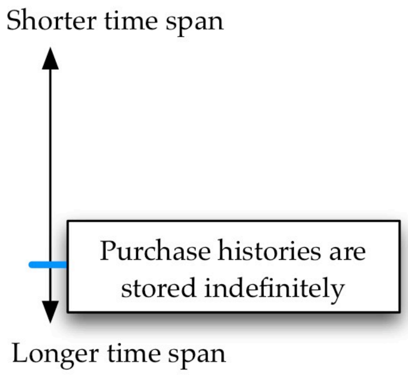


Shorter time span

Purchase history is stored indefinitely

Longer time span

Data stream 2:



Shorter time span

Purchase histories are stored indefinitely

Longer time span

EXAMPLE HIGH-RISK AND LOW-RISK FIPS IMPLEMENTATIONS

Consider two hypothetical e-commerce sites, Site A and Site B, that use single site personalization as it is described above. Site A's entire FIPs implementation consists of the following:

1. Site A discloses in its privacy policy that it collects clickstream, purchase, and user-generated data, but does not specify the purpose for doing so or the length of time that data is retained.
2. Site A applies industry-standard security protections to the data it collects.

Site A's implementation of the single site personalization practice is likely to be considered by some as high-risk. While Site A applies security protections to its data and is open about the fact that it collects data, it falls short in adhering to several other FIP principles. Site A's disclosures lack purpose specification, and Site A makes no provision for collection limitation, data quality, use limitation, individual participation, or accountability.

Site B's entire FIPs implementation consists of the following:

1. Site B explains its single site personalization practice in the context of the targeted ads it serves.
2. Site B explains its single site personalization practice to users in their profile settings page and provides on that page an easily accessible mechanism to allow users to decide not to receive targeted advertisements.
3. Site B discloses in its privacy policy that it collects clickstream, purchase, and user-generated data for the purpose of serving targeted advertisements.
4. Site B discloses in its privacy policy that it retains user data indefinitely.
5. Site B allows users to view the derived data, purchase data and user-generated data that it collects, and it allows users to make changes to the user-generated data and the derived data.
6. Site B applies industry-standard security protections to the data it collects.
7. Site B allows users to choose to delete their purchase and clickstream data after 12 months.
8. Site B undergoes regular audits to ensure that the above policies are being followed.

Site B's implementation of the single site personalization practice is likely to be considered by some as low-risk. Because this practice involves the collection and retention of directly identifiable data, it requires robust adherence to FIPs, which Site B provides.

PRACTICE 2: AD NETWORK

Description

A third-party ad network collects a user's clickstream and purchase data from visits to Web sites and uses that data to target ads to the user on other Web sites.

1. What entities are involved in collecting and using data to target the ad? Ad network

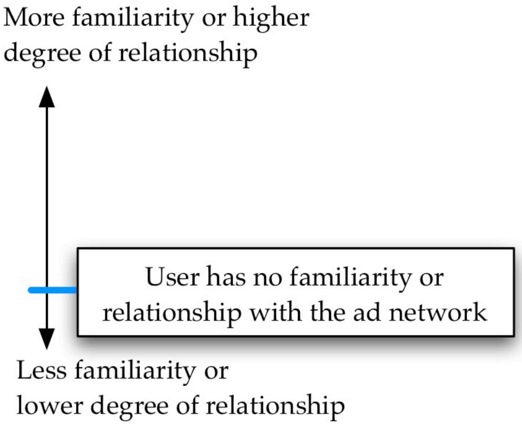
a. How many entities are involved in collecting and using data to target the ad? 1

b. For each entity, what type of entity is it? Ad network: Third party

2. For each entity involved, describe the nexus between the entity and the individual.

a. What is the user's familiarity or degree of relationship with the entity in other contexts?

Ad network:



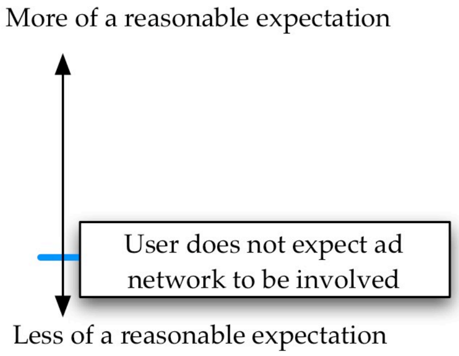
More familiarity or higher degree of relationship

User has no familiarity or relationship with the ad network

Less familiarity or lower degree of relationship

b. What is the user's reasonable expectation of the entity's involvement in ad targeting?

Ad network:



More of a reasonable expectation

User does not expect ad network to be involved

Less of a reasonable expectation

3. What data about the individual is collected and used to target the ad?

a. How many distinct data streams are collected and used to target the ad? 2

b. For each data stream, what types of data are collected and used to target the ad?

Data stream 1:

☒ Clickstream data
□ Communications content
□ Public data
□ User-generated data
☒ Purchase data
☒ Sensitive data
□ Derived data

□ Nonresident data
□ Resident data
□ Offline data
□ Mobile location data
□ Nomadic location data
□ Fixed location data

Data stream 2:

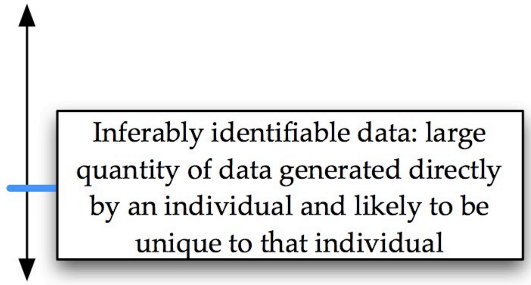
□ Clickstream data
□ Communications content
□ Public data
□ User-generated data
□ Purchase data
□ Sensitive data
☒ Derived data

□ Nonresident data
□ Resident data
□ Offline data
□ Mobile location data
□ Nomadic location data
□ Fixed location data

4. For each data stream, how identifiable is data in the stream?

Data stream 1:

Less identifiable

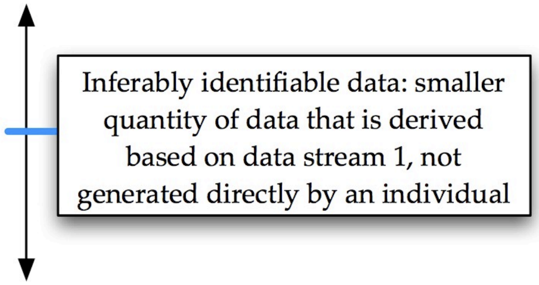


Inferably identifiable data: large quantity of data generated directly by an individual and likely to be unique to that individual

More identifiable

Data stream 2:

Less identifiable

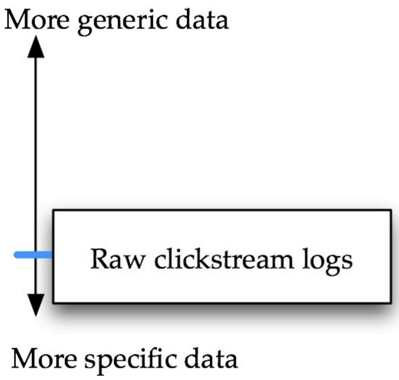


Inferably identifiable data: smaller quantity of data that is derived based on data stream 1, not generated directly by an individual

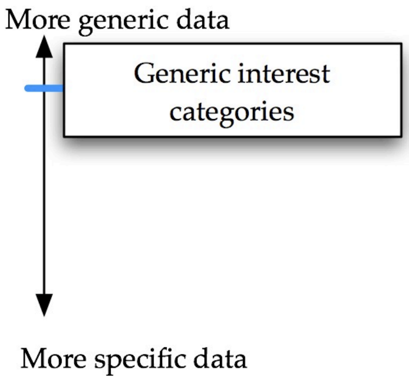
More identifiable

5. For each data stream, how specific is data in the stream?

Data stream 1:

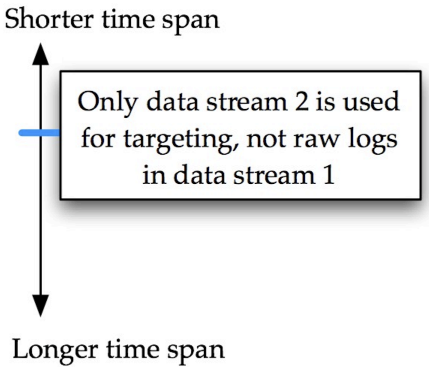


Data stream 2:

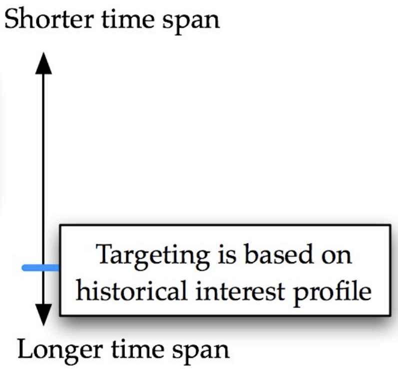


6. For each data stream, how long is data in the stream held before it is used for ad targeting?

Data stream 1:

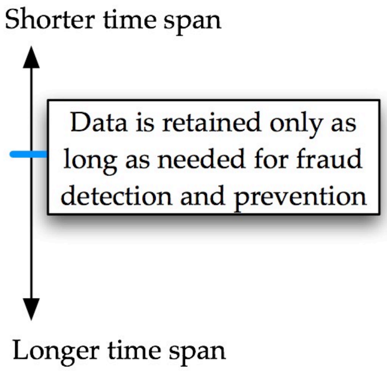

Shorter time span
Only data stream 2 is used for targeting, not raw logs in data stream 1
Longer time span

Data stream 2:

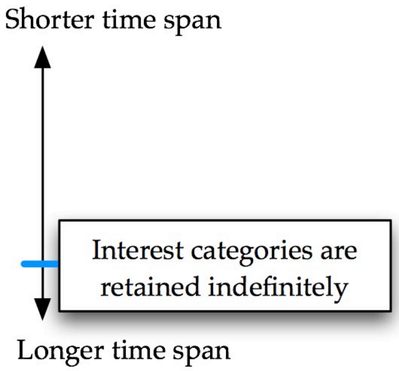


7. For each data stream, how long is data in the stream retained?

Data stream 1:


Shorter time span
Data is retained only as long as needed for fraud detection and prevention
Longer time span

Data stream 2:



EXAMPLE HIGH-RISK AND LOW-RISK FIPS IMPLEMENTATIONS

Consider two hypothetical ad networks, Ad Network A and Ad Network B, that engage in the advertising practice described above. Ad Network A's entire FIPs implementation consists of the following:

1. Ad Network A discloses in its privacy policy that it collects clickstream and purchase data for the purpose of serving targeted advertisements, but does not specify the length of time that data is retained.
2. Ad Network A requires all Web sites where it collects data to disclose in their privacy policies that the ad network data collection is happening on the sites.

Ad Network A’s implementation of the advertising practice is likely to be considered by some as high-risk. While Ad Network A is open about the fact that it collects data and requires its partners to be open about it as well, Ad Network A falls short in adhering to several other FIP principles. Ad Network A makes no provision for collection limitation, data quality, use limitation, individual participation, data security, or accountability.

Ad Network B's entire FIPs implementation consists of the following:

1. Ad Network B discloses in its privacy policy that it collects clickstream and purchase data for the purpose of serving targeted advertisements.
2. Ad Network B applies industry-standard security protections to the data it collects.
3. Ad Network B allows users to decide not to receive targeted advertisements based on the data it collects.
4. Ad Network B does not collect data about users who have opted not to receive targeted advertisements.
5. Ad Network B discloses in its privacy policy that it retains user data for one year.
6. Ad Network B requires all first-party sites where it collects data to provide prominent statements about the ad network data collection happening on the sites and what users' choices are with respect to the data collection.
7. Ad Network B allows users to view and change the derived data (i.e., the advertising profiles) it holds about them.
8. Ad Network B undergoes regular audits to ensure that the above policies are being followed.

Ad Network B’s implementation of the advertising practice is likely to be considered by some as low-risk. The data streams involved in this practice are different from those in the single site personalization practice; for example, data stream 1 is retained for a shorter amount of time, data stream 2 is more generic, and both streams are inferably identifiable. Thus, while Ad Network B’s FIPs implementation differs from Site B’s, the practice as conducted by Ad Network B could still be considered low-risk since its FIPs adherence is robust enough given the features of its advertising practice.

PRACTICE 3: SOCIAL ADS ON A SOCIAL NETWORK

Description

A social network provides the ability for users to declare each other as friends on the network. Consider two users, Alice and Bob, who are friends with each other on the social network. When Alice visits an external e-commerce site, the social network acts as a third party on that site, collecting data about the purchases Alice makes there. The social network uses this data as the basis for an ad displayed to Bob on the social network. For example, if Alice buys tickets to Gone With the Wind on CinemaTix.net, the next time Bob logs on to view his social network profile, the social network may display an ad on Bob's profile page that says, "Alice bought tickets to Gone With the Wind on CinemaTix.net."

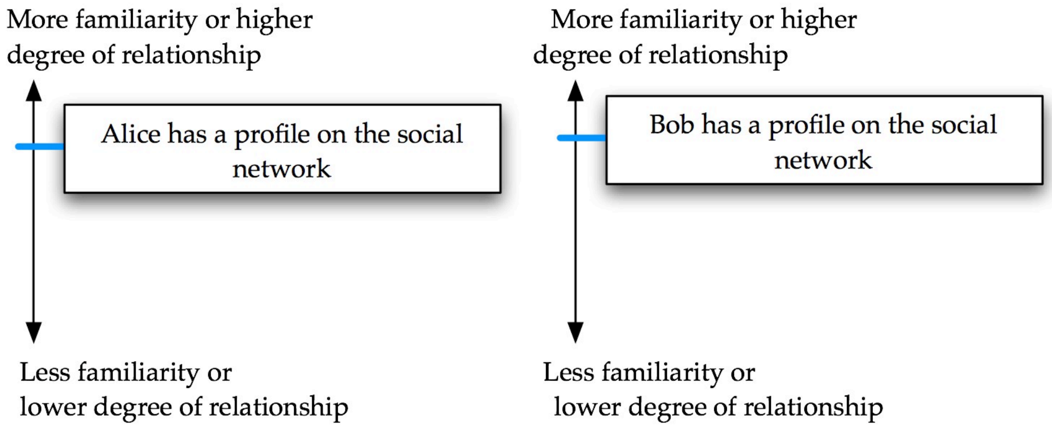
The social network may display ads about other kinds of "actions" that users take on sites external to the social network: posting a restaurant review, writing a blog comment, or sharing a recipe, for example. Under the Threshold Analysis, these actions are classified as user-generated data.

The social network also incorporates user-generated preference data from those viewing the ads. In this example, the social network provides Bob with the ability to indicate whether he likes or dislikes seeing ads about Alice's actions.

1. What entities are involved in collecting and using data to target the ad? Social network

a. How many entities are involved in collecting and using data to target the ad? 1

b. For each entity, what type of entity is it? Alice: Social network is a Third Party
Bob: Social network is a First Party



More familiarity or higher degree of relationship

Alice has a profile on the social network

Less familiarity or lower degree of relationship

More familiarity or higher degree of relationship

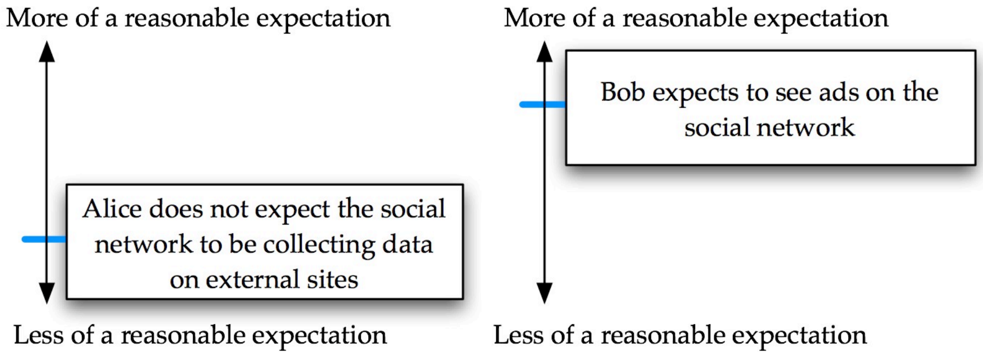
Bob has a profile on the social network

Less familiarity or lower degree of relationship

b. What is the user's reasonable expectation of the entity's involvement in ad targeting?

Social network/Alice:

Social network/Bob:



More of a reasonable expectation

Alice does not expect the social network to be collecting data on external sites

Less of a reasonable expectation

More of a reasonable expectation

Bob expects to see ads on the social network

Less of a reasonable expectation

3. What data about the individual is collected and used to target the ad?

a. How many distinct data streams are collected and used to target the ad? 1 for Alice, 1 for Bob

b. For each data stream, what types of data are collected and used to target the ad?

Alice's data stream:

□ Clickstream data
□ Communications content
□ Public data
■ User-generated data
■ Purchase data
□ Sensitive data
□ Derived data

□ Nonresident data
□ Resident data
□ Offline data
□ Mobile location data
□ Nomadic location data
□ Fixed location data

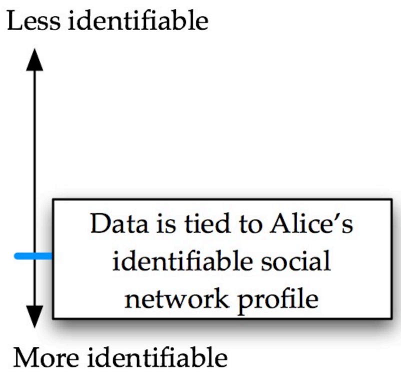
Bob's data stream:

□ Clickstream data
□ Communications content
□ Public data
■ User-generated data
□ Purchase data
□ Sensitive data
□ Derived data

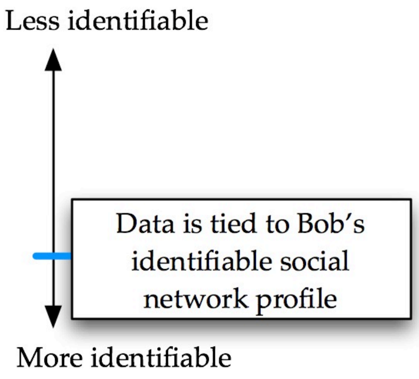
□ Nonresident data
□ Resident data
□ Offline data
□ Mobile location data
□ Nomadic location data
□ Fixed location data

4. For each data stream, how identifiable is data in the stream?

Alice's data stream:

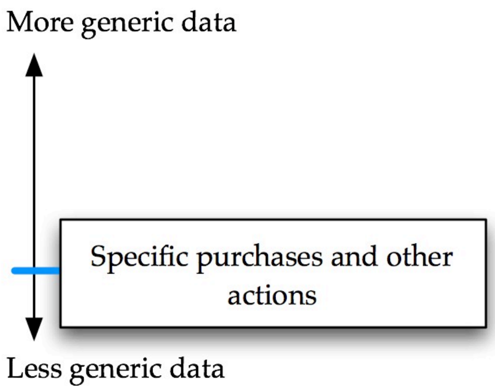


Bob's data stream:

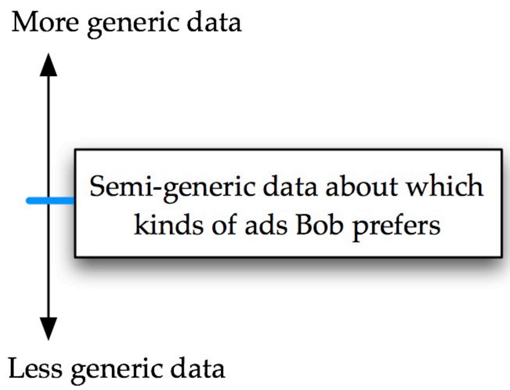


5. For each data stream, how specific is data in the stream?

Alice’s data stream:

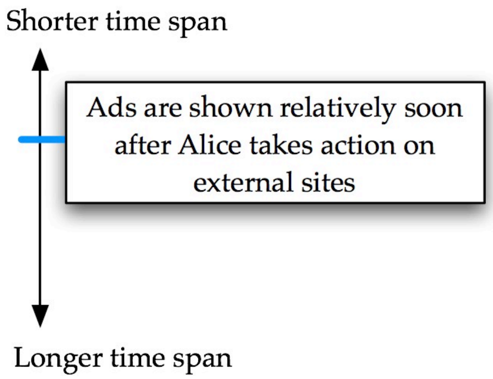


Bob’s data stream:

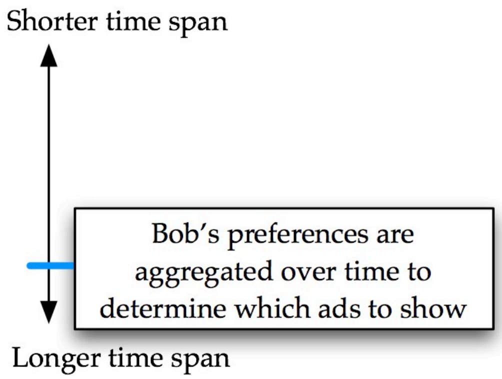


6. For each data stream, how long is data in the stream held before it is used for ad targeting?

Alice’s data stream:

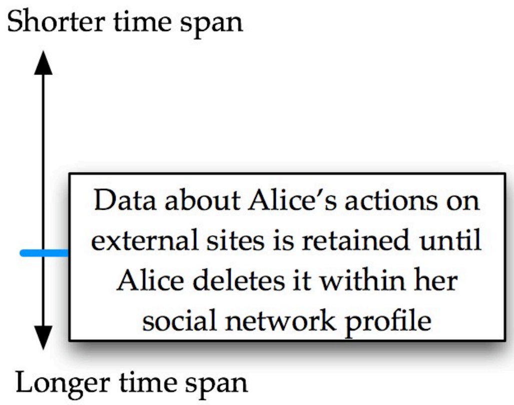

Shorter time span
Ads are shown relatively soon after Alice takes action on external sites
Longer time span

Bob’s data stream:

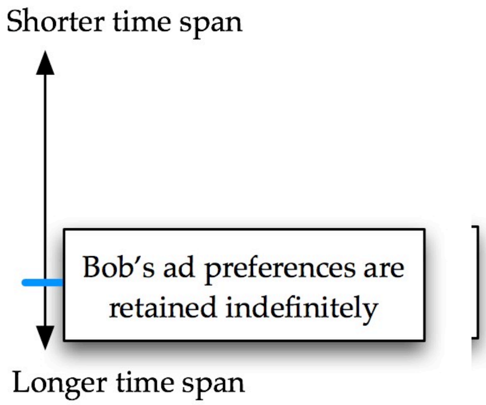

Shorter time span
Bob’s preferences are aggregated over time to determine which ads to show
Longer time span

7. For each data stream, how long is data in the stream retained?

Alice’s data stream:


Shorter time span
Data about Alice’s actions on external sites is retained until Alice deletes it within her social network profile
Longer time span

Bob’s data stream:



EXAMPLE HIGH-RISK AND LOW-RISK FIPS IMPLEMENTATIONS

Consider two hypothetical social networks, Social Network A and Social Network B, that engage in the advertising practice described above. Social Network A's entire FIPs implementation consists of the following:

1. Social Network A prominently discloses on its own site that it conducts this form of ad targeting.
2. Social Network A applies industry-standard security protections to the data it collects.
3. Social Network A provides simple controls that allow users to delete their data related to the ads served at any time.
4. Social Network A requires all Web sites where it collects data to provide prominent statements about the social network data collection happening on the sites and what users' choices are with respect to the data collection.
5. Social Network A undergoes regular audits to ensure that the above policies are being followed.
6. Social Network A displays ads about Alice's actions without first obtaining her affirmative consent to do so.

Social Network A adheres well to many FIPs, but its implementation of the advertising practice is still likely to be considered high-risk by some (from Alice's perspective). This is because the data collection and use are not limited to circumstances where social network users (like Alice) affirmatively agree to having their actions on external sites exposed as ads to their friends (like Bob) on the social network site. This deficiency with regard to these two FIP principles is significant enough for Social Network A's practice to be considered high-risk by some.

1. Social Network B's FIPs implementation is similar to Social Network A's, but the last provision is replaced by two new provisions:
2. Social Network B prominently discloses on its own site that it conducts this form of ad targeting.
3. Social Network B applies industry-standard security protections to the data it collects.
4. Social Network B provides simple controls that allow users to delete their data related to the ads served at any time.
5. Social Network B requires all Web sites where it collects data to provide prominent statements about the social network data collection

happening on the sites and what users' choices are with respect to the data collection.

6. Social Network B undergoes regular audits to ensure that the above policies are being followed.
7. Social Network B requires users (like Alice) to affirmatively consent to having their actions displayed as ads on their friends' pages, both on a global basis on a per-action basis.
8. Social Network B does not collect data about users who have not opted to have their actions used as advertisements.

By including these two provisions that Social Network A lacks, Social Network B’s FIPs implementation is likely to be considered by some as low-risk. Social Network B limits collection and use sufficiently and adheres robustly to other FIP principles.