Online Access & Security Committee
Options for the Scope of Access

Both consumers and businesses have a shared interest in the provision of reasonable access to consumer personal information. Reasonable access benefits individuals, society and business due to the openness and accountability it helps to promote. If done properly, the provision of access can also help reduce the costs to businesses and consumers of improper decision-making due to poor data quality. Moreover, increased access may help promote consumer trust and deeper customer relationships, which benefit both consumers and businesses. However, the manner in which to provide access and to what degree access should be provided are complex questions given the numerous types of non-personally identifiable and personally identifiable information, the "sensitivity" of that information, the sources of that information, and the various costs and benefits associated with providing access.

There is an extremely broad range of policy options on how access should be provided, from a very simplified "default rule" approach to a much more complex approach that subjects the scope of access to a calculation based on the sensitivity of personal data and the use of that data. We have identified three basic approaches, which we discuss in more detail below. They are: 1) the default rule approach, 2) the total access approach, and 3) the case-by-case approach.

A Default Rule Approach

Under a "default rule approach" based on the principles outlined by the BBB*OnLine* seal program, the scope of access is guided by the premise that consumers should be given as much access to their personally identifiable information (PII) as practicable. This approach would establish a default rule that PII collected online is generally accessible, with some limitations or exceptions when the cost of providing access far outweighs the benefits, and for derived data. The "default rule approach" recognizes that consumers have reason to view the information collected by businesses about them beyond being able to ensure its accuracy. Indeed, the fairly broad access rights under a "default rule approach" may promote awareness of business information practices as much as they promote accuracy. Under one theory, this broad access could affect businesses and consumers by increasing consumer awareness of the trustworthiness and responsibility of the businesses that collect information about them. Feasibly, this broad access could show the extent of information held about consumers, possibly making them wary and leading them to call for more limited collection of information. In this regard, broad access under a "default rule approach" may act to promote privacy by potentially dampening the interest of businesses in collecting more information than they need from consumers.

The over-arching rule of the "default rule approach" is that businesses should establish a mechanism whereby "personally identifiable information (PII) and "prospect information" that the business maintains with respect to an individual is made available to the individual on request.[1]

- PII (and prospect information) is information collected from an individual online (actively or passively) and is information that when associated with an individual can be used to identify him or her.[2] As an example, click-stream data is not "PII" unless it is linked to a name, email address or similarly identifying information.

- Information is *not* PII unless it is "retrievable in the ordinary course of business." Information is retrievable in the ordinary course of business if it can be retrieved by taking steps that are taken on a regular basis in the business with respect to the information, or that the organization is capable of taking with the procedures it uses on a regular basis.

- Information is *not* retrievable in the ordinary course of business if retrieval would impose an "unreasonable burden."[3] The only time a purpose or cost benefit analysis would be done would be in the rare situations where the ability to retrieve the information would be very costly or disruptive, and in that situation access could be denied if the need for the information was marginal. It is *here* that sensitivity of data, uses of data, purpose of the request, etc. would be considered.

Some other aspects of the "default rule approach" rules are:

- "reasonable terms" may be placed on access, such as frequency limits and fees, except that requests may not be limited to fewer than one request per year and charges of greater than $15 per request are not allowed;

- organizations are not required to set up new systems to maintain information beyond a time when it no longer serves the organization's purposes;

- organizations are not required to provide access to derived data or data collected from outside sources;

- steps to assure accuracy of data and processes to correct inaccuracies must be established;

- organizations have flexibility to decide *how* to make "PII" available, i.e., in what form;

- "proper identification" (undefined) may be required; and

---

[1] "Personally identifiable information" is substituted for the BBB*OnLine*'s term "individually identifiable information." "Prospect information," a term borrowed by BBB*OnLine* from the Direct Marketing Association, is information provided by a third party, such as when ordering a gift.

[2] Information collected online by others than the organization to whom the access request is made, or collected offline, is not "III." However, if "III" is merged with other non-III data, the access request would cover the merged data.

[3] This was carefully constructed language that borrowed from a concept in the Americans with Disabilities Act, which requires certain accommodations if not an "unreasonable burden," generally interpreted roughly to mean "do it unless the cost is very great and that cost far outweighs the benefits."

- there is no explicit requirement for access to be provided individuals by third party transferees.

As explained in the BBB *Online* policies:

> The term "individually identifiable information" is intended to encompass information that, when associated with an individual, can be used to identify him or her, for instance, email addresses and other information that is compiled and linked to an email address. Account, billing, and online transactional information are examples of individually identifiable information. Information need not be unique to be considered capable of identifying an individual. Consequently, addresses, telephone numbers, and dates of birth constitute individually identifiable information. Information must be capable of identifying an individual, however. Consequently, data generated by passively browsing an online site (also known as navigational or click-stream data) does not constitute individually identifiable information unless it is linked to a name, email address, or similar information that identifies an individual.
>
> In addition, the information must be information collected by the organization from the individual online. Information received by the organization, online or offline, that was collected online from the individual by others (who are not making the collection as an agent or contractor of the organization) is not itself individually identifiable information in the hands of the organization. This includes, for example, public records information in the possession of the organization that was collected online from the individual by the government agency.
> …
>
> Information is retrievable in the ordinary course of business only if it can be retrieved by taking steps that are taken on a regular basis in the conduct of the business with respect to that information or that the organization is capable of taking with the procedures it uses on a regular basis in its conduct of its business. Information is not retrievable in the ordinary course of business if retrieval would impose an unreasonable burden.
> …
>
> An organization is not required to set up any new systems to maintain information or to maintain individually identifiable information or prospect information beyond a time when it no longer serves the organization's purposes.
>
> An organization must establish a mechanism whereby, upon request and proper identification of the individual, it makes available to the individual the individually identifiable information or prospect information it maintains with respect to the individual. The information subject to this requirement tends to be, but is not limited to, (i) account or application information, for example, name, address, and level of service subscribed to, and (ii) billing information and similar data about transactions conducted online, for example, date and amount of purchase, and credit card account used.

> If an organization can not make information that it maintains available because it can not retrieve the information in the ordinary course of business, it must provide the individual w ith a reference to the provisions in its privacy notice that discuss the type of data collected, how it is used, and appropriate choices related to that data, or provide the individual with materials on these matters that are at least as complete as the in formation provided in the privacy notice.
>
> Organizations have substantial flexibility in deciding how best to make the individually identifiable information or prospect information available to the individual. For example, an organization may choose the f orm in which it discloses this information to the individual. Monthly statements from banks and credit card companies are examples of appropriate mechanisms to satisfy this disclosure obligation, even though they may reveal more than the individually iden tifiable information that the individual submitted to the organization online. The organization also determines the reasonable terms under which it will make such information available such as limits on frequency and the imposition of fees. Frequency lim its that require intervals of more than a year between requests and/or fees of more than $15 for a response to an annual request would not be reasonable except in extraordinary circumstances .

The "default rule approach" or BBB *OnLine* approach is similar t o the access principle adopted as part of the Safe Harbor discussions proposed by the U.S. Department of Commerce. The Safe Harbor was developed in response to the European Union Directive on Data Protection which, among other things, mandated that consum ers be provided reasonable access to their personal information. [4] The Safe Harbor's access principle is as follows: Individuals must have access to personal information about them that an organization holds and be able to correct, amend, or delete that i nformation where it is inaccurate, except where the burden or expense of providing access would be disproportionate to the risks to the individual's privacy in the case in question, or where the rights of persons other than the individual would be violated . Despite some language differences, the "default rule approach" and the Safe Harbor access approach are extremely similar. They both stand for the proposition that access should be provided unless the costs are too high.

Analysis of "default rule approa ch"

[4] The Directive states, in relevant part, that "Member States shall guarantee every data subject the right to obtain from the controller:
(a) without constraint at reasonable intervals and without excessive delay or expense: - confirmation as to whether or not data relating to him are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed, - communication to him in an intelligible form of the data undergoing processing and of any available information as to their source, - knowledge of the logic involved in any automatic processing of data concerning him at least in the case of the automated decisions referred to in Article 15 (1);
(b) as appropriate the rectification, erasure or blocking of data the processing of which does not comply with the provisions of this Directive, in partic ular because of the incomplete or inaccurate nature of the data;
(c) notification to third parties to whom the data have been disclosed of any rectification, erasure or blocking carried out in compliance with (b), unless this proves impossible or involves a disproportionate effort.

Proponents of this approach would argue:

1) This approach provides broad access rights and reasonably matches consumer expectations that they can access personal information collected about them. Consumers would be able to view all of the account information that they have provided and be able to correct such information if it is inaccurate.
2) Consumers would be able to view passively collected information (such as click-stream data) only when that information is linked to information that identifies them. When data is not personally identifiable, it poses much less of a privacy risk and may not need to be as accurate.
3) Businesses would not be forced to provide access if it turns out to be too costly. The growth of electronic commerce would not be burdened to as great an extent compared to that burden which could occur with the costs of providing full access.
4) Because businesses would not have to provide access to derived data, businesses could protect proprietary information. These businesses could avoid negative consequences if their competitors could otherwise use broad access to gain knowledge about how their internal processes work.
5) Compliance with this approach would likely satisfy the requirements of the European Union Directive on Data Protection.
6) The means of providing access would provide flexibility for businesses while giving consumers access without undue delay. This system would be consistent with the storage and use practices of businesses.

Opponents of this approach would argue:

1) One difficulty with the "default rule approach" lies with the determination of when the costs of providing access outweigh the benefits of that access. Specifically, who makes such a determination and how is that determination made?
2) Although the rule may be very straightforward for the majority of situations, difficulties in determining whether or not a particular business falls within the exceptions would subject web masters to becoming familiar with the latest determinations of those fringe areas. In this regard, the rule may make access unduly complex. Some examples of real world issues are:

    - Click-stream Data: Click-stream data compilation is difficult and expensive. It may be personally identifiable if a consumer becomes a customer of the commercial web site. An example might be where a consumer surfs around a web site and then decides to purchase an item or open an account. The consumer would then provide information to the web site operator that could be traced back to the click-stream data. One web session could result in numerous entries in each of these categories:

        * URL lists;
        * Web addresses;
        * Web session/duration levels;
        * Log-in or web page access information.

- Interactive Fora Data: Information entered by a consumer in a chat room, interactive forum or Webcast would be difficult and expensive to compile. This type of information can be personally identifiable if the interactive event calls for registration or a log in process. The information may be scattered throughout a broadcast or transmission, resulting in small bits of data.

- Customer Preferences Data: Preference information that is controlled by the consumer would be difficult and expensive to compile. As an example, a consumer of a commercial web site might become a customer and establish a shopping cart or a watch list for securities. The information might be changed at any time by the consumer (e.g., watch list can be revised at will). It would be very difficult and expensive to compile and provide access to such information on a historical basis.

3) Because businesses would not be required to provide access unless PII is "retrievable in the ordinary course of business," access rights could vary quite a bit from business to business, or across different types of businesses. Businesses may try to use nuances in the interpretation of "retrievable in the ordinary course of business" to avoid providing access. Potentially, a business could set up its data structures so that the data could be used to make decisions about consumers without being retrievable as a separate bit of information.
4) Consumers may have a significant interest in seeing data derived from information collected about them. As this data is what is used to make decisions based on their behavior, providing access may increase consumer awareness about what is being communicated about them and the potential impact of this information.
5) Limiting access to only that information which is collected online from that consumer does not allow the consumer to see the scope of any profiling that may be undertaken. Consumers will not be aware of what information is being used by businesses to make decisions.
6) Although this approach provides access to click-stream information when linked to PII, click-stream information attached to a Globally Unique Identifier also poses a risk to personal privacy. Consumers may expect to be able to see how they might be targeted based on this not identifiable, yet personal, collection of data.
7) The exceptions for providing access are too broad and unfairly limit individual access in favor of business interests. While rights to access should be weighed in balance with other considerations, the current access principles allow the entities least likely to consider the rights of the data subject - the data collector - to make that determination. The current access principle allows for numerous situations for refusal to access on the basis of expense or burden….[5]
8) Any fee (limited to $15 under this approach) may unduly limit the ability of consumers to access their information or it may lessen the attractiveness of accessing personal information.

[5] Commentary by the Trans Atlantic Consumer Dialogue on the Safe Harbor Access policy.

Total Access Approach

The Federal Trade Commission could also consider an expanded version of the "default rule approach" where access would also be provided to derived data[6] and data collected from both off-line and on-line environments. Under the "default rule approach," access is granted to off-line information only when that information is merged with on-line information. Under this "total access approach," access would be granted to information gathered off-line if it could be linked to information collected on-line. Furthermore, access to non-PII could also be provided if the non-PII was linked to a GUID. Under this approach, if a business has the ability to provide access, the business should provide access. Some exceptions could be allowed, such as when proprietary information would be unreasonably jeopardized. This could be characterized as more of a "total access approach." In keeping with the purpose of providing consumers as much access as possible, businesses would provide initial access for free, while charging for repetitive access requests or terminating access upon unduly repetitive access requests.

This approach would implicate the full range of costs and benefits for businesses and consumers. For businesses, this approach would lead to a substantial increase in costs, including: any required modifications or new design requirements placed on existing systems, new storage costs, new personnel costs, new legal costs and potential increased liability. Consumers would also experience additional costs, such as: pass through costs for system upgrades, new personnel, etc., potential opportunity costs of businesses not investing in new products, potential loss of privacy if someone other than the consumer wrongly access this personal information, and the potential privacy threat posed from the aggregation of personal data that would not otherwise be aggregated. On the other hand, this broad access could significantly benefit businesses. By providing greater access rights, businesses could increase the reliability and accuracy of data, could build consumer confidence and trust, could experience a public relations benefit, could make better decisions based on better data, could expand markets by giving consumers greater confidence in online privacy, and could experience greater efficiencies if they limit information collection to only what is necessary. Consumers benefits are also increased by a total access approach. Consumers might experience an enriched understanding of data collection practices, increased confidence in the online environment, more control over the accuracy of personal information, the ability to identify inaccurate data before it harms them, the ability to make better privacy decisions in the marketplace (including decisions to protect anonymity), and the ability to better police businesses for compliance with any stated policies.

Proponents would argue:

---

[6] Derived (or inferred information) has been defined by the Online Access & Security Committee as: "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)" and "deductive information inferred from detailed data which has proprietary value based upon the unique business logic applied to raw data (e.g. profile information)." Derived data is similar to credit scores in the context of credit reports.

1) Consumers should be able to access all of these types of information that are being collected about them. Without these types of information, consumers will not know the extent of profiling that is occurring. Moreover, information collected off line may pose just as much of a privacy risk as that collected on line.
2) Derived data may have a life impact and should be able to accessed by consumers.

Opponents would argue:

1) Providing this scope of access would be extremely costly to businesses and may result in too little consumer gain. Some of the other fair information practices (such as notice and choice) may be more important in protecting privacy.
2) Access to click-stream data and derived data will do little to improve its accuracy, which should be a predominant consideration when deciding access rights.
3) Access to derived data may jeopardize businesses by forcing them to disclose internal practices and proprietary information.
4) Companies in an off-line setting do not have to provide access to how they make decisions – on-line businesses should not be treated disparately.
5) Click-stream data, when not attached to PII, poses a very little privacy risk. Providing access to this information would be counterproductive because of the need to authenticate such information.

A Case-by-Case Approach

A third approach would be to treat different information differently, depending on a calculus involving the content of the information, the holder of the information, the source of the information, the likely use of the information. This approach is necessarily more complex, recognizing as it does that each different type of data raises different issues. The challenge therefore would be to develop an administrable set of rules.

Why this approach?

While an approach establishing a default rule of access enjoys easier application, it may be that it does not reflect the real purposes behind providing access. We have heard, both in the larger committee meetings and our subgroup meetings that the purpose behind providing access may be more limited than promoting consumer awareness. For example, the purpose may not be to enshrine "consumer privacy" but rather to protect data and ensure its accuracy. In fact, the purpose may be as limited as providing consumers an opportunity to correct erroneous data (and not to provide consumers an opportunity simply to know what's out there). A case by case approach may allow a more precise weighing of whether considering the nature of the data, the consumer's reasonable expectations about the data and the costs of providing access to the data, access to a particular type of data is warranted.

How would this approach work?

Essentially this approach would assign different access rights to different data. Given the many factors in the calculus, the permutations are extensive. The following is one

example of this approach (*in italics)*. Although a case-by-case approach can be very complex, the following example shows how a case-by-case approach could result in a manageable rule. The outcome of the following example is also very similar to the outcome of the "default rule approach," even though it may have involved a different analysis.

Consumers should be provided access to information about them and about their relationship with the business. Information about the consumer includes information that describes them (e.g., identity, contact information, consumer specified personal preferences), information that describes their relationship with the business (account numbers, account balances, etc.).
Information about the consumer's relationship with the business includes information that describes the history of their commercial transactions with the business (e.g., purchases, returns), and information about accounts maintained for the consumer with the business.
Consumers should only be given access to information for which it is possible to unambiguously authenticate that the person requesting access is the person the information is about.

The consumer needn't be given access to metadata used by the business solely for the purpose of facilitating an ongoing relationship with the consumer (e.g., GUID's), temporary/incidental data maintained by the business solely for the purpose of maintaining the integrity of interactions with the consumer (e.g., transaction audit records), or inferences the business has derived from other information (e.g., inferred preferences).

Type of Access

View

Consumers should be able to view all information to which they have access.

Edit

Consumers should be able to edit all information to which they have access that is not certified by the business or a 3rd party.
The business should provide a process by which consumers can challenge the correctness of the certified information and request changes to the information. The business is not obligated to change information that it believes is correct per its own certification (e.g., the record of a purchase transaction) or the certification of a 3rd party, but should provide a process by which disagreements concerning the correctness of the information can be arbitrated.

Delete

Consumers should be able to delete all consumer contributed information.
The business should provide a process by which consumers can challenge the correctness or appropriateness of information from other sources and request deletion of the information. The business is not obligated to delete 3rd party sourced or self-sourced

information that it believes is correct and appropriate to retain, but should provide a process by which disagreements concerning the accuracy and appropriateness of the information can be arbitrated.

Means of Access

Access should be provided via a means appropriate for the type of information and consistent with its storage and use by the business. If the business stores the information in online storage such that it is instantly available for use by the business (e.g., as part of an online transaction processing system or a web based e-commerce system), then instantaneous online access should be provided to consumers via an appropriate online terminal (e.g., web browser, ATM machine, telephone voice response unit).
If the business stores the information in storage for processing by batch processing systems[7] (e.g., a batch billing system), then the information should be available to consumers via a frequently (e.g., once per week) scheduled batch process (e.g., a report run at regularly scheduled intervals and mailed to the consumer).
If the business stores the information in offline storage (e.g., magnetic tapes stored offsite), then the information should be available to consumers via an ad-hoc batch process (e.g., scheduled on demand).

Cost to Consumers

There should be no charge to consumers for reasonable requests for view, edit and delete access to online information about them.
Consumer requests for access no more frequently than the rate at which the information changes under normal circumstances are considered reasonable requests for access. A business may assess a reasonable charge to cover its expenses for more frequent requests to online information.
Businesses may also assess reasonable charges to cover their expenses for batch access requests and requests to offline information.

[7] Rather than debate what is meant by "online information," I've chosen to include all information that could have been collected online or used online, even if it is no longer stored in an "online" system.

Source	***Certification***	***Minimum Level of Access***
Consumer Contributed	*Consumer Certified*	*View, Edit, Delete*
	3rd Party Certified	
	Self-Certified	
	Uncertified	
3rd Party Sourced	*Consumer Certified*	*View, Edit, Challenge*
	3rd Party Certified	*View, Challenge*
	Self-Certified	*View, Challenge*
	Uncertified	*View, Edit, Challenge*
Self-Sourced	*Consumer Certified*	*View, Edit, Challenge*
	3rd Party Certified	*View, Challenge*
	Self-Certified	*View, Challenge*
	Uncertified	*View, Edit, Challenge*

Glossary

View Access	*The ability of a consumer to examine a piece of information.*
Edit Access	*The ability of a consumer to change a piece of information.*
Delete Access	*The ability of a consumer to remove a piece of information.*
Challenge Access	*The ability of the consumer to request that a piece of information be changed or deleted (usually because the consumer considers the information incorrect or unnecessary for the business to retain).*
Consumer Certified	*Information about a consumer that the consumer has asserted is correct. For example, shopping preferences submitted by the consumer.*
3rd Party Certified	*Information about a consumer that a 3rd party has asserted is correct. For example, a medical diagnoses provided by a physician.*
Self-Certified	*Information about a consumer that the business asserts is correct. For example, the information associated with a transaction between the consumer and the business.*
Uncertified	*Information collected by the business that is not certified by the consumer, a 3rd party or the business. For example, click stream information that may or may not represent the actions of a particular consumer.*
Consumer Contributed	*Information the consumer has explicitly provided directly to the business. For example, the consumer's credit card number as entered by the consumer in the course of completing a transaction.*
3rd Party Sourced	*Information provided to the business by a 3rd party. For example, a credit report provided by a credit reporting agency.*

Self-Sourced Information collected by the business without the active participation of the consumer. For example, click stream data.

How does this approach differ from the other approaches?

It may be that much of the data gets treated similarly under each of the approaches. On the other hand, it is clear that under this third approach, there will be categories of data to which access is more limited than in the other approaches. For example, inferred data, "non-factual Data" or internal identifiers may be less accessible than under the other approaches. This approach does afford the flexibility to alter the calculus however: if the decision is to protect so called sensitive information: financial, health or relating to children, then this information, regardless of its provenance should be accessible.

Proponents would argue:

1) By allowing each type of data to be considered separately, we can undertake a more accurate balance of the propriety of providing access.
2) This approach provides a more realistic way to vindicate both consumer and business expectations.
3) This approach would depress costs to all consumers. With a broad based approach that encourages access to most data, consumers less interested are forced to bear the costs of creating the infrastructure. A narrower approach would allow costs to more fairly apportioned.

Opponents would argue:

1) There are far too many factors involved to allow a comprehensible set of rules to emerge. Moreover, many of the factors, e.g. sensitivity are difficult to assess objectively.
2) This approach does not recognize that the predominant purpose of providing access is to inform consumers of what is "out there" about them.