Subcommittee report – Entities and the ability to edit or correct

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 present some options available to address each of these issues 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. Much of the "Discussion and Debate" section is presented as open questions. These open questions were the items that most often led to the divergent views of what were the preferable options.

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

a. All entities that collect information from a data subject and actively maintain a database of consumer information that can be linked/associated with individual consumers and/or consumer households.

b. The entity the consumer reasonably believes is the Data Collector and its agents (entities acting for the Data Collector and restricted in their use and transfer of the data). Notice of transfer to other entities would be required, but access would not be required.

c. Data collector, parents, subsidiaries, and recipients including information intermediaries

Costs and Benefit Discussion

The committee felt that there were only three reasonable alternatives regarding which entities could be required to provide customers access to data maintained about them.

- Obviously, entities that don't possess the data cannot offer access to it.
- **Clearly, a company collecting information from consumers should, where such data is maintained in a form which can be linked back to an individual consumer or consumer household, make it accessible to the consumer under reasonable conditions of access, unless there is some legitimate reason for refusing (see later sections).**
- The sub-committee agreed that at a maximum, access should be provided only for information that is maintained on-line and for which the customer can practically be provided access to; e.g. information collected but not maintained would be impractical to be provided (e.g. demographic data used for determining candidates for a direct mail solicitation, but not maintained after the mailing address list is generated) would not be reasonable to provide access to. Another example would be information collected to conform to legal or regulatory or audit requirements, and maintained off-line, on tapes, or in serial files that would be difficult and costly to provide access to. As noted in many of the other comments, many members of the sub-committee thought ability to access was one factor to consider, but that there are other factors which should allow a data collector to not have to provide access (e.g. type of information, use, cost, etc.)
-

The issue, and a point of contention for the sub-committee, was whether this requirement should be extended to include the parent, and all the subsidiaries of the corporation? And whether or not the right of access should be extended to all parties with whom information has been shared,

including information intermediaries hired to assist the data collector? For example, when the customer data management function is outsourced to third parties. Some members of the sub-committee thought this extension of access to third party recipients was necessary for sufficient consumer protection. The sub-committee generally agreed that corporations should provide access to the data held by their agents (as defined above). However, several members of the sub-committee thought managing other third parties would be unduly burdensome, and that the consumers were better protected by requiring companies to provide notice of with whom they will share the information.).

Other members of the sub-committee believed the issue depended on whether the parent and/or subsidiaries are using this information. If they are, then they should make it accessible and protect it. If not, then no. With respect to "information intermediaries," it depends on how they treat and handle the data. If they use the information, view it and permanently store it then they should make it accessible and protect it. If not, then access is not required.

Other related questions:

- Should a corporation be required to provide access to all the data collected on individual consumers in an integrated fashion, even if it is not used by the corporation in this way (e.g. collected and maintained by separate corporate entities, different subsidiaries? Is it even desirable, or wouldn't the consumer, due to the privacy interests, prefer to prevent the combination of information that previously existed as separate records? Would a web page that acts as a roadmap to all the access points within a parent entity be considered acceptable access? By integrating all the information under a master web page would the benefits of ease of access outweigh the costs of creating design and engineering requirements.

- 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? In this context, how will online be defined? Does this mean any information which is reproduced in digital format, or does it mean information which is combined in a database with information collected 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 the acquiring company be required to honor the old companies past commitments made with respect to consumer privacy, or is the acquiring company merely required to provide notice of a change in privacy policy? Does this notice requirement put too big of a burden on consumers to understand when new acquisitions take place?
- Should entities be required to 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. Should the ability to access, edit or correct data vary with the use of the data?

a. Yes, no need to access, edit or correct data that is not actively used for anything, or merely maintained for system integrity, troubleshooting, or auditing.
b. Yes, only need to allow access, edit and correct data that is used to make important decisions such as financial or medical decisions, or employment decisions
c. No, the consumer should have the right to be able to access, edit or correct any data collected and maintained about them so long as that can be reasonably made accessible by the holder of the data.

Costs and Benefits Discussion:

- Many members of the sub-committee thought the use of the data should not be a factor in determining whether or not to grant a consumer the ability to access, edit or correct data maintained about them. Although the way the data is being used is an important consideration, it is a slippery slope. What is collected today and not used, might be in the future. What is considered an unimportant use or decision by some, might be considered very important by others Who should decide what decisions are "important", and what is the basis for that distinction? Furthermore, if data is not really used, or if care is not provided in ensuring its accuracy then why go through the expense of collecting and maintaining it?

-

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

a. No obligation
b. When reasonable
c. Always for all entities (including propagation back to the entity which received or supplied the data).

Costs and Benefits Discussion:

- If a consumer can show that information maintained about him/her is of error it is in the companies best interest to correct that information. However, companies will correct information when there is a market reason to do so. Is there reason to believe that the market will fail here?
- It would be desirable for a company when correcting errors to propagate these corrections to other entities, but it is recognized that the company may not be in a position to know all the entities that are currently maintaining related information about that individual, nor the state of that data (whether it has already been corrected or is in error). Once again this may be an area where it is effective to break out agents from other third parties.
- Therefore it is recommended that companies maintaining data that can be identified with an individual or household provide the affected individual(s) access to that information and the ability to correct or edit the data, if the corrections can be verified.

- 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.
- Should entities be required to disclose the source or other collectors of data they have obtained? Must records be kept of the source of information? Both would complicate interactions between entities involving data. Some members of the sub-committee believed it would be desirable to have the entities disclose the source of their data. It would facilitate corrections and other components of access. As pointed out in other areas of this document, other members of the sub-committee believed this would be burdensome and there are less restrictive alternatives available.

- Some sub-committee members pointed out 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.

4. Should the ability for a consumer to edit or correct data be determined in terms of the type of data? To answer this question, we started with the categories taken from the old Access 1 subcommittee, namely:

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.

For purposes of this discussion however, we felt it was simpler to group these into three broad classes; namely:

a. Whatever data the company maintains
b. All but inferred data, with the exception of inferred data handled under separate laws or regulations (e.g. credit loan decision)
c. Only physical contact information, online contact information, biometric identifiers, financial account identifiers, sensitive medical data, transactional data and image (or other information linked to these categories)

Costs and Benefits Discussion:

- What should be done in situations where derivations are a source of competitive advantage as in the case of credit scoring or risk assessment? **There is a case for not having to provide a customer access to inferred data as this information may be the result of a proprietary model that provides the company competitive advantage; e.g. an indicator of a customer's future purchase behavior. The only counter would be when the derived data is used to make a decision about the customer which would result in an important denial of services – e.g. granting of a loan. However, it should be noted that c**onsumers 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.

- 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? Yes, corrected information should be verifiable.
- Should consumers be able to correct any wrong information? Yes, why not? It is important for both the service provider and the consumer to work from a common base of correct information. The only caveat is that the information must be verified as correct, as we require proof that the information being corrected is wrong, and the new information is correct.
- Should users be able to correct an inference? Inferences aren't right or wrong. They are something else by their very nature, and can't be verified as right or wrong, therefore this category of information is not practical to be corrected by the user.
- What about click stream information or log data? Information could be wrong in one part per million. Providing the ability to edit or amend this information could be considerable and fantastically expensive.

- Must companies retain a record of the information that was incorrect after it has been corrected? Why would a company want to except perhaps as a record of decisions and transactions that might have been made erroneously based upon the incorrect data, prior to correction? Certainly, companies should be allowed to maintain a record of the information that was incorrect, after it has been corrected, but not required to do so. What should be done in the event that the accuracy of the data is disputed and irreconcilable? Unless there is room for reasonable doubt and disagreement (e.g. an inference), an investigation should take place?
- There is a distinction between indicating which information is incorrect and actually correcting the information. Which do we want? One can't be too careful about correcting data, we must be sure that the correcting source is authenticated and that the correct information is verifiably correct.
- Concern was expressed by several members of the sub-committee that some options would create substantial authentication hurdles (e.g. who do you give access to all the Clickstream and Navigation data connected with a particular LUI?)

5. Ease of access. This includes issues surrounding both whether access fees should be allowed, and the degree of effort required by the data access provider to ensure that the information can be easily accessed, understood and corrected by the consumer.

a. Fees
    i. Never Charge any fee. No costs should be incurred by the consumer to access their information
    ii. Selectively charge fees Nominal costs
        1) Fees commensurate with type of data being accessed.
        2) Fees commensurate with the use of data being accessed.
        3) Fees commensurate with the amount of data being accessed.
        4) Fees commensurate with frequency which a user accesses the data.
        5) Fees commensurate with the nature of the data access requirement (e.g. if the customer wants real-time access to the data when normal access is not real-time (e.g. access normally provided within 24 hours).
    iii. The service provider is free to charge any reasonable fee, but the fee must be kept within specified ceilings and floors
    iv. Always charge a fee
b. Usability of the access and correction system
    i Interface is easy-to-use, does not require any special training by a non-technical lay person; e.g. should be no harder to access than any of the services provided by the service provider.
    ii. Information is legible and intelligible (e.g. not difficult to decipher codes)
    iii. The access and correction system should both be reasonably available.

Adequate notice should be made to the consumer of what information is available for access and how to access and correct this information.

Costs and Benefits Discussion:

- Should fees be waved if there is a hardship?

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 sub-committee 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.

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?

Sub-committee members agreed 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 sub-committee 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, although even this simple integration of information might increase the vulnerability of an individual's information to compromise – e.g. now a bad guy if they can guess the password, can get access to all the customer's private information from one convenient location.. Also, such a linked page may be extremely difficult to manage for companies which regularly acquire and divest subsidiaries.

As general background on the issues raised in this document, the subcommittee recommends study of the Department of Commerce's European Union Directive on Data Protection FAQ #8. The current version of this FAQ can be found at http://www.ita.doc.gov/td/ecom/RedlinedFAQ8Access300.htm