Subcommittee report – Entities and the ability to edit or correct

Assembled by Rob Goldman (dash.com) from meeting notes with contributions from Andrew Shen (Epic) and Dan Schutzer (Citigroup)

By way of process, the subcommittee has attempted to identify important issues for the committee to consider with regard to the access and security of online consumer information as it relates to broad areas of entities and the ability for consumers to edit or correct their personal information. Further, we have tried to provide the wide spectrum of options available to address each of these issues, and finally we have tried to narrow those options to ranges that the subcommittee felt were practical to consider. Items of disagreement that subcommittee felt deserved a more complete discussion are covered in the "Discussion and Debate" section of this document.

1. Which Entities are required to provide access to data?

a. None
b. The data collector
c. Data collector and entities indicated by the data collector at the time of collection
d. Data collector and subsidiary entities
e. Data collector, subsidiary entities, and parent entity
f. Data collector, subsidiary entities, parent entity and agents
g. Whomever the consumer reasonably believes is the data collector
h. Whomever the consumer reasonably believes is the data collector + entities indicated by data collector at time of collection
i. Data collector, parents, subsidiaries, and recipients excluding information intermediaries
j. Data collector, parents, subsidiaries, and recipients including information intermediaries
k. Data collector, parents, subsidiaries and all data recipients
l. All entities

Costs and Benefit Discussion

- Obviously, entities that don't possess the data cannot offer access to it. This may make items d and e impractical.
- Items "e" onward may require some sort of aggregation and attendant aggregation costs.
- Generally speaking, the methodology of access will affect cost.
- For items "g" and h", the concept of "Reasonably believes" will be strongly effected by notice
- It is possible that costs may be managed by outsourcing customer data management functions to third parties. This is the case where an entity acts solely as an agent of the collecting company.

- What specifically is access? Would a web page that acts as a roadmap to all the access points within a parent entity be considered acceptable access?
- Should these considerations be adjusted in accordance with the origin of information? Should information obtained offline but moved online be considered separately from information obtained online?
- What should be done in the event that one company acquires another with different access policies? Which policies should apply to the combined data?
- Should entities be required disclose the source of data they have obtained? Must records be kept of the source of information? Both would complicate interactions between entities involving data.

2. To which types of information must entities provide access?

a. None
b. Whatever data the company indicates
c. Information that was positively provided by consumers
d. Information that was positively provided and passively generated by consumers
e. Information that was positively provided and passively generated by consumers and information that was collected from sources other than the subject (Transferred from other sources or publicly available)
f. Aggregated or master-file data. – This is a summary the facts contained in more detailed information. This applies to any large volume data where offering access to detail level information is impractical. This is inductive information that is derived without any intellectual property contributed by the company itself, and without the use of any proprietary algorithms or business logic.

Name	Site	Month	Count
Rob Goldman	Yahoo	March	2

g. Detailed level information – The detailed information that makes up the aggregated or masterfile data

Name	Site	Date
Rob Goldman	Yahoo	3/10/00 1:00pm
Rob Goldman	Yahoo	3/11/00 1:10pm

h. Detailed level information and derived data. – This is deductive information inferred from detailed data which has proprietary value based on unique business logic applied to raw data (e.g. Profile information)

Name	Deduction
Rob Goldman	Loyal Yahoo User

i. All information

Costs and Benefits Discussion:

- The way the data is being used is an important consideration. In the case where people are using software that collects information as a byproduct of its business function (A proxy server, for example) and this information is passively collected and not being used by the company for any purpose that affects the customer, that should be a consideration..
-
- Which order should F,G and H be in? Consumers may be more interested in information that is derived about them than they are about the detailed information that they used to derive it in the first place.
- What should be done in situations where derivations are a source of competitive advantage as in the case of credit scoring or risk assessment?
- Should different considerations apply to information that is not maintained online?
- Should information used to make "important decisions" be considered separately?

3. Is there an obligation to propagate corrections to incorrect data to other entities?

a. No obligation
b. When reasonable
c. Obligate to propagate corrections to particular entities (List taken from item #1)
    i. None
    ii. The data collector
    iii. Data collector and entities indicated by the data collector at the time of collection
    iv. Data collector and subsidiary entities
    v. Data collector, subsidiary entities, and parent entity
    vi. Data collector, subsidiary entities, parent entity and agents
    vii. Whomever the consumer reasonably believes is the data collector
    viii. Whomever the consumer reasonably believes is the data collector + entities indicated by data collector at time of collection
    ix. Data collector, parents, subsidiaries, and recipients excluding information intermediaries
    x. Data collector, parents, subsidiaries, and recipients including information intermediaries
    xi. Data collector, parents, subsidiaries and all data recipients
    xii. All entities

Costs and Benefits Discussion:

- Perhaps information of official public record should be treated differently?
- Companies will correct information when there is a market reason to do so. Is there reason to believe that the market will fail here?

4. Ability to edit or correct what types of data? (Categories taken from old Access 1 subcommittee, and do not necessarily lie on a continuum)

a. **Physical Contact Information -** Information that allows an individual to be contacted or located in the physical world -- such as phone number or address.
b. **Online Contact Information** - Information that allows an individual to be contacted or located on the Internet -- such as email. Often, this information is independent of the specific computer used to access the network. (See the category "Computer Information")
c. **Globally Unique ID (GUID) -** Non-financial identifiers issued for purposes of consistently identifying the individual across multiple entities.
d. **Locally Unique ID (LUID) -** Non-financial identifiers issued for purposes of consistently identifying the individual used by a single entity and never released to another entity association with physical contact information, online contact information, or a globally unique ID.
e. **Biometric Identifiers -** Measurable physiological and / or behavioral characteristics that can be used to verify the identity of an individual. They include fingerprints, retinal and iris scanning, hand geometry, voice patterns, facial recognition and other techniques. (Avanti -- http://www.biometric.freeserve.co.uk/whtpaper.htm)
f. **Financial Account Identifiers -** Identifiers that tie an individual to a financial instrument, account, or payment system -- such as a credit card or bank account number.
g. **Computer Information -** Information about the computer system that the individual is using to access the network -- such as the IP number, domain name, browser type or operating system.
h. **Navigation and Click-stream Data -** Data passively generated by browsing the Web site -- such as which pages are visited, and how long users stay on each page.
i. **Interactive Data -** Data actively generated from or reflecting explicit interactions with a service provider through its site -- such as queries to a search engine, or logs of account activity made on the Web.
j. **Transactional Data -** Data actively generated that reflects the purchase of products or services.
k. **Demographic and Socio-economic Data -** Data about an individual's characteristics -- such as gender, age, and income.
l. **Inferred Data -** Information attributed to an individual that is derived from other information known or associated with the individual. Imputed data can be data generated through the application of a mathematical program to known data, or it can be information such as census data that can be imputed to a range of individuals based on residence or some other trait (commonly called overlay data).
m. **Preference Data -** Data about an individual's likes and dislikes -- such as favorite color or musical tastes.

n. **Content -** The words and expressions contained in the body of a communication -- such as the text of email, bulletin board postings, or chat room communications.
o. **State Management Mechanisms -** Mechanisms for maintaining a stateful session with a user or automatically identifying users who have visited a particular site or accessed particular content previously -- such as HTTP cookies.
p. **Image -** The visual representation of an individual.

Costs and Benefits Discussion:

- There are costs and benefits to both business and consumers that must be considered here. Consumers face a higher cost in not having correct data for certain types of information (credit information vs. marketing information, for instance)
-
- Who should be allowed to edit or correct data? An authenticated user only? An authenticated user or their an agent acting on their behalf?
- Should entities requesting that information be corrected have to provide proof that the information is wrong?
- Should consumers be able to correct any wrong information?
- Should users be able to correct an inference? Inferences aren't right or wrong. They are something else by their very nature.
- What about click stream information or log data? Information could be one wrong in one part per million. Providing the ability to edit or amend this information could be considerable fantastically expensive.
- Should the use (as opposed to the type) of the information be considered when providing access or allowing consumers to correct or edit information?
- Must companies retain a record of the information that was incorrect after it has been corrected? Should they be allowed to?
- What should be done in the event that the accuracy of the data is disputed and irreconcilable? Must an investigation take place?
- There is a distinction between indicating which information is incorrect and actually correcting the information. Which do we want?

5. Ease of access

a. Fees
    i. Never Charge any fee No costs incurred
    ii. Selectively charge fees Nominal costs
        1) Fees commensurate with type
        2) Fees commensurate with use of data
        3) Fees commensurate with amount of data
        4) Fees commensurate with frequency of use access.
    iii. Fee ceilings and floors
    iv. Always charge a fee
b. Usability of the access and correction system
    v. Interface is legible
    vi. Information is intelligible
c. Notice of access and ability to edit

Costs and Benefits Discussion:

- Should fees be waved if there is a hardship?
- Can fees be required for a security need?

Discussion and Debate:

1. Access and derived data.

Some of the members believe that individuals should have the right to see data derived (given the ability to identify and authenticate users) from information collected from them. As this data is what is used to make decisions based on their behavior, it is critical in the opinion of some that this also be made available. Access to this derived data could, but does not necessarily, include the ability to review or see the algorithms used to derive such data.

Other members of the subcommittee expressed concern that providing access to derived data would affect the confidentiality of procedures companies use to make decisions and assumptions about user data. Without this confidentiality, some companies and industries would be unable to maintain their current market viability.

As a precedent, subcommittee members also believe it is important to consult with appropriate experts within the FTC and elsewhere on the issue of credit scoring. The recent move towards transparency of credit scores should inform the recommendations of the subcommittee and the full Advisory Committee.

2. Does access threaten privacy?

As many companies that are holding personal information are part of a larger corporate entity that may possess other data through different subsidiaries, would access to all the information held by the parent company necessarily bring together all this previously separated information? And, would this combining of information in itself pose an increased threat to personal privacy?

Subcommittee members agree that the goal of access is not to centralize more personal information. The most expansive interpretation of access should not have the indirect effect of creating a new file or record on an individual. Under this hypothetical expansive interpretation, the individual would have access to all available personally identifiable information *existing* at the time of the request.

However, some subcommittee members believe that these concerns should not prevent parent companies from implementing procedures increasing ease of access. One proposal made by Rob Goldman of Dash.com is to have parent companies create a central page, which would direct consumers to their various subsidiaries which may have different pieces of personal information in their own distinct records.

3. Is deletion part of access?

Some of the members of the subcommittee believe that individuals have the right to delete data that is no longer necessary to complete obligations of the businesses to the consumer. Some information currently held by companies serves no particular use and has outlived the purposes for which it was originally collected. In these situations, there should be no barrier to the removal of personally identifiable information from these databases.

Some subcommittee members have pointed out that certain categories of user data are necessary for system maintenance, network integrity, record keeping, or auditing. In considering the ability to delete data, these needs should be taken into account.